Grant agreement for: Collaborative project*

Annex I - "Description of Work"

Project acronym: EURECA

Project full title: " Enabling information re-Use by linking clinical REsearch and CAre "

Grant agreement no: 288048

Version date: 2011-12-09

Table of Contents

Part A

Workplan Tables

A1:
Project summary

Project Number [1]	288048	Project Acronym [2]	EURECA

One form per project

General information			
Project title [3]	Enabling information re-Use by linking clinical REsearch and CAre		
Starting date [4]	01/02/2012		
Duration in months [5]	42		
Call (part) identifier [6]	FP7-ICT-2011-7		
Activity code(s) most relevant to your topic [7]	:		
Free keywords [8]	semantic interoperability, bi-directional linkage, clinical research, safety, secondary use of EHR, healthcare standards, data protection, security		

Abstract [9]

EURECA aims to build an advanced, standards-based and scalable semantic integration environment enabling seamless, secure and consistent bi-directional linking of clinical research and clinical care systems to: 1.Support more effective and efficient execution of clinical research by Allowing faster eligible patient identification and enrolment in clinical trials, Providing access to the large amounts of patient data, Enabling long term follow up of patients, Avoid the current need for multiple data entry in the various clinical care. 2.)Allow data mining of longitudinal EHR data for early detection of patient safety issues related to therapies and drugs that would not become manifest in a clinical trial either due to limited sample size or to limited trial duration, 3:)Allow for faster transfer of new research findings and guidelines to the clinical setting (from bench-to-bedside), 4.) Enable healthcare professionals to extract in each patient's case the relevant data out of the overwhelmingly large amounts of heterogeneous patient data and treatment information. At the core of the project will be achieving semantic interoperability among EHR and clinical trial systems, consistent with existing standards, while managing the various sources of heterogeneity: technology, medical vocabulary, language, etc. This requires the definition of sound information models describing the EHR and the clinical trial systems, and capturing the semantics of the clinical terms by standard terminology systems. The scalability of the solution will be achieved by modularization, identifying core data subsets covering the chosen clinical domains. We demonstrate and validate concepts developed in EURECA by implementing a set of software services and tools that we deploy in the context of pilot demonstrators. EURECA will develop solutions that fulfill the data protection and security needs and the legal, ethical and regulatory requirements related to linking research and EHR data.

A2:
List of Beneficiaries

Project Number [1]	288048	Project Acronym [2]	EURECA

List of Beneficiaries

No	Name	Short name	Country	Project entry month[10]	Project exit month
1	PHILIPS ELECTRONICS NEDERLAND B.V.	Philips	Netherlands	1	42
2	FOUNDATION FOR RESEARCH AND TECHNOLOGY HELLAS	FORTH	Greece	1	42
3	Institut Jules Bordet	IJB	Belgium	1	42
4	CUSTODIX NV	CUSTODIX	Belgium	1	42
5	UNIVERSITAET DES SAARLANDES	UdS	Germany	1	42
6	THE CHANCELLOR, MASTERS AND SCHOLARS OF THE UNIVERSITY OF OXFORD	UOXF	United Kingdom	1	42
7	FRAUNHOFER-GESELLSCHAFT ZUR FOERDERUNG DER ANGEWANDTEN FORSCHUNG E.V	FhG	Germany	1	42
8	VERENIGING VOOR CHRISTELIJK HOGER ONDERWIJS WETENSCHAPPELIJK ONDERZOEK EN PATIENTENZORG	VUA	Netherlands	1	42
9	BREAST INTERNATIONAL GROUP - AISBL	BIG	Belgium	1	42
10	GOTTFRIED WILHELM LEIBNIZ UNIVERSITAET HANNOVER	LUH	Germany	1	42
11	XEROX SAS	XEROX	France	1	42
12	UNIVERSIDAD POLITECNICA DE MADRID	UPM	Spain	1	42
13	STICHTING MAASTRICHT RADIATION ONCOLOGY MAASTRO CLINIC	MAASTRO	Netherlands	1	42
14	ecancermedicalscience AG	eCancer	Switzerland	1	42
15	EUROPEAN INSTITUTE FOR HEALTH RECORDS	EUROREC	France	1	42
16	STONEROOS B.V.	SIT	Netherlands	1	42
17	GBG FORSCHUNGS GMBH	GBG	Germany	1	42
18	NATIONAL RESEARCH COUNCIL CANADA	NRC	Canada	1	42

A3:
Budget Breakdown

Project Number [1]	288048	Project Acronym [2]	EURECA

One Form per Project

Participant number in this project [11]	Participant short name	Fund. % [12]	Ind. costs [13]	Estimated eligible costs (whole duration of the project)					Requested EU contribution
				RTD / Innovation (A)	Demonstration (B)	Management (C)	Other (D)	Total A+B+C+D	
1	Philips	50.0	A	1,918,634.00	0.00	577,094.00	0.00	2,495,728.00	1,536,411.00
2	FORTH	75.0	A	834,336.00	0.00	31,200.00	0.00	865,536.00	656,952.00
3	IJB	75.0	T	652,000.00	0.00	3,000.00	0.00	655,000.00	492,000.00
4	CUSTODIX	75.0	A	1,207,750.00	0.00	81,700.00	0.00	1,289,450.00	987,512.00
5	UdS	75.0	T	624,000.00	0.00	35,600.00	0.00	659,600.00	503,600.00
6	UOXF	75.0	T	686,128.00	0.00	18,172.00	0.00	704,300.00	532,768.00
7	FhG	75.0	A	1,212,712.00	0.00	7,500.00	0.00	1,220,212.00	917,034.00
8	VUA	75.0	T	896,000.00	0.00	33,125.00	0.00	929,125.00	705,125.00
9	BIG	75.0	F	410,316.00	0.00	92,280.00	0.00	502,596.00	400,017.00
10	LUH	75.0	T	462,987.00	0.00	48,000.00	0.00	510,987.00	395,240.00
11	XEROX	50.0	A	625,500.00	0.00	0.00	0.00	625,500.00	312,750.00
12	UPM	75.0	A	804,298.00	0.00	21,067.00	0.00	825,365.00	624,290.00
13	MAASTRO	75.0	T	670,912.00	0.00	3,000.00	0.00	673,912.00	506,184.00
14	eCancer	75.0	T	418,400.00	0.00	3,000.00	0.00	421,400.00	316,800.00
15	EUROREC	75.0	F	199,350.00	0.00	0.00	0.00	199,350.00	149,512.00
16	SIT	75.0	T	344,000.00	0.00	0.00	0.00	344,000.00	258,000.00
17	GBG	75.0	T	336,720.00	0.00	0.00	0.00	336,720.00	252,540.00
18	NRC	75.0	S	139,020.00	0.00	0.00	0.00	139,020.00	104,265.00
Total				12,443,063.00	0.00	954,738.00	0.00	13,397,801.00	9,651,000.00

Note that the budget mentioned in this table is the total budget requested by the Beneficiary and associated Third Parties.

*** The following funding schemes are distinguished**

Collaborative Project (if a distinction is made in the call please state which type of Collaborative project is referred to: (i) Small of medium-scale focused research project, (ii) Large-scale integrating project, (iii) Project targeted to special groups such as SMEs and other smaller actors), Network of Excellence, Coordination Action, Support Action.

1. Project number

The project number has been assigned by the Commission as the unique identifier for your project, and it cannot be changed. The project number **should appear on each page of the grant agreement preparation documents** to prevent errors during its handling.

2. Project acronym

Use the project acronym as indicated in the submitted proposal. It cannot be changed, unless agreed during the negotiations. The same acronym **should appear on each page of the grant agreement preparation documents** to prevent errors during its handling.

3. Project title

Use the title (preferably no longer than 200 characters) as indicated in the submitted proposal. Minor corrections are possible if agreed during the preparation of the grant agreement.

4. Starting date

Unless a specific (fixed) starting date is duly justified and agreed upon during the preparation of the Grant Agreement, the project will start on the first day of the month following the entry info force of the Grant Agreement (NB : entry into force = signature by the Commission). Please note that if a fixed starting date is used, you will be required to provide a detailed justification on a separate note.

5. Duration

Insert the duration of the project in full months.

6. Call (part) identifier

The Call (part) identifier is the reference number given in the call or part of the call you were addressing, as indicated in the publication of the call in the Official Journal of the European Union. You have to use the identifier given by the Commission in the letter inviting to prepare the grant agreement.

7. Activity code

Select the activity code from the drop-down menu.

8. Free keywords

Use the free keywords from your original proposal; changes and additions are possible.

9. Abstract

10. The month at which the participant joined the consortium, month 1 marking the start date of the project, and all other start dates being relative to this start date.

11. The number allocated by the Consortium to the participant for this project.

12. Include the funding % for RTD/Innovation – either 50% or 75%

13. Indirect cost model

A: Actual Costs
S: Actual Costs Simplified Method
T: Transitional Flat rate
F :Flat Rate

Workplan Tables

Project number

288048

Project title

EURECA—Enabling information re-Use by linking clinical REsearch and CAre

Call (part) identifier

FP7-ICT-2011-7

Funding scheme

Collaborative project

WT1
List of work packages

Project Number [1]	288048	Project Acronym [2]	EURECA

LIST OF WORK PACKAGES (WP)

WP Number [53]	WP Title	Type of activity [54]	Lead beneficiary number [55]	Person-months [56]	Start month [57]	End month [58]
WP 1	User needs	RTD	5	78.00	1	30
WP 2	Architecture, Standards and Integration	RTD	4	100.00	1	42
WP 3	Information extraction and data access	RTD	11	103.00	1	38
WP 4	Semantic interoperability	RTD	12	181.00	1	42
WP 5	Data mining and knowledge discovery	RTD	6	99.00	1	42
WP 6	Applications, semantic reasoning and decision support	RTD	8	147.00	1	42
WP 7	Ethics, Legislation, Privacy and Security	RTD	10	68.00	1	42
WP 8	Q&A, Evaluation and Validation	RTD	2	56.00	6	42
WP 9	Models, deployment and clinical pilots	RTD	1	197.00	1	42
WP 10	Knowledge management	RTD	9	81.00	1	42
WP 11	Project Management	MGT	1	56.00	1	42
			Total	1,166.00		

WT2:
List of Deliverables

Project Number [1]	288048	Project Acronym [2]	EURECA

List of Deliverables - to be submitted for review to EC

Delive-rable Number [61]	Deliverable Title	WP number [53]	Lead benefi-ciary number	Estimated indicative person-months	Nature [62]	Dissemi-nation level [63]	Delivery date [64]
D1.1	User needs and specifications for the EURECA environment and software services	1	5	19.00	R	PU	6
D1.2	Definition of relevant user scenarios based on input from users	1	3	19.00	R	PU	9
D1.3	Report on state of the art on relevant external knowledge and data sources and on reusable tools	1	13	20.00	R	PU	12
D1.4	Consolidation of the user needs, use-case development and requirements analysis	1	5	20.00	R	PU	18
D2.1	State of the art report on standards	2	2	16.00	R	PU	7
D2.2	Initial EURECA architecture	2	4	17.00	R	PU	12
D2.3	Initial report on the interface layer towards external systems	2	1	16.00	R	PU	18
D2.4	Integration guidelines: Interoperability Framework and Integration Profiles	2	4	17.00	R	PU	24
D2.5	EURECA architecture and interface layer update	2	12	17.00	R	PU	30

WT2:
List of Deliverables

Delive-rable Number [61]	Deliverable Title	WP number [53]	Lead benefi-ciary number	Estimated indicative person-months	Nature [62]	Dissemi-nation level [63]	Delivery date [64]
D2.6	Final report on the EURECA architecture, reference implementation and integration	2	4	17.00	R	PU	42
D3.1	Initial prototype for concept extraction out of EHR free text	3	11	18.00	P	PU	12
D3.2	Initial prototype for relation identification between concepts	3	11	19.00	P	PU	18
D3.3	Service for uniform access to clinical trial data and other external sources	3	1	16.00	P	PU	18
D3.4	Recommendations for extended minimal set of data representing clinical trials	3	3	15.00	R	PU	24
D3.5	Refined IE prototypes based on evaluation with the users	3	11	15.00	P	PU	30
D3.6	Data model for clinical trial data repository	3	18	20.00	O	PU	38
D4.1	Requirements analysis and selection of the initial clinical scenarios for core datasets	4	3	31.00	R	PU	7
D4.2	Initial proposal for the core datasets	4	1	30.00	R	PU	18
D4.3	Initial proposal for the mapping formalism and mappings to	4	12	30.00	R	PU	24

WT2:
List of Deliverables

Delive-rable Number [61]	Deliverable Title	WP number [53]	Lead benefi-ciary number	Estimated indicative person-months	Nature [62]	Dissemi-nation level [63]	Delivery date [64]
	EHR and CT models						
D4.4	Initial prototype of the semantic interoperability framework	4	12	30.00	P	PU	32
D4.5	Extension of the core data sets	4	7	30.00	O	PU	36
D4.6	Final prototype of the semantic interoperability platform	4	12	30.00	P	PU	40
D5.1	Requirements analysis and knowledge discovery scenarios	5	7	16.00	R	PU	7
D5.2	State of the art review of existing methods and tools for hypothesis generation and association stud	5	6	16.00	R	PU	12
D5.3	Initial prototype of the generic knowledge discovery framework	5	7	16.00	P	PU	24
D5.4	Initial services for hypothesis generation and association studies for safety risks and new research	5	6	17.00	P	PU	28
D5.5	Refined generic knowledge discovery framework an services based on evaluation with users	5	7	17.00	P	PU	36
D5.6	Validation of the knowledge discovery services and framework	5	6	17.00	R	PU	40

WT2:
List of Deliverables

Delive-rable Number [61]	Deliverable Title	WP number [53]	Lead benefi-ciary number	Estimated indicative person-months	Nature [62]	Dissemi-nation level [63]	Delivery date [64]
D6.1	Formalization of eligibility conditions of CT and a patient recruitment method	6	1	20.00	R	PU	9
D6.2	Initial prototype of the EURECA patient recruitment service for CTs	6	8	21.00	P	PU	18
D6.3	Initial prototype of EURECA safety service based on EHR data	6	1	21.00	P	PU	18
D6.4	Initial prototype of the EURECA contextualization service	6	8	21.00	P	PU	24
D6.5	Initial prototype of the EURECA safety service including PHR data	6	16	21.00	P	PU	32
D6.6	Method for incorporation of research findings into medical guidelines	6	8	21.00	R	PU	36
D6.7	Refined services based on evaluation with the users and validation	6	8	22.00	P	PU	40
D7.1	Initial EURECA legal and ethical requirements	7	10	13.00	R	PU	11
D7.2	Initial EURECA security and privacy services	7	4	13.00	P	PU	12
D7.3	EURECA security and privacy services	7	4	14.00	P	PU	24
D7.4	Analysis of the most important	7	10	14.00	R	PU	30

WT2:
List of Deliverables

Delive-rable Number [61]	Deliverable Title	WP number [53]	Lead benefi-ciary number	Estimated indicative person-months	Nature [62]	Dissemi-nation level [63]	Delivery date [64]
	IP issues within EURECA						
D7.5	The EURECA data protection framework	7	4	14.00	R	PU	42
D8.1	Evaluation and validation procedures for the EURECA environment	8	2	9.00	R	PU	11
D8.2	Specifications of the evaluation and validation scenarios for the different EURECA components	8	1	9.00	R	PU	24
D8.3	Report on evaluation and validation of EURECA components	8	2	9.00	R	PU	28
D8.4	Specifications of the evaluation and validation scenarios and demonstrators for the clinical pilots	8	2	9.00	R	PU	30
D8.5	Report on the evaluation and validation of the EURECA environment and services	8	2	10.00	R	PU	40
D8.6	Report on the user workshops at clinical sites	8	14	10.00	R	PU	42
D9.1	Report on the development environment and on the available test data	9	3	28.00	R	PU	9
D9.2	Canonical models of EHRs and CT systems	9	1	28.00	O	PU	12

WT2:
List of Deliverables

Delive-rable Number [61]	Deliverable Title	WP number [53]	Lead benefi-ciary number	Estimated indicative person-months	Nature [62]	Dissemi-nation level [63]	Delivery date [64]
D9.3	Initial proposal for the mapping formalism and mappings to EHR and CT models	9	12	28.00	R	PU	24
D9.4	Solution providing uniform access to sources	9	7	28.00	P	PU	28
D9.5	Report on preparation of the deployment environment for the clinical pilots	9	2	28.00	R	PU	32
D9.6	Guidelines and recommendations for joining the EURECA environment	9	4	28.00	R	PU	36
D9.7	Validation of the EURECA technologies at pilot sites	9	1	29.00	P	RE	40
D10.1	Communication portal Twiki	10	9	8.00	O	CO	3
D10.2	External project website	10	14	8.00	O	PU	6
D10.3	Initial dissemination plan	10	9	8.00	R	RE	9
D10.4	Initial exploitation plan	10	13	8.00	R	RE	12
D10.5	Project newsletter	10	14	9.00	O	PU	12
D10.6	Report on the EURECA workshop/ launching event	10	14	8.00	R	PU	32
D10.7	One-stop-shop at eCancer for all interested in recruitment, participation, exploitation and applicat	10	14	8.00	O	PU	36

WT2:
List of Deliverables

Delive-rable Number [61]	Deliverable Title	WP number [53]	Lead benefi-ciary number	Estimated indicative person-months	Nature [62]	Dissemi-nation level [63]	Delivery date [64]
D10.8	Final dissemination plan	10	9	8.00	R	PU	36
D10.9	Final exploitation plan including sustainability plan	10	13	8.00	R	PU	42
D10.10	The EURECA certification programme	10	15	8.00	O	PU	42
D11.1	Public summary of project	11	1	28.00	O	PU	1
D11.2	Internal project website	11	1	28.00	O	CO	3
			Total	1,166.00			

WT3:
Work package description

Project Number [1]	288048	Project Acronym [2]	EURECA

One form per Work Package

Work package number [53]	WP1	Type of activity [54]	RTD
Work package title	User needs		
Start month	1		
End month	30		
Lead beneficiary number [55]	5		

Objectives

WP1 will elaborate on the user needs and on the user and technical requirements for the proposed environment. This will provide the user perspective of the project and will address the technological requirements (in conjunction with all other work packages) from the user and domain application standpoint. In this WP, user scenarios will be built, analyzed and detailed into user requirements.
The main objectives of this WP are therefore to identify the users and their needs, to define and prioritize comprehensive user scenarios based on which to develop use cases and extract technical requirements, and to define regulatory requirements to which compliance needs to be ensured.
The definition of requirements on basis of scenarios will be based on the approach from the European Commission's funded ESPRIT 21903 'CREWS' (Cooperative Requirements Engineering With Scenarios) long-term research project. We will use a simplified version of the process in order to extract requirements from scenarios, defined as "facts describing an existing system and its environment including the behaviors of agents and sufficient context information to allow discovery and validation of system requirements". The requirements engineering process can be decomposed into three activities :
1. elicit requirements from various individual sources;
2. insure that the needs of all users are consistent and feasible; and
3. validate that the requirements so derived are an accurate reflection of user needs.
This model implies a sequential ordering to the activities, with elicitation done once at the very beginning of the process. In reality, though, the process is iterative, with these activities revisited many times. Thus, while requirements elicitation consists of the earliest activities in the requirements engineering process, it cannot be divorced from the subsequent activities. Elicitation will likely iterate with these other activities during requirements development. The techniques often used during the requirements elicitation phase of a project include interviews, scenarios, soft systems methods, prototyping, observations and social analysis, and requirements reuse. The complexity of the domain which is addressed by the EURECA project necessitated that a spiral process of requirements analysis, elicitation, documentation and validation is adopted. Specific techniques have also been selected for the elicitation, negotiation and agreement of requirements as well as their validation. These techniques are scenarios and prototyping.

Description of work and role of partners

T1.1 State-of-the-Art Review
This task will review current EHR and clinical trial systems, tools and software for the integration or linking of clinical care and clinical research information, registries for clinical trials and databases for clinical trial results, treatment guidelines, clinical trial guidelines, repositories of clinical, bio-molecular, and medication information, etc. Special care will be taken in order to ensure that progress and achievements from previous and running EU funded projects will be incorporated if possible. We will also evaluate open source solutions and their uptake.
T1.2 User Needs and Requirements Analysis
In this task we will carry out interviews with potential stakeholders that will have direct or indirect influence/interaction with the EURECA environment and software services. We will consider a) user needs for clinicians (including surgeons, radiation oncologists, medical oncologists, organ specialists, etc.), b) user needs for basic researchers, and c) user needs in the pharmaceutical research. System requirements analysis will be carried out at all levels and the design goals of the system will be specified. Next to the EURECA clinical

partners, we will collect requirements from wider communities of users represented by the clinical research networks supporting EURECA and by the pharmaceutical industry (through our Pharma Advisory Board).
T1.3 Scenario based Requirements
This task will be specifically concerned with the consolidation of the user needs. In this task clinical scenarios from the EURECA clinical partners will be elaborated upon. The initially defined scenarios will be refined based on the user needs. During the project, new scenarios will be developed iteratively to express user needs and requirements which are not covered by the scenarios already defined.
Task 1.4: Definition of the relevant use cases and requirements analysis
This task will be specifically concerned with the consolidation of the user needs, while also taking into account all the legal and ethical requirements. We will define and further develop use cases and analyse the user requirements to support the research, design and development of prototypes in all technical work packages.

Person-Months per Participant

Participant number [10]	Participant short name [11]	Person-months per participant
1	Philips	2.00
2	FORTH	2.00
3	IJB	8.00
4	CUSTODIX	2.00
5	UdS	12.00
6	UOXF	6.00
7	FhG	4.00
8	VUA	2.00
9	BIG	8.00
10	LUH	2.00
11	XEROX	2.00
12	UPM	4.00
13	MAASTRO	6.00
14	eCancer	6.00
15	EUROREC	2.00
17	GBG	8.00
18	NRC	2.00
	Total	78.00

List of deliverables

Delive-rable Number [61]	Deliverable Title	Lead benefi-ciary number	Estimated indicative person-months	Nature [62]	Dissemi-nation level [63]	Delivery date [64]
D1.1	User needs and specifications for the EURECA environment and software services	5	19.00	R	PU	6

WT3:
Work package description

List of deliverables

Delive-rable Number [61]	Deliverable Title	Lead benefi-ciary number	Estimated indicative person-months	Nature [62]	Dissemi-nation level [63]	Delivery date [64]
D1.2	Definition of relevant user scenarios based on input from users	3	19.00	R	PU	9
D1.3	Report on state of the art on relevant external knowledge and data sources and on reusable tools	13	20.00	R	PU	12
D1.4	Consolidation of the user needs, use-case development and requirements analysis	5	20.00	R	PU	18
		Total	78.00			

Description of deliverables

D1.1) User needs and specifications for the EURECA environment and software services: [month 6]

D1.2) Definition of relevant user scenarios based on input from users: [month 9]

D1.3) Report on state of the art on relevant external knowledge and data sources and on reusable tools: [month 12]

D1.4) Consolidation of the user needs, use-case development and requirements analysis: [month 18]

Schedule of relevant Milestones

Milestone number [59]	Milestone name	Lead benefi-ciary number	Delivery date from Annex I [60]	Comments
MS2	Requirements and specifications of the EURECA environment	5	9	

WT3:
Work package description

Project Number [1]	288048	Project Acronym [2]	EURECA

One form per Work Package

Work package number [53]	WP2	Type of activity [54]	RTD
Work package title	Architecture, Standards and Integration		
Start month	1		
End month	42		
Lead beneficiary number [55]	4		

Objectives

This work package will define the architecture of the EURECA platform, provide a reference implementation and oversee the overall EURECA platform integration and operation.
The architecture will be defined based on the use scenarios and requirements analysis of WP1. This work package envisages an open, scalable SOA (Service Oriented Architecture) platform and will therefore focus on interoperability and interfacing. The main objective of this WP is to define an interoperability framework, in which the tools and services created by the technical WPs can be fit and subsequently provide a reference implementation of this framework.
Integration of developed modules and services will be guided by the definition of integration profiles. This approach reuses existing systems and services whenever possible (speeding up initial deployment), ensures extensibility on the long run, allows for integration of (future) third party solutions and will enable interfacing with other platforms from similar initiatives and rely on each other for scale and critical mass (see Section 1.2.24).
The adoption of international standards is critical to the creation of this platform; hence a task is dedicated to the identification, evaluation and selection of appropriate standards.
Finally, an iterative approach will be taken to definition and implementation of the architecture.

Description of work and role of partners

Task 2.1 Overall Architecture definition
This task will consolidate the workflows defined by the WP1 scenarios and the results of the user requirements analysis relevant for the overall architecture.
The EURECA architecture will be defined according to the principles of a loosely coupled standardised Service Oriented Architecture. The architecture definition impacts most implementation work within the project and is (on the technical level) key to the success of the project and sustainability beyond the end-date. Major technical decision will therefore be taken by the Technical Board, which will include expert architects of all key EURECA partners.

Task 2.2 Interoperability Framework and definition of Integration Profile
This task will focus on analyzing the requirements to provide an interoperability framework (as foundation to the architecture). The framework should ensure scalability and extensibility of the EURECA platform, and provide a basis for interoperability between services and components that already exist and that will be newly developed within the project.
Integration profiles will be defined, enabling the composition of the EURECA components into end-end solutions answering the user needs.

Task 2.3 Definition of a standards-based Interface Layer towards external Systems
The goal of the interface layer is to hide the heterogeneity of the external EHR, CDW or CDMS data sources for the EURECA services. The interface layer will be defined as a set of abstracted interfaces (e.g. providing functional access through standard EURECA APIs, providing direct data access, etc.).
This detailed work plan for creating the interface layer will be defined by the chosen interoperability standards, the approaches taken in the work packages dealing with semantic interoperability, data mining and end-user applications (WP4, WP5, WP6) and the capabilities of the source systems connected in the clinical pilots. This work will include (but is not limited to) analyzing the abstraction layer requirements and design with respect to:

• Providing data access towards the external data sources, taking into account the information and knowledge models used by the data sources and the EURECA services.
• Exposing functionalities of the source systems, e.g. existing query functionality in an EHR or other relevant functionality such as the IHE RFD and CRD industry standard.
• Analyzing the need for “intermediate” (shadow) extracts of source data as part of the interface layer.
The work will be performed in close cooperation with WP4.

Task 2.4 Overall Security Architecture
In line with the overall view on the EURECA architecture, security functionality should not be provided by a monolith solution (fixed proprietary toolset).
The overall security architecture will build upon widely accepted security standards (e.g. SAML, Liberty-Alliance, WS-*, PKIX, etc.) in order to allow for integration of different security solutions into the EURECA platform. This will ensure that choices with respect to security technology will not hamper connectivity of new organisations and systems to the EURECA platform (which is in reality unfortunately often the case).
The security architecture will need to support integrating tools and services into EURECA which provide functionality for:
• Identity Management (of platform users)
• Authentication and Authorisation (platform authorisation at the service level)
• Trust (establishing trust between service providers)
• Audit (user and data subject or data-set centric)
• Provenance
• Specialized privacy protection solutions

The work will be performed in close cooperation with WP7.

Task 2.5 Standards selection
The adoption of international standards is critical in the creation of this platform. This activity will identify, evaluate (based on regulatory or legislative endorsement, technical valor, scope of adoption and strength of community support) and select standards relevant for EURECA.
The tasks will a.o. deal with identifying the standards that are required by relevant regulatory bodies such as EMA and FDA; examine and select information exchange standards (e.g. IHE, HL7, CDISC BRIDG, CDASH); examine and select security and privacy standards (e.g. SAML, Liberty-Alliance, WS-*, PKIX, ISO TS 25237, SAFE-BioPharma); etc.

Task 2.6 Implementation, Systems Integration and Operations
This task is dedicated to the actual development of the reference implementation of the EURECA architecture. The task will also oversee the integration of services and tools produced in the project
This work will be performed in close cooperation with WP9.

Person-Months per Participant

Participant number [10]	Participant short name [11]	Person-months per participant
1	Philips	14.00
2	FORTH	16.00
3	IJB	2.00
4	CUSTODIX	33.00
5	UdS	4.00
7	FhG	11.00
11	XEROX	2.00
12	UPM	13.00
13	MAASTRO	2.00
15	EUROREC	2.00

WT3:
Work package description

Person-Months per Participant

Participant number [10]	Participant short name [11]	Person-months per participant
16	SIT	1.00
	Total	100.00

List of deliverables

Delive-rable Number [61]	Deliverable Title	Lead benefi-ciary number	Estimated indicative person-months	Nature [62]	Dissemi-nation level [63]	Delivery date [64]
D2.1	State of the art report on standards	2	16.00	R	PU	7
D2.2	Initial EURECA architecture	4	17.00	R	PU	12
D2.3	Initial report on the interface layer towards external systems	1	16.00	R	PU	18
D2.4	Integration guidelines: Interoperability Framework and Integration Profiles	4	17.00	R	PU	24
D2.5	EURECA architecture and interface layer update	12	17.00	R	PU	30
D2.6	Final report on the EURECA architecture, reference implementation and integration	4	17.00	R	PU	42
		Total	100.00			

Description of deliverables

D2.1) State of the art report on standards: [month 7]

D2.2) Initial EURECA architecture: [month 12]

D2.3) Initial report on the interface layer towards external systems: [month 18]

D2.4) Integration guidelines: Interoperability Framework and Integration Profiles: [month 24]

D2.5) EURECA architecture and interface layer update: [month 30]

D2.6) Final report on the EURECA architecture, reference implementation and integration: [month 42]

Schedule of relevant Milestones

Milestone number [59]	Milestone name	Lead benefi-ciary number	Delivery date from Annex I [60]	Comments
MS4	Initial EURECA architecture	4	12	
MS7	Operational EURECA architecture for validation	4	24	

WT3:
Work package description

Project Number [1]	288048	Project Acronym [2]	EURECA

One form per Work Package

Work package number [53]	WP3	Type of activity [54]	RTD
Work package title	Information extraction and data access		
Start month	1		
End month	38		
Lead beneficiary number [55]	11		

Objectives

Our main goal in this WP is to use NLP techniques to support the effort of the domain experts and IT experts to build comprehensive, standards-based information models of the sources, including the information currently only available in free text reports. This extraction process will be gradual and based on the identified core dataset for that domain, i.e. search with a clear goal in mind, not a random search in an infinite space. The idea is to gradually understand the data available in the free text report, find, classify and annotate the relevant concepts and also find the relations when possible.
We will provide tools to perform syntactic and semantic analysis of textual data available in the EHR and Clinical Trial systems available in the EURECA environment. Information Extraction in this context will go far beyond simple pattern matching techniques. Pattern matching techniques are not enough if we want to make a step towards knowledge discovery and reasoning. Information Extraction (IE) usually consists of two main tasks: concept identification and event identification (i.e. relations between concepts). However, in order to obtain accurate results in both concept and event identification, complex linguistic phenomena have to be considered.
From a broad perspective, another aim of this work package is to supplement the integration of clinical trial management systems with EHRs and with sources of external information in support of patient enrolment in clinical trials and of patient safety by facilitating clinical research. In concrete terms, it is also the objective of this work package to develop uniform interfaces for clinical trial repositories and for clinical trial data and other sources or relevant information, such as computerized clinical guidelines.
In addition to enabling usage of the high-level information about trials contained in online registries, it was also recognized that providing access to the raw data collected in trials would facilitate and accelerate clinical research. Thus another objective of this work package is to support the sharing of clinical trial data by designing a generic data model for these data. In addition we propose to investigate the definition of a core set of data that should be submitted to such a repository to consider that a clinical trial data set is “minimally complete”.

Description of work and role of partners

Task 3.1 Review of existent tools and selection of relevant approaches and tools to fit the different EURECA systems and use cases
We believe that instead of a one-fits-all solution, a variety of methods and tools (concept annotators, chunkers, entity recognizers, etc.) need to be combined for each available system to obtain best performance. High quality data extraction can help us build accurate information models of the sources, improve the quality of the data, and ensure high quality semantic linkage among the various systems available at the clinical sites. It is the role of this task to evaluate existing NLP tools and solutions and to evaluate the needs of the different clinical systems, and to select the best suitable combination for each case.

Task 3.2 Concept identification in EHR free text data
The discovery of the underlying information model contained in free text reports requires additional effort that stars with identifying the relevant concepts
1) Concept recognition
Recognizing concepts of interest present in free text is a common task in the processing of medical textual data. Concept recognition is very dependent on the chosen terminologies used in the project as the concepts labels that are retained. Concept recognition is thus strongly linked with decisions taken in WP4: Semantic interoperability.
2) Negated and hedged information processing

Negated information is often present in medical text, especially in EHR information where a negated fact or negated exam results can be extremely informative regarding the state of the patient. Negation is a well-known and difficult problem to handle in NLP. In the medical domain, previous work on this topic has been performed, with interesting results that we believe can still be improved. We intend to use refined NLP techniques based on deep linguistic knowledge for both negated and hedged information processing.
3) Temporal information processing
Temporal information is another element that has to be considered for refined information extraction tasks in the medical domain. Xerox already developed a temporal processing module which is embedded within the general purpose parser. We intend to enrich the temporal processing module to other kinds of medical documents available in the EURECA project and to support the EURECA clinical scenarios. Once again the temporal dimension in IE is a necessary first step for knowledge discovery and reasoning.

Task 3.3 Identification of links (relations) among concepts
Once concepts are detected and annotated with the correct background information (negation, modality, time), an important aspect is to be able to see how these concepts are related and the kind of relations holding between them. Concepts and relations build complex events. An event is thus a frame in which participants consist of terms of interest which are linked according to given types of relations. These links are supported by syntactic dependencies holding between concepts. These dependencies are extracted using NLP techniques. We propose to adapt the Xerox parser to the EURECA domain. The parser computes syntactic and semantic dependencies between the linguistic units (words or terms) present in texts. As the approach adapted is a dependency-based approach, long distance relations between textual fragments can also be calculated. In the EURECA context, events of interest have first to be defined (which kind of relations, which kind of entities participate to the relations). The parser will then be adapted accordingly in order to match occurrences of the defined events in texts.

Task 3.4 Implement a service and an API for the query of clinical trial repositories and other sources such as computerized clinical guidelines
In this task a service federating online repositories and the corresponding API will be developed. The API will allow querying online repositories using the EURECA semantic core. Some aspects of the API will be developed jointly with WP4 and WP9, in order to ensure that a seamless interoperability with other EURECA components is achieved and that the requirements for the pilots are covered (e.g. for the automatic identification of trials relevant for a patient).

Task 3.5 Provide recommendation for extended minimal set of data representing clinical trials
Considering that the “science-oriented” elements composing the WHO minimal trial information is of limited used for meta-analysis and hypothesis generation and validation, a survey will be conducted among clinical partners in the project (and potentially in the wider community) to identify a set of data that could be collected to extend the WHO minimal set in order to provide a substantial research benefit at a small data collection cost. Simple demographics elements are obvious candidates, other data elements usually collected in clinical trials (such as disease-specific biomarkers) may also extend the data set. In this task we also propose a data model for the representation of data collected in the context of clinical trials in view of their utilization for clinical research. The model will be designed to provide an easy (if not transparent) access to the data both from the viewpoint of syntax and of the semantics. The data model will be designed considering the performance issues that may be posed by the increasing usage of high-throughput technologies in clinical trials (e.g. microarrays or UHTS techniques).

Person-Months per Participant

Participant number [10]	Participant short name [11]	Person-months per participant
1	Philips	10.00
2	FORTH	6.00
3	IJB	10.00
4	CUSTODIX	4.00
5	UdS	3.00

WT3:
Work package description

Person-Months per Participant

Participant number [10]	Participant short name [11]	Person-months per participant
6	UOXF	6.00
7	FhG	2.00
8	VUA	2.00
11	XEROX	35.00
12	UPM	4.00
13	MAASTRO	3.00
18	NRC	18.00
	Total	103.00

List of deliverables

Delive-rable Number [61]	Deliverable Title	Lead benefi-ciary number	Estimated indicative person-months	Nature [62]	Dissemi-nation level [63]	Delivery date [64]
D3.1	Initial prototype for concept extraction out of EHR free text	11	18.00	P	PU	12
D3.2	Initial prototype for relation identification between concepts	11	19.00	P	PU	18
D3.3	Service for uniform access to clinical trial data and other external sources	1	16.00	P	PU	18
D3.4	Recommendations for extended minimal set of data representing clinical trials	3	15.00	R	PU	24
D3.5	Refined IE prototypes based on evaluation with the users	11	15.00	P	PU	30
D3.6	Data model for clinical trial data repository	18	20.00	O	PU	38
		Total	103.00			

Description of deliverables

D3.1) Initial prototype for concept extraction out of EHR free text: [month 12]

D3.2) Initial prototype for relation identification between concepts: [month 18]

D3.3) Service for uniform access to clinical trial data and other external sources: [month 18]

D3.4) Recommendations for extended minimal set of data representing clinical trials: [month 24]

D3.5) Refined IE prototypes based on evaluation with the users: [month 30]

D3.6) Data model for clinical trial data repository: [month 38]

Schedule of relevant Milestones

Milestone number [59]	Milestone name	Lead beneficiary number	Delivery date from Annex I [60]	Comments
MS8	Initial evaluation of the EURECA interoperability environment and of the EURECA services	2	32	
MS9	Validation of the EURECA interoperability environment and services.	1	40	

WT3:
Work package description

Project Number [1]	288048	Project Acronym [2]	EURECA

One form per Work Package

Work package number [53]	WP4	Type of activity [54]	RTD
Work package title	Semantic interoperability		
Start month	1		
End month	42		
Lead beneficiary number [55]	12		

Objectives

The main objective of this work package is to provide a platform which will provide the necessary level of semantic interoperability between the EHR systems and the clinical trial systems. This platform will provide a set of data services such as data extraction, appropriate data transformation, translation etc. These data services will be composed into end-user EURECA service such as “Find clinical trial”, “Suggest patients for trial enrolment”, etc.
In the core of the EURECA semantic interoperability platform lie the standard-based semantic core datasets which are linked to the canonical information models that represent the EHR systems and the CT systems respectively. This semantic core datasets together with the devised mappings will enable the linkage between the patient data residing in the EHR and that in the clinical trial systems.
The existing language heterogeneity will also be addressed by translation/mapping of the core data set into the languages used in the information systems for the primary data capture (when a translation of the standard terminologies does not exist in that language).
The EURECA semantic interoperability platform will enable implementation of several scenarios related to the identified issues in patient recruitment and safety in clinical trials, and to enabling large scale data mining and epidemiology studies. Additionally, the increased semantic interoperability and a proper semantic access both to EHR and clinical trial data will also allow to develop a new generation of clinical decision support systems where computerized guidelines can be directly applied to the extracted patient data, helping also the clinical practitioners improve the patient safety and outcomes. Special care will be taken to assure that the entire process of data extraction and linkage is in line with all applicable legal and ethical requirements.
In order to generalize the EURECA approach, we will also conduct a study on extending the extracted core datasets towards other clinical domains and applications (e.g. other types of cancer and other disease domains such as cardiology).

Description of work and role of partners

Task 4.1: Semantic Core Dataset
In this task, we will extract for each domain of focus a well defined set of domain concepts that sufficiently describe the semantics of the chosen clinical domain. This semantic data set will be mapped to concepts from relevant existing standardized terminologies e.g. well established and widely used ontologies such as SNOMED CT, ICD-10, etc. The semantic data set will be validated in concrete use cases, for the different EHR and clinical trial systems available at the clinical and clinical trial sites.
The semantic core dataset is an essential prerequisite to machine processable access to both EHR and Clinical trial data. Concepts in the dataset will have their unique identifiers, well understood meaning as well as a set of synonyms they can be referred as. Addressing the anticipated language heterogeneities, we will work out together with the clinical experts a translation of the core dataset to the languages that are used for the primary data capture (when a mapping to a standard terminology does not exist).
We define the core datasets in a modular, scalable way, describing the clinical areas relevant for our scenarios. We will also conduct a study demonstrating possible extensions of the semantic core dataset approach to other clinical domains.

Task 4.2: Mapping formalism and mappings between the core data set and EHR and CT models
In this task, we intend to identify the requirements for mappings that bridge a semantic core data set with the information models representing the EHR systems and the clinical trial systems. These information models will

be designed in WP9; they provide a canonical view, reflecting the content and the structure of the respective information management system. The proposed mapping formalism should be able to mitigate the foreseen structural and contextual differences between the core dataset and the information models. We will use this formalism to instantiate the necessary schema-level mappings that will be executed by the semantic interoperability platform during the data extraction process.

Task 4.3: The semantic interoperability platform
In this task, we will build a prototype of the EURECA semantic interoperability platform. This platform will utilize the semantic core dataset as well as the schema-level mappings that link to the EHR and CT information models. The platform will be able to execute these mappings during the data extraction phase, instantiating thus the semantic concepts with patient data and/or clinical trial data. The semantic interoperability platform will be the core of the EURECA services, enabling linkage between the patient data in the EHR and the clinical trial systems.
The increased semantic interoperability provided by this platform will also enable new applications in the clinical decision support domain where computerized guidelines can be directly applied to the patient data, helping the clinical practice staff to make the right decision with respect to patient safety and outcomes.

Person-Months per Participant

Participant number [10]	Participant short name [11]	Person-months per participant
1	Philips	18.00
2	FORTH	12.00
3	IJB	16.00
4	CUSTODIX	8.00
5	UdS	16.00
6	UOXF	10.00
7	FhG	18.00
8	VUA	16.00
9	BIG	2.00
11	XEROX	5.00
12	UPM	32.00
13	MAASTRO	8.00
15	EUROREC	1.00
16	SIT	11.00
17	GBG	8.00
	Total	181.00

List of deliverables

Delive-rable Number [61]	Deliverable Title	Lead benefi-ciary number	Estimated indicative person-months	Nature [62]	Dissemi-nation level [63]	Delivery date [64]
D4.1	Requirements analysis and selection of the initial clinical scenarios for core datasets	3	31.00	R	PU	7

WT3:
Work package description

List of deliverables

Delive-rable Number [61]	Deliverable Title	Lead benefi-ciary number	Estimated indicative person-months	Nature [62]	Dissemi-nation level [63]	Delivery date [64]
D4.2	Initial proposal for the core datasets	1	30.00	R	PU	18
D4.3	Initial proposal for the mapping formalism and mappings to EHR and CT models	12	30.00	R	PU	24
D4.4	Initial prototype of the semantic interoperability framework	12	30.00	P	PU	32
D4.5	Extension of the core data sets	7	30.00	O	PU	36
D4.6	Final prototype of the semantic interoperability platform	12	30.00	P	PU	40
		Total	181.00			

Description of deliverables

D4.1) Requirements analysis and selection of the initial clinical scenarios for core datasets: [month 7]

D4.2) Initial proposal for the core datasets: [month 18]

D4.3) Initial proposal for the mapping formalism and mappings to EHR and CT models: [month 24]

D4.4) Initial prototype of the semantic interoperability framework: [month 32]

D4.5) Extension of the core data sets: [month 36]

D4.6) Final prototype of the semantic interoperability platform: [month 40]

Schedule of relevant Milestones

Milestone number [59]	Milestone name	Lead benefi-ciary number	Delivery date from Annex I [60]	Comments
MS6	Specification of the core datasets	12	18	
MS8	Initial evaluation of the EURECA interoperability environment and of the EURECA services	2	32	
MS9	Validation of the EURECA interoperability environment and services.	1	40	

WT3:
Work package description

Project Number [1]	288048	Project Acronym [2]	EURECA

One form per Work Package

Work package number [53]	WP5	Type of activity [54]	RTD
Work package title	Data mining and knowledge discovery		
Start month	1		
End month	42		
Lead beneficiary number [55]	6		

Objectives

The main objective of this work package is to build the EURECA data mining and knowledge discovery services to support the identified user scenarios with a data mining focus. These tools and services will be integrated into an advanced framework addressing both clinical research and care (also supporting this way WP6: Application, semantic reasoning and decision support).
Tools enabling the discovery of new, clinically relevant knowledge will be implemented. Tools for the automatic analysis of biomedical data in the form of data mining and statistical procedures will be implemented. Next to the integration of existing tools, research activities will concentrate on machine learning algorithms for the discovery of long-term patterns of low frequency. This will allow the discovery of patient safety issues that are not frequent enough to be found within the scope of clinical trials. This work package will also investigate the automatic generation of research hypotheses by statistical pattern recognition. The development of new approaches for similarity learning on partly structured domains will allow to support the researcher with help for the selection of data sets for meta analysis, and support of the transfer of analytic solutions to new trials and datasets.

Description of work and role of partners

Task 5.1: Generic knowledge discovery framework
In this task, a generic framework will be set up that allows to scalably and flexibly execute knowledge discovery tasks over all relevant data types. In collaboration with WP3, a flexible architecture will allow to deploy knowledge discovery tools for heavily distributed, parallel or long-term monitoring tasks, as well as for local data analysis. In particular, the framework will support to specify data analysis tasks on a conceptual (workflow) level, which hides all the implementation complexities from the user, and enable the automatic transfer to efficient implementations (which may differ for local vs. distributed analysis and one-time analysis vs. long-term monitoring). Modularization and the implementation of generic data pre-processing and analysis services, combined with intelligent analysis and exploitation of meta data, will facilitate the re-use of solutions, both inside and outside of EURECA.
Task 5.2: Data mining for patient safety
The goal of this task is to develop knowledge discovery tools that combine care (EHR) and research (CT) data to support data mining and association studies to detect patient safety issues and to identify relevant biomarkers that accurately predict risks in patients. New hypotheses can be generated to be validated in large (genomic) studies.
Research-oriented literature and trial-results mining tools will also be developed and/or integrated in this task, to facilitate the research process, for instance to identify and automatically suggest papers in relation to the findings of a specific study.

Task 5.3: Secondary use of data for research
This task will concentrate on the needs of the clinical researcher by implementing tools that are necessary for the analysis of clinical data. This includes tools for the efficient discovery of patterns of low frequency in large populations of patients, or the detection of concept drift for the discovery of newly developing patient safety issues. In addition, intelligent tools to support the setup of trials will be developed, such as tools to support the meta-analysis of trial data or the quick generation and testing of new hypotheses for trials on existing EHR data.

WT3:
Work package description

Person-Months per Participant

Participant number [10]	Participant short name [11]	Person-months per participant
1	Philips	8.00
2	FORTH	10.00
5	UdS	4.00
6	UOXF	30.00
7	FhG	26.00
8	VUA	2.00
11	XEROX	4.00
12	UPM	6.00
13	MAASTRO	6.00
16	SIT	3.00
	Total	99.00

List of deliverables

Delive-rable Number [61]	Deliverable Title	Lead benefi-ciary number	Estimated indicative person-months	Nature [62]	Dissemi-nation level [63]	Delivery date [64]
D5.1	Requirements analysis and knowledge discovery scenarios	7	16.00	R	PU	7
D5.2	State of the art review of existing methods and tools for hypothesis generation and association stud	6	16.00	R	PU	12
D5.3	Initial prototype of the generic knowledge discovery framework	7	16.00	P	PU	24
D5.4	Initial services for hypothesis generation and association studies for safety risks and new research	6	17.00	P	PU	28
D5.5	Refined generic knowledge discovery framework an services based on evaluation with users	7	17.00	P	PU	36
D5.6	Validation of the knowledge discovery services and framework	6	17.00	R	PU	40
		Total	99.00			

Description of deliverables

D5.1) Requirements analysis and knowledge discovery scenarios: [month 7]

D5.2) State of the art review of existing methods and tools for hypothesis generation and association stud: [month 12]

D5.3) Initial prototype of the generic knowledge discovery framework: [month 24]

D5.4) Initial services for hypothesis generation and association studies for safety risks and new research: [month 28]

D5.5) Refined generic knowledge discovery framework an services based on evaluation with users: [month 36]

D5.6) Validation of the knowledge discovery services and framework: [month 40]

Schedule of relevant Milestones

Milestone number [59]	Milestone name	Lead benefi-ciary number	Delivery date from Annex I [60]	Comments
MS8	Initial evaluation of the EURECA interoperability environment and of the EURECA services	2	32	
MS9	Validation of the EURECA interoperability environment and services.	1	40	

WT3:
Work package description

Project Number [1]	288048	Project Acronym [2]	EURECA

One form per Work Package

Work package number [53]	WP6	Type of activity [54]	RTD
Work package title	Applications, semantic reasoning and decision support		
Start month	1		
End month	42		
Lead beneficiary number [55]	8		

Objectives

This WP will implement the software services for more efficient enrolment of patients into clinical trials and better trial execution, and for improved detection and reporting of safety issues.
In this WP, tools able to bring new research results and latest clinical guidelines to the point of care will also be developed. We will apply techniques and implementations from the field of information retrieval and search engines, which will also be adapted to the case of structured and semi-structured data which is available in some of the EURECA scenarios. Using these techniques, the tools will support the clinician with an automatic identification of literature and clinical research results relevant in the context of a patient case, based on both structured and unstructured EHR data, and of relevant clinical guidelines. This approach will exploit live feedback from the system users in its recommendations, and hence will allow generalizing the identification of relevant guidelines with a soft matching approach, which can adapt itself to the current context and person.
Specific objectives of this WP are:
• More efficient recruitment of eligible patients for clinical trials and supporting protocol feasibility evaluation and interactive improvement
• To improve medical care by personalisation of information to address specific information needs, and contextualization to a patient case
• Faster transfer of new research findings into guidelines
• More efficient detection and reporting of safety issues, also reducing the need for multiple data entry
• Use of PHR data for early identification of safety issues and for improved assessment of benefits versus risks of drugs and treatments

Description of work and role of partners

Task 6.1: Software services to support patient recruitment and protocol feasibility
This task captures the development of solutions for the automatic identification of eligible patients for clinical trials, and for interactive protocol design and feasibility.
There are a large number of running clinical trials and the tasks of identifying eligible patients for a specific trial, or of a suitable trial for a patient is not trivial. To be able to automatically verify that a patient satisfies the eligibility criteria of the clinical trial requires a formalisation step from the eligibility condition such that it can be matched with particular patient data. The terminology used in the clinical trials and the patient data have to be aligned. This task will rely on the semantic interoperability layer enabling the linkage among concepts in the CT system to concepts in the EHR system (through the relevant EURECA core data set) This service will enable faster identification of eligible patients for entering a clinical trial, and results in a more efficient recruitment.
Task 6.2: Software services supporting the identification and reporting of SAE/SUSAR based on EHR and PHR data
In this task we will develop a service/tool to support identification and reporting of serious side effects in clinical trials (SAE/SUSARs) according to the specification provided by WP1. The models for the SAE/SUSAR reports will be mapped to the EURECA core semantic data set to achieve semantic linkage within the EURECA environment. That enables the reporting module to extract the required data from patient records, which are integrated into the semantic interoperability platform, and to identify side effects and when applicable generate automatic reports. This task aims to enable early detection of potential safety issues and to avoid multiple data entry.
We will also pay attention to the patient empowerment trend in health care and take into account the patient involvement into the care cycle by integrating patient managed data sources (user reporting). This data will be

acquired by uniform access to PHR and we will demonstrate this use case in the context of the Microsoft's PHR system.
Task 6.3: Adaptation of computerized clinical guidelines based on latest results and clinical decision support
In this task, we will investigate how existing clinical guidelines (e.g. the national breast cancer guidelines) can be linked to the EURECA platform. We will create a model of the chosen clinical guidelines and align it with the core EURECA dataset enabling thus the semantic interoperability platform not only to bridge the EHR and clinical trial systems, but also to demonstrate that the increased level of semantic interoperability brings substantial benefits to the clinical decision support applications which are often driven by computerized clinical guidelines.
Next, we will investigate how to transfer new research findings into guidelines. These research findings can be stronger evidence or even new evidence. Both can have an effect on the care decision making based on the guideline. In this task we develop a method how to incorporate those new research results in the existing guidelines. This enables to accelerate the improvement of standard care.
Task 6.4: Contextualization of relevant research results and literature
This task aims at personalisation of clinical care based on several important sources of knowledge and information (clinical trial repositories, results repositories, guidelines, literature, other sources of biomedical data and knowledge). We will identify and combine relevant information for a particular patient out of several sources, for instance identifying patient relevant results from clinical trials, or from biomedical data bases.
In this task tools will be developed that provide recommendations concerning literature, guidelines, or completed trials relevant for a patient case under investigation. Alerts concerning possible safety issues may also be provided. The recommendation and alerting functionality will be contextualized and adaptive, such that experts, general practitioners and medical students can receive different kinds of alarms and recommendations which are appropriate for them. In addition, the mechanism for recommendations and alerts will also be adaptive to consider the requirements of different hospitals, and of country-specific healthcare regulations.

Person-Months per Participant

Participant number [10]	Participant short name [11]	Person-months per participant
1	Philips	23.00
2	FORTH	8.00
3	IJB	8.00
5	UdS	14.00
7	FhG	8.00
8	VUA	38.00
10	LUH	4.00
12	UPM	10.00
13	MAASTRO	8.00
14	eCancer	4.00
16	SIT	19.00
18	NRC	3.00
	Total	147.00

WT3:
Work package description

List of deliverables

Delive-rable Number [61]	Deliverable Title	Lead benefi-ciary number	Estimated indicative person-months	Nature [62]	Dissemi-nation level [63]	Delivery date [64]
D6.1	Formalization of eligibility conditions of CT and a patient recruitment method	1	20.00	R	PU	9
D6.2	Initial prototype of the EURECA patient recruitment service for CTs	8	21.00	P	PU	18
D6.3	Initial prototype of EURECA safety service based on EHR data	1	21.00	P	PU	18
D6.4	Initial prototype of the EURECA contextualization service	8	21.00	P	PU	24
D6.5	Initial prototype of the EURECA safety service including PHR data	16	21.00	P	PU	32
D6.6	Method for incorporation of research findings into medical guidelines	8	21.00	R	PU	36
D6.7	Refined services based on evaluation with the users and validation	8	22.00	P	PU	40
		Total	147.00			

Description of deliverables

D6.1) Formalization of eligibility conditions of CT and a patient recruitment method: [month 9]

D6.2) Initial prototype of the EURECA patient recruitment service for CTs: [month 18]

D6.3) Initial prototype of EURECA safety service based on EHR data: [month 18]

D6.4) Initial prototype of the EURECA contextualization service: [month 24]

D6.5) Initial prototype of the EURECA safety service including PHR data: [month 32]

D6.6) Method for incorporation of research findings into medical guidelines: [month 36]

D6.7) Refined services based on evaluation with the users and validation: [month 40]

Schedule of relevant Milestones

Milestone number [59]	Milestone name	Lead benefi-ciary number	Delivery date from Annex I [60]	Comments
MS8	Initial evaluation of the EURECA interoperability environment and of the EURECA services	2	32	
MS9	Validation of the EURECA interoperability environment and services.	1	40	

WT3:
Work package description

Project Number [1]	288048	Project Acronym [2]	EURECA

One form per Work Package

Work package number [53]	WP7	Type of activity [54]	RTD
Work package title	Ethics, Legislation, Privacy and Security		
Start month	1		
End month	42		
Lead beneficiary number [55]	10		

Objectives

The goal of WP7 is to guarantee that the research outcome of EURECA is in line with current European legislation and ethical guidelines. WP7 combines ethical, legal (IT Law) and technical (e-security, e-privacy) expertise in order to define a framework and provide a set of tools and services to allow the EURECA applications to achieving regulatory compliance with minimal effort.
Focus lays on data protection issues, including topics such as informed consent and purpose of use. EURECA data flows will be examined and technically “enhanced” with a data protection building blocks (tools and services) designed and developed in this work package.
WP7 solutions will fit the EURECA security architecture.

Description of work and role of partners

T7.1: The EURECA Data Protection Framework (eDPF)
In this task WP7 will create a Data Protection Framework to guarantee EURECA’s compliance with European data protection regulations and ethical rules. The aim of the EURECA Data Protection/Ethical framework is to enable researchers to do their research without being hindered by legal restrictions, but at the same time take care of the concerned persons’ privacy. Therefore the EURECA Data Protection/Ethical Framework has to be composed of both legal and technical means to balance the needs for advancing medical research and protection of individual rights (privacy).
A researcher should comply with current (complex) legislation automatically by following the policies of the framework to be set up. Task 7.1 includes:

1. Analysis of the relevant ethical and legal regulation.
The processing of personal data is subject to a large number of different regulations already on the International and on the European level. In a first step this WP will define the relevant legal and ethical guidelines of the project, with special regard to the access to EHRs, the purpose of use with respect to treatment on the one and research on the other hand and the identification of the data controller in such networks. Only International and European Regulations shall be considered for this analysis excluding expressly National Regulations.

2. Establishment of an EURECA internal Data Protection Authority
This WP will be responsible for the establishment and guidance of a EURECA internal data protection authority taking care of data protection issues within the project. The EURECA data protection authority will be responsible for the data processing within the project, safeguard the projects compliance with current data protection legislation.
Such a DPA typically needs to be an independent legal body (capable of signing contracts). EURECA will be able to rely on the Center of Data Protection (CDP) for this matter. The CDP was founded as non-profit organization as spin-off of the ACGT project exactly for these purposes (and serves different EU projects). Several EURECA partners are involved as member in the CDP.

3. Informed consent
Uniform informed consent forms for participating patients will be created and provided to clinical partners. Furthermore WP7 will analyze the legal and ethical requirements to incorporate the informed consent into the framework through a “Consent Management Service”, which is considered to be a specialized policy based authorization service.

T7.2: EURECA Security Services
This task is responsible for the implementation of the security services required for EURECA which will be fit into the security architecture designed in WP2. Solutions will need to be provided for dealing with: Identity Management, Authentication and Authorisation, trust, audit and provenance. Where possible, existing (industry standard) solutions will be used (e.g. XAML PDP engines, Shibboleth IdP, ZXID IdP, etc.). This task will deal with a number of specific topics.

• This task will research means to technically enforce the eDPF. For one, the use of policy based authorization services to address the issue of translating complex legal rule sets (for “access to” or “processing of” highly sensitive data) into computer interpretable authorisation decisions will be researched. Suitability of existing policy based authorisation mechanisms, i.e. policies and associated decision engines such as defined by XACML , PERMIS , GAS , etc. will be examined. Extensions to the most appropriate mechanism(s) will be proposed to address the specific eDPF. Need for support can be anticipated for things like the “notion of datasets” in policies and data bound concepts such as “purpose of use” and “conditions on use” leading into research on data-protection related meta-data (e.g. sticky policies).

• An important task is the research for a solution to automate “patient consent” management and integrate it with the overall policy evaluation mechanism. A “Consent Management Service” that guarantees integrity of consent directives and correctly combines them to avoid conflicting preferences, will be researched and developed. The idea is to incorporate this service into the EURECA framework, as a specialised policy based authorization service.

• Appropriate auditing mechanisms fulfilling the eDPF needs will be researched. EURECA will develop audit mechanisms that can provide user-centric and data-centric audit trails in a distributed environment. Auditing will be extended with provenance capabilities according to the needs expressed by the eDPF. For this proposed models such as the Open Provenance Model will be examined.

T 7.3: Privacy Enabled Processes and Services
Within EURECA different processes will “implement” the project goals such as providing bidirectional patient recruitment, long time follow-up of patients, etc. This task is dedicated to ensure that the associated dataflows, which handle particularly sensitive information (medical data), are compliant with the legal and ethical requirements defined in T7.1.
For example, the task encompasses research on feasibility of solutions such as: concentrating the processing of sensitive micro-data at the data source side (and only exporting aggregated result sets – which decreases privacy risks).
Further, research will concentrate on generic building blocks which can be included in the EURECA processes. Although it will only become clear which specific blocks are useful once the EURECA flows are defined, they will most certainly include:
• De-identification (anonymisation & pseudonymisation) tools
o E.g. for the export of data for secondary use, for the creation of “shadow” EHRs which will be used for connecting care centers to EURECA
o EURECA can build on previous work such as CAT, the Custodix Anonymisation Tool (developed within the ACGT project), which takes a generic approach towards de-identification of various data sources.,
o The NLP research done in EURECA will also be applied to further work on de-identification of free text.
• Pseudonym management & patient matching
o E.g. for longitudinal and multicenter follow-up of patients.
These components will be fit into the EURECA processes as services where possible (e.g. deployed by a Trusted Third Party which governs the data protection related aspects of the process). This is in line with the EURECA vision of building a loosely coupled service oriented architecture.

T7. 4: Intellectual Property Rights (IPR)
This task will deliver guidelines for intellectual property (IP) issues of the data, databases and computer programs to be used in EURECA. Therefore IP legislation, e.g. copyright law, trademark law, patent law, will be studied on a European and international level. Legal advice will be given to partners regarding license agreements for software tools used in EURECA. For software developed in EURECA legal advice will be provided as well. The ownership of patient data and results will also be clarified. Legal advice in IP issues will furthermore be provided regarding the exploitation and dissemination of EURECA results, e.g. access rights, rights of use etc.

WT3:
Work package description

Person-Months per Participant

Participant number [10]	Participant short name [11]	Person-months per participant
1	Philips	2.00
3	IJB	2.00
4	CUSTODIX	28.00
9	BIG	8.00
10	LUH	28.00
	Total	68.00

List of deliverables

Delive-rable Number [61]	Deliverable Title	Lead benefi-ciary number	Estimated indicative person-months	Nature [62]	Dissemi-nation level [63]	Delivery date [64]
D7.1	Initial EURECA legal and ethical requirements	10	13.00	R	PU	11
D7.2	Initial EURECA security and privacy services	4	13.00	P	PU	12
D7.3	EURECA security and privacy services	4	14.00	P	PU	24
D7.4	Analysis of the most important IP issues within EURECA	10	14.00	R	PU	30
D7.5	The EURECA data protection framework	4	14.00	R	PU	42
		Total	68.00			

Description of deliverables

D7.1) Initial EURECA legal and ethical requirements: [month 11]

D7.2) Initial EURECA security and privacy services: [month 12]

D7.3) EURECA security and privacy services: [month 24]

D7.4) Analysis of the most important IP issues within EURECA: [month 30]

D7.5) The EURECA data protection framework: [month 42]

Schedule of relevant Milestones

Milestone number [59]	Milestone name	Lead benefi-ciary number	Delivery date from Annex I [60]	Comments
MS3	Initial EURECA legal and ethical requirements	10	11	

Schedule of relevant Milestones

Milestone number [59]	Milestone name	Lead beneficiary number	Delivery date from Annex I [60]	Comments
MS8	Initial evaluation of the EURECA interoperability environment and of the EURECA services	2	32	
MS9	Validation of the EURECA interoperability environment and services.	1	40	

WT3:
Work package description

Project Number [1]	288048	Project Acronym [2]	EURECA

One form per Work Package

Work package number [53]	WP8	Type of activity [54]	RTD
Work package title	Q&A, Evaluation and Validation		
Start month	6		
End month	42		
Lead beneficiary number [55]	2		

Objectives

This work package aims at answering the two questions, “Does the software do the right things?” which addresses the adequacy of the software for its intended goal, thus requiring an evaluation by its intended users, and “Does the software do the things right?” which stresses the need to validate its behavior from the viewpoint of performance and of correctness of results.
Thus, considering the user needs as described in WP1 and the corresponding intended pilots as described in WP 9, this work package will identify specific application objectives to be tested, define the evaluation criteria and devise monitoring procedures to be executed by the involved stakeholder groups. Special care will be taken to involve the biomedical, clinical and pharma end-user community as early as possible in the evaluation and validation effort. Technical validation will be conducted in tight collaboration with WP2 – WP7 and WP9, and the procedures for the assessment of the adequacy of treatment of personal data will be established jointly with WP7. Adequate personal-data treatment is of special importance in the EURECA project as foreseen pilots will involve real clinical data.
Specifically the objectives of the work package are:
• To formulate evaluation criteria, validation procedures, and feedback report guidelines
• To coordinate the specifications of test (validation) cases and demonstrators
• To coordinate evaluation and validation activities
• To write evaluation reports on the components of the system and on the integrated EURECA environment.

Description of work and role of partners

Task 8.1: Formulate evaluation criteria, validation procedures and feedback report guidelines
In this task the procedures for the evaluation and validation (E+V) activities will be established. In general evaluation criteria will be continuously adapted to the current state of development of the environment, considering the end-user scenarios and clinical pilots as general guideline. Usability, user-friendliness, clarity of on-line documentation, speed and robustness will be key criteria in the evaluation process.
Quantitative measures of the benefits of the project as a whole (e.g. time saved in the recruitment of patients) will also be developed in Task 8.1.
The validation of the platform will essentially be conducted by the design of and execution of test cases with known results, those will adapted to the specificities of the software issued in each work package.
The outcome of this task will be a set of procedures to guarantee a proper monitoring of the adequacy of the software developed in EURECA to its intended goal, as well as guidelines describing the feedback procedure to developers.

Task 8.2: Coordinate specifications of test scenarios and of demonstrators
In this task, minimal scenarios based on the user needs expressed in WP1 will be elaborated to test targeted software components. This includes notably the identification and preparation of relevant test data, as well as the preparation of standalone validation results to which the outcome of the execution of EURECA components can be compared to.
In addition, complete demonstrators of clinical relevance will be designed, jointly with the clinical partners of the project, to illustrate the progress of the project during reviews and to serve as evaluation material. As much as possible, this material will be reused as external demonstration and training material in the context of WP10.

Task 8.3: Coordinate evaluation and validation activities and reporting

Evaluation workshops involving software developers and users will be held periodically during the development of the EURECA environment. Evaluation groups will be provided with the criteria and validation procedures defined in Task T8.1 and will execute the scenarios designed in Task T8.2 to provide feedback to the developers.
Feedback reports will suggest possible improvements, modifications, and additional functionalities.
In this task, WP8 will report to management board and steering groups on progresses achieved and remaining problems to be addressed, on the extent to which the EURECA environment meets the identified evaluation criteria and will a provide comprehensive overview of key demonstration results.

Task 8.4: User workshops at clinical sites
For the effective execution of the worked planned, the WP will also focus on delivering the required training to end users, through structured user workshops to be held at each clinical site. The results of the training will be reported.

Person-Months per Participant

Participant number [10]	Participant short name [11]	Person-months per participant
1	Philips	3.00
2	FORTH	10.00
3	IJB	2.00
4	CUSTODIX	4.00
5	UdS	2.00
6	UOXF	8.00
7	FhG	3.00
8	VUA	2.00
9	BIG	4.00
10	LUH	2.00
13	MAASTRO	2.00
14	eCancer	10.00
17	GBG	4.00
	Total	56.00

List of deliverables

Delive-rable Number [61]	Deliverable Title	Lead benefi-ciary number	Estimated indicative person-months	Nature [62]	Dissemi-nation level [63]	Delivery date [64]
D8.1	Evaluation and validation procedures for the EURECA environment	2	9.00	R	PU	11
D8.2	Specifications of the evaluation and validation scenarios for the different EURECA components	1	9.00	R	PU	24
D8.3	Report on evaluation and validation of EURECA components	2	9.00	R	PU	28

WT3:
Work package description

List of deliverables

Delive-rable Number [61]	Deliverable Title	Lead benefi-ciary number	Estimated indicative person-months	Nature [62]	Dissemi-nation level [63]	Delivery date [64]
D8.4	Specifications of the evaluation and validation scenarios and demonstrators for the clinical pilots	2	9.00	R	PU	30
D8.5	Report on the evaluation and validation of the EURECA environment and services	2	10.00	R	PU	40
D8.6	Report on the user workshops at clinical sites	14	10.00	R	PU	42
		Total	56.00			

Description of deliverables

D8.1) Evaluation and validation procedures for the EURECA environment: [month 11]

D8.2) Specifications of the evaluation and validation scenarios for the different EURECA components: [month 24]

D8.3) Report on evaluation and validation of EURECA components: [month 28]

D8.4) Specifications of the evaluation and validation scenarios and demonstrators for the clinical pilots: [month 30]

D8.5) Report on the evaluation and validation of the EURECA environment and services: [month 40]

D8.6) Report on the user workshops at clinical sites: [month 42]

Schedule of relevant Milestones

Milestone number [59]	Milestone name	Lead benefi-ciary number	Delivery date from Annex I [60]	Comments
MS8	Initial evaluation of the EURECA interoperability environment and of the EURECA services	2	32	
MS9	Validation of the EURECA interoperability environment and services.	1	40	

WT3:
Work package description

Project Number [1]	288048	Project Acronym [2]	EURECA

One form per Work Package

Work package number [53]	WP9	Type of activity [54]	RTD
Work package title	Models, deployment and clinical pilots		
Start month	1		
End month	42		
Lead beneficiary number [55]	1		

Objectives

The first and crucial objective of this WP is to coordinate the efforts with the technical staff and the IT departments of all EURECA pilot sites, so that the Consortium receives all information required for developing the information models representing the EHR systems and the Clinical Trial systems of the pilots sites, and all the data necessary for the testing and validation of the interoperability framework and of the EURECA tools and services.
The second objective of WP9 is to build the canonical information models for EHR and CT systems making use of suitable standardized formalisms, such as CEN13606 (RM and archetypes), and HL7 v3. Additionally, the work package will evaluate existing mapping formalisms and tools that would enable seamless data linkage between the canonical model and the underlying implementation of the sources. Next to this, a solution will be provided to enable access to the various implementations of the sources via the uniform interface of the canonical model.
WP9 will also collaborate with WP4 to develop the appropriate mappings between the canonical models and the EURECA core data sets.
Another objective of the WP is to prepare the technical and procedural infrastructure – in compliance with the defined legal and security framework of the project – for the installation of EURECA technologies and tools for their extensive evaluation and validation.
The final objective of the WP is to organize together with WP8 the effective evaluation and validation activities of the EURECA technologies – once these are ready and delivered by the other project WPs.

Description of work and role of partners

T9.1: Building the EURECA development and testing environment
The objective of this task is to coordinate all efforts that need to take place locally at each and every pilot site – early enough in the project's implementation period – to build the development environment (e.g. "surrogate" databases), and to provide access to suitable schema- and instance-level datasets to be used by the EURECA prototypes.
T9.2 Canonical information models of EHR and CT systems
In this task the information models representing the EHR systems and the clinical trial systems of the pilot sites are developed. These information models are required input to WP4, so that it can proceed with the identification of the requirements for mappings that bridge the semantic core data sets with the information models representing the EHR systems and the clinical trial systems. The models developed will provide a canonical view, reflecting the content and the structure of the respective information management systems. This task will also involve the collaboration with WP4 to develop mappings between the canonical models and the core data sets.
T9.3 Uniform access to the source data
This task will evaluate existing mapping formalisms tools that can enable seamless data linkage between the canonical model and the underlying implementation of the sources. It will also develop a solution to enable access to the various technical implementations of the sources via the uniform interface of the canonical model. Additionally, it will provide uniform interfaces towards the architectural framework.
T9.4 Deployment Environment
The objective of this task is to prepare the technical and procedural environment – in compliance with the defined legal and security framework of the project – for the installation of EURECA technologies and tools for

their extensive evaluation and validation. It is also its responsibility to design, oversee and execute all activities for preparing the clinical pilots for the validation work.
T9.5 Validation activities at Clinical Pilots
The objective of this task is to carry out together with WP8 the validation of the EURECA environment at the clinical sites, according to the validation methodology defined in WP8, and the scenarios defined in WP1 and elaborated in WP8.
T9.6 Guidelines for compliance and use
This task will coordinate, based on contribution from all technical workpackages and on the lessons learnt during development, deployment and validation, the elaboration of a set of guidelines including minimum requirements and steps to be followed by an external clinical organization or service developer in order to become EURECA-compliant, to be able to join the EURECA interoperability environment and/or to make use of our services.

Person-Months per Participant

Participant number [10]	Participant short name [11]	Person-months per participant
1	Philips	20.00
2	FORTH	18.00
3	IJB	20.00
4	CUSTODIX	21.00
5	UdS	12.00
6	UOXF	20.00
7	FhG	20.00
8	VUA	8.00
9	BIG	2.00
10	LUH	2.00
12	UPM	20.00
13	MAASTRO	14.00
15	EUROREC	6.00
16	SIT	6.00
17	GBG	8.00
	Total	197.00

List of deliverables

Delive-rable Number [61]	Deliverable Title	Lead benefi-ciary number	Estimated indicative person-months	Nature [62]	Dissemi-nation level [63]	Delivery date [64]
D9.1	Report on the development environment and on the available test data	3	28.00	R	PU	9
D9.2	Canonical models of EHRs and CT systems	1	28.00	O	PU	12

WT3:
Work package description

List of deliverables

Delive-rable Number [61]	Deliverable Title	Lead benefi-ciary number	Estimated indicative person-months	Nature [62]	Dissemi-nation level [63]	Delivery date [64]
D9.3	Initial proposal for the mapping formalism and mappings to EHR and CT models	12	28.00	R	PU	24
D9.4	Solution providing uniform access to sources	7	28.00	P	PU	28
D9.5	Report on preparation of the deployment environment for the clinical pilots	2	28.00	R	PU	32
D9.6	Guidelines and recommendations for joining the EURECA environment	4	28.00	R	PU	36
D9.7	Validation of the EURECA technologies at pilot sites	1	29.00	P	RE	40
		Total	197.00			

Description of deliverables

D9.1) Report on the development environment and on the available test data: [month 9]

D9.2) Canonical models of EHRs and CT systems: [month 12]

D9.3) Initial proposal for the mapping formalism and mappings to EHR and CT models: [month 24]

D9.4) Solution providing uniform access to sources: [month 28]

D9.5) Report on preparation of the deployment environment for the clinical pilots: [month 32]

D9.6) Guidelines and recommendations for joining the EURECA environment: [month 36]

D9.7) Validation of the EURECA technologies at pilot sites: [month 40]

Schedule of relevant Milestones

Milestone number [59]	Milestone name	Lead benefi-ciary number	Delivery date from Annex I [60]	Comments
MS5	Initial information models	1	18	
MS8	Initial evaluation of the EURECA interoperability environment and of the EURECA services	2	32	
MS9	Validation of the EURECA interoperability environment and services.	1	40	

WT3:
Work package description

Project Number [1]	288048	Project Acronym [2]	EURECA

One form per Work Package

Work package number [53]	WP10	Type of activity [54]	RTD
Work package title	Knowledge management		
Start month	1		
End month	42		
Lead beneficiary number [55]	9		

Objectives

The objectives of WP10 are:
• To exchange information within the project consortium
• To exploit and disseminate the results of the project
• To manage the generation of intellectual property and to contribute to standardization activities
This work package has a multiple role: It has the responsibility to investigate and support the exploitability of the EURECA results and to propose realistic exploitation models, to disseminate the project's results to the user community, to contribute to standards whenever applicable and to protect the IPR of the project partners.
Important exploitation goals of EURECA are the sustainability of our solutions beyond the duration of the project and the use of the project outcomes by stakeholders outside of the consortium.
We believe that a successful dissemination of our results is key to the wide-scale uptake of our solutions by the user communities. All project partners (especially academic partners and research centers) will present findings at high-level conferences and workshops, as well as publish research results in premium peer-reviewed journals. Industrial partners will complement dissemination of results through the realization of proof-of-concept prototypes, field tests, and demonstrations at conferences and professional exhibitions.
Exploitation
Our exploitation goal is the use of project outcomes by third party stakeholders from clinical research and care, and also from the bio-pharmaceutical industry. We have linked to several large clinical user groups that expressed interest to support us with clinical scenarios and in the requirements collection phase, but also to test and use the EURECA services. We will also address these user groups to provide a prioritization of the EURECA tools and services according to their expected benefits for each user group.
During the project we will examine together with our users the various workflows in which the EURECA services can be deployed and provide that information to the user community. We will also investigate ways to maintain the results of EURECA beyond the end of the project.
WP10 will also manage the collaboration with the bio-pharmaceutical industry, which may lead us to propose exploitation models able to maintain the sustainability of the EURECA outcomes, while still providing solutions for the non-profit community as well. All EURECA partners will contribute to the exploitation tasks.
Finally, in order to promote the reuse of our interoperability framework in new open collaborations we will set up a certification system of EURECA compliant services (similar to caBIG) and provide support to users who would like to develop based on our reference architecture, join our environment, make use of our interoperability layer and services, and provide services on top of EURECA. The goal of EURECA is to provide an open architecture specification (interoperability framework) that can be used by collaborative initiatives to set easily up their data integration projects, in a way that such collaborations can easily interconnect with EURECA compliant data sources, re-use EURECA compliant services, etc. The certification task will be carried out by EuroRec.

Dissemination
The WP will set up a unique protected website, placed within ecancer.eu for posting of all significant results and notices from the intramural work of each WP. This involves interacting and understanding the nature of the obstacles and their ultimate solutions as they evolve. This requires close collaboration with all other WPs. Through eCancer the project outcomes will be advertised and evaluated using established cohorts of clinicians, pharma and patient representatives.
The rules of engagement on communication to stakeholders, internal and external will be agreed and implemented from the first quarter of the project. eCancer will advertise the existence, aims and progress of the project, agreed with all partners, by publication on ecancer.eu, editorials and advertorials in key newspapers, lay

and technical, and promotions at cancer conferences in Europe and the rest of the world. The potential market for the deliverables of this project will not be restricted to Europe and the USA, as it is clear that the Far East and Australasia are fast becoming a growth market for clinical research and trials.
This WP will also address the patient-related community, and aim to describe relevant project results in language understood by patients and their carers. The patient preferences of styles of presentation of clinical information (e.g. clinical trial information to support enrolment) will be evaluated.

Description of work and role of partners

Task 10.1: Dissemination
Task leader: eCancer
T10.1.1 Dissemination plans
A comprehensive Dissemination Plan (D10.1 and D10.6) will delineate the overall strategy and provide a "road-map" for dissemination of information and knowledge generated by the project, including the definition of communication strategies, the target groups, communication tools to be used (some are which are enumerated below) and their expected impact. It will also foresee the way in which these communications activities will be evaluated.
T10.1.2 Web sites, communication &collaboration
The exchange of information within the project consortium will be supported by setting up a server like TWIKI (D10.1). In addition, regular project meetings and internal workshops will be held allowing for knowledge sharing and collaborative work.
The dissemination of the results will be supported by setting up a website (D10.2) dedicated to the project, which will contain the latest news, activities, and results of the project, as well as by producing a project flyer. A project-specific newsletter will also be produced on a bi-annual basis (D10.5)
T10.1.3 Partnership Programmes
It is a well established practice for public/academic groups developing and utilising advanced technology for practical applications to create Partnership Programmes to reach out to well defined communities of potential beneficiaries. These programmes permit very efficient customised channels for communication in both directions, eliciting advice and sharing issues and technologies. This task will also evaluate the rates of adoption of multimedia dissemination tools by all stakeholders.

T10.1.4 Events organisation, participation and presentation
All project partners (especially academic partners and research centers) will present findings at high-level conferences and workshops, as well as publish research results in premium peer-reviewed journals. Industrial partners will complement dissemination of results through the realization of proof-of-concept prototypes, field tests, and demonstrations at conferences and professional exhibitions.
Finally, a key element of this task is the organisation of a launching event/workshop close to the end of the project that will aim at fostering the implementation of the platform by providing introductory "hands on" training to potential end-users. This workshop will be held in the context of a major European cancer conference (ECCO, ESMO or EBCC) to reach the largest public possible.
Task 10.2: Exploitation
Task leader: MAASTRO
Task 10.2.1 Adoption of the EURECA solutions
The objective of this task is to build on the existing clinical networks supporting EURECA (BIG, GBG, EORTC, EuroSarc, METOXIA, SIOP) to support the adoption of the EURECA infrastructure in several European cancer research communities. The starting point will be the use of the platform in the context of trials led by the EURECA clinical partners. The system would then gradually be integrated in other programs and expanded beyond, to other interested user communities. These different stages will be detailed in the exploitation plans (D10.4 and D10.9).

Task 10.2.2 Ensuring project sustainability
The long term sustainability of the EURECA solutions, i.e. beyond the FP7 funding period, will also be addressed in the exploitation plans (D10.4 and D10.9). This task will investigate suitable business models and will propose practical solutions to cover the costs of maintenance of the infrastructure such as the definition of different type of access (e.g. free access for EURECA partners and paying for outsiders) or sponsoring.
This task will also assess the market potential of the EURECA environment and of the EURECA software services and tools. This includes investigating together with our pharma collaborators the potential of EURECA to add value to the pharmaceutical industry.

Task 10.2.3 Educational activities
The academic partners will enrich their educational activities with the knowledge generated within the project. This will be achieved for example by proposing and supervising student semester and diploma projects related to the project. In addition, it is also expected that the project will attract prospective PhD students.
The industrial partners will try to introduce the know-how developed in the project into pre-existing research and development projects, in order to launch new products.

Task 10.2.3 Certification of EURECA compliance
EURECA aims to provide an open architecture specification (interoperability framework) which will allow collaborative initiatives to setup medical data integration projects with minimal effort. Successful exploitation beyond the project end requires governance mechanisms to be put in place sustaining the EURECA architectural concept. One key element is setting up a certification programme for “EURECA compliant” services. The “EURECA compliant” label will guarantee system integrators that labelled services can easily be deployed in a EURECA compliant environment and that interconnecting with EURECA compliant data sources requires no additional development.This programme will be set-up and managed by the EuroRec Institute of whom certification and quality labeling is the core business.
Task 10.3 Standardisation
Task leader: EuroRec
The project will seek active participation in some of the initiatives targeting standards in the semantic interoperability area (e.g. contribute to existing initiatives, to the NoE on semantic interoperability) and in data sharing and interoperability, or if appropriate develop additional ones. Contribution to standardization activities will be in turn linked to technology transfer aiming at an efficient exploitation and reuse of technology.
Task 10.4 Intellectual Property
Task leader: Philips
According to the innovative nature of this work, all partners in the consortium will examine the requirement for IPR protection for new inventions conceived on various aspects of the research work. Intellectual property will be protected by patent filing or publication. In addition, patent searches will be done regularly to keep the project up to date with the state of the art and to be alert to strategic opportunities. The project will also carry out benchmark and competitive analysis. The appropriate departments of the industrial partners (in particular the Philips Research Library, the Intellectual Property Rights and Standardization department (IP&S) and the Knowledge Database of Philips Research) will play an active role in assisting the project with filing patents and with patent searches.

Person-Months per Participant

Participant number [10]	Participant short name [11]	Person-months per participant
1	Philips	6.00
2	FORTH	2.00
3	IJB	2.00
4	CUSTODIX	6.00
5	UdS	4.00
6	UOXF	2.00
7	FhG	3.00
8	VUA	2.00
9	BIG	16.00
10	LUH	2.00
11	XEROX	2.00
12	UPM	2.00
13	MAASTRO	6.00

WT3:
Work package description

Person-Months per Participant

Participant number [10]	Participant short name [11]	Person-months per participant
14	eCancer	20.00
15	EUROREC	4.00
17	GBG	2.00
	Total	81.00

List of deliverables

Delive-rable Number [61]	Deliverable Title	Lead benefi-ciary number	Estimated indicative person-months	Nature [62]	Dissemi-nation level [63]	Delivery date [64]
D10.1	Communication portal Twiki	9	8.00	O	CO	3
D10.2	External project website	14	8.00	O	PU	6
D10.3	Initial dissemination plan	9	8.00	R	RE	9
D10.4	Initial exploitation plan	13	8.00	R	RE	12
D10.5	Project newsletter	14	9.00	O	PU	12
D10.6	Report on the EURECA workshop/launching event	14	8.00	R	PU	32
D10.7	One-stop-shop at eCancer for all interested in recruitment, participation, exploitation and applicat	14	8.00	O	PU	36
D10.8	Final dissemination plan	9	8.00	R	PU	36
D10.9	Final exploitation plan including sustainability plan	13	8.00	R	PU	42
D10.10	The EURECA certification programme	15	8.00	O	PU	42
		Total	81.00			

Description of deliverables

D10.1) Communication portal Twiki: [month 3]

D10.2) External project website: [month 6]

D10.3) Initial dissemination plan: [month 9]

D10.4) Initial exploitation plan: [month 12]

D10.5) Project newsletter: [month 12]

D10.6) Report on the EURECA workshop/launching event: [month 32]

D10.7) One-stop-shop at eCancer for all interested in recruitment, participation, exploitation and applicat: [month 36]

D10.8) Final dissemination plan: [month 36]

D10.9) Final exploitation plan including sustainability plan: [month 42]

D10.10) The EURECA certification programme: [month 42]

WT3:
Work package description

Schedule of relevant Milestones

Milestone number [59]	Milestone name	Lead benefi-ciary number	Delivery date from Annex I [60]	Comments
MS1	Formation of boards and committees	9	2	
MS10	Final EURECA exploitation plan	9	42	

WT3:
Work package description

Project Number [1]	288048	Project Acronym [2]	EURECA

One form per Work Package

Work package number [53]	WP11	Type of activity [54]	MGT
Work package title	Project Management		
Start month	1		
End month	42		
Lead beneficiary number [55]	1		

Objectives

The objectives of this work package are:
• To coordinate the project and the related activities
• To coordinate the work between the EURECA work packages
• To coordinate the day by day management activities
• To manage the partners, to collect regular progress information
• Management of communication between all stakeholders
• Acting as the primary contact to the commission with respect to legal, financial and administrative tasks

Description of work and role of partners

Task 11.1 Management guidelines:
• A set of project management guidelines will be agreed at the very beginning of the project.
• Regular meetings will be managed of the participants, the work packages and the Project Management Committee.
• A detailed planning will be maintained.
• Coordination of the GA and chairing its meetings
• Quality Assurance: Validation of the project results and assessment and evaluation of project success with respect to the objectives of the project
• Risk Management: Identification of the project risks and development of a contingency plan

Task 11.2 Organizing the project kick off meeting
• A project kick off meeting will be organized. Aim is to have all people present who are working on the project. This must ensure a fast and effective start of the project.

Task 11.3 Relations with EU
• The coordinator Philips will manage the legal, financial and administrative relations with the commission (like contracts, reporting, management reports etc.).
• Technical contacts will be established by all partners as part of their regular work package.
• Besides the deliverables we will report containing the EC guidelines (Periodic Progress Reports and other reports mentioned in the EC contract).

Task 11.4 Project website
A website, with restricted use for the consortium members only, will be set-up to help proper dissemination of the project's results among the consortium members

Person-Months per Participant

Participant number [10]	Participant short name [11]	Person-months per participant
1	Philips	42.00
2	FORTH	2.00
4	CUSTODIX	2.00

WT3:
Work package description

Person-Months per Participant

Participant number [10]	Participant short name [11]	Person-months per participant
5	UdS	2.00
6	UOXF	2.00
8	VUA	2.00
9	BIG	2.00
12	UPM	2.00
	Total	56.00

List of deliverables

Delive-rable Number [61]	Deliverable Title	Lead benefi-ciary number	Estimated indicative person-months	Nature [62]	Dissemi-nation level [63]	Delivery date [64]
D11.1	Public summary of project	1	28.00	O	PU	1
D11.2	Internal project website	1	28.00	O	CO	3
		Total	56.00			

Description of deliverables

D11.1) Public summary of project: [month 1]

D11.2) Internal project website: [month 3]

Schedule of relevant Milestones

Milestone number [59]	Milestone name	Lead benefi-ciary number	Delivery date from Annex I [60]	Comments
MS1	Formation of boards and committees	9	2	
MS2	Requirements and specifications of the EURECA environment	5	9	
MS3	Initial EURECA legal and ethical requirements	10	11	
MS4	Initial EURECA architecture	4	12	
MS5	Initial information models	1	18	
MS6	Specification of the core datasets	12	18	
MS7	Operational EURECA architecture for validation	4	24	
MS8	Initial evaluation of the EURECA interoperability environment and of the EURECA services	2	32	

WT3:
Work package description

Schedule of relevant Milestones

Milestone number [59]	Milestone name	Lead benefi-ciary number	Delivery date from Annex I [60]	Comments
MS9	Validation of the EURECA interoperability environment and services.	1	40	
MS10	Final EURECA exploitation plan	9	42	

WT4:
List of Milestones

Project Number [1]	288048	Project Acronym [2]	EURECA

List and Schedule of Milestones

Milestone number [59]	Milestone name	WP number [53]	Lead beneficiary number	Delivery date from Annex I [60]	Comments
MS1	Formation of boards and committees	WP10, WP11	9	2	
MS2	Requirements and specifications of the EURECA environment	WP1, WP11	5	9	
MS3	Initial EURECA legal and ethical requirements	WP7, WP11	10	11	
MS4	Initial EURECA architecture	WP2, WP11	4	12	
MS5	Initial information models	WP9, WP11	1	18	
MS6	Specification of the core datasets	WP4, WP11	12	18	
MS7	Operational EURECA architecture for validation	WP2, WP11	4	24	
MS8	Initial evaluation of the EURECA interoperability environment and of the EURECA services	WP3, WP4, WP5, WP6, WP7, WP8, WP9, WP11	2	32	
MS9	Validation of the EURECA interoperability environment and services.	WP3, WP4, WP5, WP6, WP7, WP8, WP9, WP11	1	40	
MS10	Final EURECA exploitation plan	WP10, WP11	9	42	

WT5:
Tentative schedule of Project Reviews

Project Number [1]	288048	Project Acronym [2]	EURECA

Tentative schedule of Project Reviews

Review number [65]	Tentative timing	Planned venue of review	Comments, if any
RV 1	12	Brussels	
RV 2	24	Brussels	
RV 3	42	Brussels	

WT6:
Project Effort by Beneficiary and Work Package

Project Number [1]	288048	Project Acronym [2]	EURECA

Indicative efforts (man-months) per Beneficiary per Work Package

Beneficiary number and short-name	WP 1	WP 2	WP 3	WP 4	WP 5	WP 6	WP 7	WP 8	WP 9	WP 10	WP 11	Total per Beneficiary
1 - Philips	2.00	14.00	10.00	18.00	8.00	23.00	2.00	3.00	20.00	6.00	42.00	148.00
2 - FORTH	2.00	16.00	6.00	12.00	10.00	8.00	0.00	10.00	18.00	2.00	2.00	86.00
3 - IJB	8.00	2.00	10.00	16.00	0.00	8.00	2.00	2.00	20.00	2.00	0.00	70.00
4 - CUSTODIX	2.00	33.00	4.00	8.00	0.00	0.00	28.00	4.00	21.00	6.00	2.00	108.00
5 - UdS	12.00	4.00	3.00	16.00	4.00	14.00	0.00	2.00	12.00	4.00	2.00	73.00
6 - UOXF	6.00	0.00	6.00	10.00	30.00	0.00	0.00	8.00	20.00	2.00	2.00	84.00
7 - FhG	4.00	11.00	2.00	18.00	26.00	8.00	0.00	3.00	20.00	3.00	0.00	95.00
8 - VUA	2.00	0.00	2.00	16.00	2.00	38.00	0.00	2.00	8.00	2.00	2.00	74.00
9 - BIG	8.00	0.00	0.00	2.00	0.00	0.00	8.00	4.00	2.00	16.00	2.00	42.00
10 - LUH	2.00	0.00	0.00	0.00	0.00	4.00	28.00	2.00	2.00	2.00	0.00	40.00
11 - XEROX	2.00	2.00	35.00	5.00	4.00	0.00	0.00	0.00	0.00	2.00	0.00	50.00
12 - UPM	4.00	13.00	4.00	32.00	6.00	10.00	0.00	0.00	20.00	2.00	2.00	93.00
13 - MAASTRO	6.00	2.00	3.00	8.00	6.00	8.00	0.00	2.00	14.00	6.00	0.00	55.00
14 - eCancer	6.00	0.00	0.00	0.00	0.00	4.00	0.00	10.00	0.00	20.00	0.00	40.00
15 - EUROREC	2.00	2.00	0.00	1.00	0.00	0.00	0.00	0.00	6.00	4.00	0.00	15.00
16 - SIT	0.00	1.00	0.00	11.00	3.00	19.00	0.00	0.00	6.00	0.00	0.00	40.00
17 - GBG	8.00	0.00	0.00	8.00	0.00	0.00	0.00	4.00	8.00	2.00	0.00	30.00
18 - NRC	2.00	0.00	18.00	0.00	0.00	3.00	0.00	0.00	0.00	0.00	0.00	23.00
Total	78.00	100.00	103.00	181.00	99.00	147.00	68.00	56.00	197.00	81.00	56.00	1,166.00

WT7:
Project Effort by Activity type per Beneficiary

Project Number [1]	288048	Project Acronym [2]	EURECA

Indicative efforts per Activity Type per Beneficiary

Activity type	Part. 1 Philips	Part. 2 FORTH	Part. 3 IJB	Part. 4 CUSTODI	Part. 5 UdS	Part. 6 UOXF	Part. 7 FhG	Part. 8 VUA	Part. 9 BIG	Part. 10 LUH	Part. 11 XEROX	Part. 12 UPM	Part. 13 MAASTRO	Part. 14 eCancer
1. RTD/Innovation activities														
WP 1	2.00	2.00	8.00	2.00	12.00	6.00	4.00	2.00	8.00	2.00	2.00	4.00	6.00	6.00
WP 2	14.00	16.00	2.00	33.00	4.00	0.00	11.00	0.00	0.00	0.00	2.00	13.00	2.00	0.00
WP 3	10.00	6.00	10.00	4.00	3.00	6.00	2.00	2.00	0.00	0.00	35.00	4.00	3.00	0.00
WP 4	18.00	12.00	16.00	8.00	16.00	10.00	18.00	16.00	2.00	0.00	5.00	32.00	8.00	0.00
WP 5	8.00	10.00	0.00	0.00	4.00	30.00	26.00	2.00	0.00	0.00	4.00	6.00	6.00	0.00
WP 6	23.00	8.00	8.00	0.00	14.00	0.00	8.00	38.00	0.00	4.00	0.00	10.00	8.00	4.00
WP 7	2.00	0.00	2.00	28.00	0.00	0.00	0.00	0.00	8.00	28.00	0.00	0.00	0.00	0.00
WP 8	3.00	10.00	2.00	4.00	2.00	8.00	3.00	2.00	4.00	2.00	0.00	0.00	2.00	10.00
WP 9	20.00	18.00	20.00	21.00	12.00	20.00	20.00	8.00	2.00	2.00	0.00	20.00	14.00	0.00
WP 10	6.00	2.00	2.00	6.00	4.00	2.00	3.00	2.00	16.00	2.00	2.00	2.00	6.00	20.00
Total Research	106.00	84.00	70.00	106.00	71.00	82.00	95.00	72.00	40.00	40.00	50.00	91.00	55.00	40.00
2. Demonstration activities														
Total Demo	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3. Consortium Management activities														
WP 11	42.00	2.00	0.00	2.00	2.00	2.00	0.00	2.00	2.00	0.00	0.00	2.00	0.00	0.00
Total Management	42.00	2.00	0.00	2.00	2.00	2.00	0.00	2.00	2.00	0.00	0.00	2.00	0.00	0.00
4. Other activities														
Total other	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

WT7:
Project Effort by Activity type per Beneficiary

Total	148.00	86.00	70.00	108.00	73.00	84.00	95.00	74.00	42.00	40.00	50.00	93.00	55.00	40.00

WT7:
Project Effort by Activity type per Beneficiary

Activity type	Part. 15 EUROREC	Part. 16 SIT	Part. 17 GBG	Part. 18 NRC	Total
1. RTD/Innovation activities					
WP 1	2.00	0.00	8.00	2.00	78.00
WP 2	2.00	1.00	0.00	0.00	100.00
WP 3	0.00	0.00	0.00	18.00	103.00
WP 4	1.00	11.00	8.00	0.00	181.00
WP 5	0.00	3.00	0.00	0.00	99.00
WP 6	0.00	19.00	0.00	3.00	147.00
WP 7	0.00	0.00	0.00	0.00	68.00
WP 8	0.00	0.00	4.00	0.00	56.00
WP 9	6.00	6.00	8.00	0.00	197.00
WP 10	4.00	0.00	2.00	0.00	81.00
Total Research	15.00	40.00	30.00	23.00	1,110.00
2. Demonstration activities					
Total Demo	0.00	0.00	0.00	0.00	0.00
3. Consortium Management activities					
WP 11	0.00	0.00	0.00	0.00	56.00
Total Management	0.00	0.00	0.00	0.00	56.00
4. Other activities					
Total other	0.00	0.00	0.00	0.00	0.00
Total	15.00	40.00	30.00	23.00	1,166.00

WT8:
Project Effort and costs

Project Number [1]	288048	Project Acronym [2]	EURECA

Project efforts and costs

Benefi-ciary number	Beneficiary short name	Estimated eligible costs (whole duration of the project)						Requested EU contribution (€)
		Effort (PM)	Personnel costs (€)	Subcontracting (€)	Other Direct costs (€)	Indirect costs OR lump sum, flat-rate or scale-of-unit (€)	Total costs	
1	Philips	148.00	1,370,398.00	9,000.00	27,963.00	1,088,367.00	2,495,728.00	1,536,411.00
2	FORTH	86.00	413,200.00	5,400.00	54,396.00	392,540.00	865,536.00	656,952.00
3	IJB	70.00	399,500.00	3,000.00	8,000.00	244,500.00	655,000.00	492,000.00
4	CUSTODIX	108.00	821,000.00	4,000.00	95,000.00	369,450.00	1,289,450.00	987,512.00
5	UdS	73.00	365,000.00	2,000.00	46,000.00	246,600.00	659,600.00	503,600.00
6	UOXF	84.00	417,336.00	2,276.00	21,429.00	263,259.00	704,300.00	532,768.00
7	FhG	95.00	623,229.00	7,500.00	32,000.00	557,483.00	1,220,212.00	917,034.00
8	VUA	74.00	393,341.00	5,000.00	30,000.00	500,784.00	929,125.00	705,125.00
9	BIG	42.00	309,400.00	3,000.00	106,930.00	83,266.00	502,596.00	400,017.00
10	LUH	40.00	259,160.00	0.00	60,207.00	191,620.00	510,987.00	395,240.00
11	XEROX	50.00	375,000.00	0.00	12,000.00	238,500.00	625,500.00	312,750.00
12	UPM	93.00	431,055.00	4,000.00	49,000.00	341,310.00	825,365.00	624,290.00
13	MAASTRO	55.00	364,320.00	3,000.00	55,000.00	251,592.00	673,912.00	506,184.00
14	eCancer	40.00	200,000.00	3,000.00	61,500.00	156,900.00	421,400.00	316,800.00
15	EUROREC	15.00	148,125.00	0.00	18,000.00	33,225.00	199,350.00	149,512.00
16	SIT	40.00	200,000.00	0.00	15,000.00	129,000.00	344,000.00	258,000.00
17	GBG	30.00	191,550.00	0.00	18,900.00	126,270.00	336,720.00	252,540.00
18	NRC	23.00	95,850.00	0.00	20,000.00	23,170.00	139,020.00	104,265.00
	Total	1,166.00	7,377,464.00	51,176.00	731,325.00	5,237,836.00	13,397,801.00	9,651,000.00

1. Project number

The project number has been assigned by the Commission as the unique identifier for your project. It cannot be changed. The project number **should appear on each page of the grant agreement preparation documents (part A and part B)** to prevent errors during its handling.

2. Project acronym

Use the project acronym as given in the submitted proposal. It cannot be changed unless agreed so during the negotiations. The same acronym **should appear on each page of the grant agreement preparation documents (part A and part B)** to prevent errors during its handling.

53. Work Package number

Work package number: WP1, WP2, WP3, ..., WPn

54. Type of activity

For all FP7 projects each work package must relate to one (and only one) of the following possible types of activity (only if applicable for the chosen funding scheme – must correspond to the GPF Form Ax.v):

• **RTD/INNO =** Research and technological development including scientific coordination - applicable for Collaborative Projects and Networks of Excellence

• **DEM =** Demonstration - applicable for collaborative projects and Research for the Benefit of Specific Groups

• **MGT =** Management of the consortium - applicable for all funding schemes

• **OTHER =** Other specific activities, applicable for all funding schemes

• **COORD =** Coordination activities – applicable only for CAs

• **SUPP =** Support activities – applicable only for SAs

55. Lead beneficiary number

Number of the beneficiary leading the work in this work package.

56. Person-months per work package

The total number of person-months allocated to each work package.

57. Start month

Relative start date for the work in the specific work packages, month 1 marking the start date of the project, and all other start dates being relative to this start date.

58. End month

Relative end date, month 1 marking the start date of the project, and all end dates being relative to this start date.

59. Milestone number

Milestone number:MS1, MS2, …, MSn

60. Delivery date for Milestone

Month in which the milestone will be achieved. Month 1 marking the start date of the project, and all delivery dates being relative to this start date.

61. Deliverable number

Deliverable numbers in order of delivery dates: D1 – Dn

62. Nature

Please indicate the nature of the deliverable using one of the following codes

R = Report, **P** = Prototype, **D** = Demonstrator, **O** = Other

63. Dissemination level

Please indicate the dissemination level using one of the following codes:

• **PU =** Public

• **PP =** Restricted to other programme participants (including the Commission Services)

• **RE =** Restricted to a group specified by the consortium (including the Commission Services)

• **CO =** Confidential, only for members of the consortium (including the Commission Services)

• **Restreint UE =** Classified with the classification level "Restreint UE" according to Commission Decision 2001/844 and amendments

• **Confidentiel UE =** Classified with the mention of the classification level "Confidentiel UE" according to Commission Decision 2001/844 and amendments

• **Secret UE =** Classified with the mention of the classification level "Secret UE" according to Commission Decision 2001/844 and amendments

64. Delivery date for Deliverable

Month in which the deliverables will be available. Month 1 marking the start date of the project, and all delivery dates being relative to this start date

65. Review number

Review number: RV1, RV2, ..., RVn

66. Tentative timing of reviews

Month after which the review will take place. Month 1 marking the start date of the project, and all delivery dates being relative to this start date.

67. Person-months per Deliverable

The total number of person-month allocated to each deliverable.

PART B

COLLABORATIVE PROJECT

Contents

1. Concept and objectives, progress beyond state-of-the-art, S/T methodology and work plan

1.1. Concept and project objective(s)

1.1.1. S&T Objectives

EURECA Rationale

From academic medical research centres to community hospitals and other stakeholders, the healthcare industry continues to improve its capabilities for electronic data capture. Despite these advances, a gap remains in the ability of these IT systems to deliver knowledge and insight back to the very researchers and clinicians they are intended to support.

Additionally, there is a widening knowledge gap between the care provided in top research clinical sites and standard care sites, resulting in large differences in treatments and outcomes. In this context, the need to bring the latest therapy options validated in clinical research to each and every hospital must be addressed before being able to significantly reduce the numbers of patients that receive suboptimal treatment (e.g. overtreatment, wrong dose, etc.), or the wrong treatment. There are currently very few mechanisms and formally established channels for transferring the best practices to clinicians and the current dissemination means are insufficient.

The current separation between clinical research and clinical care makes the detection of many serious patient safety issues difficult. Serious side effects[1],[2] of therapy and drugs that appear outside a clinical trial either due to a low incidence or a late onset are very difficult to detect and to explain in the absence of a feedback loop from standard care to research.

Despite identified benefits[3],[4], the secondary use of care data for research, quality assurance and patient safety is still rarely supported. Main barriers to enabling secondary use of data are the lack of interoperability, common standards and terminologies[5].

Challenges for Patient Safety

The fact is that we will never know for sure whether a drug is entirely safe when it gets to the market. A main problem persists: access to the samples and to data that could help us to study rare adverse events:

- Usually, the size of safety databases that are generated during drug development are too small to do a good job in picking up rare safety events
- The largest safety information is produced when a drug is on the market

Our current tools do not allow us to effectively capture this information and capitalize on its potential.

"The Transition from Pre-Clinical to Clinical Application of Safety Related Genomics"
Felix W. Frueh, CDER/FDA
IOM/FDA Emerging Safety Science Workshop, 2007.

The recruitment rate of patients into clinical trials is currently low despite recognized better outcomes and the enrolment process is slow and inefficient, involving redundant data entry, inconsistencies and several manual verification steps. Additionally, the progress in treating rare diseases is currently hampered by the fact that there is not enough data to support research.

Even when sufficient data was available, the acceleration of discoveries in basic sciences would not necessarily change at the same speed the way patients are treated. As discoveries can only be implemented after proper prospective validation, supporting a faster enrolment of patients in clinical trials would also have a direct impact on the speed of improvement of the standard of care.

[1] Breast Cancer Surveillance Practices Among Women Previously Treated With Chest Radiation, Oeffinger et al. *JAMA.*2009; 301: 404-414.

[2] Breast Cancer after Childhood cancer: A Report from the Childhood Cancer Survivor Study, http://www.annals.org/cgi/reprint/141/8/590.pdf.

[3] Electronic medical records for clinical research: application to the identification of heart failure, Pakhomov et al., Am J Managed Care, 2007

[4] Toward a national framework for the secondary use of health data: an American Medical Informatics Association white paper, Safran C. et al., JAMIA, 2007

[5] Adding value to the electronic health record through secondary use of data for quality assurance, research, and surveillance, Hersh W., Am J Managed Care, 2007

EURECA Main Disease Domain

The second largest cause of death in the EU and the US, and estimated to have become the leading cause of death worldwide by 2010[6], cancer is a genetic disease that requires complex stratification and patient management and access to large volumes of information and knowledge. Despite large investments in cancer research, the breakthrough results are scarce and the improvement of patient outcome is slow.

A large problem in both basic and clinical research is the lack of sufficient data, while the large amounts of patient data collected in clinical care in the EHR systems are seldom properly accessible for secondary use in research. Progress in rare types of cancer is very difficult when sufficient data is not available for analysis and for building new research hypotheses and improving the treatment protocols. On the other hand, the volume of data collected for a cancer patient in the context of care has increased tremendously. As more and more data is being used for current clinical practice, it becomes increasingly important to preserve all that data and to use it for research but also for the future benefit of the patient, in the light of new discoveries. Preserving the data is especially meaningful as storing and maintaining it is significantly cheaper than re-acquiring it, and because valuable new insights could be derived from old data.

> With genomics discoveries relating to common chronic diseases, numerous genetic tests may emerge that hold promise for significant changes in the delivery of health care, particularly in preventive medicine and in tailoring drug treatment.
> The primary care workforce, which will be required to be on the front lines of the integration of genomics into the regular practice of medicine, feels woefully underprepared to do so.
> The most prominent of these [barriers] include health professionals' lack of basic knowledge about genetics and their lack of confidence in interpreting familial patterns of disease, which limits their ability to appropriately counsel their patients, order and accurately interpret genetic tests, and refer their patients for genetics consultation.
>
> "Large gap between Genomic Medicine and Clinical practice",
> JAMA, March 2008.

Resistance to classical (chemo- and radiotherapy) and targeted therapy regimens in cancer is still a huge challenge. With the recent insight into the differences at molecular level among different cancer types and with the translation of the underlying wealth of molecular biology knowledge towards the clinic it has become clear that most –if not all- cancers require sub-classification.

The great diversity in response to therapy between patients is both due to molecular tumour heterogeneity and to differences between the genetic makeup of the patients. A lot of effort should be dedicated to the identification of biomarkers able to predict a patient's response to a specific drug or treatment, the optimal dose to be administered, and the likelihood that the patient will develop serious side-effects during the clinical trial or beyond its duration.

EHR patient data, linked to research results on pharmacogenomics and on the genetic variation of cellular response and resistance to treatment in the host and the tumour may allow for improving, adapting, and refining the standard treatment plans, taking into account both the disease stratification and the genetic profile of each individual patient.

Due to its incidence, its complexity in terms of data collected, patient stratification and therapy options, and its need for a multi-disciplinary approach to diagnosis and treatment, we believe that the oncology domain would benefit greatly from improved semantic interoperability and the ability to reuse for research the vast amounts of data collected within care.

Additionally, taking into account the wide use of therapies with long lasting, serious (sometimes life-threatening) and often unknown side effects, the cancer domain has a very strong need for tools supporting improved safety and the early detection of health risks.

EURECA aim and approach

The goal of the EURECA project is to enable seamless, secure, scalable and consistent linkage of healthcare information residing in EHR systems with information in clinical research information systems, such as clinical trial systems, supporting the two currently separated worlds of clinical research and clinical practice to connect and benefit from each other.

Main barriers of secondary use of EHR data for research and of enabling a consistent feedback loop to care are the lack of common technology standards and concept terminologies. While solving the interoperability issue in healthcare at generic level is not a realistic approach[7], EURECA aims at semantic interoperability on domains of concepts (i.e. describing specific clinical areas). We start from disease- and treatment-related sets of concepts in the oncology domain and demonstrate our solution in concrete clinical scenarios. On top

[6] The World Cancer Report, The International Agency for Research on Cancer, 2008

[7] Semantic Interoperability for Better Health and Safer Healthcare, SemanticHEALTH Report, pp 12-13, 2009

of the achieved semantic interoperability we build software services and tools to support more efficient research, better care and improved patient safety. In the second part of the project we also evaluate the extension of the core sets of concepts to other disease domains.
The approach taken in EURECA is to rely when possible on existing initiatives and previous efforts in terminology development and standardization. We will demonstrate the viability of the solutions developed by implementing a set of loosely-coupled interconnected services/modules that we will deploy in the context of several pilot demonstrators in the cancer area, at the sites of the healthcare organizations participating in EURECA.

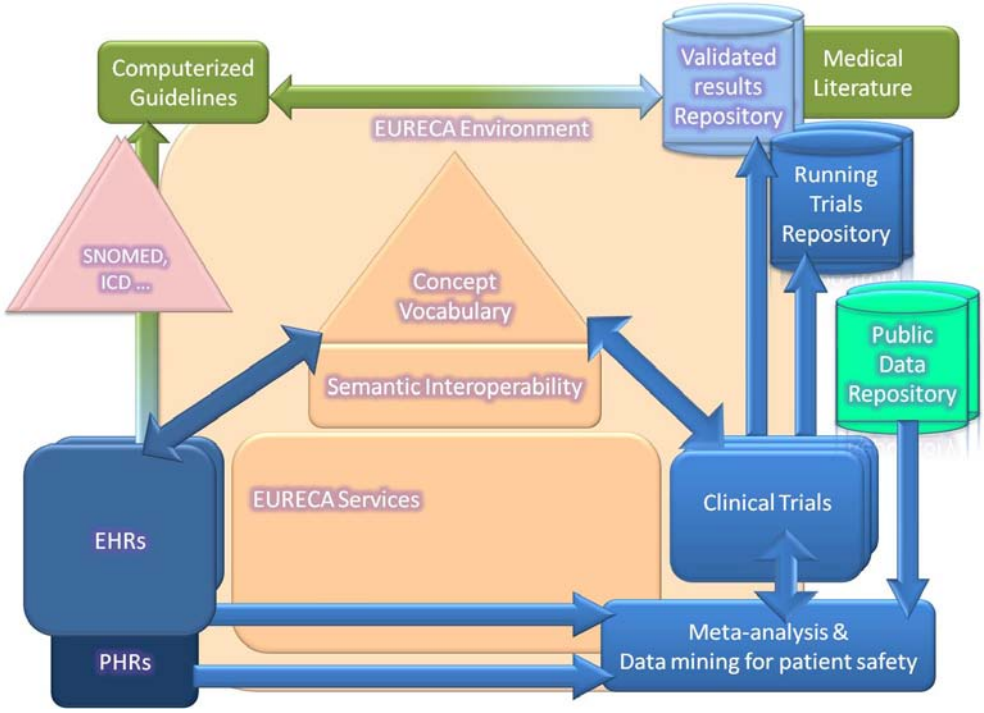

Figure 1 An overview of the EURECA environment and of the relevant systems. The interoperability framework implemented by EURECA will facilitate the urgently needed interconnection between the clinical trial systems and EHR systems.

The EURECA environment aims to provide several software services that help to securely interconnect the clinical trial systems (CTS) and the electronic health record systems (EHRs). This will bring several benefits among which early detection of patient safety issues and more efficient recruitment of eligible patients. Consistent linkage between CTS and EHRs will also help to significantly reduce the need for double data entry, which is currently often common practice.
Figure 1 describes the main components of the EURECA environment and Figure 2 depicts the envisioned services that will help to interconnect the two currently mostly disconnected worlds: the CTS and EHRs. Using the SemanticHEALTH classification of semantic interoperability (SIOp)[7] we can observe that the current level of SIOp between CTS and EHRs is somewhere between level 0 (no interoperability at all) and level 1 (syntactic interoperability). In order to achieve the above objectives we will have to increase the SIOp level to at least 2b (bidirectional semantic interoperability of meaningful fragments) or 3 (full semantic interoperability, sharable context) on specific clinical areas.
The essential steps for achieving this SIOp improvement include the definition of sound information models describing the clinical trial systems, building on existing research results when possible[8]. Electronic health records too need to be properly modelled; to that end we will adopt the appropriate state-of-the-art representation formalisms such as HL7 CDA, the openEHR Reference Model, ISO/EN 13606, etc.

[8] G.Weiler et al., Ontology Based Data Management Systems for post-genomic clinical Trials within an European Grid Infrastructure for Cancer Research, Proc of the 29th Annual Int. Conf. of the IEEE EMBS, 2007

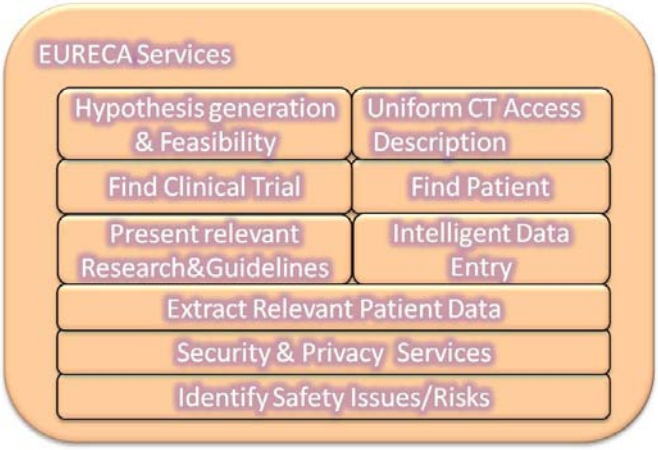

Figure 2 An overview of the EURECA software services.

The semantics of the clinical terms should be captured by standard terminology systems such as SNOMED CT, ICD, LOINC. The scalability of the solution needs to be achieved by modularization, e.g. instead of aiming at inclusion of the complete SNOMED terminology (more than 300 thousand concepts) we will identify a core subset that covers the chosen clinical domain. Such core data set shall be validated both by clinical and knowledge engineering experts to assure proper coverage and soundness. In the process of identifying the core data set and the corresponding mapping tools, care will be taken to allow to easily extend the core data set, should the inclusion of new concepts become necessary. This will be demonstrated by studies in other clinical domains and by extending core datasets covering more restrictive scenarios to wider scenarios (e.g. see section 1.2.24 – INTEGRATE project)
Utilizing the core data set we will devise a mapping system (Figure 3) between the information models of CTS and those of EHRs. These mappings will too be verified by our clinical experts and validated in properly chosen scenarios/use cases focusing on those with a high potential for patient safety improvement and more efficient and effective research (described in section 1.2.22).

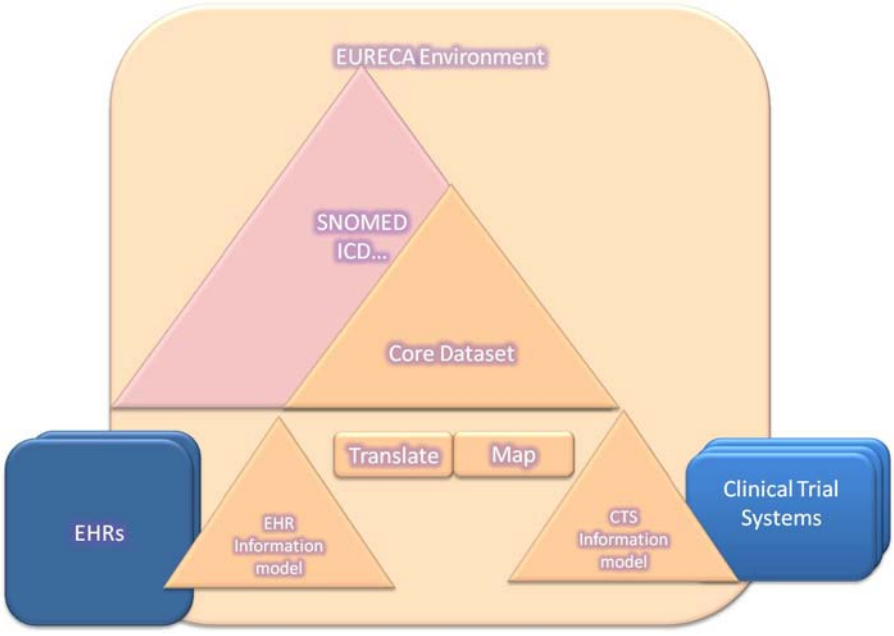

Figure 3 EURECA mapping components for achieving semantic interoperability

For the scenarios focused on early detection of safety risks and on long term follow-up of patients beyond the end of a clinical trial we will provide uniform access to Patient Health Record information, enabling the use of patient-generated information to improve the safety and quality of care.

The EURECA infrastructure will enable new ways of conducting meta-analysis over larger population samples, including long term patient follow-up as well as the results of previous clinical trials including those with negative results in order to be able to avoid known “dead-ends”.
The need and usefulness of tools that help to push forward the knowledge of risks associated with unique patient characteristics and treatment patterns has been well recognized by the EC[9]. EURECA will enable data mining for patient safety over the combined data from electronic medical health records and clinical research databases, thus also allowing to extend the scope of data mining to a currently untapped resource which is the EHR system. Additionally, the value of patient-acquired data out of the Patient Health Record in such a scenario will be assessed. Utilizing the EURECA framework, we plan to develop a proof of concept of such new data mining components.
Our clinical collaborators will bring to the project their extensive expertise in the research and treatment of breast cancer (BIG, Oxford, IJB, GBG), bone sarcoma (Oxford), paediatric oncology (UdS), radiation oncology (MAASTRO). They will provide clinical scenarios, domain knowledge for the building of the clinical information models and of the core datasets, clinical research and care data, and will act as pilot sites for the EURECA tools and software services contributing to the development of the interoperability framework through their experienced ICT departments.

EURECA Objectives

Enabling semantic interoperability among EHR and clinical trial systems
Achieving semantic interoperability among EHR and clinical trial systems is at the core of the EURECA project, as it is the basis for enabling many of the software services and tools developed in the project.
From the technology viewpoint, in order to provide an efficient, robust and semantically interoperable solution, one needs to move from plain keyword matching to a combined approach where keywords are mapped to higher level concepts with clearly defined semantics. Such concepts are usually organized in concept hierarchies and include domain specific attributes and relations. Reasoning at this level, rather than at keyword level, is expected to enable us to move from error-prone lexical matching to more robust semantic-aware solutions.

> … regardless of the type of vision one may develop, semantic interoperability is not a phenomenon to be expected over night.[..] The SemanticHEALTH is characterized by a large number of changes at both the technical and the use case level. Note however, that even in this vision, no full semantic interoperability or a complete harmonization of either EHR models or terminologies can be expected.
>
> “Semantic Interoperability for Better Health and Safer Healthcare”
> *SemanticHEALTH Report, 2009*

Linking clinical care information in the EHR systems with research data in the clinical trial systems will be achieved in EURECA by exposing the EHR information in a canonical clinical care model leveraging on existing standards and vocabularies, exposing the clinical trial information in a canonical clinical trial model, again leveraging on existing standards and vocabularies, and reaching semantic interoperability through the definition of a model and a core set of concepts that provide a mapping between the clinical care (EHR) model and the clinical trial model. To enable scalability and usability the development of the core set of concepts will be modular, per (sub-)disease domain. The use of supported terminology standards, such as SNOMED CT, ICD-10 and LOINC, will facilitate in the future easy linkage to clinical information systems adhering to those standards.
The semantic core data set will be validated in concrete use cases, with the different EHR and clinical trial systems available at the clinical care and clinical trial sites participating in EURECA. Handling the anticipated language heterogeneities represents an important challenge we plan to address.
As some of the relevant data in the EHR is stored in free text format, its standards-based seamless reuse relies on the accurate extraction of the important concepts and their relations. The task of building canonical information models of the EHR and CT systems becomes more complex when we also deal with unstructured documents in those systems. The discovery of the underlying information model contained in free text reports requires additional effort to identify the relevant concepts and their relations and to understand their meaning. We will achieve this gradually, with the support of a mix of NLP techniques used to answer concrete questions. Whenever possible, existing tools will be used.
For the free text documents we will carry out information extraction based on our defined core dataset, with a clear search goal and not a random search in an infinite space. We start by scanning the text with concepts from the core dataset (tagging, synonyms, annotations, etc.) and progress to relations and meaning. As data and model quality is essential, and any error in the NLP task would propagate further into the model and into

[9] V.N. Stroetmann et al., Impact of ICT on Patient Safety and Risk Management, eHealth for Safety Report, 2007

the application, we do not aim at a fully automated NLP process (that would replace the expert); instead, our goal is to make the tasks of the modelling and domain experts efficient and manageable.
This objective will be initially achieved by milestone M8 (T0+32) and validation will be carried out by milestone M9 (T0+40).

Enabling secondary use of care data for research
EURECA aims to support more effective and efficient execution of clinical research by providing access – in a legally compliant and secure manner – to the large amounts of patient data collected in the EHR systems to be used for new hypotheses building and testing (e.g. to benefit rare diseases), cohort studies, as well as protocol feasibility.
Association studies on large volumes of EHR data can also reveal serious side effects to drugs and therapies and link those with patient characteristics (generate new research hypotheses for biomarkers identification). Data mining of the EHR data can be carried out based directly on the information model of the source or based on the EURECA core set of concepts when it is assumed that the end user does not need detailed knowledge of the source. We will demonstrate this secondary use in concrete scenarios and provide uniform access interfaces to the data and tools/services developed to support the execution of the scenarios. Our services will also provide uniform access to external repositories of data and knowledge relevant in the discovery scenarios.
Another important re-use of EHR data is to support patient recruitment and to test trial feasibility in a data-driven protocol design process. Semantic interoperability between EHR and CT systems can support faster recruitment of patients in clinical trials, but also help to improve protocols to target the right patient population, define the inclusion/exclusion criteria, identify sites with sufficient patients, etc. This can lead to more efficient enrolment, better protocols, and fewer protocol amendments (e.g. criteria that are too narrow can be detected earlier).
This objective will be achieved by milestone M9 (T0+40).

Enabling efficient recruitment by matching relevant patient data with eligibility criteria from clinical trials
Good semantic interoperability among EHR and clinical trial systems can be exploited to allow more efficient patient enrolment in clinical trials, which is a need shared by both clinical care and clinical research as it would both improve care and speed up research. The eligibility and exclusion criteria of running clinical trials described in the Clinical Report Forms can be matched based on the EURECA core set of concepts with patient data in the EHR to find eligible patients for clinical trials, and in the context of a patient case to find relevant clinical trials. Additionally, we will provide support for efficient protocol design and refinement, enabling the interactive automatic evaluation of the various criteria based on EHR data.
Semantic interoperability between EHR and CT systems also enables us to provide solutions for patient recruitment that help avoid double data entry: establishing a single source for each data item, automatic storing of clinical trial eligibility criteria into the EHR and using the EHR data for automatic Electronic Data Capture (EDC).
This objective will be achieved by milestone M9 (T0+40).

Enabling long term follow up of patients beyond the end of a clinical trial, for better research and improved safety
It is relevant for clinical research to follow the patients long after the end of the trial, to monitor survival, recurrence, and serious side effects of treatments. Next to important benefits for the patient, this can also contribute to the generation of new research hypotheses and the identification of safety risks. EURECA will provide solutions to access the relevant information while adhering to all applicable privacy and security requirements. Relevant information can be extracted from the EHR based on the information model of the source and linked through the EURECA core set of concepts to clinical research concepts. Uniform access to patient-generated PHR data will add a new dimension to this objective: Leveraging this new source of information can support better (more accurate) assessment of the risks and benefits of drugs and treatments, at different time granularity (e.g. patient may describe side effects and symptoms when they happen), in a different context (patient at home) and on a time span beyond the duration of a clinical trial.
This objective will be achieved by milestone M9 (T0+40).

Improving efficiency by reducing the need for multiple data entry in the information exchange between clinical research and clinical care
We aim to avoid double data entry in care and research by establishing a primary source for the various types of data and providing the necessary data-feed loops to the EHR and the clinical trial systems (e.g. automatically filling in clinical trial eligibility criteria in the EHR based on the clinical trial CRFs) and preserving

compliance with workflows in clinical care and research. Additionally, EURECA will enable a consistent and efficient reporting of SAEs and SUSARs (adverse events in clinical trials) and address the current issues of multiple and inconsistent reporting.
EURECA tools/software services helping avoid multiple data entry are:

- Recruitment (e.g. through automatically filling in clinical trial eligibility criteria in the EHR based on the clinical trial CRFs, and automatic EDC prefilling based on EHR data)
- Safety issue detection and reporting (primary source, single reporting)

In this objective, we include evaluation of the use of PHR data as potential source for detecting safety issues. This objective will be achieved by milestone M9 (T0+40).

Exposing a uniform presentation of clinical trial information, validated clinical trial results and other relevant external knowledge and data resources

There are currently two types of databases collecting clinical trial-related data. Clinical trial registry databases are online catalogues of hypothesis-testing clinical trials conducted on human subjects. Clinical trial results databases are online repositories containing the results of clinical trials.

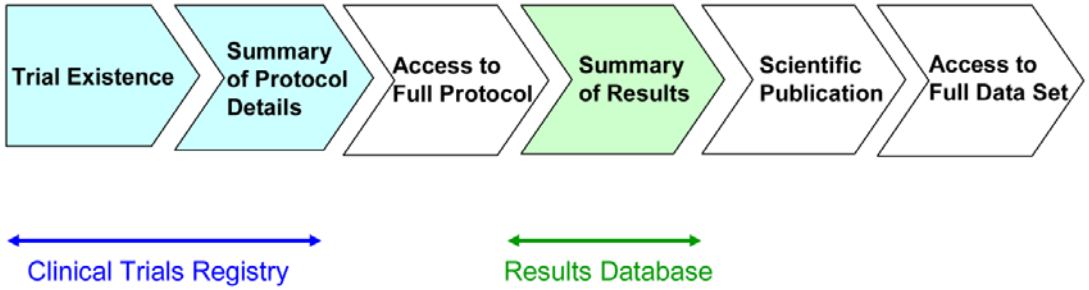

Figure 4 Information available in the clinical trial registry and results database[10]

EURECA will provide uniform interfaces to access external repositories of clinical trial information and results. Based on existing initiatives and emerging standards such as WHO ICTRP[11] we will select a core set of information items that need to be provided to comprehensively describe a clinical trial (compliant with the set of 20 key elements specified by WHO) and its results (including the case of trials with negative results) and findings, propose an information model and build a prototype repository to instantiate this information model. Additionally, we will link to external relevant clinical trial registries and clinical trial databases to be used by services in the EURECA environment.

- **Repository of running trials and public data repositories**

  We will identify relevant existing external data repositories and provide uniform access to them. When available, EURECA will utilize the existing repository(ies) of running clinical trials including their focus/aim, patient inclusion criteria, and suitable location (taking into account the geographic proximity to the patients) . EURECA will also provide access to public biomedical databases, the combination of those with the EHR and clinical trial data will be of interest to broader meta-analysis and data mining for patient safety. Important for our scenarios will be the linkage to PHR systems to enable the reuse of patient-authored information.

- **Repository of validated results**

  We also plan to identify those repositories which are not yet (widely) available and propose a suitable data model of the information that they should contain. This for instance includes a repository of validated results of completed clinical trials. We believe that structured access to validated results of clinical trials will bring tremendous benefits in speeding up the knowledge transfer and the rate of adoption of the latest research advances by the clinical practice.
  Positive results are currently being published in journals and conference papers and the data which was used during the research is required to be made public so that the repeatability of the findings can be verified and new discoveries are made possible from the existing data. Despite these positive trends, the solutions that are currently used for sharing the published results and the underlying data do not allow for efficient machine processing and identification. By structuring these results (as well as data when appropriate), the information access will become easier and the impact on further research larger, e.g.,

[10] D.A. Zarin and T. Tse, Moving Towards Transparency of Clinical Trials, Science. 2008 March 7; 319(5868): 1340–1342
[11] World Health Organization's International Clinical Trials Registry Platform, http://www.who.int/ictrp/en

by avoiding duplication in hypothesis generation, enabling efficient and effective data analysis. The research community will be able to take advantage of the uniform access to validated results of the completed trials, enabling efficient meta-analysis and accelerating the process of hypothesis generation.
If such repository also includes the negative results of clinical trials which are currently hardly ever published/shared, researchers can learn from the past by avoiding the already explored "dead-ends", focusing their efforts and resources mainly on new promising directions. We are aware that building, but also maintaining and contributing to such repository will be a considerable effort which should be incentivized, but we also believe that this is one of the clear examples when a moderate investment in improving the sub-optimal status-quo will deliver considerable benefits and help to achieve in the near future more by better utilization of the research resources and available data.

- **Computerized Guidelines**

Uniform access to validated results will accelerate the development and deployment of decision support systems. Proper implementation and use of clinical decision support systems is regarded as an important recommendation for reducing the frequency and consequences of errors in medical care[12]. Having access to such repository will also help to make the process of creation of clinical guidelines much more systematic and ICT tools will be much more applicable in this, currently mainly manual, process. As a result, the clinical guidelines will become computerized from the very beginning and this will be another major enabler for a new generation of decision support systems which will be able to bring these guidelines to the point of care and ensure that the standard of care will be consistently improving.

- **Warehouses of clinical trial data, data models, ontologies, etc.**

EURECA aims to build on existing initiatives focusing on standardization and sharing of data, models (e.g. risk models, predictive models of response, etc.) and knowledge in the clinical domain. We plan to achieve uniform access to relevant repositories and infrastructures. Partners in the project are currently active in several such initiatives in Europe and beyond, and we intend to build bi-directional links to those initiatives in order to achieve a critical mass. Interoperability is by definition a matter of critical mass, openness and team work and cannot be achieved in isolation.

This objective will be achieved by milestone M8 (T0+32).

Extracting relevant clinical information from the EHR and contextualizing it to the patient case

A proper semantic link between EHR and clinical trial systems will enable, among others, an efficient generation and adaptation of computerized clinical guidelines such that they reflect the latest validated research results. As a result, standard clinical care can use ICT tools to quickly adopt and personalize these guidelines, mapping them to the actual patient records, identifying optimal treatment plans based on the extracted relevant information from the EHR. These technological and conceptual advancements will lay the foundations for a new generation of clinical decision support systems that are both semantically aware and up-to-date with the latest validated research results.
This objective will be achieved by milestone M9 (T0+32).

Building solutions for faster transfer and dissemination of relevant research results and literature to the care

EURECA aims to research the contextualization of validated research data and literature to a patient case, to be used in standard care. Relevant external data, such as references to articles, applicable literature, guidelines, protocols and standardized descriptions of validated trials, will be selected based on the current patient case under investigation. In this process we will use the information model of the source EHR, the EURECA core of concepts and the external sources of information to which EURECA provides unifo0rm access, such as computerized clinical guidelines, external research public databases and repositories of clinical trial information and validated results.
This objective will be achieved by milestone M9 (T0+40).

People making decisions about healthcare need access to knowledge derived from the findings of clinical trial research,

"Reporting the findings of clinical trials: a discussion paper"
Ghersi D et al., Bulletin of the World Health Organization, 2008.

[12] D.W. Bates et al., Reducing the Frequency of Errors in Medicine Using Information Technology, Journal of the American Medical Informatics Association, 2001

Prototyping and validation of the EURECA services/modules and tools
The capability of the EURECA environment (services and tools) to achieve the above mentioned objectives will be demonstrated through prototypes implementing realistic clinical scenarios. These scenarios will also allow us to demonstrate the interoperability of the EURECA services.
The prototypes will be first deployed at the sites of the EURECA clinical partners and validated with our clinical partners. In a second stage, we will also promote and validate our solutions within the large user groups that expressed interest in EURECA and with which we have set up close collaboration.
Additionally, in order to gain momentum and community support, we are committed to provide an open source reference architecture, disclose code as open source, and to ensure low entry barrier into EURECA for relevant open source systems and tools.
This objective will be initially achieved by milestone M8 (T0+32), while the final validation is achieved by milestone M9(T0+40).

Clinical scenarios
We will demonstrate the software services and tools of the EURECA environment and the achieved semantic interoperability among EHR and clinical trial systems in several clinical scenarios, detailed further in the document (Section 1.2.22 and WP1). In all the scenarios full adherence to all legal, security and privacy requirements will be ensured.
The scenarios will be defined by milestone M2 (T0+9), the legal requirements will be available by M3 (T0+11) and the validation will be achieved by milestone M9 (T0+40).

1.1.2. Relevance to the ICT Call Topic

The EURECA proposal focuses on **Objective 5.3: Patient Guidance Services (PGS), safety and healthcare record information reuse, target outcome b) Tools and environments enabling the re-use of electronic health records**. Table 1 indicates how the EURECA proposal is addressing the individual topics of the objective of the call. In this section we briefly present the main EURECA objectives aligned with the call objectives.

Table 1 – Fit of EURECA Topics to Work Programme Topics

Topics of Objective 5.3 b)	EURECA Topics
Development of an advanced environment for clinical research that enables seamless, secure and consistent integration or linking of clinical care information in EHRs with information in CT systems.	• Bi-directional linking of clinical care (EHR) and research (CTMS and various research databases) • Semantic interoperability among EHR and clinical trial systems, consistent with existing standards, while managing the various sources of heterogeneity: technology solution, medical vocabulary, language, etc. • The semantic layer will support advanced end-user applications such as patient recruitment, detection of serious side effects, automatic extraction of relevant data for research to avoid multiple data entry, contextualization of information to a patient case, etc. • Definition of sound information models describing the EHR and the clinical trial systems. • Capture the semantics of the clinical terms by standard terminology systems such as SNOMED CT, ICD, LOINC. • Scalability of the solution by modularization, identifying a core subset covering the chosen clinical domain in supported clinical scenarios. The core data sets are defined together with domain experts. Extension to new clinical domains or scenarios is achieved by the definition of suitable core data sets describing the new domains/scenarios. • Implementation of loosely-coupled interoperable services/modules and tools. Provide consistent, uniform interfaces among services/modules enabling flexible composition of modules into an advanced and scalable environment that seamlessly addresses the user needs. • All services/modules and tools developed will be fully compliant with privacy and security needs.
Avoid double data entry	• Propose a primary source for data items • Model and formalize eligibility criteria using relevant standards • Automatically export clinical trial eligibility criteria contained in the Clinical Report Forms to the EHR, automatically fill in relevant data from the EHR into Clinical Report Forms • Single reporting of serious safety issues by integration of care systems, research systems and the PHR
Assist in identification of patients for clinical trials	• Improve patient recruitment and trial search by automatic identification of potential candidates for a clinical trial: o Find patients for clinical trial o Find relevant trials for patients
Enable early detection of potential patient safety issues	Enable long term follow up of patients, beyond the end of a trial Enable data mining on EHR data to detect safety issues that could not be found within the scope of a clinical trial Support consistent and efficient reporting of adverse events (SUSARs, SAEs) Enable uniform access to PHR data to detect safety issues relying on patient reports Enable contextualization of information in clinical care to a patient case, including linkage with relevant information out of external sources.

Research will focus on the areas of improving semantic interoperability between EHR and CT systems. This will include the definition and the validation of core data sets that enable scalable and standardised linking with EHR repositories.	• Achieving semantic interoperability among EHR and clinical trial systems is at the core of the EURECA project • From the technology viewpoint, in order to provide an efficient, robust and semantically interoperable solution, one needs to move from plain keyword matching to a combined approach where keywords are mapped to higher level concepts with clearly defined semantics. Such concepts are usually organized in concept hierarchies and include domain specific attributes and relations. Reasoning at this level, rather than at keyword level, is expected to enable us to move from error-prone lexical matching to more robust semantic-aware solutions. • We reach semantic interoperability through the definition of a model and a core set of concepts that describe a specific domain and provide a mapping between the clinical care (EHR) information model and the clinical trial model. To enable scalability and usability the development of the core set of concepts will be modular, per (sub-)disease domains. The use of supported terminology standards, such as SNOMED CT, ICD-10 and LOINC, will facilitate in the future easy linkage to clinical information systems adhering to those standards. • The semantic core data set will be validated in concrete use cases, with the different EHR and clinical trial systems available at the clinical care and clinical trial sites participating in EURECA.
Proposals will address data protection and security needs	• All services/modules and tools developed will be fully compliant with privacy and security needs • The consortium includes experts in data protection and security • Work package 9 defines several tasks regarding data protection and security, including building a consent service. • We build on expertise and results from the ACGT project, where privacy and security were thoroughly addressed[13],[14].
Be fully compliant with all applicable legislation	• The consortium includes experts in IT law • WP 7 defines several tasks focusing on relevant legal aspects emerging from linking EHR and clinical trial data. • We build on expertise and results from the ACGT project, where legal aspects concerning handling and sharing patient data across clinical trials were addressed in several tasks[15]. • EURECA can rely on the Center for Data Protection (CDP)[13] non-profit organisation to act as independent Data Protection Authority for EURECA where needed.

[13] www.privacypeople.org/
[14] http://www.eurorec.org/files/filesPublic/ehrworkshop/2007/20071012-EFPIA-Custodix.ppt
[15] http://www.eu-acgt.org/uploads/media/ACGT_D10.2_IRI_Final_01.pdf

Be fully compliant with best practice	• The expertise of our clinical partners, top research institutions in the oncology area but also large hospitals treating large numbers of cancer patients in and outside clinical trials, will ensure full compliance with best practices. • External sources of relevant information, such as clinical guidelines, will be used to contextualize EHR information for a patient case. • EURECA will support the use of standard terminology systems such as SNOMED CT, ICD, LOINC. • We will provide uniform access to trial information and propose uniform reporting of clinical trial information and validated results , taking into account existing initiatives and standardization processes (e.g. led by WHO[16], or in the US by NIH[17]). • The project will link with several large user groups and research organizations which will participate to user requirements and evaluation and validation of our solutions. • Results will be fully compliant with regulatory requirements (e.g. those set by the FDA – as our clinical partners and the pharma work on global scale).

Research results should be validated in well defined use cases with a high potential for improving patient safety in the clinical research and epidemiology fields.	• Demonstrate our EURECA environment and the achieved semantic interoperability among EHR and clinical trial systems in several well-defined clinical scenarios at the participating clinical sites:
	  • Long term follow up of patients, beyond the end of a clinical trial
	  • Data mining of EHR data to detect safety issues that could not be found within the scope of a clinical trial
	  • Identification of potential candidates for clinical trials
	  • Consistent and efficient reporting of adverse events
	  • Contextualization of information in clinical care to a patient case

Resources are to be targeted to use and complete the common shared info-structure that will be established by the PCP under the governance of the NoE on semantic interoperability	• We have established a Technical Board, managed by a Lead Architect. This Technical Board, including representatives of all partners involved in ICT R&D work, will have the responsibility to collaborate with the NoE and to other related projects on semantic interoperability and to define the contribution to the common tasks. • To support our contribution to the common shared info-structure we have reserved a budget to cover for additional activities unaccounted for in the EURECA description of work.

In both healthcare and the pharmaceutical industry there is an urgent need to combine higher performance with cost awareness and increased efficiency. The EURECA research aims to support these goals by enabling access to relevant longitudinal data from clinical practice to be used for research and automatic selection of potentially eligible patients from clinical trials, and by supporting increased patient safety through improved detection, reliable and efficient reporting, and reduction of serious patient health risks, side effects of therapies and medical errors. To enable scalability, EURECA proposes technology solutions based on loosely-coupled flexible services and tools, interconnected by well-defined standard interfaces, and making use of well-established terminology standards such as SNOMED, ICD, UMLS and LOINC.

Within the limited duration and size of clinical trials it is not possible to detect adverse effects that are either rare or manifest themselves beyond the trial end. In this context, long-term reliable follow up of patients beyond the trial duration is required, while access to longitudinal patient data from the standard care repositories could enable research to identify health risks, and serious drug and treatment side effects.

The EURECA project aims to bi-directionally link clinical care information present in the EHR systems with clinical research data stored in clinical trial systems. In achieving the required semantic interoperability among EHR systems and clinical trial systems we intend to extract a well defined set of relevant domain concepts that describe the semantics of the chosen clinical domain. This semantic data set will be mapped to concepts from existing standardized vocabularies (e.g. well established and widely used clinical terminologies

[16] http://www.who.int/ictrp/
[17] http://www.clinicaltrials.gov

such as SNOMED CT, ICD-10) to foster scalable semantic interoperability not only among the clinical partners within the project, but also towards other healthcare and research organizations adhering to the adopted standards.
The semantic data set will be validated in concrete scenarios, for the different EHR and clinical trial systems available at the clinical sites from Belgium, Germany, Switzerland, and the UK participating to EURECA. Handling the anticipated language heterogeneities is another important challenge we will address.
EURECA will build on our expertise and results achieved in the ACGT**Error! Bookmark not defined.** project nd the links established with several large projects[18],[19],[20] in the oncology domain will enable us to test and deliver our solutions to a large user base. The initial core sets of concepts (the EURECA core data sets) will represent various diseases in the oncology domain, but the services and tools developed will be generic and applicable to other disease domains by building new core sets of concepts describing those domains. We will also evaluate the extension of the breast cancer core data set that is developed in the FP7 INTEGRATE project, to represent the set of concepts relevant for the repository of data collected within the neo-adjuvant clinical trials in breast cancer led by the Breast International Group. Next to ramping up our work on the development of core datasets, this experiment will allow us to evaluate the scalability of our approach. The EURECA core datasets will be developed in a modular, scalable way to cover the different areas in oncology that are relevant for our clinical partners. In the second part of the project we will evaluate the applicability of our solutions to other clinical domains.

1.1.3. Quantification of Results

The EURECA results will be measured in the following areas matching the goals of the call.
Development of an advanced environment for clinical research that enables seamless, secure and consistent integration or linking of clinical care information in electronic health records (EHR/PHR) with information in clinical trial systems.

We will validate our semantic interoperability environment in concrete scenarios such as trial recruitment, detection of safety risks, reporting of serious side effects, cohort studies, etc. To demonstrate secure linkage and detect issues, an audit trail will be implemented for all access and data operations. **During the validation we aim to achieve 100% secure transactions.**

Seamlessness of linkage will be demonstrated by providing transparent access to end-users in both types of systems and eliminating the need for double data entry in selected scenarios (**reduction by at least a factor of 3**).

Consistency will be demonstrated during the validation of our interoperability layer by **full semantic linkage between the two types of systems for our core datasets**.

Results are expected to help health professionals avoid double data entry, assist in identification of patients for clinical trials and enable early detection of potential patient safety issues.

- There are several steps in the execution of a clinical trial (e.g. eligible patient identification, entry of Clinical Report Forms, entry of lab results, adverse events reporting, etc.) in which the same data needs to be entered twice, in the EHR system at the clinical side and in the clinical trial system at the research site. We will study the current clinical workflow in clinical trials run by our clinical partners and show how the use of the EURECA services mitigate the need for multiple data entry at several points along the workflow. . As concrete case study, we will perform a “simulation study”, i.e. we will evaluate for concrete clinical trials that have been already completed by our clinical partners how many CRF fields describing information that also resides in the EHR can be automatically filled in. Based on our preliminary analysis, we expect that the EURECA solution will provide, in specific clinical scenarios, **a reduction of the avoidable duplicative data entries by about a factor of 3.**
- Adverse-events reporting currently requires several steps of data entry describing the same event, which often results in inefficiency and inconsistency. The EURECA solutions will eliminate this need while supporting the execution of the needed clinical workflow (**decrease of double data entry by 100%**).
- Currently, registering patients into clinical trials and finding eligible trials for patients require manual search and clinicians may be overwhelmed by the number of CTs and the exclusion and eligibility criteria. The EURECA service, relying on the EURECA core set of concepts describing the disease sub-domain, will provide suggestions of potentially eligible patients according to the available eligibility and exclusion

[18] Continuation project of EuroBoNeT, http://www.eurobonet.eu
[19] http://www.euroxy.info
[20] http://www.iss.uio.no/forskning/eumargins/news/media/2008-07-02.html

criteria. This service will be demonstrated in the context of concrete clinical trials, with realistic test data set including longitudinal EHR and clinical trial data. By automatic matching, we expect to **reduce the search space with respect to the number of patients, CTs and exclusion/inclusion criteria** that need to be manually reviewed **to approximately 20% of the original search space**.

- Several EURECA scenarios aim at improving patient safety by enabling long term follow up, data mining of EHR data, detection and reporting of safety risks and serious side effects, and contextualization of information to a patient case. We will select concrete use cases in the disease domains chosen for the core set of concepts enabling semantic interoperability and carry out validation with realistic data sets. To validate the EURECA services in those scenarios access to longitudinal patient records combined with matching clinical trial data will be used. **For the same number of occurred safety issues, we estimate a reduction of the time spent on reporting safety issues by a factor of 3.**
- Additionally, we will evaluate how interoperability to PHR information can help early detection of serious side effects and safety risks.

Research will focus on the areas of improving semantic interoperability between EHR and clinical research systems. This will include the definition and validation of core data sets that enable scalable and standardised linking with EHR repositories.

We will build canonical information models representing the EHR systems and the clinical trial systems deployed at clinical sites participating in EURECA. The core set of concepts enabling the linkage of the two types of systems will be built in a modular, scalable way, starting with a contained domain, e.g. breast cancer. The core data set will be linked to existing standards such as SNOMED CT. The linkage will be first demonstrated in the context of a single language (i.e. English to English) and a single institution, and then extended (scaled) to include other institutions and languages corresponding to the clinical sites which are partners in EURECA. Subsequently, we will extend the scope to other selected clinical domains relevant for the clinical partners of EURECA. While the main focus of the clinical experts in EURECA is the research and treatment of cancer, the commitment of the IT departments of the large academic hospitals they represent will ensure that our pilots will be adopted and extended to other disease domains and implemented hospital-wide. By defining new core datasets and extending the existing ones, the interoperability environment could be used in any other clinical domain.

We will consider to have reached the desired level of semantic interoperability when our interoperability solution can support the implementation of the EURECA services as described, in the context of the defined clinical scenarios and for the selected clinical sub-domains. The ability to demonstrate the scenarios in concrete use cases and with real data also represents a validation of the core set of concepts and of the information models built.

Proposals will address data protection and security needs and be fully compliant with all applicable legislation as well as best practice.

The data protection and security needs related to the execution of the EURECA clinical scenarios will be evaluated in the project and solutions will be devised to address those needs. The data flow through the EURECA environment will be in full compliance with all privacy and security requirements identified. (**100% compliance**)

The legal requirements regarding patient data management emerging for the linkage of EHR and clinical trial systems data, and from the execution of the EURECA clinical scenarios will be studied in a European context. The EURECA services/modules and tools will be validated at the clinical sites in compliance with the applicable legislation identified. (**100% compliance**)

Research results should be validated in use cases with a high potential for improving patient safety in the clinical research and epidemiology fields.

The following EURECA scenarios aiming at research and epidemiology directly impact patient safety:

- Enable seamless long term follow up of patients, beyond the end of a clinical trial
- Efficient execution of data mining and association studies on EHR data to detect safety issues that could not be found within the scope of a clinical trial
- Late effects of radiotherapy and side effects of treatment that become manifest after a period of time larger than the duration of the clinical trial
- Severe side effects of treatment that are rare and cannot be detected within the limited population of a clinical trial
- Generate research hypotheses and retrospective validation for rare diseases based on EHR data
- Early, consistent and efficient reporting of adverse events (SUSARs, SAEs)

- Enable interoperability to PHR systems, providing access to patient-authored data, which can be used for early detection of serious side effects

We will validate the EURECA environment in the above scenarios, in concrete use cases provided by our clinical partners, with realistic data sets, in the context of the disease domains represented in our core sets of concepts. In every case, "WP8: Q&A, evaluation and validation" will develop a detailed evaluation and validation framework which will enable us to quantify the anticipated benefits of our novel solutions and services. These quantified results will be reported and also presented in international conferences and journals.

1.2. State of the art and beyond

1.2.1. Progress beyond the state-of-the-art

EURECA will advance the state-of-the-art in various ways. **The foundation of the advances is set by providing a semantic interoperability layer over the EHR and CT systems.** Canonical information models will be built for the EHR and CT systems and the incorporation of a focussed core data set (reusing existing initiatives and standards as much as possible) will help deliver semantic interoperability. Technically, the definition of the semantic interoperability layer will be a huge leap forward. The emphasis will be on a targeted, scalable contribution: Determining what subset of existing concepts and terminologies out of available standards are applicable, and coping with the fact that those will not be complete and not completely sound.

The semantic interoperability layer will greatly advance the currently tedious patient recruitment for clinical trials. It enables the development of a service which automatically identifies eligible patients based on their EHR data. And once patients are enrolled in a clinical trial, multiple data entry is avoided by relying on the semantic interoperability layer to identify and re-use applicable clinical data elements residing in the EHR.

The main advances over the state-of-the-art targeted by the EURECA project in relation to the topics of the project call are presented in the table below. Other areas in which EURECA will deliver substantial improvements are also described in this section.

Table 1.2 – Advances over the State-of-the-Art

Topic	State-of-the-Art	EURECA goals
Consistent integration of EHR systems and CT systems	Distinct worlds of clinical research and care. At each clinical research site, several non-integrated systems coexist.	Semantically interoperable linkage of clinical care and clinical research systems on clinical domains.
Leverage on the integration need	Many low scale efforts towards integration and interoperability. The different initiatives create no synergy, there is no critical mass. Current approaches do not scale and often diverge.	Propose a scalable, open, consistent interoperability framework, with a low-entry barrier, able to leverage on existing efforts and create the necessary critical mass. Develop solutions together with large user groups and establish collaboration with several similar initiatives to join forces and reach impact.
Avoiding double data entry	The disparity of EHR systems and clinical trial systems often requires multiple data entries of the same information, leading to sub-optimal resource utilization and increasing potential inconsistencies and errors	EURECA will provide a platform where existing relevant clinical data can be extracted from the EHR systems, populating the corresponding entries clinical trial systems.
Automatic identification of patients for clinical trials	Eligible patients are identified mostly manually which often results in low clinical trial enrolment.	The EURECA services enable automatic identification of eligible patients whenever possible. When not all inclusion criteria can be directly obtained from the EHR, EURECA will prune the search space of eligible patients by applying the exclusion criteria, reducing thus the required effort for eligibility identification.

Topic	State-of-the-Art	EURECA goals
Automatic identification of clinical trials for patients (care setting)	Selection of appropriate trials is in the responsibility of the treating physician and requires manual search and criteria matching which can lead to low clinical trial enrolment	In a setting that is symmetrical to the identification of patients for clinical trials, the EURECA system will enable the automatic identification of suitable trials for patients.
Semantic core data sets	Huge medical terminologies (400,000 concepts) not completely sound and consistent, hard to apply to particular semantic interoperability problems.	Identifying relevant and sound subsets of exiting terminologies that cover domains with high potential for improving semantic interoperability in defined use cases with high potential for addressing patient safety issues and improving research and care.
Early detection of patient safety issues	Patient safety issues are detected very late, no practical way exists to access the heterogeneous landscape of existing EHR systems for research purposes and those systems are disconnected from the clinical research systems. Lack of availability of tools to detect safety issues in EHR data.	Data mining tools/services enabling association studies to detect safety issues, combined with seamless linkage to clinical trial systems to enable the detection of side effects that are rare or with late onset. Automated data mining techniques allow to implement services that actively monitor EHR data for patient safety issues instead of only retrospective analyses
Reporting of safety issues	Inefficient reporting requiring multiple data entry in distinct systems and by different people. Often reporting is inconsistent among actors in research and care and errors occur.	Seamless integrated reporting based on semantic core dataset and standard terminology. Less errors due to single data entry.
Information extraction	Free text reports are common in EHRs and efforts to structure all information have failed. NLP tools are not yet accurate enough.	A pragmatic and advanced information extraction process will be applied, driven by the defined core data sets, to extract out of free text relevant concepts and their relations.
Efficient and effective use of clinical guidelines	Information contained in EHR and clinical trial systems is seldom linked to computerized clinical guidelines.	Contextualized application of computerized clinical guidelines and care pathway to individual patients allows to implement a single point of information access for the treating physician. This supports the clinical staff in selecting an optimal treatment, especially in critical situations with high information overload.
Patient Health Record	Low use of PHR, in isolated, fragmented scenarios (mainly patient portals).	Link through our interoperability framework EHR and CT systems to PHR in an integrated scenario for early detection of safety issues.
Security and privacy	Implementation of data and resource access control through authorization services in Service Oriented Architectures. This is a complex approach which needs simplification.	EURECA aims to technically enforce and govern the legal framework through policy based authorization services, innovative in extending authorization services to enforce the complex data protection policies which originate from the legal implications connected to crossing the domain of care and clinical trials.
Consent Management	Ad-hoc patient consent solutions, mostly deployed locally in a confined environment. Cross institution data sharing is hence very limited (the obtained consent usually governs only local use cases).	Sound patient consent service framework across various collaborating parties. EURECA will research consent in a European perspective, covering both care and research (trials). Part of the innovation lays in researching novel scenario's like automated patient recruitment and unsolved issues such as complex individual consent (i.e. leading the way to patient-empowerment).

Topic	State-of-the-Art	EURECA goals
Improved access to outcomes of clinical trials	Non-computable heterogeneous description of clinical trials Literature biased towards positive results reporting; lack of easily accessible and exhaustive survey of clinical-trials and associated outcomes.	Standardized, computable trial descriptions through federation of online trial registries and uniform presentation of their contents facilitates the definition of new trials and minimizes loss of resources (e.g. avoiding investigating multiple times research "dead-ends", and repeating prior research).
Standardizing clinical trial results	No standard for the representation of detailed clinical trial results. Results published on heterogeneous supports. Retrieved data require a time-consuming work of curation before being usable.	Availability of a proper representation of clinical trial data reduces the time spent on curation, minimizes errors and increases the time available for proper research. Consistent data representation is facilitating meta-analyses.
Contextualized access to research results	Validated research results and research literature need to be manually searched and reviewed.	Research results and literature brought at the point of care, contextualized to the patient case under review.
Selection of relevant information for care episode	Overwhelming amounts of health information from the care clinical system need to be reviewed for a patient even when not relevant for the current health episode.	Information prioritized by relevance to the patient case (current health episode) and presented in a structured way.

EURECA will greatly improve patient safety. The semantic interoperability layer will enable data mining tools and services to monitor patient safety issues by supplying data mining tools with transparent, standardized access to validated and semantically consistent clinical data. This will allow to regularly analyze data on a much larger scale than what is currently possible inside of clinical trials. Thus, patient safety issues can be spotted which did not show up during a clinical trial – for instance due to the rareness of the issue or to late onset.

The innovative use of patient data needs to comply with the applicable legislation concerning patient consent and privacy. Therefore, effort is allocated within EURECA to **research the legal implications when crossing the domain of care and clinical trials, and to provide a technical framework which allows and enforces implementation of services according the legal framework.**

The nature of issues related to semantic interoperability as well as the envisioned use of the EURECA platform calls for adopting a service oriented paradigm at architectural level which enables to build the required solution out of flexible, loosely-coupled set of composable services. The term service oriented architecture (SOA) has been adopted by the main ICT actors and by the main standardization groups. We envision that EURECA will contribute to the relevant initiatives by expanding the state of the art with a proper semantic description of healthcare-related software services identified by the EURECA scenarios.

1.2.2. EHR and clinical research integration

Integrating EHRs and clinical trial data would provide a significant step towards increasing the efficiency and effectiveness of clinical trials[21]. Examples[22] of how EHR's could accelerate Clinical Trials are:

Trial Step	*EHR potential role*
Study setup	• Query EHR database to establish number of potential study candidates. • Incorporate study manual or special instructions into EHR "clinical content" for study encounters
Study enrolment	• Implement study screening parameters into patient registration and scheduling. • Query EHR database to contact/recruit potential candidates and notify the patient's provider(s) of potential study eligibility.
Study	• Incorporate study-specific data capture as part of routine clinical care / clinical

[21] Meeting the Challenges of Patient Recruitment. A Role for Electronic Health Records, C. Ohmann and W. Kuchinke, Int. J. Pharm. Med., 2007

[22] Integrating Electronic Health Records and Clinical Trials: An Examination of Pragmatic Issues, Michael Kahn

execution	documentation workflows • Auto-populate study data elements into care report forms from other parts of the EHR database • Embed study-specific data requirements (case record forms) as special tabs/documentation templates using structured data entry. • Implement rules/alerts to ensure compliance with study data collection requirements • Create range checks and structured documentation checks to ensure valid data entry
Submission & Reporting	• Provide data extraction formats that support data exchange standards • Document and report adverse events
Evidence based review	• Assess congruence of new findings and existing evidence with current practice and outcomes (incorporate into meta-analyses) • Submit findings to electronic trial banks using published standards.
Evidence based clinical care	• Implement study findings as clinical documentation, orders sets, point-of-care rules/alerts • Monitor changes in care and outcomes in response to evidence-based clinical decision support • Provide easy access to detailed clinical care data for motivating new clinical trial hypotheses

The ideal environment provides non-redundant systems and processes that allow the use of patient electronic health data for clinical research in a way that meets data protection, regulatory, and ethical research requirements and minimizes the challenges of clinical research in for healthcare professionals[23]. However, several barriers need to be taken before the benefits can be reaped. Unfortunately, the EHR market is highly fragmented[24,25] with global technology majors, EHR vendors and local systems developers all vying for a market share, and lacks standardization. In addition, a majority of large (academic) medical centers have implemented home grown or locally customized solutions. Due to their research emphasis, large academic centres place high value in their data, focus on efficiency and integration, and invest significant amounts in building customized ICT solutions. This is also the case at our clinical partners.

Home-grown and/or customized solutions and large IT crews are the norm among the very large clinical and research centres (e.g. MDA[26], VA[27]), while the vendor market is and will probably remain very fragmented and non-dominant in the clinical research environment. EHR vendors adhere to very few health information standards and standardized terminologies[28]. Additionally, many legacy systems and large prior investments make the wide adoption of a single commercial off the shelf EHR system an unlikely scenario. In this context, the singular solution towards large scale semantic interoperability is offered by global collaboration, and the adoption of common models and of standards-based solutions to link the various systems, within and across organizations, while being able to deal with a wide variety of local and customized infrastructures.

Only a very open, standards-based, scalable approach, with a low adoption barrier and significant critical mass, can ensure the success of a semantic interoperability solution. This is the approach taken by the EURECA architecture, as detailed further in this document. We will also elaborate a set of guidelines clearly describing all the necessary steps that would enable a new organization to join our environment.

At The University of Texas M.D. Anderson Cancer (Houston, Texas), with its very high patient volumes and commitment to research-based treatments and education, we were not able to identify an “off the shelf” electronic health record (EHR) system that would scale up and suit the needs of the patients, staff, and institution. In 2003, the decision was made to build on the foundation of a home-grown system called Clinic Station.

Additionally, we will follow the major open source initiatives in the EHR and CT systems area. For instance, UOXF has deployed a system based on OpenClinica[29], which will become part of our semantic interoperability environment. An important goal is to reduce to minimum the entry barrier for EHR solutions that are open and based on widely adopted standards.

Our commitment to open solutions is also demonstrated by the outcomes of the ACGT project that include an open source, ontology based Trial Builder and Management System (ObTiMA). In EURECA the

[23] EuroRec Electronic Health Records for Clinical Research Functional Profile, Version 1.0, www.eclinicalforum.org/main/ehrcrproject

[24] Electronic Health Record (EHR) Data: Modernizing the Pharmaceutical Process, Deloitte, 2009

[25] Strategic Analysis of the Key Healthcare IT modernization Initiatives in Europe, Frost & Sullivan, 2005

[26] http://www.jopasco.org/content/3/3/172.full

[27] http://www.healthcareitnews.com/news/va-good-model-ehr-systems-and-implementation

[28] http://www.ncrr.nih.gov/pubications/informatics/ehr.pdf

[29] www.openclinica.org

interoperability scenario at the UdS will focus on linking through the EURECA framework the local home-grown EHR and the ObTiMA system.

High profile initiatives, such as caBIG[30] and Sage Bionetworks[31], have as a mission to bring together researchers and their data and knowledge, build tools and infrastructures enabling collaboration, support reuse of data, models and tools, and promote common standards and interoperability on a large scale. EURECA will collaborate with these initiatives, contribute to common standards and promote our solutions in these large networks of users. We will rely on already established close collaborations (e.g. Sage Bionetworks is an advisor in the INTEGRATE project and Philips, BIG and IJB are members in several working groups of Sage Commons[32], the users and contributors community of Sage Bionetworks).

We are aware that the US-based solutions will not be directly applicable to the EU market, where a specific issue is a very high fragmentation and heterogeneity with respect to: approach to integration and interoperability, implementation of standards, use of standard terminologies, legislation, regulations, language, EHR vendor market (often country specific). Therefore, our goal is to develop solutions closely fitting the EU context, while at the same time achieving synergy at global scale by contributing the EU-specific requirements to these large global initiatives, collaborating on standards, models and terminologies, linking our frameworks, reusing suitable parts of their solutions, and contributing back our solutions. Next to our large scale collaboration at EU-level with clinical research networks and European interoperability initiatives, this will enable EURECA to gain the needed momentum.

Integration requirements

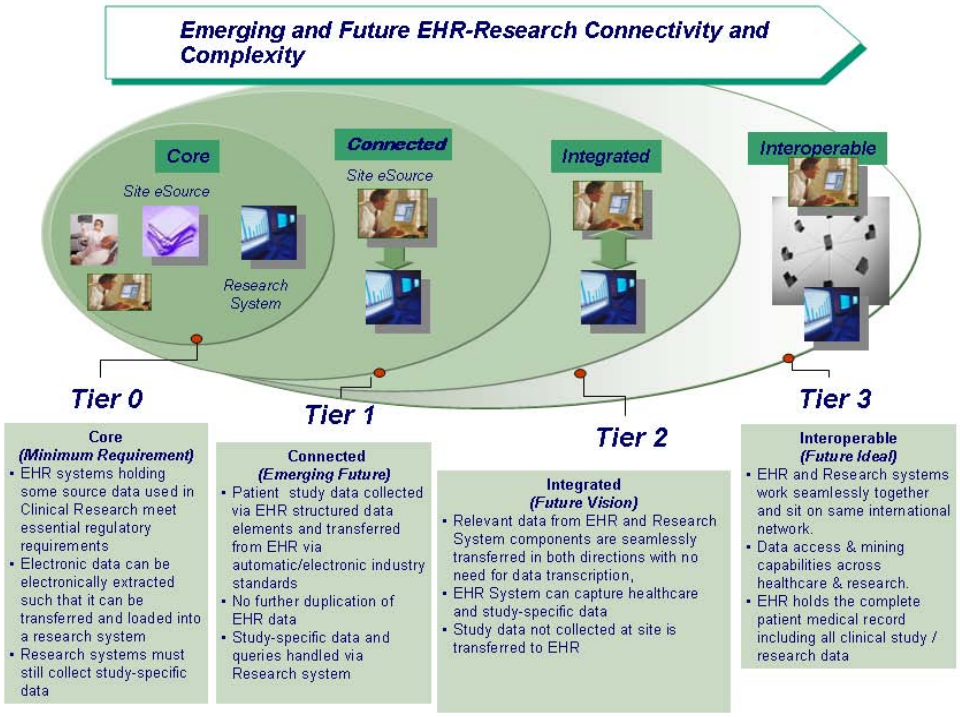

Figure 5 Vision of emerging EHR connectivity by EuroRec[33]

The eClinical Forum has produced a set of user requirements for EHR Systems providing source for clinical research[34], ensuring that data protection, patient privacy and regulatory research requirements are met.

[30] http://cabig.cancer.gov/
[31] http://www.sagebase.org/
[32] http://sagebase.org/commons/background.php
[33] www.eurorec.org
[34] "EHRCR User Requirements Document", eClinical Forum, 2010, www.eclinicalforum.org/main/ehrcrproject

Figure 5 shows EuroRec's integration roadmap, from core requirements for a reliable data source today toward future probable and ideal states requiring additional functionality.
CDISC has performed an analysis35 to investigate the use of electronic technology in the context of existing regulations for the collection of electronic source data (including that from eDiaries, EHR, EDC) in clinical trials for regulatory submission by leveraging the CDISC standards, in particular the Operational Data Model (ODM). In this analysis, twelve requirements for conducting regulated clinical research using electronic source (eSource) data collection in the context of existing regulations are formulated:

- An instrument used to capture source data shall ensure that the data are captured as specified within the protocol.
- Source data shall be Accurate, Legible, Contemporaneous, Original, Attributable, Complete and Consistent.
- An audit trail shall be maintained as part of the source documents for the original creation and subsequent modification of all source data.
- The storage of source documents shall provide for their ready retrieval.
- The investigator shall maintain the original source document or a certified copy.
- Source data shall only be modified with the knowledge or approval of the investigator.
- Source documents and data shall be protected from destruction.
- The source document shall allow for accurate copies to be made.
- Source documents shall be protected against unauthorized access.
- The sponsor shall not have exclusive control of a source document.
- The location of source documents and the associated source data shall be clearly identified at all points within the capture process.
- When source data are copied, the process used shall ensure that the copy is an exact copy preserving all of the data and metadata of the original

In addition, three potential scenarios are described which include the use of electronic health record systems (EHR), and associated benefits of standards.

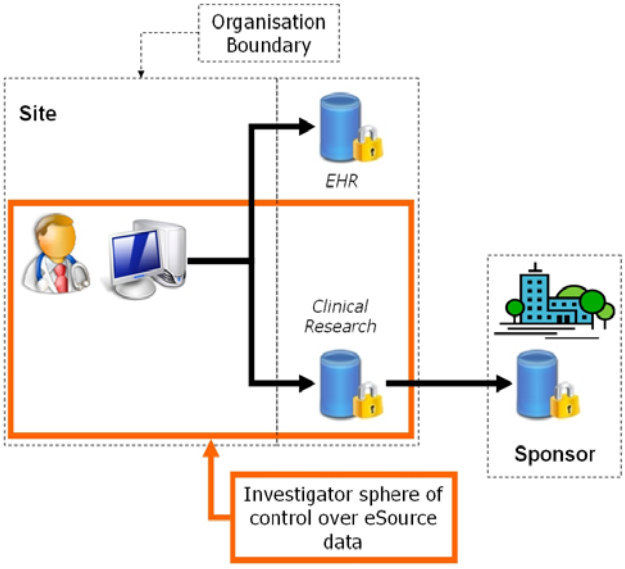

Figure 6 Single source concept

In the *Single Source Concept* scenario[36] (see Figure 6), data are entered into an electronic source document at the site (typically as an interface to the electronic health record (EHR) system but conceivably as an interface to an EDC system). All the eSource data can flow into the EHR database, while the clinical trial data (as identified by the protocol) can be simultaneously passed into eSource repository and passed onwards to the clinical trial database. (Typically, the EHR has been configured to collect data for a particular trial). CDISC concluded that this scenario is not the ideal future methodology to facilitate clinical research by investigators.

[35] http://www.cdisc.org/esdi-document
[36] As defined by CDISC.

However, it does offer a viable means for data to be entered just once for multiple purposes (research, patient care, safety surveillance, etc.) within the context of existing regulations.

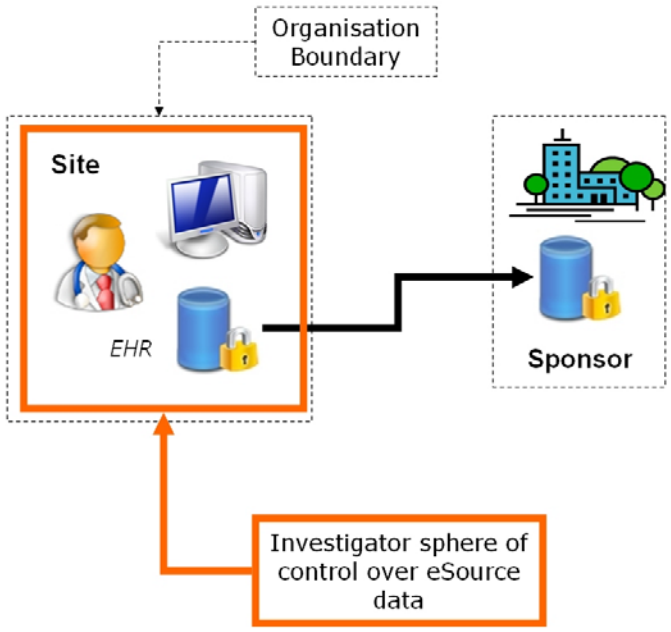

Figure 7 Direct Extraction from Electronic Health Records

In the *Direct Extraction from Electronic Health Records* scenario (see Figure 7), extraction of EHR data for clinical research is described. Unfortunately, direct extraction of EHR data is only compliant with the current regulations if the complete EHR can be validated in compliance with FDA's 21 CFR part 11. This typically puts too much strain on the care processes to be a viable option.

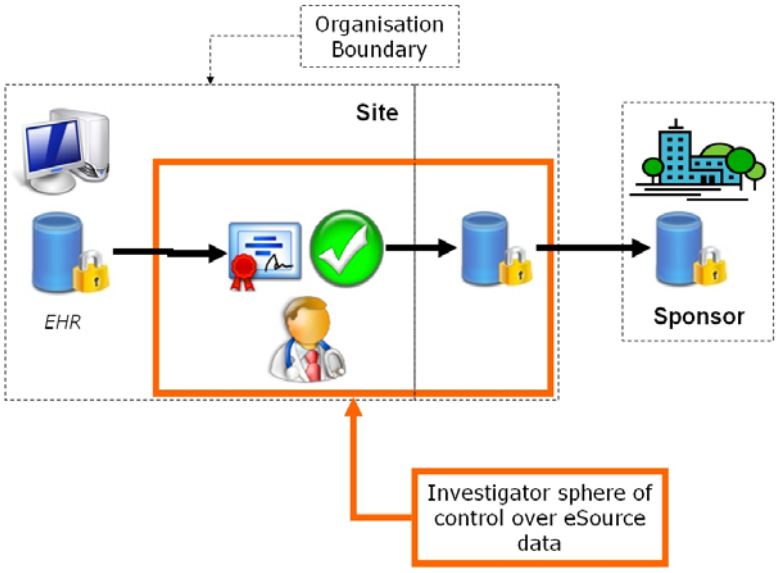

Figure 8 Extraction and Investigator Verification

To alleviate this problem an additional process step is introduced in the *Extraction and Investigator Verification (Electronic Health Records)* scenario (see Figure 8), where the investigator verifies that the extracted data, for clinical research use, accurately reflect the source data for that subject before it is included as part of the clinical trial data record.

The *Extraction and Investigator Verification* scenario has been demonstrated in a pilot[37].
A considerable amount of data on the eCRF was covered by data that had to be entered for standard care. For screening only 52% of the data items on the eCRF were not covered, and for chemotherapy only 31%. In addition a huge time saving was reported for personnel engaged in the trial (compared to a paper-based trial). The time spent per week went down from 375 minutes in the paper based scenario to 98 minutes in the integrated EHR-EDC scenario.

A Survey on Clinical Trial Barriers

A survey of almost 6,000 people with cancer conducted in 2000 took a look at why so few adults participate in cancer clinical trials. Some of the highlights included:

- About 85 percent of people with cancer were either unaware or unsure that participation in clinical trials was an option, though about 75 percent of these people said they would have been willing to enrol had they known it was possible.
- Of those who were aware of the clinical trial option, most declined to participate because they believed common myths about clinical trials.

They either thought that:

- The medical treatment they would receive in a clinical trial would be less effective than standard care.
- They might get a placebo.
- They would be treated like a "guinea pig.
- Their insurance company would not cover costs.

People who received treatment through a clinical trial found it to be a very positive experience:

- Ninety-seven percent said they were treated with dignity and respect and that the quality of care they received was "excellent" or "good."
- Eighty-six percent said their treatment was covered by insurance.

CURRENT CHALLENGES IN CLINICAL TRIAL PATIENT RECRUITMENT AND ENROLMENT
Genevieve Frank,Clinical Research Associate -ICON
http://www.socra.org/pdf/200402_Current_Challenges_Recruitment_Enrolment.pdf

Open source EHR's

For institutions not yet in possession of an EHR, an open source EHR is also a possibility. A variety of open source EHR's can be found online[38,39], in various states of development. It should be carefully assessed whether the EHR provides the required functional components (given the patient population, types of clinical care, and research to be supported by the EHR). In addition, it should be assessed whether the open source community has enough momentum to provide the required support and sustainability.
The United States Department of Veterans Affairs have constructed an enterprise-wise electronic health record, the *Veterans Health Information Systems and Technology Architecture* (VistA). Due to the uptake in VA hospitals, VistA is one of the most widely used EHR's (though predominantly deployed in the USA). The VistA system is public domain, and various derivatives exists, such as openVistA[40] and WorldVistA[41]. VA's EHR solution is also the model for Indian Health Service's RPMS EHR, having a high potential for wide adoption in emerging markets.
Another open source example is PatientOS[42], which is an EHR for small hospitals and clinics. Scheduling, Orders, Meds, Pharmacy, Clinical Doc, HL7, Billing are among the supported functional components. In addition, commercial support is available. Finally, OpenEMR[43] is an open source electronic medical record and medical practice management software. OpenEMR includes electronic health records, scheduling software, insurance billing, accounting, fine-grained access controls, with commercial support possible.
If the need arises in the EURECA project, the use of one of the open source EHR's in the test bed can be assessed.

1.2.3. Improve patient recruitment and trial search

In Europe, a number of new regulations, directives and corresponding guidelines have been issued over the last decade. In particular, the EU Directive 2001/20/EC on Implementing Good Clinical Practice in the conduct of clinical trials on medicinal products for human use has dramatically affected the regulation, oversight and practice of clinical research in Europe. The intention of the Directive was to protect patients and make the European pharmaceutical industry more competitive by ensuring that all Member States had the

[37] EHR and EDC Integration in Reality. Applied Clinical Trials, 2009
[38] http://en.wikipedia.org/wiki/List_of_open_source_healthcare_software
[39] http://www.medfloss.org/node/161
[40] http://medsphere.org/index.jspa
[41] http://worldvista.org
[42] http://www.patientos.org/
[43] http://www.oemr.org/

same rules. The weight of the directive at least doubled the costs of clinical trials[44] and increased the administrative burden so much that many academic researchers could no longer perform such trials.

Also, transposition of the Directive into the legislature of Member States has led States to make additional changes, with often tougher requirements on trial designs. The variety of interpretations of the Directive by Member States has thus resulted in many countries having different rules, hence increasing rather than decreasing disharmonies between EU Member States[45]. As a consequence, Sweden, for instance, has seen a 25% decrease in academic clinical trials, Ireland 60% and Poland a staggering 90% reduction[46]. This is also an issue that needs to be quickly addressed.

A standard, computable representation of CTs is required to uniformly identify and recruit potential trial participants from multiple, geographically dispersed sites of care that are often small physician practices. The general need to uniformly identify individuals who are suitable candidates for trials is a pressing one. Trial recruitment accounts for about 30% to 40% of trial costs, only 1 of 20 patients approached for potential enrolment is eventually enrolled, and a majority of trials have recruitment delays[47]. Numerous publications have pointed out the difficulties of identifying and recruiting trial participants[48,49,50] .

One very promising solution to the problem is to *identify such persons through information from electronic medical records systems created as part of the medical care process*[51]. However, to do so, trial eligibility criteria and other relevant aspects must be expressed in standard terms that can facilitate automated searching for such patients within a system of electronic records.

1.2.4. Improve patient recruitment through linkage to the EHR systems

While it is considered that improved patient enrolment represents one of the largest opportunities to eliminate delays in clinical trials, the recruitment rate of patients in adult oncology is very low, estimated to be below 5% of newly diagnosed cancer patients[52].
Several studies demonstrated that the integration of EHR and clinical trial systems has a large potential to increase recruitment rates, however there are important prerequisites to enable this application, such as provision of adequate patient data, availability of technical solutions and dual interoperability between the worlds of medical care and clinical research[53].
At the same time, access to the eligibility criteria is not sufficient as their evaluation is also influenced by the perception of the physicians and by their awareness of a suitable clinical trial at the point of care, when the treatment is decided. Low recruitment can also be a consequence of the physician not remembering the protocols, lacking time or not being able to access the appropriate information about the trial[54]. Therefore, it would be beneficial to be able to present information on open trials of potential relevance for an individual patient at the point of care.
Checking the eligibility criteria using computer-based screening represents a promising approach, as often the exclusion and inclusion criteria are very complex. However previous such implementations used separate data entry and were not integrated with the EHR systems[55].
Extracting data related to the patient eligibility from the EHR system is currently considered challenging, as the system needs to be integrated, comprehensive, and well structured. In general, in order to enable the use

[44] J. Hear, R. Sullivan. The impact of the 'Clinical Trials' directive on the cost and conduct of non-commercial cancer trials in the UK. European Journal of Cancer, vol. 43, pp. 8-13, 2007
[45] Editorial. Striking the right balance between privacy and public good. Lancet, vol. 367, pp. 275, 2006
[46] R. Hoey. The EU clinical trials directive: 3 years on. Lancet, vol. 369, pp. 1777-8, 2007
[47] Mcdonald AM, Knight RC, Campbell MK, et al. What influences recruitment to randomized controlled trials? A review of trials funded by two UK funding agencies. Trials 2006;7:9
[48] Warren JM, Golley RK, Collins CE, et al. Randomised controlled trials in overweight children: practicalities and realities. Int J Pediatr Obes 2007;2:73– 85
[49] Kingry C, Bastien A, Booth G, et al. ACCORD study group recruitment strategies in the Action to Control Cardiovascular Risk in Diabetes (ACCORD) trial. Am J Cardiol 2007;99:68i–79i.
[50] Fransen GA, van Marrewijk CJ, Mujakovic S, et al. Pragmatic trials in primary care. Methodological challenges and solutions demonstrated by the DIAMOND-study. BMC Med Res Methodol 2007;7:16
[51] Mcdonald AM, Knight RC, Campbell MK, et al. What influences recruitment to randomized controlled trials? A review of trials funded by two UK funding agencies. Trials 2006;7:9
[52] Rate limiting factors in recruitment of patients to clinical trials in cancer research: descriptive study, P. Corrie et al., BMJ, 2003
[53] Meeting the Challenges of Patient Recruitment. A role for Electronic Health Records, C. Ohmann and W. Kuchinke, Int. J. Pharm. Med., 2007
[54] Clinical trial accrual among new cancer patients at a community-based cancer center, R.S. Go et al., Cancer, 2006.
[55] Selection of patients for clinical trials: an interactive web-based system, E. Fink et al., Artif. Intel. Med,, 2004

of EHR data for research issues related to the reliability and completeness of databases, variability in medical practice, pervasiveness of unstructured text, lack of specificity of patient data, and lack of support to search across many records and data sources need to be resolved. Additionally, the solutions need to comply with privacy, security and legal requirements.

1.2.5. Improve patient recruitment through a clinical alert system

Physician participation is critical to the successes of most clinical trial recruitment efforts. Not only do clinicians play a vital role by identifying potentially eligible subjects, but patients are much more likely to participate in a study if their physician has suggested it to them[56].

Unfortunately, the identification and recruitment of eligible patients during the course of busy clinical practice can be difficult. In order to successfully recruit patients, physicians engaging in traditional recruitment have to remember which local clinical trials are active, recall the trial details in order to determine patient eligibility, take time to explain the trial details to potentially eligible patients, and often take more time to perform other recruitment activities. Doing all of this while also attempting to provide the individual patient with good care during a short clinic visit can be difficult, at best. Current privacy regulations add further challenges to overcome in solving this problem.

The Clinical Trial Alert Approach

Efforts have been reported in the literature to develop an EHR-based Clinical Trial Alert (CTA) approach. The approach was designed to overcome many of the known obstacles to trial recruitment by physicians while complying with current privacy regulations[57].

The reported results of the application of the CTA approach in a large, multi-center, NIH sponsored type 2 diabetes mellitus clinical trial, the Action to Control Cardiovascular Risk in Diabetes (ACCORD) study were very encouraging, as shown below in Figure 3.

Parameter	**Clinical Trial Alert System**	
	Before	**After**
no. of referrals ***(per month)***	5.7	59.5
no. of enrolments ***(per month)***	2.9	6.0

Figure 9 Impact on patient recruitment by a clinical trial alert system[58]

1.2.6. Safety reporting in clinical trials

Collected adverse events and effects are reported by investigators through different forms and with different delays:

- Paper and/or electronic flow in case of occurrence of non-serious criteria included in CRF and eCRF and collected during monitoring visit. Reporting is done at the end of the study
- Paper and/or electronic flow in case of occurrence of serious criteria on the base of an expedited report. Reporting is carried out through an alert form within a short delay (24-48h).

Sponsors send different information to the different protagonists. For example, in Belgium the following messages are sent:

- SUSARs are sent as expedited reports to the principal Ethical Committee which has to forward its analysis to the local Ethical Committee
- Investigators receive as expedited reports all worldwide SAE

[56] Siminoff LA, Zhang A, Colabianchi N, Sturm CM, Shen Q: Factors that predict the referral of breast cancer patients onto clinical trials by their surgeons and medical oncologists. J Clin Oncol 2000, 18(6):1203-1211

[57] P.J. Embi et al., Development of an electronic health record-based Clinical Trial Alert system to enhance recruitment at the point of care. AMIA Annu Symp Proc 2005:231-235.

[58] P.J. Embi et al., Arch Intern Med 2005;165:2272-2277

- The periodic line listings (quarterly, semester) aggregate all the summarized data for each study
- The annual report aggregates all the data regarding a project including all the concerned studies.

Different formats and delays are applied in different situations with a common goal: To generate alerts and useful information to protect the patients. The constantly problematical issue is also the reconciliation between collected data, sent data, received data through different ways and formats, including compliance in case of delay of transmission and delay of submission. In EURECA we do not have to define new standards or new safety databases but we have to set up a system to identify the well exchanged information and perform an easier and secure reconciliation process. The usefulness of the completed system would be to improve the data flow between the different protagonists to improve the safety of the drugs for the security of the patients.

In clinical trials a Clinical Trial Management System manages the collection of clinical, biomedical, imaging and other trial specific and relevant data. In ACGT, a Clinical Trial Management system has been developed named 'Ontology based Trial Management System of ACGT' (ObTiMA) and consists of the ACGT Trial Builder – composed of a Trial Outline Builder (TOB) and the CRF Creator (CC) -, the Clinical Data Management System and a Repository. ObTiMA can be used in two different ways. With the help of ObTiMA a new trial can be build up with all its components and ObTiMA can be used as a Remote Data Entry (RDE) system for managing a patient in a clinical trial including data entry and regarding all administrative aspects for patients enrolled in a trial.

The ACGT Trial Builder supports the concept of reusable components. There are basic modules proposed such as a module for patient specific data (age, gender, date of diagnosis), a module for histology, etc., besides trial specific modules.

In EURECA, a SAE/SUSAR reporting module will be integrated in the context of ObTiMA. The module will use the Common Terminology Criteria for Adverse Events (CTCAE v3.0)[59] of the NCI, which is a descriptive terminology. CTCAE v3.0 grading is by definition a 5-point scale generally corresponding to mild, moderate, severe, life-threatening and death. This grading system inherently places a value on the importance of an event, although there is not necessarily proportionality among grades (a Grade 2 is not necessarily twice as bad as a Grade 1). Some adverse events (AEs) are difficult to fit into this schema, but altering the general guidelines of severity scaling would render the system useless for comparing results between trials, an important purpose of the system. Not all grades are appropriate for all AEs. A second way to classify SAEs and SUSARS will be developed according to the MedDRA classification system. The International Conference on Harmonization (ICH) develops requirements for drug regulatory reporting globally and has chosen MedDRA (Medical Dictionary for Regulatory Activities) terminology to be the standard. To facilitate data transfer to regulatory agencies, CTCAE v3.0 terms are mapped to the current MedDRA version. The mapping can be found on the CTEP (Cancer Therapy Evaluation Program) Homepage[60].

1.2.7. Improving interoperability between EHR and clinical research systems

In successfully addressing the above problems, one must first resolve a key issue. That is the current lack of interoperability (both syntactic and semantic) of the diverse systems that exist in the clinical care versus clinical research setting. The problem is schematically shown in Figure 6. There are significant differences between the clinical care and the research domains with respect to:

- targets
- regulations
- processes, actors, roles
- ontologies, classifications, terminologies
- different and distributed databases
- software solutions

[59] http://www.fda.gov/cder/cancer/toxicityframe.htm
[60] http://www.ctep.cancer.gov

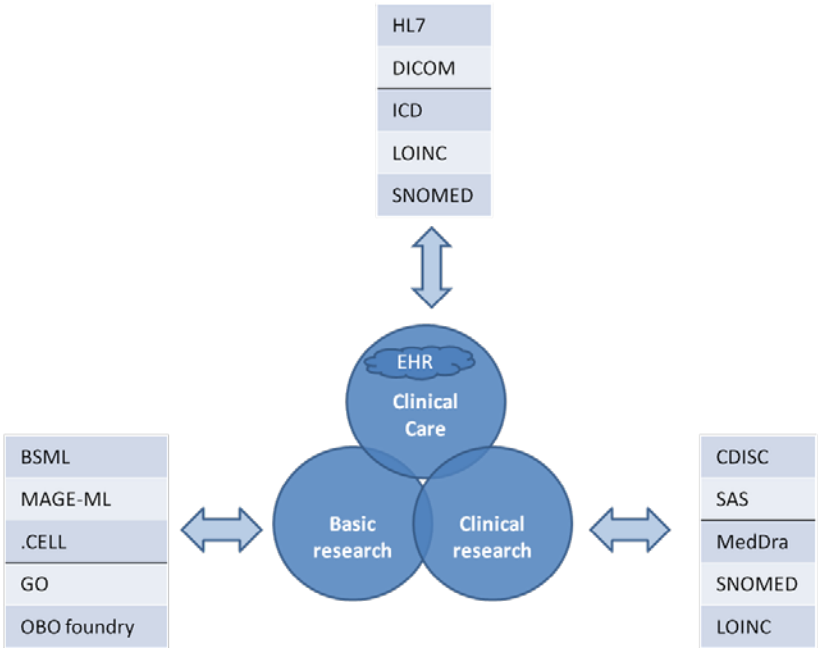

Figure 10 The different worlds and standards in clinical care, basic and clinical research

Also, a standard computable representation of CTs is a prerequisite for interoperability among clinical trial management systems, clinical information systems, and decision support systems[61]. The need for interoperability arises from the development of electronic clinical trials management and trial registration systems that require a structured representation of a clinical trial to accomplish their functions[62].

The issue of interoperability emerges when it is desired to move information from one system to another. For example, it is desirable to be able to register a trial in one system and be able to subsequently disseminate that trial across multiple registration systems. The originating system must be able to construct a message containing a description of a trial that is understood in a computational sense by the receiving system. This requires both syntactic interoperability in which the structure of the message itself is understood and semantic interoperability in which there is a commonality of meaning in the contents of the message structure.

Similarly, both semantic and syntactic interoperation are required for exchanging information among clinical trial managements systems, as may be needed when combining systems from several vendors to run a single trial, for exchanging information with electronic medical record systems, and with decision support systems that may be separate from the electronic medical record. The challenges of interoperation are heightened in practice-based research where electronic health record penetration is still relatively low and where the data exchange infrastructure is absent (e.g., regional health information organizations).

1.2.8. Medical vocabularies

In recent years, a lot of effort has been put into the development of vocabularies in order to assist in the semantic interoperability in the medical domain. Each of these vocabularies have been developed with a particular focus in mind.

The development and maintenance of vocabularies is a gigantic undertaking. Therefore, existing vocabularies will be assessed, from a soundness and completeness view, an applicability view, as well as expected uptake into practice. EURECA will take a pragmatic approach by leveraging on existing vocabularies and using those subsets that are relevant to the task at hand.

[61] Brandt CA, Sun K, Charpentier P, Nadkarni PM. Integration of web-based and PC-based clinical research databases. Methods Inf Med 2004;43:287–95

[62] Kuchinke W, Wiegelmann S, Verplancke P, Ohmann C. Extended cooperation in clinical studies through exchange of CDISC metadata between different study software solutions. Methods Inf Med 2006;45:441– 6

MedDRA[63], for example, focuses on the regulatory process of drug development and is a medical vocabulary that is used by regulatory bodies and the regulated biopharmaceutical industry for data entry, retrieval, evaluation and display. Therefore, medDRA is used in clinical trials for reporting adverse events.

SNOMED CT (Systematized Nomenclature of Medicine-Clinical Terms)[64], is a clinical vocabulary focussed on accurately recording health care encounters and the associated electronic health information exchange. Although SNOMED CT is sometimes criticized, its significant uptake in clinical practice, such as its use in HL7 messaging, cannot be ignored.

MeSH (Medical Subject Headings)[65] is a vocabulary maintained by the National Library of Medicine. It is used for indexing biomedical articles for the MEDLINE/PubMED databases and thus can be used to access scientific medical literature.

LOINC: The purpose of Logical Observation Identifiers Names and Codes (LOINC)[66] is to facilitate the exchange and pooling of results for clinical care, outcomes management, and research. LOINC is a database and a universal standard, LOINC codes are universal identifiers for laboratory and other clinical observations. The *laboratory* portion of the LOINC database contains categories of chemistry, hematology, serology, microbiology (including parasitology and virology), toxicology; as well as categories for drugs and the cell counts, antibiotic susceptibilities, and more. The *clinical* portion of the LOINC database includes entries for vital signs, hemodynamics, intake/output, EKG, obstetric ultrasound, cardiac echo, urologic imaging, gastroendoscopic procedures, pulmonary ventilator management, selected survey instruments (e.g. Glascow Coma Score, PHQ-9 depression scale, CMS-required patient assessment instruments), and other clinical observations. LOINC has been translated into various languages (and dialects), e.g. English, Chinese, Portuguese, Estonian, French, German, Greek, Italian, Korean, Spanish, etc. LOINC is one of the standards for use in U.S. Federal Government systems for the electronic exchange of clinical health information. It is a preferred code set for HL7 for laboratory test names in transactions between health care facilities, laboratories, laboratory testing devices, and public health authorities.

ICD: The International Classification of Diseases (ICD)[67] (see footnote 67) is the international standard diagnostic classification for all general epidemiological, many health management purposes and clinical use. These include the analysis of the general health situation of population groups and monitoring of the incidence and prevalence of diseases and other health problems in relation to other variables such as the characteristics and circumstances of the individuals affected, reimbursement, resource allocation, quality and guidelines. It is used to classify diseases and other health problems recorded on many types of health and vital records including death certificates and health records. In addition to enabling the storage and retrieval of diagnostic information for clinical, epidemiological and quality purposes, these records also provide the basis for the compilation of national mortality and morbidity statistics by WHO Member States.

ICD-O: The International Classification of Diseases for Oncology (ICD-O)[68] is a disease specific extension of ICD. ICD-O is used principally in tumour or cancer registries for coding the site (topography) and the histology (morphology) of neoplasms, usually obtained from a pathology report. It supports a multi-axial classification of the site, morphology, behaviour, and grading of neoplasms.

Terminology servers provide interfaces for querying and accessing terminologies. By providing a terminology service, consistent access to terminology resources can be provided across organizations. HL7 has specified the (minimum set of) common functionality (by means of an API) that an external terminology must provide for use with HL7 in the HL7 *Common Terminology Services* specification. This decouples the terminology from the terminology service.

The following thematic areas are considered in scope for CTS version 2[69].

- ***Administration***. This is a set of functionality that provides the ability to manage content as part of a terminology service. Administration functions include the ability to load terminologies, export terminologies, activate terminologies, and retire terminologies. These functions are generally protected and accessible by service administrators with appropriate authorization.
- ***Search / Query.*** This is a set of functionality that provides the ability to find concepts based on some search criteria. This includes restrictions to specific associations or other attributes of the terminology, including navigation of associations for result sets. This represents the primary utility for using terminology content in a number of application contexts.

[63] http://www.meddramsso.com
[64] http://www.ihtsdo.org/SNOMED CT/
[65] http://www.nlm.nih.gov/mesh/
[66] http://loinc.org/background
[67] http://www.who.int/classifications/icd/en/
[68] http://www.who.int/classifications/icd/adaptations/oncology/en/
[69] http://www.lexgrid.org/LexGrid/downloads/CTS/cts2/HL7_Common_Terminology_Services_2_Service_Functional_Model_%28SFM%29.htm

- ***Authoring / Maintenance.*** This is a set of functionality that provides the ability to create and maintain content. From a terminology service perspective, this would include the appropriate APIs to add, change, or delete concepts and associations. This would also include the processing of change events from various terminology providers.
- ***Associations.*** This is a set of functionality that provides the ability to map concepts and the concept's associated attributes from a source terminology to a concept in a target terminology, or create relationships between concepts within a single code system.

caBIG has developed LexGrid[70] (Lexical Grid), providing support for a distributed network of lexical resources such as terminologies and ontologies via standards-based tools, storage formats, and access/update mechanisms.

1.2.9. Standards

CDISC (The Clinical Data Interchange Standards Consortium)[71] focuses on clinical research and is an organisation that aims to develop and support global, platform-independent data standards that enable information system interoperability to improve medical research and related areas of healthcare. The FDA encourages the usage of CDISC standards for (clinical trial) result reporting.
Currently, CDISC is the leading standards development organization for the medical research domain. Its mission is to develop industry standards to support acquisition, exchange, submission and archiving of clinical trials data for medical and biopharmaceutical product development. The various standards are based on the CDISC Operational Data Model (ODM) standards. The standards wildly vary in maturity and uptake into practice. According a communication (Feb 2010) from CDISC, the FDA's Center for Drug Evaluation and Research (CDER) and Center for Biologics Evaluation and Research (CBER) give the following advice:

- Use CDISC Standards (SDTM and ADaM) NOW in eSubmissions to CDER.
- If you want high quality eSubmissions (which will be reviewed better and more quickly), start with CDISC CDASH for your case report forms when you plan your study.
- The expected transport format for the foreseeable future is SASXPT with Define.xml.

In order to facilitate an efficient submission of trial results to regulatory bodies, CDISC has defined the Study Data Tabulation Model (SDTM). The **SDTM** is a general framework describing the organization of the information that is collected during clinical trials. The SDTM consists of a set of clinical data file specifications and underlying guidelines. These different file structures are referred to as domains. Each domain describes a type of data associated with clinical trials, such as demographics, vital signs or adverse events. CDISC also provides a standard for running a clinical trial, namely the Operational Data Modelling (ODM) standard. ODM supports interchange between applications used in collecting, managing, analysing and archiving data. ODM provides a format for representing study metadata, study data and administrative data associated with a clinical trial. Various clinical trial management systems like SAS[72], Oracle clinical[73] and ObTiMA[74] support the ODM standard for data exchange.
ODM provides a format for representing study metadata, study data and administrative data associated with a clinical trial. It represents the data that would be transferred between different software systems during a trial, or archived after a trial. There are two types of ODM files: snapshots files and transactional files. A snapshot file is completely self-contained, containing the current state of the source database. A transactional file shows, for each included entity, both the latest state of the source database, and (optionally) some prior states of that entity in the source database. The transactional files can provide an audit trail.
Clinical Data Acquisition Standards Harmonization (**CDASH**) is focused on data collection (as opposed to data reporting). The CDASH project identifies the basic data collection fields needed from a clinical, scientific and regulatory perspective to enable more efficient data collection at the investigative sites. The CDASH data collection fields (or variables) can be mapped to the SDTM structure. When the data are identical between the two standards, the SDTMIG variable names are presented in the CDASH domain tables. In cases where the data are not identical, CDASH has suggested new variable names. The CDASH recommendation also includes some data collection fields that are not included in the SDTMIG, these collection fields are intended to assist in the cleaning of data and in confirming that no data are missing. In some instances the optimal data collection method conflicts with the SDTM for reporting data, in these cases additional transformations and derivations may be needed to create the final SDTM compliant datasets.

[70] https://cabig-kc.nci.nih.gov/Vocab/KC/index.php/LexGrid
[71] http://www.cdisc.org/.
[72] http://www.sas.com/industry/pharma/cdisc
[73] http://www.oracle.com/industries/life_sciences/oracle-clinical.html
[74] http://www.obtima.net

Terminology applicable to CDASH data collection fields is either in production or under development by the CDISC Terminology Team. Production terminology is published by the National Cancer Institute's Enterprise Vocabulary Services (NCI EVS). The code list is referenced in the Definition column in the domain tables when a CDASH field has associated controlled terminology.
The main benefit of CDASH is standardizing the definition for the data that is collected over multiple studies, reducing the production time for CRF design and training time. CDASH encapsulates related variables is domains. The following domains are defined:

- Adverse Event (AE) - Events. For non-solicited or pre-specified adverse events.
- Comments (CO) – Special Purpose. Solicited comments are defined as those entered in free-text data collection fields (such as "Specify" or "Please comment") intentionally included on the CRFs. Unsolicited comments are those comments entered outside of pre-defined data collection fields.
- Prior and Concomitant Medications (CM) - Interventions. The same basic data collection variables should be collected for all medications (Prior, General Concomitant Medications and Medications of Interest). It is assumed that additional fields will be added as applicable for each specific Medication of Interest.
- Demographics (DM) – Special Purpose.
- Disposition(DS) - Events. Both disposition events and protocol milestones.
- Drug Accountability (DA) - findings. The Drug Accountability variables are sometimes used to calculate the subject's compliance with the study treatment, however, in most study designs and depending on the drug under study, this may not provide the most accurate information as medication that is not returned may not necessarily have been consumed by the subject, thereby giving a false estimate of compliance. In addition, the SDTMIG standard separates drug accountability from compliance and treats each differently.
- ECG Test Results (EG)- Findings.
- Exposure(EX) - Interventions. The Exposure domain model records the details of a subject's exposure to protocol-specified study treatment. Study treatment may be any intervention that is prospectively defined as a test material within a study, and is typically but not always supplied to the subject. Examples include but are not limited to placebo, active comparator, and study treatment. Treatments that are not protocol-specified should be recorded in the Concomitant Medications (CM) domain.
- Inclusion / Exclusion Criteria Not Met (IE) - Findings.
- Laboratory Test Results (LB) - Findings.
- Medical History (MH) - Events. Coding using MedDRA, though not required by the FDA, is recommended. Coding of Medical History is recommended because it provides methodology to match Medical History to specific adverse events, makes it easier tomine data for potential relationship to past treatments and to the safety profile, as well as a providing a means to help identify unexpected safety concerns.
- Physical Examination (PE) - Findings. The scope of the PE domain tables is limited to general physical examinations as part of overall safety data collection.
- Protocol Deviations (DV) - Events. The CDASH data collection standard for Protocol Deviations maps to the SDTMIG DV domain. However, the DV Domain Team recommends avoiding the creation of a Protocol Deviations CRF (individual sponsors can determine whether it is needed for their particular company). If a sponsor decides to use a Protocol Deviations CRF, the DV Domain Team felt the sponsor should not rely on this CRF as the only source of protocol deviation information for a study. Rather, they should also utilize monitoring, data review and programming tools to assess whether there were protocol deviations in the study that may affect the usefulness of the datasets for analysis of efficacy and safety. By utilizing this information a sponsor can then decide which method is best for their company.
- Subject Characteristics (SC) - findings. Subject Characteristics is for data not collected in other domains that is subject-related, and collected only once per subject.Substance Use (SU) - Interventions. E.g. drug use, cigarettes, caffeine, alcohol.
- Vital Signs (VS) - Findings.

Each domain lists all of the common variables for that domain along with a definition and implementation instructions. Each variable is either Highly Recommended (a data collection field that should be on the CRF, e.g. due to a regulatory requirements), Recommended/Conditional (a data collection field that should be collected on the CRF for specific cases or to address TA requirements (may be recorded elsewhere in the

CRF or from other data collection sources) or Optional (a data collection field that is available for use if needed).
CDISC is currently working on the **SHARE**[75] project to build an electronic repository of metadata that can be easily and readily consumed by organizations and their computer applications, where the NCI is providing the technology base. CDISC SHARE is a global, accessible, electronic library, which through advanced technology, enables precise and standardized data element definitions that can be used within applications and across studies to improve biomedical research and its link with healthcare. In the first instance, CDISC SHARE will contain the existing CDISC standards, such as CDASH and SDTM, providing machine readable element (variables) within those standards. This will allow a range of applications used within organisations to automatically access those definitions. However, the environment will also allow for the development of such definitions, in particular new therapeutic area domains along with the associated terminology. These new domains will also be available by the same electronic access mechanisms. CDISC intends to partner with other organizations to ensure that the SHARE content supports the vision of 'enter once for multiple purposes', including biomedical research and related areas such as public health, quality and safety. However, currently there is not much material available for review or use.
HL7 (Health Level 7)[76], on the other hand, focuses on standardization for the clinical and administrative domain. HL7 provides standards for interoperability that improve care delivery, optimize workflow, reduce ambiguity and enhance knowledge transfer among all of our stakeholders, including healthcare providers, government agencies, the vendor community, fellow SDOs and patients. With version 3, HL7 moves away from merely a syntax specification and uses object oriented principles to provide semantic interoperability. It provides a standard clinical model and common terminologies. Often, already existing terminologies (e.g. SNOMED CT) are referenced.
In order to achieve complete technical and semantic interoperability, existing standards have to be harmonized and bridged. In the US major consortia have been formed to provide semantic interoperability between the standards (e.g. BRIDG covering CDISC, HL7 (RCRIM), FDA, NCI) and to provide core sets of data collection fields (CDISC CDASH). Furthermore, efforts exploring the potential to improve interoperability between the electronic health record and electronic data capture (e.g. CDISC eSDI, eClinical Forum/PhRMA EDC/eSource Taskforce) have been initiated.
BRIDG[77] is a domain analysis model representing protocol-driven biomedical/clinical research. It was developed to provide an overarching model that could readily be understood by domain experts and would provide the basis for harmonization among standards within the clinical research domain and between biomedical/clinical research and healthcare. The model emerged from a collaborative effort among clinical trial experts from CDISC, the US National Institutes of Health (NIH)/National Cancer Institute (NCI), the Food and Drug Administration (FDA), Health Level Seven (HL7), and other volunteers. The modelling effort is using the HL7 Development Framework (HDF). This structured information model is being used to support development of data interchange standards and technology solutions that will enable harmonization between the biomedical/clinical research and healthcare arenas. The BRIDG Model serves to bridge standards, as well as organizations and various communities, including academic research institutions and pharmaceutical product development organizations and related service and technology providers.
openEHR is an open, detailed, and tested specification for a comprehensive interoperable health information computing platform for the EHR. It is based on a two level methodology concerning the development of information systems that separates the semantics of information and knowledge into two levels. The Reference Model corresponds to the information level and consists of a relatively small number of non-volatile abstract concepts. At the knowledge level models defining domain concepts by expressing constraints on instances of the underlying Reference Model are built, called archetypes. This approach may enable the building of future-proof information systems, avoiding the common problems of single-layer systems such as complexity, large numbers of concepts, high rate of change and high maintenance costs. Additionally, by separating the technology from the domain-specific part of the system, the archetype-based approach may bring the benefit of empowering the domain-specialists (e.g. clinicians) to define models that closely suit their requirements[78]. The CEN 13606 standard is a subset of the full openEHR specification.
EURECA aims to technically enforce and govern the legal framework through policy based authorization services. Implementation of data and resource access control through authorization services in Service Oriented Architectures is not uncommon. The approach has some obvious advantages, such as simplification of access management (leading to reduced cost and more security). EURECA is innovative in extending authorization services to enforce the complex data protection policies which originate from the legal

[75] http://www.cdisc.org/cdisc-share
[76] http://www.hl7.org/
[77] http://www.cdisc.org/bridg
[78] Archetypes: Constraint-based Domain Models for Future-proof Information Systems, T. Beale, http://www.deepthought.com.au

implications connected to crossing the domain of care and clinical trials. This includes incorporating the "notion of datasets" and data bound concepts such as "purpose of use" and "conditions on use" (privacy-metadata, sticky policies[79]) and dealing with cross-enterprise policy interoperability (cf. recent work on XSPA[80]).
Furthermore, EURECA aims to incorporate automated consent management into the framework. A number of patient consent tools exist[81], but they mainly deal with consent in a care setting only and focus typically on achieving HIPAA compliance (US legislation). Many solutions are ad-hoc, but attention is turning towards more generic solutions[82], such as through the "Basic Patient Privacy Consents (BPPC)" profile[83] on which IHE is working and the HL7 efforts for establishing a consent related vocabulary[84]. EURECA will research consent in a European perspective, covering both care and research (trials). Part of the innovation lays in researching novel scenario's like automated patient recruitment and unsolved issues such as complex individual consent (i.e. leading the way to patient-empowerment).

Relevant IHE profiles
IHE[85] is an initiative by healthcare professionals and industry to improve the way computer systems in healthcare share information. IHE promotes the coordinated use of established standards such as DICOM and HL7 to address specific clinical need in support of optimal patient care. Systems developed in accordance with IHE communicate with one another better, are easier to implement, and enable care providers to use information more effectively.
IHE has specified a *Quality, Research and Public Health* domain which contains the *Clinical Research Document (CRD)* profile[86] (currently proposed for trial implementation). CRD leverages the *Retrieve Form for Data Capture (RFD)* profile[87], which provides a method for gathering data within a user's current application to meet the requirements of an external system. The profile specifies retrieval of forms from a form source, display and completion of a form, and return of instance data from the display application to the source application, and form archiving. CRD leverages RFD (amongst others) to select or generate forms based on context (e.g. location in the workflow) and to pre-populate forms.

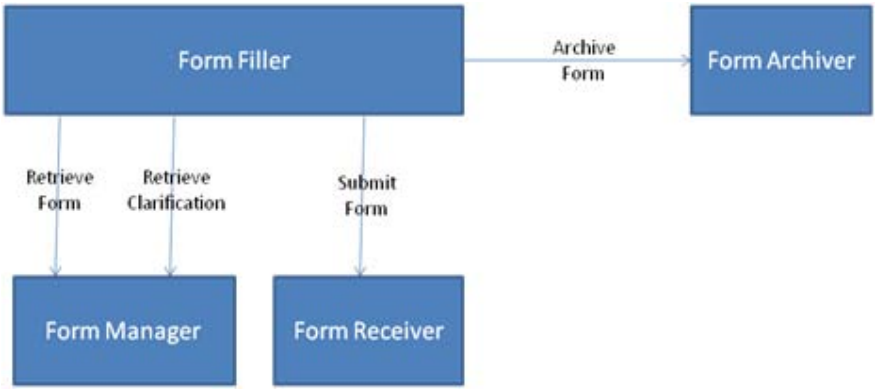

Figure 11 RFD Actor Diagram

Figure 11 shows the actors and the possible transaction defined in RFD. The Form Fillter can retrieve forms from the Form Manager. To retrieve a form, a specific form can be requested (by means of an id) or workflow information can be provided which the Form Manager can use to select or generate a form. In addition, context data can be sent to the Form Manager, which allows the Form Manager to pre-population the selected form (partially). The Form Filler can retrieve clarifications (specific for the organization of the Form

[79] Chadwick D, Lievens S. Enforcing "Sticky" Security Policies throughout a Distributed Application. http://www.cs.kuleuven.be/conference/MidSec2008/sticky.pdf.
[80] OASIS Cross-Enterprise Security and Privacy Authorization (XSPA). http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=xspa
[81] http://dialogmedical.com
[82] Privacy Must Take Precedence. Psychiatr News April 18, 2008. Volume 43, Number 8, page 7. http://pn.psychiatryonline.org/cgi/content/full/43/8/7-a
[83] Basic Patient Privacy Consents - http://wiki.ihe.net/index.php?title=BPPC
[84] HL7 Consent Related Vocabulary ConfidentialityCodes Recommendation, http://lists.oasisopen.org/archives/xacml-demotech/200712/doc00003.doc
[85] http://www.ihe.net/
[86] http://www.ihe.net/Technical_Framework/upload/IHE_QRPH_Suppl_CRD_Rev2-2_TI_2010-08-30.pdf
[87] http://www.ihe.net/Technical_Framework/upload/IHE_ITI_Suppl_RFD_Rev2-1_TI_2010-08-10.pdf

Filler) for a specific form. The Form Receiver receives completed (or partially completed) forms and the Form Archiver stores the completed (or partially completed) forms for archival purposes.

1.2.10. Security, Privacy and Consent Management

EURECA will ensure that its solution is in line with governing legislation, industry regulations and best practice (ethical), and will therefore address several topics with respect to data protection (security & privacy).

Identity Management (IM)

EURECA will base its core security solutions as much as possible on existing standards and implementations, integrating them into the EURECA architecture. The state-of-the-art with respect to identity management and authentication is largely defined by a number of (open) implementations and a set of protocols and open standards, including: the OASIS Security Assertion Markup Language (SAML), the Liberty Alliance standards (mainly ID-FF Liberty Identity Federation and ID-WSF Liberty Identity Web Services), the WS-* security stack (WS-Security, WS-Trust and WS-Federation). Relevant (open source) implementations include: Shibboleth[88] project, OpenAM (formerly OpenSSO), JA-SIG Central Authentication Service (CAS),ZXID, etc.

Policy Based Authorisation, Consent and Audit

EURECA aims to technically enforce the legal framework through policy based authorization services. Implementation of data and resource access control through authorization services in Service Oriented Architectures is not uncommon. This approach has some obvious advantages, such as simplification of access management (leading to reduced cost and more security). EURECA will innovate by extending authorization services to enforce the complex data protection policies which originate from the legal implications connected to crossing the domain of care and clinical trials. This will include making systems aware of “logical datasets” ,introducing data bound concepts such as “purpose of use” and “conditions on use” (privacy-metadata, sticky policies[89]) and dealing with cross-enterprise policy interoperability (cf. recent work on XSPA[90]).

Furthermore, EURECA aims to incorporate automated consent management into the overall security framework. Very little solutions are available and those that exist are often proprietary. However standardization initiatives exist[91]: for example the “Basic Patient Privacy Consents (BPPC)” profile[92] of IHE is working and the HL-7 Security Domain analysis Model (DAM)[93].

Current auditing mechanisms typically log individual events in isolation per application or computer system. In practice, this makes it very hard to use audit logs for monitoring purposes (detection of security breaches) or to reconstruct the logical chain of events after an incident involving multiple systems (which may use different coding schemes, logging formats and identifiers). Furthermore, in an environment mainly concerned with handling personal data (such as EURECA), audit trails need to be readily accessible in a user-centric and data-centric way. Few generic solutions and standards exist with respect to audit logging and provenance in service oriented architectures (or distributed applications in general). State-of-the-art includes initiatives such as ISO standards 13606 Part 4 (Health informatics -- Electronic health record communication -- Part 4: Security), ISO 27789 (Audit trails for electronic health records, draft), IETF Syslog, IHE ATNA (Audit Trail and Node Authentication), the Open Provenance Model[94].

Much “state-of-the art” on the mentioned topics has been and is still being generated in the European projects MASTER[95] (which work towards embedding trust and security in service infrastructures) and TAS3[96] (Trusted Architecture for Securely Shared Services). The latter also involves EURECA partners.

Improving on common practice, the EURECA security architecture aims to provide a “unified” solution to the issues described above, offering regulatory compliance (privacy & security) “by design” to organizations building upon the EURECA Architecture.

De-identification

[88] http://shibboleth.internet2.edu/

[89] Chadwick D, Lievens S. Enforcing "Sticky" Security Policies throughout a Distributed Application. http://www.cs.kuleuven.be/conference/MidSec2008/sticky.pdf.

[90] OASIS Cross-Enterprise Security and Privacy Authorization (XSPA). http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=xspa

[91] Privacy Must Take Precedence. Psychiatr News April 18, 2008. Volume 43, Number 8, page 7. http://pn.psychiatryonline.org/cgi/content/full/43/8/7-a

[92] Basic Patient Privacy Consents - http://wiki.ihe.net/index.php?title=BPPC

[93] http://gforge.hl7.org/gf/project/security/docman/?subdir=137

[94] Open Provenance Model v1.1. http://openprovenance.org, soon to be published in Future Generation Computer Systems

[95] MASTER - Managing assurance, security and trust for services – http://www.master-fp7.eu/ - MASTER aims to provide methodologies and infrastructure that facilitate monitoring, enforcement, and auditing of security compliance,in service oriented environments.

[96] http://www.tas3.eu/

Dealing with individual patient records in a clinical trial and study environment requires a range of dedicated data protection services in order to address data privacy requirements. One measure is to work with anonymised data wherever possible. However, true de-identification of individual person records is a hard problem. The state-of-the-art[97] related to de-identification, re-identification and anonymity calculation[98,99,100,101,102] shows that there are no fully satisfying practical solutions available.
The problem escalates in application domains where individual line data is required (such as the clinical research domain), where it is completely impossible to achieve “provable security”. The ACGT project has proposed a solution where inherent issues that remain with de-identified microdata are covered by the overall governance and security framework (concept of “de-facto anonymous data”)[103]. It has also produced a tool for data de-identification focussing on practical aspects of de-identification (not covering e.g. automatic privacy risk estimation). The tool, CAT (Custodix Anonymisation Tool), is rather a modular framework which can be adjusted to the particular needs of a project.
EURECA will build on this expertise and on the results produced by the ACGT project with respect to data protection.
Finally, partners involved in EURECA have been contributing to standards in the area of de-identification and pseudonymisation (ISO TS 25237).

1.2.11. Need for Standardized, Computable Trial Descriptions

Evidence-based medicine is a laudable but still impractical ideal. Finding, retrieving, interpreting, and applying just a single article to a particular clinical case can take a prohibitive amount of time and effort.

Though clinical trials are our best source of clinical evidence, it is logistically difficult to practice medicine based on the results of the most up-to-date clinical trials. Frontline practitioners simply do not have the time nor the expertise to formally synthesize evidence from the clinical literature, a synthesis which is increasingly done with a statistical technique called meta-analysis.

What is required is the standardization of the description of clinical studies and trials and the standardized reporting of such studies.
In Europe, the recent reporting recommendations for tumor marker prognostic studies (REMARK) statement[104] is making an effort to standardize the reporting of tumor prognostic marker studies, while CONSORT specifies trial standards for publication[105]. Nevertheless, significant efforts are still required towards this direction.

Research findings are most valid and generalizable when they arise from an appropriately designed randomized clinical trial (CT), performed in the setting in which the research is to be applied.

As a result, a standardized, computable representation of trials is needed to support activities throughout a trial’s lifecycle, from a trial’s execution, to comparing and combining trial results for metaanalysis, to applying trial results to clinical care. As reviewed herein, many of the existing research modelling efforts have focused on representing trials mostly for execution, whereas the needs of evidence-based medicine focus more on comparing, combining, and applying trial results.

Below we summarize the needs for a computable representation of CTs:

---

[97] With respect to the anonymisation of micro data (data pertaining to discernible individuals), numerous approaches and techniques have been proposed and studied over the last decades. These approaches usually encompass a combination of segmentation, perturbation, aggregation and (local) suppression techniques. The main objective is to maximize the data content level (usability of the data) while minimizing the re-identification risk with respect to the individuals involved. The underlying morale is that privacy guarantees should be mathematically provable.

[98] Samarati, P., Sweeney, L.: Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression. Technical report, SRI International (1998)

[99] A. Machanavajjhala, J. Gehrke, D. Kifer, and M. Venkitasubramaniam. l-diversity: Privacy beyond k-anonymity. In Proc. 22nd Intnl. Conf. Data Engg. (ICDE), page 24, 2006.

[100] Li, N., Li, T., Venkatasubramanian, S.: t-closeness: Privacy beyond k-anonymity and l-diversity. In: Proc. IEEE Int. Conf. Data Eng. (ICDE), Istanbul, Turkey (April 2007) page 106-115.

[101] D. Rebollo-Monedero, J. Forné, and J. Domingo-Ferrer. From t-Closeness to PRAM and Noise Addition Via Information Theory. Lecture Notes In Computer Science; Vol. 5262. In Proceedings of the UNESCO Chair in data privacy international conference on Privacy in Statistical Databases. Istanbul, Turkey, Pages: 100 – 112, 2008.

[102] µ-argus manual, http://neon.vb.cbs.nl/casc/Software/MuManual4.2.pdf

[103] A data protection framework for trans-European genetic research projects. Claerhout B, Forgó N, Krügel T, Arning M, De Moor G. 2008, Stud Health Technol Inform, pp. 67-72.

[104] LM. McShane et al. Reporting recommendations for tumor marker prognostic studies (REMARK). J Natl Cancer Inst, vol. 97, pp. 1180 –1184, 2005

[105] http://www.consort-statement.org

- A standard representation of CTs is necessary to clearly and accurately communicate the structure of a trial for uniform implementation at multiple sites. One of the challenges in such multisite trials is consistent implementation, when numerous individuals at the different sites are charged with executing the trial. Inconsistencies can arise, not necessarily from deliberate deviations from the trial's protocol but from different understandings of the protocol's elements. Consistency is supported by a common understanding of the relevant aspects of the trial. Such a common understanding is facilitated by communication of a shared standard representation of the trial's elements, and is especially important in practice-based trials where site investigators are often less familiar with conducting clinical research.
- Aside from needing a standard representation of CTs to help run a trial, such a representation is essential for combining results from multiple heterogeneous CTs in a meta-analysis, where small differences in trial design and outcome measures may lead to inaccuracy in the overall effect estimate[106]. The ability to determine which elements of two or more trials are similar and which are different is critical to detecting such differences. Without a standard method of representing the components of a trial, it is necessary to depend entirely on the interpretations of readers regarding the comparability of trial elements. There is an overlapping and equally important issue of the standard representation and reporting of clinical data for the purposes of comparing the results of multiple clinical studies[107].
- Integral to the task of conducting a systematic review of CTs is the need to objectively evaluate the quality of the trials. For this task, it is important to be able to ***understand the design elements of a given trial and be able to compare it with others of known quality***. These comparisons require identification and description of trial components such as treatment allocation strategies, in clear and unambiguous terms, to make valid judgments about the overall trial quality. The lack of a standard representation of trial design features impedes this process by making it more difficult to locate and characterize the important elements of a trial that are used in critical appraisals of trial evidence. A standard, computable representation would improve the ability to evaluate the quality of CTs and provide a basis for doing so in an automated fashion.
- The need for automated search and retrieval of clinical trial descriptions and their results spans trial design, execution, and matching of trial evidence to patients. ***Fundamental to the formulation of clinical trials and the use of trial outcomes to patient treatment decisions is the ability to locate and retrieve relevant existing clinical trials***. In the modern electronic world, search and retrieval strategies are based on the computable characteristics (metadata) of published trials, such as Medical Subject Headings classifications and terms embedded in the text (among many such methods of tagging or categorizing a given published work). These classification schemes assume that there is some underlying, commonly understood and accepted description that characterizes all such clinical trials. Although numerous efforts are addressing this issue, a commonly accepted and widely used representation is still largely lacking. As a consequence, the process of searching out applicable CTs continues to be inhibited both in terms of accuracy and efficiency.

1.2.12. Existing Efforts to Create Standard CT Models

Several organizations, including Clinical Data Interchange Standards Consortium (CDISC), Health Level Seven (HL7), the National Cancer Institute, and the World Health Organization (WHO), have recognized the need for and are developing standard trial representations or models. The CDISC aims ". . . to develop and support global, platform independent data standards that enable information system interoperability to improve medical research and related areas of health care."

Their efforts focus on clinical trials that support submissions to regulatory agencies such as the Food and Drug Administration and hence tend to be focused on pharmaceutical interventions and on meeting regulatory requirements as expressed in the rules and regulations of the U.S. Food and Drug Administration. In partnership with HL7, CDISC has undertaken a similar effort through its Regulated Clinical Research Information Management committee that has attempted to characterize clinical trials and the information they generate in terms of the Reference Implementation Model Version 3.0 (RIM 3.0)[108].

The CONSORT Statement[105] is intended to improve the reporting of a randomized controlled trial (RCT), enabling readers to understand a trial's design, conduct, analysis and interpretation, and to assess the

[106] L. Wood et al., Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. BMJ 2008 Mar 15;336:601–5. Epub Mar 3, 2008.
[107] JR. Deitzer et al., Coverage of clinical trials tasks in existing ontologies. AMIA Annu Symp Proc 2006:903
[108] HL7 Version 3 Resources. Available at: http://www.hl7.org.au/HL7-V3-Resrcs.htm.

validity of its results. It emphasizes that this can only be achieved through complete transparency from authors.

At the same time the Cancer Biomedical Informatics Grid (caBIG) of the National Cancer Institute commissioned the development of the Clinical Trials Object Data System to address the problems of combining the results of multiple clinical trials of cancer treatments to increase the speed of discovering of new methods of treatment. In Europe the ACGT project is developing a CDISC compliant Clinical Trial Design and Management System based on a shared ontology on Cancer, called ObTiMA (Ontology based Trial Management system). This semantic compliance of data collected in multiple CTs will, thus, enable the process of cross-trial data integration (i.e. meta-trials).

Recognizing the commonality of these efforts, the involved parties are working to harmonize their various efforts into a single standard domain model of regulated clinical research that shares a common set of terms spanning CDISC, HL7, and all caBIG, among others. This ongoing work is incorporated into the Biomedical Research Integrated Domain Group (BRIDG) model that represents the process and results of protocol-driven biomedical/clinical research[109]. That model is serving as a focus for trial model standardization worldwide.
In addition the Integrating the Healthcare Enterprise (IHE) initiative is developing specifications of integration profiles based on actors that interact through transactions, using the HL7 messaging standard. One relevant such specification is the integration profile "***Retrieve Form for Data Capture (RFD***)" whose objective is to gather data with a user's current application to meet the requirements of an external system.

1.2.13. Developing Clinical Trial Registries

As of September 27, 2008, the US Food and Drug Administration Amendments Act of 2007 (FDAAA) requires that clinical trials results be made publicly available on the Internet through an expanded "registry and results data bank"[110].

Under FDAAA, enrolment and outcomes data from trials of drugs, biologics, and devices (excluding phase I trials) must appear in an open repository associated with the trial's registration, generally within a year of the trial's completion, whether or not these results have been published. The new law is innovative in bridging the gap between a clinical trial's registration at inception (now an established requirement for publication) and the public archiving of its final peer-reviewed report.

The clinical trial protocols registered must be consistent with the standards issued by the World Health Organization and the International Committee of Medical Journal Editors.

For each trial falling within its scope, the law requires the posting of a table of "demographic and baseline characteristics" of the study participants, as well as a "table of values for each of the primary and secondary outcome measures for each arm of the clinical trial, including the results of scientifically appropriate tests of the statistical significance." Safety outcomes must be posted as of 2009, and further information may be required in future years. These are not just recommendations; the law imposes fines of up to US$10,000 per day for non-compliance[111].

Clinical trial registry databases are online catalogues of hypothesis-testing clinical trials conducted on human subjects. Information about the trial, such as the drug being tested and purpose of the study, is placed in an online registry before the trial begins, and remains available regardless of whether or not the trial is completed or published. Such registries have three broad purposes. First, they should lead to a reduction in publication bias, because the scientific community is made aware that a trial is planned and non-completion or non-publication can be detected. Second, they describe the main features of the study, such as the outcome variables and the study duration, in an attempt to ensure that the study accords with its originally stated purposes and methods. Third, for both patients and investigators, they facilitate recruitment into clinical trials and the transparency aims to improve the patients' confidence in clinical trials participation. Clinical trial results databases are online repositories for the results of clinical trials whether or not they are published in the medical literature. Results databases permit the review of all completed studies on a topic by academics, regulatory bodies, public interest groups, and study participants. If a results database is constructed properly,

[109] C. Weng et al., User-centered semantic harmonization: a case study. J Biomed Inform 2007;40:353– 64
[110] United States Code (2008) US Public Law 110-85: Food and Drug Administration Amendments Act of 2007, http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=110_cong_public_laws&docid=f:publ085.110. Accessed 8 March 2009
[111] The PLoS Medicine Editors (2008) Next stop, don't block the doors: Opening up access to clinical trials results. PLoS Med 5(7): e160. doi:10.1371/journal.pmed.0050160

it can also facilitate a meta-analysis (a statistical combining of similar studies) to formally evaluate safety and efficacy[112].

A large number of clinical trials registries currently exist, with a varied degree of standardization of their content. References and descriptions of these registries are collected, through online searches and a lot of manual work, in clinical trials registries databases such as the Systematic Review resource[113].

One of the earliest efforts is the National Library of Medicine's ClinicalTrials.gov database[114] which laid the groundwork for implementation of the International Committee of Medical Journal Editors' policy requiring the public registration of clinical trials to be considered for publication[115].

Subsequently the WHO established a global network of international, national, and regional trial registries. These registers collaborate in capturing the 20-item Trial Registration Data Set of design and administrative information to be publicly registered before the enrolment of the first participant for all trials worldwide[116].

To minimize the need for multiple data entry and to facilitate data reuse for multiple purposes, such as health systems planning, meta-analysis, and patient recruitment, the WHO is working with CDISC to define a standard XML model for interchanging the WHO Trial Registration Data Set among the Register's Network, and this model is incorporated into BRIDG[117].

The NLM is currently expanding the registry database and building the results database, which will also include links to existing results such as advisory committee information and journal publications. However, it is currently unclear whether the NLM will have the resources and the authority to monitor and enforce the accuracy of the trial registration information and results posted on their ClinicalTrails.gov site. At present the correctness of the data submitted to the registry is the responsibility of the sponsor and of the investigators. [Yan J., All U.S. Clinical Trials Must Now be Registered, Psychiatric News, 43(14), 2008]. In a recent study, Public Citizen[118] evaluated most of the private registries reviewed as inconsistent with the WHO's 20 key-element standard.

From September 2009 the reporting of safety results has also become mandatory and include the rates of all serious adverse events, adverse events that occurred in more than 5 percent of subjects in any treatment arm, and adverse events in the active-treatment group that are at least twice as frequent as those in the placebo arm.

> The pharmaceutical industry finds itself plagued by drug safety concerns and scandal as prescription drugs continue to make headline news around the world. Currently, the general public has a high level of distrust regarding pharma companies. This lack of trust is largely based on the limited availability of timely information regarding the clinical development process. The steady drop in clinical trial enrollment by patients is also a likely reflection of this general lack of faith. The bottom line is that healthcare professionals and consumers are demanding broader access to trial results—both positive and negative.
>
> *The Era of Clinical Trial Registries*
> *L. Grimes, Campbell Alliance, 2005*

1.2.14. The pharmaceutical industry position on reporting clinical trials

The creation of trial registries was a response to the public demand for balanced and timely information on clinical trials and their results, both positive and negative. These repositories are not meant to offer interpretations or conclusions but to present objective and comprehensive data. On the other hand, from the perspective of pharmaceutical companies the requirement to register all clinical trials and to provide public access to the information generates additional costs and potentially reduces their competitive advantage when intellectual property cannot be properly safeguarded.

As a response to the public pressure and the new legislation the pharmaceutical industry issued a joint position document in 2005, followed by an update in November 2008, in which they recognized the health benefits of making clinical trial data more widely available. At the same time they emphasized on the need to maintain protection for individual privacy, intellectual property and contract rights. Their 2008 document "Joint Position on the Disclosure of Clinical Trial Information via Clinical

[112] A Policy Study of Clinical Trial Registries and Results Databases, Public Citizen, 2007
[113] Clinical Trials Registries Database, ssrc.tums.ac.ir/SystematicReview/CTRDB.asp
[114] McCray AT, Ide NC. Design and implementation of a national clinical trials registry. J Am Med Inform Assoc 2000;7:313–23
[115] Laine C, Horton R, DeAngelis CD, et al. Clinical trial registration: looking back and moving ahead. JAMA 2007;298:93– 4
[116] Registration Data Set (Version 1.0), International Clinical Trials Registry Platform (ICTRP). Available at: http://www.who.int/ictrp/network/trds/en/index.html Accessed 6 March 2009
[117] Wood L, Egger M, Gluud LL, et al. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. BMJ 2008 Mar 15;336:601–5. Epub Mar 3, 2008
[118] www.citizen.org

Trial Registries and Databases"[119] stated commitment with respect to providing input to both clinical trial registries and clinical trial results databases.
For new and ongoing clinical trials the data to be registered is described and the time frame for filling in the data is also specified. In principle they support the Minimum Trial Registration Data Set published by WHO, but specify that the trial sponsor can reserve the right to delay the disclosure of sensitive data elements, such as primary outcome, key secondary outcomes, intervention name, target sample size, and official scientific study of the trial.
The clinical trial results database maintains the summary results of completed clinical trials. The results of all confirmatory clinical trials conducted on an approved medicinal product that is commercially available should be posted no later than one year after the commercialization of the product, regardless of the outcome of the trial. On the other hand, the document states no obligations for trials not related to a commercial product. It is only encouraged that trial results for an investigational product that has failed in development which have significant medical importance should still be posted.
To avoid multiple postings and ensure transparency each trial listed in a registry should have a unique identifier with which the trial could be tracked through multiple databases, including clinical trial results databases.

1.2.15. External data and knowledge sources

In the bioinformatics field, an extremely large body of relevant external data resources is available. During the project, an approach to making relevant data resources available to the platform will be developed. Based on the use cases and scenarios, a selection of resources will be made available within the platform. The available data spans a wide range of molecular types and interactions, both on a content level (the "raw" data) as on a description level (various ontologies, vocabularies and literature). Examples here are nucleotide databases, expression databases, protein databases, and pathway databases together with their associated standards, as can be found (aggregated) on for instance the National Center for Biotechnology Information portal[120] and the European Bioinformatics Institute portal[121]. A wide variety of ontologies can be found at National Center for Biomedical Ontology's Bioportal[122] and at "The Open Biological and Biomedical Ontologies" site[123] (OBOFoundry). For example, the Gene Onology (GO)[124] is an often used ontology, which is part of the OBOFoundry. GO provides a controlled vocabulary of terms for describing gene product characteristics and gene product annotation data, allowing for consistent descriptions in different databases. The ontology is organized using three principles: cellular component, biological process and molecular function. In this way, the use of GO terms for instance facilitates uniform queries across collaborating databases. Another interesting resource available is the vast body of knowledge locked up in the literature. An example, PubMed[125], provides a feely available service which exposes MEDLINE[126], the National Library of Medicine's premier bibliographic database covering the fields of medicine, nursing, dentistry, veterinary medicine, the health care system, and the preclinical sciences.
Another type of external data sources are drug-related databases. There are for instance drug discovery databases and drug information databases. An example of a drug discovery database is ChEMBL[127], a database of bioactive drug-like small molecules containing 2-D structures, calculated properties and abstracted bioactivities. The data is abstracted and curated from the primary scientific literature, and cover a significant fraction of the SAR and discovery of modern drugs. Additional data on clinical progress of compounds is being integrated into ChEMBL at the current time.
There are various drug information databases available, typically targeting either the clinician or patient. They typically provide information on for instance synonyms, manufacturer brand names, therapeutic information, dosage, pharmacokinetics, cautions, interactions, adverse effects and drug interactions. Some examples of available products are Thomson Reuters Micromedex[128], VIDAL Expert[129], Metabolism & Transport Drug Interaction Database[130] and Medi-span Clinical[131].

[119] http://www.clinicaltrials.ifpma.org/fileadmin/files/pdfs/EN/Revised_Joint_Industry_Position_Nov_2008.pdf
[120] http://www.ncbi.nlm.nih.gov/sites/gquery?itool=toolbar
[121] http://www.ebi.ac.uk/
[122] http://bioportal.bioontology.org/
[123] http://www.obofoundry.org/
[124] http://www.geneontology.org/
[125] http://www.ncbi.nlm.nih.gov/pubmed
[126] http://www.nlm.nih.gov/databases/databases_medline.html
[127] http://www.ebi.ac.uk/chembldb/index.php
[128] http://www.micromedex.com/products/hcs/
[129] http://www.vidal.fr/professionals-products/vidal-expert
[130] http://www.druginteractioninfo.org/Home.aspx
[131] http://www.medi-span.com/medispan-clinical.aspx

One of the types of external data sources which are online available is data on cell lines. A cell line is a group of cells adapted to proliferate indefinitely in a laboratory setting. A cell line originates from cells from a specific tissue and organism[132]. Often the cells have been adapted as the original cells were not immortalized (and therefore could not proliferate in-vitro indefinitely). There are various cell lines available (e.g. for purchase). An example of a set of cell lines is NCI's "NCI-60 DTP Human Tumor Cell Line Screen"[133], developed as a in-vitro anticancer drugs discovery screen in the 1980's. The screen has provided a wealth of information about cell growth inhibition and tumor-cell kill[134]. Various data resources on this screen are available online[135], for instance thousands of molecular targets[136] (including protein levels, RNA measurements, mutation status and enzyme activity levels) have been measured.

1.2.16. Relevant external initiatives – caBIG

An important relevant initiative is the cancer Biomedical Informatics Grid (caBIG)[137], launched by the National Cancer Institute and maintained by the Center for Biomedical Informatics and Information Technology (CBIIT). The goals of caBIG are to connect scientists and practitioners through a shareable and interoperable infrastructure, develop standard rules and a common language to more easily share information, and build or adapt tools for collecting, analyzing, integrating, and disseminating information associated with cancer research and care.

The basic infrastructure consists of two main parts. The *Cancer Common Ontologic Representation Environment* (caCORE)[138] provides functionality to ensure interoperability. For instance, it includes a service for managing vocabularies (EVS) and a service for managing metadata (caDSR), along with a software development kit for easy access to these services. Typically, (object-oriented) information models are registered in caDSR and their meaning is linked to vocabularies stored in EVS. Furthermore, caBIG provides a technical platform and infrastructure (caGRID) based on GRID technology[139]. This allows for secure sharing of resources (for instance computational or data resources) . These two parts form the basis for the more functional services, where typically a grid service is exposed expressed in terms of the caCORE.

caBIG provides various tools which further build on this basic infrastructure. Four *domain workspaces* were formed to focus development[140]:

- *Clinical Trial Management Systems*: Develops a comprehensive set of modular, interoperable and standards based tools designed to meet the diverse clinical trials management needs of the Cancer Center community.
- *Integrative Cancer Research Workspace*: Produces modular and interoperable tools and interfaces that provide for integration between biomedical informatics applications and data. This will ultimately enable translational and integrative research by providing for the integration of clinical and basic research data.
- *In Vivo Imaging Workspace*: Creates, optimizes and validates tools and methods to extract meaning from and share imaging data.
- *Tissue Banks and Pathology Tools Workspace*: Develops a set of tools to inventory, track, mine, and visualize biospecimens and related annotations from geographically dispersed repositories.

The EURECA project will assess the relevant parts from caBIG from an adopt, adapt and interoperate perspective. Next to the functional requirements, the components should also fulfil various nonfunctional requirements (as will be specified during the use case elaboration), such as costs (e.g. various tools rely on Oracle Clinical), ease of deployment and maintenance, and performance.

Next, we highlight some relevant parts of caBIG. **caArray** provides an open array data management system which allows federation of multiple installations. **caTissue** is a biorepository tool for biospecimen inventory management, tracking, and annotation. **geWorkbench** provides an innovative, open-source software platform for genomic data integration, bringing together analysis and visualization tools for gene expression, sequences, protein structures, pathways, and other biomedical data. **Cancer Central Clinical Database** (C3D) is a clinical trials data management system. C3D collects clinical trial data using standard case report

[132] http://www.molecularstation.com/wiki/Cell_line
[133] http://dtp.nci.nih.gov/branches/btb/ivclsp.html
[134] "The NCI60 human tumour cell line anticancer drug screen", Robert H. Shoemaker, Nature Reviews Cancer 6, 813-823 (October 2006)
[135] http://dtp.nci.nih.gov/webdata.html
[136] http://dtp.nci.nih.gov/mtargets/mt_index.html
[137] https://cabig.nci.nih.gov/
[138] https://cabig.nci.nih.gov/concepts/caCORE_overview?pid=primary.2006-07-07.4911641845&sid=caCORE_Overview&status=True
[139] http://www.globus.org/toolkit/
[140] https://cabig.nci.nih.gov/workspaces/domain

forms (CRFs) based on common data elements (CDEs). C3D utilizes security procedures to protect patient confidentiality and maintain an audit trail as required by FDA regulations. **The Clinical Connector**[141] provides a semantically integrated service layer via caGRID that allows C3D adopters to expose functions within C3D (Oracle Clinical). The exposed service layer uses a BRIDG based model and defines service operations that could be implemented by other CTMS systems. The first service allows external applications to load laboratory test result data into C3D study structures for registered patients. The second service allows external applications to enroll patients onto a C3D hosted study. Exposing additional C3D functionality is being planned.

For our project, we will assess whether there are appropriate offerings from caBIG in terms of standards, ontologies or tools. As various partners collaborate with non-European partners, emphasis will be put on standardization, harmonization and standards compliance of the solutions.

1.2.17. Service oriented architecture, web service semantics and standards

The objective of WP2 within EURECA is to produce an Open Reference Architecture Specification for the services and end user tools that the project will deliver, test and evaluate. The Architectural Specifications should integrate concepts from existing standards, models and architectures, while extending and refining them where appropriate and required.
There are many specifications, standards and proprietary technologies which are relevant to the description of ICT infrastructure components and their properties. This section presents a short selection which, while not exhaustive, covers some of the most prominent activities.

Web Services Modeling Ontology (WSMO) - Web Services Modeling Language (WSML)
Web Service Modeling Ontology (WSMO)[142] is an ontology for semantically describing Semantic Web Services. It is a model for the description of semantic web services that tries to overcome the limit of the existing technologies for the service description, in particular OWL-S. Web Service Modeling Language (WSML)[143] is a language that formalizes the WSMO. It uses well-known logic formalisms, namely, Description Logics, First-Order Logic and Logic Programming, in order to enable the description of various aspects related to Semantic Web Services. It consists of a number of language variants with different underlying logic formalisms. The conceptual grounding of WSMO is based on the Web Service Modeling Framework (WSMF)[144].

OWL-based Web Service Ontology (OWL-S)
OWL-S[145] is an OWL based Ontology, within the OWL-based framework of the Semantic Web, for describing Web services. OWL-S ontology is also sometime considered as a language for describing services, reflecting the fact that it provides a standard vocabulary that can be used together with the other aspects of the OWL description language to create service descriptions It will enable users and software agents to automatically discover, invoke, compose, and monitor Web resources offering services, under specified constraints. OWL-S supplies Web service providers with a core set of mark-up language constructs for describing the properties and capabilities of their web services in unambiguous, computer-interpretable form.

Web Service Semantics - WSDL-S
The Web Services Semantics - WSDL-S[146] specification is a W3C Member Submission that defines how to add semantic information to WSDL documents.
WSDL-S tries to overcome the lack of semantics in WSDL by adding new extensibility elements to the WSDL standards to annotate the semantic of web services. Each service description refers one or more Semantic Model. A semantic model captures the terms and concepts used to describe and represent an area of knowledge or some part of the world, including a software system. A semantic model usually includes concepts in the domain of interest, relationships among them, their properties, and their values. WSDL-S provide mechanisms to annotate the service and its inputs, outputs and operations. Additionally, it provides mechanisms to specify and annotate preconditions and effects of Web Services. These preconditions and effects together with the semantic annotations of inputs and outputs can enable automation of the process of service discovery.

Semantic Annotation for WSDL (SAWSDL)
In 2006, the W3C created a charter for the Semantic Annotation of Web Services (SAWSDL[147]), which used WSDL-S[148] as its primary input. SAWSDL became a W3C candidate recommendation in January 2007.

[141] https://cabig.nci.nih.gov/tools/C3DClinicalConnector
[142] Web Service Modeling Ontology, http://www.wsmo.org/
[143] Web Service Modeling Language, http://www.wsmo.org/wsml
[144] Web Service Modeling Framework, http://www.w3.org/2005/04/FSWS/Submissions/1/wsmo_position_paper.html#fensel
[145] OWL-S: Semantic Markup for Web Services, http://www.w3.org/Submission/OWL-S/
[146] Web Service Semantics - WSDL-S, ,http://www.w3.org/Submission/WSDL-S/
[147] SAWSDL,http://www.w3.org/TR/sawsdl

SAWSDL defines mechanisms using which semantic annotations can be added to WSDL components. SAWSDL defines how to add semantic annotations to various parts of a WSDL document such as input and output message structures, interfaces and operations. The extension attributes defined in this specification fit within the WSDL 2.0 extensibility framework. It provides mechanisms by which concepts from the semantic models that are defined either within or outside the WSDL document can be referenced from within WSDL components as annotations.The annotations on schema types can be used during Web service discovery and composition. In addition, SAWSDL defines an annotation mechanism for specifying the data mapping of XML Schema types to and from an ontology; such mappings could be used during invocation, particularly when mediation is required. To accomplish semantic annotation, SAWSDL defines extension attributes that can be applied both to WSDL elements and to XML Schema elements.

1.2.18. Personal Health Records

Although various PHR offerings have been available for quite some time, the recent trend of a more patient driven healthcare together with PHR product offering from big companies (e.g. Microsoft HealthVault, Google Health) has raised patient awareness, causing a significant rise in the uptake. Traditionally, two types of PHR systems could be distinguished, approaching the problem from two sides. The first type is an extension of the healthcare institute's electronic medical record. It provides a view of the data of a patient in a particular healthcare institute. Typically, the patient cannot add or modify any data (see Figure 12-a). The second type is the standalone PHR. The standalone PHR allows patients to enter health information, but typically is not interconnected with the information systems of healthcare systems (see Figure 12-b).
Lately, this difference has been diluted by products and services of independent vendors. These offerings allow different healthcare institutes to provide patient data and allow patients to manage, add and annotate their data (see Figure 12-c).

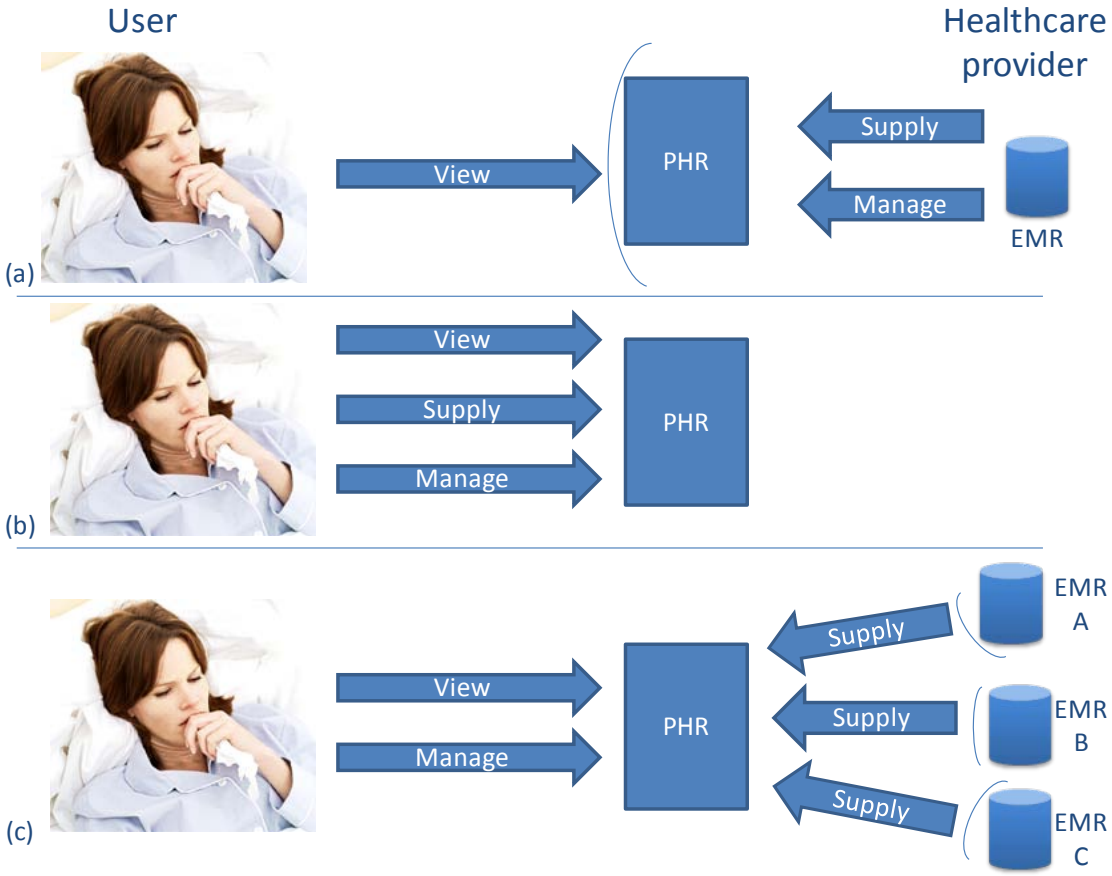

Figure 12 PHR types

The current state-of-the art allows data to enter the PHR in a variety of ways, often leveraging on healthcare standards such as HL7 CDA and CCR. Not only data coming from a healthcare institute can enter the PHR,

[148] WSDL-S,http://www.w3.org/Submission/WSDL-S/

but also different sources are often supported such as measurements from a patient's device (e.g. a weighing scale). Where the current state of the art falls short is in explaining the data contained in the PHR to the patient. The primary goal of the data might not be to inform the patient, but the data could be generated for a different goal (e.g. to assist a clinician to make a proper diagnosis). When such data enters the PHR, it might be too complex and technical to be of real value to the patient.
In[149], three early PHR implementations are reviewed (MyChart, a vendor created, clinic hosted solution; PatientSite, a self built hospital hosted solution; and Indivo, a self built, institution-neutral hosted service). A number of challenges for 2008 and beyond were identified: Challenge #3 "*Patients may want to integrate knowledge sources on the internet with their PHR"*, challenge #4 "*Patients with specific diseases may want to connect to communities of others with similar diagnosis*" and challenge #5 "*Patients may wish to participate in clinical trials, post market pharmaceutical vigilance, or public health surveillance via their PHR*". "*Our experience to date at three institutions demonstrates that personal health records which share data among patients and providers can successfully be deployed but require careful attention to policy around privacy, security, data stewardship and personal control*".

Prominent PHRs

In this section, four PHR solutions are highlighted to illustrate current state of the art product offerings: Google Health, Indivo, Microsoft HealthVault and Tolven.

Google Health

Google Health[150] provides an online platform to store and manage health information online. Various healthcare institutes allow exporting medical records to Google Health[151]. The platform enables third-party services and tools to interoperate with Google Health. The user manages the privacy profile and decides what permissions are granted to third-party services. In the Google Health model, the user creates an account with the third-party services, and subsequently links this account to his Google Health account. At the moment of writing, the use of Google Health is restricted to the USA.

Indivo

Indivo[152] is a free and open-source product providing PHR functionality. Indivo is used (amongst others) by the Children's Hospital Boston and Dossia (a Personal health record service offered by various large employers in the United States). Indivo's differentiating characteristics are the focus on open standards and its open-source character.

Microsoft HealthVault

Microsoft HealthVault[153] provides a central repository for healthcare data. Just as Google Health, most of the functionality is provided by third-party services and tools. Also here, the user manages the privacy profile. In the Microsoft HealthVault model, the user uses his Microsoft HealthVault account to authenticate with the third-party services.

Tolven

Tolven[154] is an open-source solution is specifically designed to act as the primary repository for all clinical and administrative data for a healthcare enterprise. It can map existing data to Tolvens core data model. Tolven uses standards such as HL7 data types and structures, UMLS coding and a Continuity of Care Record index. Tolven's Electronic Health Solution consists of an electronic Personal Health Record and an electronic Clinician Health Record. Tolven is successfully used for primary clinical data input at the participating sites by the TRANSCEND[155] system developed to support the I-SPY multi-centric clinical trials[156] in the US (led by UCSF under the governance of FNIH).

In EURECA we will investigate how extending standards-based semantic interoperability towards PHR systems can contribute to improved safety. Our scenario will be implemented with the Tolven open source system. This proof of concept can of course be extended to other PHR systems by applying the same standards-based modular approach to semantic interoperability. Tolven is licensed under the LGPL (Lesser General Public License), and uses standards such as HL7 data types and structures, UMLS coding and a Continuity of Care Record index, which is in line with EURECA's standards-based approach with respect to semantic interoperability. The use of an open source system under a permissive license and which adheres

[149] J Halamka et al., "*Early experiences with Personal Health Records"*, Journal of the American Medical Informatics Association, volume 15, number 1, 2008.
[150] http://www.google.com/health
[151] https://www.google.com/health/directory?cat=importrecords
[152] http://indivohealth.org/
[153] http://www.healthvault.com/
[154] http://www.tolven.org
[155] http://cabig.cancer.gov/resources/newsletter/issueXXVIII/action.asp
[156] http://ispy2.org/

to adopted standards in the clinical domain will facilitate reuse and the further adoption of our solutions by a wide community.

1.2.19. Natural Language Processing and Information Extraction

Introduction of general domain Information Extraction

In general domain IE the following tasks are usually considered. The first task is Named Entity Recognition (NER) which consists in identifying textual fragments corresponding to semantically relevant concepts. General domain NER usually extracts proper names like person names, organization names, geographic names with a high degree of accuracy[157] for a state-of-the-art showing the different techniques that are used and the accuracy of current systems).

The second task often performed in IE is called fact extraction. It consists to the detection of relationships holding between Named Entities. For instance, the fact "works_in(X,Y)" captures relations between NE of type Person and of type Organization (represented respectively by X and Y) and denotes that a given person (instantiating X) works in a given company (instantiating Y). A system performing such a fact extraction should be able to deal with linguistic variations like "X works in Y", "Y is the workplace of X" etc.

Finally, event extraction is also an important and more complex goal in IE. An event can be seen as a specific situation in which several facts and entities can be involved. It usually describes a state-of-affair of interest for a given domain. The typical illustration for event extraction is the extraction of terrorist attacks events from news which was the domain chosen by the MUC-3 and MUC-4 conferences for information extraction[158]. Terrorist attacks are represented by templates in which slots have to be filled by information conveyed in the text, as for instance the perpetrator of the attack, the place of the attack, the injuries etc. For a general overview of IE, see MUC-7 conference task descriptions and papers[159].

Information Extraction in the medical domain

Extraction of relations between concepts (fact extraction) is the second level of information one may want to extract. The SemRep system[160], developed at the National Library of Medicine and used for different specific applications, extracts medically relevant semantic relations such as drug interactions, medical treatments etc. The system is a rule-based symbolic system. It uses shallow syntactic processing and depends on MetaMap[161] for the extraction of concepts.

In the medical domain, a considerable amount of heterogeneous textual documents are produced, such as clinical guidelines, scientific medical literature, and documents entered in clinical information systems. Currently, vast amounts of useful information reside in unstructured clinical text and serves a critical role in administration, clinical practice, as well as research. In the context of the EHR, next to structured data, one often needs to deal with free text reports. Since clinical text became available, the natural language processing community has been motivated to develop automated ways to extract meaningful concepts such as diagnoses, procedures, and medications out of free text reports[162], and their relations in order to enable automated processing and semantic reasoning on that data.

With this amount of medical literature on one side, and the recent advances in Natural Language Processing (NLP) aiming at facilitating the access and understanding of document content on the other side, medical NLP has become an important subfield of Natural Language Processing. [Demner-Fushman et al. 09][163] reviews NLP techniques and applications dedicated to Clinical Decision Support. Different kinds of applications for medical NLP have been developed. Among them, question-answering applications as described for instance in [Lee et al. 06][164]. Text summarization in the medical domain is also promising, taking into account the large amount of documents that are produced in the domain. For instance [Elhadad

[157] Ehrmann M., Les Entités Nommées, de la linguistique au TAL : statut théorique et méthodes de désambiguïsation. Thèse de Doctorat, Université Paris VII, June 2008

[158] Sundheim, B. M. Overview of the fourth Message Understanding Evaluation and Conference. (http://clair.si.umich.edu/clair/anthology/query.cgi?type=Paper&id=M92-1001)

[159] http://www-nlpir.nist.gov/related_projects/muc/proceedings/muc_7_toc.html

[160] Rindflesch T., Fiszman, M., Lubbus, B. Semantic interpretation for the biomedical research literature. Cheh H (eds). Medical informatics: knowledge management and data mining in biomedicine. Springer Science and Business Media, pp 299-422. (2004)

[161] Aronson, A.R. Effective mapping of biomedical text to the UMLS metathesaurus: the MetaMap program .2001. AMIA Symposium Hanley & Belfus, 2001: 17—21

[162] Meystre SM, Savova GK, Kipper-Schuler KC, Hurdle JF. Extracting information from textual documents in the electronic health record: a review of recent research. Yearb Med Inform. 2008:128-44.

[163] Demner-Fushman D., Chapman, W. W., McDonald C. J. What can Natural Language Processing do for Clinical Decision Support ? Journal of Biomedical Inform. 2009 October; 42(5): 760–772 (2009)

[164] M., Cimino, J. , Zhu, H. R., Sable, C., Shanker, V., Ely, J., Yu, H., Beyond Information Retrieval – Medical Question Answering. AMIA Annual Symp Proceedings. (2006)

06][165] presents a user-oriented multi-document processing system that produces textual summaries of medical findings according to patient history. In a very different way, [Aramaki et al. 09][166] developed a system which converts medical texts into table structures producing thus a structured version of discharge summaries.
The IE subtasks described previously are also performed and adapted in the medical domain. Instead of general entities like Person, Place, etc., we extract medical concepts or medical terms. While in the first case a given typeset (such as for instance Drug names, Bacteria names etc.) is defined a priori for a given application purpose, in the second case the system extracts all medical terms found in an input text and replace them with their codes in one or more given terminologies. A good illustration for concept recognition is given by the challenge organized by i2b2[167] for which one subtask consists in the automatic extraction of occurrences of concepts in medical tests, medical problems and treatments in clinical data (see description of the concept recognition task for 2010 challenge[168].
Automatic assignment of terminological codes is illustrated by the medical NLP challenge organized by the CMC in 2007[169], where different systems were evaluated in the automatic assignment of ICD-9CM clinical codes. For instance, [Cramer et al. 09][170] describe their system behaviour in the Computational Medicine Center's challenge for the automatic assignment of ICD-9-CM code in free text radiology reports. Results are compared both to other systems and manual annotation. Other systems map fragments of text to UMLS concepts[161] or to MESH terms[171].

Existing medical NLP tools
Some systems have been previously mentioned for information extraction in the medical domain. Here we provide main characteristics of some of the best known tools used in the domain.
MedLEE (Medical Language Extraction and Encoding System)[172] was developed at Columbia University. It is a text processor that extracts and structures clinical information from textual reports and translates the information to terms in a controlled vocabulary (mapping to concepts of the UMLS). Clinical information can then be accessed by further automated procedures. It has been used in radiology, discharge summaries, sign out notes, pathology reports, electrocardiogram reports, and echocardiogram reports, and can readily be ported to other clinical domains.
MedIE[173] was developed at Drexel University. It extracts terms and relations from English medical text. This system determines terms that are syntactically or semantically related to each other. A Link Grammar Parser is used to find syntactically relations, while relations predefined in the terminology (UMLS) help to find semantically relations. This approach tries to combine existing lexical-semantic resources and uses GATE for shallow Natural Language Processing (tokenization and tagging).
For the French language, we will rely in the EURECA project on existing linguistic processors that have been developed for the general purpose domain and have shown high accuracy in different evaluation campaigns (ESTER for NE for French, PASSAGE for French Parsing). XIP (Xerox Incremental parser)[182] has been developed at Xerox for more than a decade and provides a deep linguistic analysis based on dependency parsing. The dependency approach enables to establish syntactic and semantic links between possibly distant linguistic units (words or complex terms). This links are labelled either with syntactic functions like subject, object, etc. or with semantic relations that are computed using both previously extracted syntactic functions and lexical semantics information attached to the domain. The linguistic representation of XIP output is very rich and carries information about verbal tenses, aspect, thematic roles. This enables to integrate in the core linguistic dependency processing, extensions dealing with temporality, modality which is crucial to capture hedged information.
First experiments in adapting XIP to the medical domain for French have been performed in the context of the French government-funded ALADIN-DTH project aiming at the automatic detection of Hospital acquired

[165] Elhadad, N. User-Sensitive Text Summarization: Application to the Medical Domain. PhD thesis, Columbia University, 2006.
[166] TEXT2TABLE: Medical Text Summarization System based on Named Entity Recognition and Modality Identification
[167] https://www.i2b2.org/index.html
[168] https://www.i2b2.org/NLP/Relations/Documentation.php
[169] http://www.computationalmedicine.org/challenge/index.php
[170] Crammer K., Dredze, M., Ganchev K., and Talukdar P.P., Automatic Code Assignment to Medical Text. Proceeding BioNLP '07 Proceedings of the Workshop on BioNLP 2007: Biological, Translational, and Clinical Language Processing
[171] Cooper, G. F., Miller R.A. An experiment comparing lexical and statistical methods for extracting MeSH terms from clinical free text.Journal of the American Medical Informatics Association (JAMIA), v5(1) pp. 62--75 (1998)
[172] http://lucid.cpmc.columbia.edu/medlee/
[173] Zhou, X., Han, H., Chankai, I., Prestrud, A., Brooks, A. Approaches to text mining for clinical medical records. Proceedings of the 2006 ACM symposium on Applied computing, Dijon, France, pp. 235 – 239 (2006)

infections from processing patient discharge summaries of different care units. The project is still on-going but first results are very promising[174].
In EURECA, we expect that the adaptation of Xerox deep linguistic analysis tools to the oncology domain will produce very accurate results for the information extraction task in both EHR and clinical trial data. We want to take advantage of the rich linguistic analysis we obtain to tackle problems such as negation, modality and temporality processing which have influence in the accuracy of information extraction tasks. This kind of problems have been addressed[175] [176] [179], but still a lot remains to do in this area, especially in the French language as most of the experiment and work on the subject has been performed for English.

IE in the EURECA project
A common characteristic of these applications is the need to access the content and meaning of medical documents. In the context of EURECA, we are interested in Information Extraction (IE) task in order to facilitate the access to information enclosed in unstructured textual data.
We will provide tools to perform syntactic and semantic analysis of textual data available in the project. Information Extraction in this context will go far beyond simple pattern matching techniques. Pattern matching techniques are not enough if we want to make a step towards knowledge discovery and reasoning. Information Extraction (IE) usually consists in two main tasks: concept identification and event identification (i.e. relations between concepts). However, in order to obtain accurate results in both concept- and event-identification, complex linguistic phenomena have to be considered. We distinguish the following steps for IE:
1) Concept recognition
Recognizing concepts of interest present in texts is a common task in the processing of medical textual data and is a known task. For instance [Friedman & al. 2004][177] describe a system using advanced Natural Language Processing (NLP) techniques in order to map clinical documents to UMLS code. A precision and recall of 89% and 84% respectively are obtain for term extraction. Concept recognition is very dependent on the chosen terminologies used as the concepts labels that are retained.
But simple term extraction without taking into account contextual information remains unsatisfactory for accurate results. An important subtask in IE consists in detecting the status of the information conveyed by the textual data, more specifically detect if information is negated or speculative.
2) Negated and hedged information processing
Negated information is often present in medical documents, especially in EHR where a negated fact or negated exam results can be extremely informative regarding the patient state. Negation is a well-known and difficult problem to handle in NLP. In the medical domain, previous work on this topic has been performed, with interesting results that we believe can still be improved. For instance, the Negfinder system presented in [Mutalik et al. 2001][178] uses a set of negative triggers together with a lexer and a context-free grammar to detect negated concepts in medical texts. In the same line a more recent work presented in [Harkema et al. 2009][179] presents a system dealing with negative information in clinical reports. Another syntax based method is presented in [Gindl et al. 2008][180]. This approach is linguistically oriented and generalizes over the previous approaches. However, problems such a double negation are not properly handled.
In a similar way, hedged information can also be extremely misleading in an information extraction task. Hedged information consists of assertions that are not factual. It is estimated that MEDLINE contains about 11% of sentences carrying speculative information. Obviously this information has to be treated in a different way than truly asserted facts.

[174] Hagège, C., Marchal, P., Gicquel Q., Darmoni, S., Pereira S., Metzger, M-H.: Linguistic and Temporal Processing for Discovering Hospital Acquired Infection from Patient Records. In Proceedings of the 2nd International Workshop on Knowledge Representation for Health Care (KR4HC-2010), Lisbon, Portugal. (2010).
[175] Morante, R., Daelemans W.: A metalearning approach to processing the scope of negation.In: Proceedings of CoNLL 2009, pp 21--29 (2009)
[176] Morante, R.: Descriptive analysis of negation cues in biomedical texts In: Proceedings of the Seventh International Language Resources and Evaluation (LREC'10), pp. 1429--1436.
[177] Friedman C., Shagina L., Lussier Y., Hripcsak, G.: Automated Encoding of Clinical Documents Based on Natural Language Processing. Journal of the American Medical Informatics Association (JAMIA), v11(5) pp. 392–402 (2004).
[178] Mutalik, P.G., Deshpande, A., Nadkarni, P.M.: Use of General-purpose Negation Detection to Augment Concept Indexing of Medical Documents: A Quantitative Study Using the UMLS. Journal of the American Medical Informatics Association (JAMIA), v8(6), pp. 598--609 (2001).
[179] Harkema, H., Dowling, J. N., Thomblade, T., Chapman W.: ConText: An Algorithm For Determinining Negation, Experiencer, and Temporal Status from Clinical Reports. Journal of Biomedical Informatics, 42, pp. 839--851 (2009).
[180] Gindl, S., Kaiser K., Miksh S.: Syntactical Negation Detection in Clinical Practice Guidelines. In Andersen, S.K.; Klein, G.O.; Schulz, S.; Aarts, J.; Mazzoleni, M.C. (eds.) eHealth Beyond the Horizon – Get IT There. Proc. of the 21st International Congress of the European Federation for Medical Informatics (MIE 2008), Göteborg, Sweden, pp. 187--192, IOS Press, May 2008.

As with negation, the main problem when dealing with hedged information is to determine the scope of the hedged or negation cues. Some work in this domain uses machine learning techniques to detect in a sentence the textual elements in the scope of a hedged cue[181]. Results are not fully satisfactory.
We intend to use refined NLP techniques based on deep linguistic knowledge for both negated and hedged information processing. An existing robust parser (XIP) developed at Xerox[182] is currently adapted to the medical domain, more specifically patient discharge summaries from different unit cares of three different French hospitals. This adaptation is performed in the context of a French National Project (ALADIN) and current developments show that our generic linguistic is adaptable to medical text processing.

3) Temporal information processing

Temporal information is another element that has to be considered for refined information extraction tasks in the medical domain. Previous work on temporal processing in medical literature has been performed. For instance, in [Zhou et al. 2008][183] the TimeText system is presented. This system performs temporal processing of clinical discharge summaries and the evaluation of the temporal relations between events extracted by the system (before, after or during relations).
Xerox already developed a temporal processing module which is embedded within the general purpose parser[184]. This module detects temporal blocks in patient discharge summaries coming from different units of French hospitals. In this specific application, temporal coordinates are associated to infection occurrences in order to help to determine if these infections are acquired in hospital or not.
Once again the temporal dimension in IE is a necessary first step for knowledge discovery and reasoning.

4) Links between concepts

Once concepts are detected and annotated with the correct background information (negation, modality, time), an important aspect is to be able to see how these concepts are related and to identify the kind of relations holding between them. Concepts and relations build complex events. An event is thus a frame in which participants consist of terms of interest which are linked according to given types of relations. These links are supported by syntactic dependencies that hold between the concepts. These dependencies are extracted using NLP techniques.

1.2.20. Data mining

Distributed Data Mining

Many tools and algorithms for the analysis of clinical data exist in standardized toolkits such as Biomoby[185] and R[186]. Unfortunately, in practice it is often not straight-forward to scale data analysis tools by one or more orders of magnitude, as may be necessary in the envisioned EURECA scenarios. New approaches like the map-reduce paradigm promise to solve the scalability problem, but require a radically new way of re-implementing existing tools. First steps have also been made with respect to the integration of knowledge discovery tools in distributed environments such as the Grid[187]. However, despite all of this new approaches, it is still up to the researcher to collect and integrate data, take care of administrative and privacy issues, and take care of problems of data size and computational complexity. Hence, a clear need for some automatic support in executing data analysis exists.
An innovation that will be developed in EURECA will be to integrate support into a data mining framework for automatically and transparently mapping data mining processes from the conceptual level to the appropriate implementation. Based on an appropriate meta data and modelling of (1) the data mining process, (2) available data, (3) available resources, and (4) relevant policies, and appropriate monitoring of data mining process execution, the system will be semi-automatically configurable to very diverse scenarios, ranging from

[181] Morante, R.: Descriptive analysis of negation cues in biomedical texts In: Proceedings of the Seventh International Language Resources and Evaluation (LREC'10), pp. 1429—1436, Valletta, Malta (2010).
[182] Aït-Mokhtar, S., Chanod, J.P., Roux, C.: Robustness beyond Shallowness: Incremental Deep Parsing. Natural Language Engineering, 8, pp.121--144 (2002).
[183] Zhou, L., Parsons, S., Hripcsak, G.: The Evaluation of a Temporal Reasoning System in Processing Clinical Discharge Summaries. Journal of the American Medical Informatics Association (JAMIA), v5(1) pp. 99--106 (2008).
[184] Hagège, C., Marchal, P., Gicquel Q., Darmoni, S., Pereira S., Metzger, M-H.: Linguistic and Temporal Processing for Discovering Hospital Acquired Infection from Patient Records. In Proceedings of the 2nd International Workshop on Knowledge Representation for Health Care (KR4HC-2010), Lisbon, Portugal. (2010).
[185] M. D. Wilkinson, M. Links, BioMOBY: an open-source biological web services proposal, Briefings in Bioinformatics 3:4 (2002), 331-341
[186] R Development Core Team, R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, 2008, http://r-project.org.
[187] D. Wegener, Dennis et al., GridR: An R-based tool for scientific data analysis in grid environments. Future Generation Computer Systems : The International Journal of Grid Computing: Theory, Methods and Applications, 2008.

quick local data analysis at a hospital site to the monitoring of complex data from a multi-centric trial over the course of several years, all without any changes on the user side.

Pattern Recognition on Data Streams

One particular challenging task in EURECA will be the efficient long-term monitoring of patterns in clinical data. An example of a clinically relevant pattern could be the co-occurence of a new drug with an unexpected increases or decreases in mortality. Such a generic search for patterns requires a solution which

- can efficiently scan for a high number of patterns in parallel
- can operate on very large sets of data
- is able to detect patterns which may change over time
- is able to process user feedback about the interestingness of pattern in order to focus on new, clinically relevant patterns
- is readily applicable by end-users without extensive data mining experience

While approaches exist for each of these requirements, an overall approach that effectively deals with all of these requirements is lacking. Existing monitoring solutions are usually not very efficient, with the effect that only very basic statistical procedure can be used. In the field of data stream mining[203] efficient solutions exists, but often only for specific analytic problems. In EURECA, more generic implementations will be tested based on recent developments in distributed data mining[204]. In the end, this will allow to follow more patients over longer periods of time, while looking out for more possible risk patterns.

1.2.21. Semantic reasoning and semantic interoperability

Executive summary: technology is available for representing the knowledge models required for EURECA's purposes, as well as for manipulating this knowledge ("inference") at sufficiently large scales. However, substantial progress is still needed at the level of *semantic* interoperability between the highly heterogeneous knowledge sources that EURECA will have to deal with.

Representation and inference: Two decades of foundational research on Description Logics and database logics (e.g. DataLog), and a decade of standardization work by the World Wide Web Consortium (W3C) has by now resulted in widely adopted standards for representation of web-enabled data and knowledge: Resource Description Framework (RDF) is the de-facto serialization and is by now finding wide deployment in such diverse sectors as life sciences (HCLS working group), bio-medicine (GO, PubMed), the media (BBC, CNN), government (US, UK), and others.

This widely adopted serialisation format is used to serialise knowledge expressed in a variety of logics, ranging from the very simple (RDF Schema) to highly expressive Description Logics and a large number of possibilities in between (e.g. Description Logic Programming and others). These logics are all well formalised, their properties are mathematically well understood, and they are of very different computational complexity. Furthermore, well engineered engines have been implemented for a large spectrum of these logics. Many of these implementations have started as academic prototypes, but have in the meantime become mature commercial products (Jena, Sesame, OWLIM, Virtuoso, 4Store, Racer PRO, Pellet and many others).

These engines implement both inference (i.e. deriving conclusions that are universally valid given the validity of the axioms) and the computationally simpler querying (i.e. deriving new facts about a particular situation given a set of facts describing that particular situation).

Due to the W3C standardisation work, it is possible to move smoothly between the different logics, and even to combine axioms expressed in different logics. This so-called "layering" is possible through exploiting the widely adopted serialisation syntax and the compatible semantics that these logics have been given.

As a result of EU funded efforts such as LarKC (the Large Knowledge Collider)[188], engines for these different logics are now available for a variety of scales and a variety of logics. Available scalability ranges over 7 orders of magnitude, varying from hundreds of axioms and instances to billions of axioms and instances for sub-second response times.

Semantic interoperability can be defined as the ability of two or more computer systems to exchange information in such a way that the meaning of that information can be automatically interpreted by the receiving system accurately enough to produce useful results to the end users of both systems. According to a recent US report[189], some approaches to semantic interoperability are successfully applied in limited settings. However, many of the increasingly sophisticated needs of clinicians and researchers are not, and will not be, met by these approaches, since the realization of the goals of modern translational medicine requires semantic interoperability which spans scientific domains and national boundaries. With the vast amount of data becoming available and exchanged, the challenge now is to ensure that transmitted data are

[188] http://www.lark.eu
[189] http://ontology.buffalo.edu/medo/Semantic_Interoperability.pdf

understood not only by the human beings on both ends of the IT communication channel, but also by computer systems and their associated software. There is a growing need for systems that are able to act and react automatically to changes in the data repositories to which they have access. Only thus will we have the opportunity to avoid data overload on the side of the end user. Currently many health and life science databases, including ontologies, terminologies and electronic health records (EHRs), are organized in ways that only fulfill the needs of the original designers, but have little chance of bringing benefits to the research community at large. Thus, to give only one example, resources designed to support semantic interoperability in the experimental biology and clinical trial domains do not support interoperation with counterpart resources developed in the contexts of hospital care and general practice

Linking between vocabularies, establishing semantic connections between decoupled content, is a key element in transparent access to heterogeneous data sources, both by allowing cross-discipline querying through vocabulary-based query translation, and by driving more traditional data integration. Ontology mapping, as the most commonly used technique for this task, is a very active research field, which has its own yearly workshops (OM, OEAI), and for which recently an excellent textbook was published (Euzenat & Shvaiko). The field has made much progress in recent years, but is far from solved, with current recall and precision rates not surpassing the 80% range.

Semantic interoperability between multiple data sources (such as patient-records, guidelines and clinical trials) requires not only interoperability at the vocabulary level ("ontology matching"), but also requires to establish matches between the individual objects that occur in these datasets. In well curated collections, care is taken that each digital object has a single unique identifier. Such an identifier is then used throughout the collection to denote the object, to attach metadata to the object, to pass the object between applications, etc. This assumption of 'a single unique identifier per object' no longer holds in heterogeneous and distributed collections, and can no longer be 'legislated away'. Not having a unique identifier per object causes significant problems: can metadata attached to different identifiers for the same objects be combined? EU funded efforts such as OKKAM[190] have provided infrastructure that can be exploited for the purposes of EURECA.

1.2.22. The EURECA clinical pilots

Our clinical partners have preliminarily selected seven specific clinical scenarios to be used for the evaluation and validation of the EURECA semantic interoperability environment and software services/tools. These scenarios will also enable us to evaluate the performance our solutions compared with current practice.

As mentioned before, our clinical partners have home-grown EHR systems and a heterogeneous landscape of clinical research systems. This is widely common among large European clinical research/academic centres, where in order to support research large investments in ICT infrastructures and tools were necessary. Most commercial EHR systems currently provide little support for clinical research and a low level of integration with legacy systems.

Each of the defined scenarios include the use of the EURECA interoperability framework and of a variety of EURECA services in different configurations. Each EURECA software service will be validated in at least one scenario.

1. A clinical-research oriented scenario proposed by the Maastricht Radiation Oncology clinic aims at exploiting external clinical data jointly with the in-house EHR and CDMS data in order to validate a predictive model of radiation therapy outcomes (survival, oesophagitis and radiation-induced lung damage) which was derived from findings at MAASTRO. This "scientific validation" scenario will in turn be used to validate the EURECA infrastructure in that it will require: a) setting up a prototyping environment at the hospital EHR (with the associated legal and technical constraints), b) building an information model for the EHR and CDMS systems, c) building a mapping of the site-specific information model to the EURECA core set, d) carrying out coding and structuring of data, e) deploying prototypes in clinical setting and f) preparing and executing the data mining validation scheme for the predictive model of radiation outcome. This scenario will be used to assess various EURECA metrics, such as the fraction of data retrievable from EHR to CDMS, or the fraction of automatically identified candidate patients that could actually be successfully enrolled in a trial.
2. A patient-recruitment scenario based on the EHR infrastructure of Jules Bordet Institute, demonstrated in the context of a multicentric clinical trial led by the Breast International Group. This scenario will involve implementation steps similar to scenario 1, in an environment involving an EHR system (Oribase) with a language (French) potentially different from the clinical-trial systems (English or French). Special care will thus be put on cross-language semantic mappings. Next to interoperability and efficient recruitment, this scenario we will also demonstrate automatic storing of

[190] http://project.okkam.org

clinical trial eligibility criteria into the EHR (for automated recruitment or criteria checking) and the use of the EHR for automatic Electronic Data Capture. The steps required will be: a) identifying the core eligibility criteria for adjuvant breast cancer clinical trials, b) mapping the core eligibility criteria with the corresponding elements of the EHR (with the associated legal and technical constraints), c) coding the elements extracted from the EHR, d) importing the criteria in the CDMS system.

3. A patient-recruitment scenario based on the infrastructure of GBG/ Luisenkrankenhaus Düsseldorf demonstrated in the context of a multicentric clinical trial. This scenario will involve implementation steps similar to scenario 1, but in an environment involving an EHR system with a language (German) potentially different from the clinical-trial systems (English or German). Special care will thus be put on cross-language semantic mappings. As above, in this scenario we will also demonstrate automatic storing of clinical trial eligibility criteria into the EHR and use of the EHR for automatic Electronic Data Capture. This scenario may be combined with scenario 2 in a multicentric setting.
4. A scenario at Oxford University, concerning large-scale data-mining of genomic trials and care data to generate and test research hypotheses with the aim to maximize patient benefit (refined treatment selection for patient) while minimizing risk through better identification of risk factors (risk of SAE, late effect of therapy) and earlier disease detection. The Oxford University scenario is in the context of the EuroSarc trial network (under evaluation, follow-up project of EuroBoNeT) on translational and clinical research in rare bone carcinoma. Selection of patients for treatment will be based on gene-expression signatures and other biomarkers. Research aspects include: a) Correlating gene expression profiling of osteosarcoma, Ewings sarcoma, chordoma and Giant cell tumours of bone with patient health records including late effects, b) Improving identification of patients for clinical trials across Europe, c) Using patient health records to evaluate diagnostic delays, treatment plans and late outcomes for patients with rare bone sarcomas, e.g. effects of referral pathways, chemotherapy, radiotherapy, radiological surveillance and trial therapy.
A clinical-care-oriented application will be developed in the identification of treatment options and relevant clinical trials for patients with rare bone sarcomas and aged over 40.
5. A breast cancer translational research scenario at UOXF: CRUK Molecular Oncology groups are involved in laboratory work and retrospective clinical studies where prognostic molecular signatures and single markers associated to key cancer pathways (e.g. hypoxia, angiogenesis) are derived, validated and assessed for their prognostic potential. PhaseI-II trials, and small drug-response studies, are conducted in the unit. The scenario involves the validation of these signatures, and laboratory findings, by exploiting external clinical trial data jointly with the in-house data. The UOXF is participating to the construction of a repository of on-going trials in breast cancer in the FP7 P-Medicine project, thus this resource could be directly exploited in EURECA both for validation of the developed signatures and also could be mined jointly to the in-house data. The requirements for this scenario are similar to those of scenario 1 with the additional use of genomic and genetic databases and services, pathology and immunohistochemistry databases from pathology. The databases/information systems involved in this scenario are:
    a. The ORH Trust Patient Administration System is used to register patient and record their outpatient and inpatient events, time on waiting lists etc; from that the administration of clinics is undertaken, coders also record diagnoses and procedures undertaken, for central reporting of activity and performance.
    b. A Laboratory Information management System (LIMS) is used for results like bloods and radiology reports.
    c. For cancer patients, the clinical annotations and discharge summaries are outsourced, and available through an electronic system.
    d. An electronic chemotherapy prescribing system.
    e. clinical trial unit electronic case report software with facility of GCP compliance based on OpenClinica
    f. pathology database: in-house database build on file-maker pro
    g. Custodianship of samples LabVantage SAPPHIRE^{TM} (commercial software)
    h. Genomic research: in-house databases, in-house and external web services and repositories of data/workflows
6. A safety scenario demonstrating the effect of interoperability on avoiding multiple data entry when reporting SAE/SUSAR events, based on the EHR infrastructure of Saarland University Hospital. This scenario will be implemented an environment involving an EHR system with a language (German) different from the clinical-trial system (English). Special care will thus be put on cross-language

semantic mappings. Our semantic interoperability layer will link the hospital's home-grown EHR and ObTiMA, the open source Clinical Trial Design and Data Management System developed in the ACGT project.

7. A scenario showing the use of PHR data to improve safety, implemented with the Tolven PHR system, integrated in a safety-detection and reporting scenario (see UdS scenario 6) or a data mining application for identification of safety risks (see scenario 4).
8. Protocol feasibility and hypothesis generation scenarios in the context of a multi-centric breast cancer clinical trial of the Breast International Group. This pilot will be carried out together with IJB and GBG/ Luisenkrankenhaus Düsseldorf.
9. A trial execution scenario in the context of a multi-centric breast cancer clinical trial of the Breast International Group and of the German Breast Group. This pilot will be carried out together with IJB and Luisenkrankenhaus Düsseldorf. We demonstrate in the context of a previous trial that using EURECA solutions in this multicentric setting would avoid multiple data entry and collect data faster, so make the data and trial management faster.

1.2.23. The EURECA baseline

Data availability

Our experience teaches us that the timely availability of a sufficient volume (and variety) of clinical data is essential to the success of ICT research projects in the healthcare and clinical research domain. In order to ramp up the project in the shortest time we have made sure that data will be available from the beginning of the project. This data will be used in the development of our tools and in evaluation and validation.

- We already have access to significant amounts of anonymized patient data from clinical trials, in full compliance with the legal, ethical and security requirements. This data was collected and made available during the ACGT**Error! Bookmark not defined.** project following all the required regulatory steps.
- Moreover, a database of all clinical trials approved by IJB Ethics Committee is maintained by IJB including a description of their eligibility criteria, and will serve as a starting point for providing models for clinical trial description.
- The database of the Translational Research Unit at the Cancer Research UK Oxford Cancer Centre will provide models for translational clinical studies description.

Additionally, already before the start of the project our clinical partners have committed to give EURECA access to the data of the following completed trials and studies:

- Safety Study of Individualised Radiation Dose Determination for Lung Cancer Patients.
  Sponsor: Maastricht Radiation Oncology
  Information provided by: Maastricht Radiation Oncology
  ClinicalTrials.gov Identifier: NCT00181545
- A clinical trial testing Rapamycin, an mTOR-inhibitor, in combination with preoperative radiotherapy in operable rectum cancer: A phase I-II study.
  Sponsor: Maastricht Radiation Oncology
  Information provided by: Maastricht Radiation Oncology
  ClinicalTrials.gov Identifier: NCT00409994
- Prospective multicentre and randomized trial for children with nephroblastoma (SIOP 93-01/GPOH)
  Sponsor: Saarland University/ Paediatric oncology
  Information provided by: Saarland University/ Paediatric oncology
  ClinicalTrials.gov Identifier for United Kingdom only: NCT00003804
- Adjuvant Chemotherapy With Sequential or Concurrent Anthracycline and Docetaxel: BIG 02 – 98 Randomized Trial[191]
  Sponsor: Breast International Group
  Information provided by: Breast International Group
- Predicting the efficacy of anthracyclines in breast cancer patients: Neoadjuvant TOP trial[192]
  Sponsor: Breast International Group
  Information provided by: Breast International Group
  ClinicalTrials.gov Identifier: NCT00162812

[191] JNCI, Vol. 100, issue 2, 2008.
[192] http://www.asco.org/ascov2/Meetings/Abstracts?&vmview=abst_detail_view&confID=65&abstractID=33916

- Phase I/II trials in Ewing sarcoma, including combination studies based in major centres for rare sarcoma in Oxford, Leiden, Bologna and Munster. .
  Information provided by: Oxford University
- Phase II trial in chondrosarcoma, based in the above centres and extending to EORTC centres
  Information provided by: Oxford University
- Phase IIa trials of MePACT in metastatic osteosarcoma.
  Information provided by: Oxford University

It is important to note that for all of the above all necessary administrative, legal and ethical steps have already been cleared out and the data can be used in our project. Additional data will become available during the project. Our clinical partners are committed to also provide access to EHR data, following a legislation-compliant process similar to the one developed and used in ACGT.
The core data sets developed in the project will be tailored to the clinical needs and research interests of our clinical partners, examples of which may include breast cancer, bone sarcoma, nephroblastoma, tumor hypoxia, and lung and colon with focus on radiotherapy delivery. The development of core data sets for several cancer types will allow us to validate the generality and the scalability of the approach. While some core data sets will be comprehensive (e.g. breast cancer), others will be limited to the relevant EURECA application (e.g. lung cancer with respect to radiotherapy). We will provide tools supporting the users to easily extend a particular core data set to cover new scenarios that suit their needs.

Prior interoperability results of the consortium
One of the partners of the consortium, **MAASTRO**, has already performed experiments with interoperability scenarios. The successful outcomes of these experiments support the feasibility of EURECA's envisioned approach.

- In one experiment[193] an in-hospital Computer Aided Theragnostics data warehouse had been constructed, aggregating a wide variety of in-hospital clinical content. The data warehouse aggregates data from the EHR, the Radiotherapy PACS, and a Record and Verify system (to provide delivered treatment parameters) in order to collect care data for research purposes. For the benchmark, two sample sets of variables are setup for two specific diseases, namely non-small cell lung cancer (NSCLC) and rectal cancer. Collection times and data inconsistencies between manual collection and automated extraction were determined for both cancers. The average collection time for the NSCLC data using manual extraction was 10.4±2.1 min per case, compared to 4.3±1.1 min per case for the automated extraction ($p<0.001$). For rectal cancer, automated extraction improved the collection time from 16.5±4.1 min. to 6.8±2.4 min. per case ($p<0.001$). In 3.2% of the data collected for NSCLC and 5.3% for rectal cancer, there was a discrepancy between the variable values found with the manual and automated extraction. In most cases we found that it was the automated extraction that returned the correct data values. The results shows that using a data-warehouse improves data quality and shortens data-collection time significantly, compared to a manual process of data collection.
- In another experiment (to be published), an ad-hoc data sharing infrastructure is set up between the radiotherapy departments of the Policlinico Universitario Agostino Gemelli in Rome and the Maastro Clinic in Maastricht. The infrastructure – based on open source tools – currently allows the exchange of three types of information. Clinical data (e.g. demographics, TNM-stage, date of diagnosis, histopathology) and outcome data, diagnostic imaging data (e.g. diagnostic and follow-up PET, CT and MR imaging), and Radiotherapy treatment data (cone beam CT's, orthogonal EPID imaging, delivered fractions). The data is imported into a research database, which contains a patient centric data model (containing medical data and imaging metadata). A pseudonymisation scheme is used to ensure patient privacy, while maintaining the patient – data link for the site providing patient care. Italian local terms have been mapped to SNOMED Clinical Terms to break the language barrier. In November 2010, the research database contained the data of over 2300 patients, with over 5000 de-identified imaging studies. The infrastructure is currently actively used in the Thunder trial (NCT00969657), will be used in the "Radiomics in NSCLC" trial (NIH U01-CA143062-02), and is used to test hypotheses on historical data and assists in patient selection estimation in order to formulate new prospective clinical trials (hypothesis generation & feasibility).

These experiments show the impact of EURECA's vision on clinical research and healthcare - improving data quality, improving data collection time (greatly reducing the load of the clinical personnel), and improving

[193] "Advantages of a data warehouse for radiotherapy research", L Persoon, S Nijsten, Ph Lambin, A Dekker, Proceedings of the XVIth International Conference on the Use of Computers in Radiation Therapy, 2010.

protocol feasibility and clinical trial hypothesis generation – and at the same time indicate the feasibility of the suggested approach.
Another partner in the consortium, **IJB** has designed and implemented Oribase, their own Electronic Health Record software, focusing from the start on semantic interoperability norms, in particular HL7 and Clinical Document Architecture (CDA) formats in a production environment. Additionally, they have carried out work on the development of models for EHR data, collected at various key points in the patient's clinical process, that were coded and normalized using international coding systems (as LOINC, SNOMED CT, ICD9 and ICDO) for domains relevant to screening for inclusion a in a clinical trial and adverse event reporting. This work has enabled them to experiment with semantic standards-based linkage among different components in their hospital environment. Being convinced of the benefits of such an approach, IJB aim to extend it to the scale of their entire clinical care and research environment and to contribute to the systematic development of a scalable and consistent interoperability framework based on widely used standards and on standardized terminologies. Next to increasing the efficiency within their institute, the EURECA semantic interoperability framework will enable them to improve the efficiency of their numerous research collaborations with other European hospitals.
Several of the clinical and technology partners of EURECA have jointly carried out the ACGT project (**FORTH, Philips, UdS, FhG, Custodix, UOXF, UPM**). ACGT has been an Integrating Project (IP) funded in the 6th Framework Program of the European Commission under the Action Line "Integrated biomedical information for better health". Its high level objective has been the development of methods and systems for improved medical knowledge discovery and understanding through integration of biomedical information (e.g. using modeling, visualization, data mining and grid technologies). The over-arching objective of the ACGT project has been the provision of a unified technological infrastructure which facilitates the seamless and secure access and analysis of multilevel clinico-genomic data enriched with high performing knowledge discovery operations and services.
During the course of its life, the project has defined a detailed architectural blueprint and has developed, tested and validated a range of novel technologies, including:

- A domain-specific Master Ontology (MO) on Cancer, built on established theoretical foundations and taking into account current initiatives, existing standard data representation models, and reference ontologies. The ACGT MO is meant to constitute a reference ontology for the cancer domain, and its objective is to enable semantic data integration. The ACGT Master Ontology (ACGT-MO) is implemented in OWLDL, and can be freely downloaded from http://www.ifomis.org/acgt.
- A biomedical infrastructure offering seamless mediation services for sharing data and data processing methods and tools; a key element of this infrastructure is a Semantic Mediation (SM) layer and services providing clients with a seamless interface for integrated querying of heterogeneous data sources and standards for exposing the properties of local sources in a federated environment.
- An open source, ontology based Trial Builder and Management System (ObTiMA), supporting efficient design and execution of multicentric post-genomic clinical trials, and enabling researchers in performing cross trial analysis. In supporting the whole life cycle of a clinical trial, ObTiMA utilizes the features provided by the ACGT-MO and the ACGT-SM.
- Advanced security tools including anonymisation and pseudonymisation of personal data in accordance to European legal and ethical regulations; In doing so, the project has established the Center for Data Protection (CDP) –a non-profit organization under Belgian law. The CDP is a fully operational entity data protection related services (e.g. act as data controller or contact point for patients to enforce their rights) to European research projects dealing with medical data. The CDP is committed to support EURECA with its services where necessary.
- A large number of bioinformatics data analysis and literature mining services that supported the ACGT ontology-based semantic mediation layer and its services and improved complex knowledge discovery processes; Service discovery and orchestration was made available in ACGT via an easy to use, web-enabled workflow editor enabling efficient design and controlled sharing of discovery driven eSience workflows.

All these results constitute relevant expertise for EURECA and will allow us a quick start. This past work convinced us of the feasibility of achieving semantic interoperability in the healthcare domain in a modular, scalable way, based on standards and widely used ontologies/terminologies. Our belief in the feasibility of semantic linkage based on core sets of concepts (**core data sets**) out of widely-used, community-built ontologies/terminologies such as SNOMED has originated in ACGT. This semantic interoperability based on a core set of concepts describing a domain is also at the centre of our approach in the FP7 INTEGRATE project described in section 1.2.24.

Data infrastructures, semantic technologies and interoperability

In various communities, effort is invested in the semantic annotation of data. EURECA, however, has the ambition to encompass a multitude of domains in the biomedical field. It should span clinical care, basic research and clinical research. Each of these domains has a variety of terminologies, standards and ontologies, which should ultimately be interoperable in order to bring maximum benefit to healthcare.

For instance, various partners of this consortium have gained a lot of experience in the ACGT project with the integration of heterogeneous databases used in a clinical trial context. In ACGT, the ACGT Master Ontology (http://www.ifomis.org/wiki/ACGT_Master_Ontology_%28MO%29) was developed to describe the domain of cancer research and management, and this ontology has successfully been used to integrate various heterogeneous data sources. The fundamental "operational assumption" in ACGT has been that "data is left at the point it is produced" and integrated as and when needed. EURECA requires a different approach with respect to data harmonization and management. Relevant to this is SNOMED CT, which is used in the clinical practice as terminology in electronic healthcare systems to allow appropriate retention, processing and exchange of unambiguous clinical records.

In order to provide seamless data access, syntactic and semantic integration needs to take place.

Syntactic data integration handles differences in formats and mechanisms of data access, the fact that information can be represented in different ways, using different terms and identifiers. In ACGT we have implemented syntactic data access services to access relational databases, DICOM image repositories and BASE microarray databases. Relational databases can also be used to access data not yet stored in a relational database, but that can be mapped to the relational data model. This holds for data collected in files of various formats, such as Excel files, plain text files, XML files, etc. The ACGT data access services export the structure of the database using a common data model, together with possible query limitations of the data source. An RDF Schema of the data resources is exported on demand. Clients may use this information for constructing queries, e.g. the semantic mapping editor uses this schema to provide the mapping to the ACGT Master Ontology. Finally, the data access services enforce the data source access policy, and audit access to data sources. The suitability of the services developed within ACGT to access relevant external data sources will be evaluated in EURECA, and they will be extended when necessary. However, we expect that new services will need to be developed to access data sources of types not relevant within ACGT, when these are identified as essential by the EURECA clinical scenarios.

With the dramatic growth of data generation within the biomedical field, information gathering has become one of the main challenges to translate research results to clinical practice. During the last years, projects such as INFOGENMED (2002-2004) or ACGT (2006-2010) have worked on infrastructures for heterogeneous biomedical data integration. For each scenario, a comprehensive analysis of syntactic and semantic heterogeneities has been carried out. Web-based applications and standard ontologies and vocabularies have been applied to provide homogenous data retrieval interfaces to access heterogeneous and decentralized data. Although certain software developments are reused for each situation, a highly dynamic field as biomedicine requires a constant update and adaptation to new requirements and technologies for each scenario.

Within the EURECA project, previous systems of data integration will be adapted and updated to the new scenarios and technologies available, introducing new approaches whenever it is necessary.

The first phase will be to analyse new technologies available and requirements of the project, including common heterogeneities and how are they solved. Previous developments from the ACGT project will be the basis from which to introduce new approaches to the EURECA solutions:

- The semantic mediator, responsible for receiving user queries, automatically generating the corresponding queries of each data source and unifying the results to be presented to the user. Originally, it has been envisioned to work over an OGSA-DAI GRID framework, RDF wrappers of database schemas and SPARQL queries. We will analyse the suitability of reusing components of the core of the system, according to the final requirements and user needs. In addition, variations within the mapping format will require modifications in the query translation process.
- The mapping format / API, for formally defining a correspondence among physical objects of a data source and the standard ontology selected within the project. The format was designed to store correspondences between pairs of RDF schemas. The mapping format and its associated java API will be evaluated for reuse in EURECA. If relevant, they will be adapted and enhanced using innovative approaches for each particular task to be defined with other WPs.

EURECA will base its approach on reusing already existing ontologies, vocabularies and terminologies to leverage the effort already invested. To enable scalability and usability the development of the core set of concepts will be modular, per (sub-)disease domain, covering the wealth of data from post-genomic clinical trials in breast cancer that constitute our initial domain. The use of supported

terminology standards, such as SNOMED CT, ICD-10 and LOINC, will facilitate in the future easy linkage to other clinical systems adhering to those standards.

1.2.24. Synergy and close collaboration with relevant initiatives

The EHR4CR project

The EHR4CR project (**Electronic Health Records for Clinical Research**) is a 4 year project funded under the Innovative Medicines Initiative (IMI)[194] that will kick off in the first quarter of 2011. The EHR4CR consortium consists of 11 pharmaceutical companies (members of EFPIA) and 22 public partners (academia, hospitals and SMEs).

The EHR4CR (Electronic Health Records for Clinical Research) project aims to design and demonstrate a scalable and cost-effective approach to interoperability between Electronic Health Record systems (EHRs) and Clinical Research through multiple but unified initiatives across different therapeutic areas, with varying local and national stakeholders and across several countries under various legal frameworks. This unified approach will be made possible by the development of both an EHR4CR business model and an EHR4CR platform. This platform will:

- Support the feasibility, exploration, design and execution of clinical studies and long-term surveillance of patient populations;
- Enable trial eligibility and recruitment criteria to be expressed in ways that permit searching for relevant patients across distributed EHR systems, and initiate confidentially participation requests via the patients' authorised clinicians;
- Provide harmonised access to multiple heterogeneous and distributed clinical (EHR) systems and integration with existing clinical trials infrastructure products (e.g. EDC systems);
- Facilitate improvements of data quality to enable routine clinical data to contribute to clinical trials, and importantly vice versa, thereby reducing redundant data capture.

The platform will be implemented as a common set of tools and services that will be able to integrate the whole lifecycle of clinical studies with heterogeneous clinical systems, including data extraction and aggregation, de-identification and linkage, security, and conformance to regulatory requirements. Working closely with the 11 EFPIA partners, the consortium will address 4 clinical research scenarios with services to be developed for:

- Clinical protocol feasibility
- Patient identification and recruitment (from the perspective of a pharmaceutical company initiated trial)
- Clinical trial execution (re-use of EHR data for Electronic Data Capture)
- Adverse event reporting

EURECA – EHR4CR link

It is clear from the above that the EHR4CR project addresses for a large part the same topic as the ICT-2010.5.3.b "Tools and environments enabling the re-use of electronic health records" Work Programme.

The main scientific challenge for both projects is the same: the semantic integration of EHR and clinical data management systems data (allowing for increased re-use of EHR data). However, the application scope is different in the two projects. The EHR4CR solution focuses strictly on the re-use of EHR data for research of pharmaceutical companies (drug clinical trials). The objectives of the project are relatively narrow, focussing on four clear, well aligned application scenarios.

EHR4CR is of great importance because of the impact that its outcome can have because of the direct cooperation of the pharmaceutical industry. Proven solutions coming out of the EHR4CR project are certain to be picked up by the industry because they will have been designed according to their requirements. Moreover, business development is a very important aspect within the project.

[194] The Innovative Medicines Initiative (http://www.imi.europa.eu/) is a unique public-private partnership designed by the European Commission and European Federation of Pharmaceutical Industries and Associations (EFPIA). It is a pan-European collaboration that brings together large biopharmaceutical companies, small- and medium-sized enterprises (SMEs), patient organisations, academia, hospitals and public authorities. The initaive aims to accelerate the discovery and development of better medicines by removing bottlenecks in the drug development process. It focuses on creating better methods and tools that improve and enhance the drug development process, rather than on developing specific, new medicines.

The EURECA project addresses a broader set of scenarios, focussing on applicability of its infrastructure in the whole clinical research domain (beyond drug trials for the pharmaceutical industry) and even within the care domain. Next to enabling more effective and efficient execution of clinical research, EURECA addresses solutions for early detection of patient safety issues, faster transfer of new research findings and guidelines to the clinical setting and providing healthcare professionals with information management (extraction of relevant information in a patient's context).
In fact, the projects are largely complementary. EHR4CR addresses a narrow set of application scenarios, and focuses in piloting a commercial environment (introducing a huge procedural load). EURECA aims to serve a wider community, focusing more on a broader set of scientific and technical challenges.
The EHR4CR partner in charge of the EHR4CR technical work package group (the cluster of all technical WPs in the project) and more in particular of the overall EHR4CR architecture, fulfils a similar role in EURECA as leader of the architectural design WP and as Lead Architect within the Technical Board. This situation provides the unique opportunity to ensure that both technical solutions (platform architecture) developed by respectively EHR4CR and EURECA are at least interoperable, and at best based on the same integration framework (allowing service to be interchanged between projects).
This link on the technical management level could significantly increase the impact and value of the project work of both projects. Additionally, the chairman of the EHR4CR advisory board (Manolis Tsiknakis) is involved in the EURECA project in a majority of the technical tasks through one of the EURECA important partners. In this position, he will be able to further stimulate the exchange of knowledge between both projects.
Obviously the scientific work of the two projects would benefit from a closer cooperation and joint work on overlapping aspect could free up resources for more in depth scientific work on other areas in the projects. In particular, the EURECA project could benefit from the leveraging power of the pharmaceutical industry, EHR4CR can benefit from the broader scope of the EURECA project. The strong link on the technical management level will increase the synergy of the projects and can considerably reduce work overlap (increasing Return On Investment).
The clear benefits created by this link between EURECA and EHR4CR has also been recognised by the pharmaceutical industry itself. The three key persons responsible for scientific and technical work in EHR4CR working on the EFPIA[195] side (from: Amgen, Sanofi-Aventis and Johnson and Johnson) will be active in EURECA as member of the Pharmaceutical Advisory Board. They will ensure that EURECA is aware of the requirements of the pharmaceutical industry and at the same time they will ensure that EHR4CR can maximally benefit from EURECA findings.

The INTEGRATE project

Driving Excellence in Integrative Cancer Research through Innovative Biomedical Infrastructures

The goal of the FP7 INTEGRATE project (funded under the ICT for Health programme) is to develop flexible infrastructure components and tools for data and knowledge sharing and large scale collaboration in biomedical research, to bring together heterogeneous multi-scale biomedical data generated through standard and novel technologies within post-genomic clinical trials and seamlessly link to existing research and clinical infrastructures.

INTEGRATE brings together a wide multi-disciplinary community of biomedical and clinical researchers committed to work together, to establish common methodologies and clinical protocols, to collaboratively build predictive models, carry out research and select the most suitable integrative workflows. This lasting translational research infrastructure will provide a robust, secure and flexible IT solution that will facilitate collaboration between cancer research centres and other relevant stakeholders such as basic research organizations and pharmaceutical companies, and will be validated in the context of the NeoBIG research programme of BIG.

THE INTEGRATE VISION

Within the VPH context the vision of INTEGRATE is to empower the clinical researcher with the unique opportunity to access breast-cancer multi-scale data from 29 European cancer centers and develop breast cancer predictive model methodologies related to specific clinical questions for optimizing therapy, identifying high-risk patients, etc.
Within INTEGRATE the clinical researcher will be able to retrieve temporal (e.g. before and after therapy), multi-level (e.g. from microarray to MRI/PET), data from specific population groups (e.g. postmenopausal women), with specific characteristics (e.g. that have received specific therapy regime) and then extract/develop predictive biomarkers/models (e.g. based combination of imaging/genetic biomarkers) that could answer a question such as "can these models/biomarkers help predict the specific therapy outcome for a patient in order to avoid unnecessary/costly treatment?"

[195] www.efpia.org

The lack of comprehensive, properly annotated datasets is a recognized issue in the biomedical research community. INTEGRATE aims to contribute to tackling this issue by proposing methodologies and building data sets and models based on data collected in multicentric clinical trials. On top of the data sharing infrastructure, INTEGRATE will provide tools enabling biomedical researchers to share new data, collaboratively work on the datasets and reuse analysis tools and algorithms. The project will give support for building and linking coherent and comprehensive datasets, for format conversion and annotation, and build repositories for the various types of data available from post-genomic clinical trials, for the predictive models and for the annotation and versioning of data and models.
To support ease of use and efficiency for clinical research and reduce the distance between research and clinical practice, INTEGRATE also builds semantic interoperability with existing research and clinical infrastructures relevant in the context of the NeoBIG[196] program. Similarly to EURECA, this is based on relevant standards and terminologies and on building a core data set, i.e. a well defined and consistent set of relevant domain concepts that describe the semantics of the chosen clinical domain and application (in the case of INTEGRATE, neoadjuvant clinical trials in breast cancer).

Through the INTEGRATE environment, global comprehensive datasets and predictive models out of the innovative NeoBIG trials will become available to the biomedical community. INTEGRATE closely collaborates with an important similar initiative in the US, Sage Bionetworks[31], to work together on infrastructure, tools, models, standards, and to share data.

Philips is the coordinator of INTEGRATE, and BIG, IJB, Custodix, FORTH and UPM are also partners in INTEGRATE. Together, they will ensure full interoperability between the two infrastructures and the link to Sage Bionetworks (Philips, BIG and IJB are participants in several Sage working groups).

EURECA – INTEGRATE link

There are several ways in which EURECA and INTEGRATE will benefit from each other:

- The core data set defined in INTEGRATE to cover the concepts relevant in the neoadjuvant trials in breast cancer will be based on the same standard terminologies relevant for EURECA. Not only reusing and extending the core dataset of INTEGATE to the wider breast cancer context will enable us a quick start in EURECA, but it will also give us the opportunity to test and validate the extensibility of the core data set-based approach.
- By bringing INTEGRATE into the EURECA architectural framework, the comprehensive datasets and predictive models out of the NeoBIG clinical trials that INTEGRATE builds, will become available within the EURECA environment. These unique datasets will be of high value for both the bio-medical research community and the biopharmaceutical industry and will constitute an additional incentive to join the EURECA environment and build upon our open reference infrastructure.
- INTEGRATE's sharing and collaboration tools will also become available to users in the EURECA framework helping to increase the momentum of both projects: Enhancing the EURECA environment with relevant sharing and collaboration tools for clinical researchers, and increasing the adoption of those tools in the user communities.
- INTEGRATE will provide in EURECA access to the large global network of bio-medical researchers brought together by Sage Bionetworks.
- By achieving interoperability between EURECA and INTEGRATE, also the comprehensive datasets and models build by the Sage Commons will become available to the EU clinical research communities in a semantically interoperable way. Currently Sage Commons bring together a vast community of biomedical researchers from the US and Europe to build and share common standards, data models, comprehensive datasets, and methodologies for the acquisition, processing and representation of data sets in a way that can support reuse.
- INTEGRATE shares the same goal of sustainability as EURECA and addresses the same communities of users. The two projects can work together to develop a common business model and pursue a shared exploitation path.

The euroCAT and duCAT projects
Euregional Computer Assisted Theragnostics
Cancer research has shown that the best cancer treatment is a personalized treatment in which the specific characteristics of a patient, a tumor and even areas within a tumor are taken into account. However,

[196] http://www.breastinternationalgroup.org/Research/NeoBIG.aspx

knowledge on how to identify the best treatment for an individual patient with lung cancer is largely lacking. For this, reliable personalized predictive models are needed that can predict what outcome a certain treatment will have in an individual patient. Successful cancer treatment thus requires an individual approach, in which diagnostic and treatment modalities are chosen according to the characteristics of an individual patient and tumor. The euroCAT[197] project (running since September 2010) is an Euregional Computer Assisted Theragnostics network funded by INTERREG IV. euroCAT intends to develop an Euregional IT platform for clinical cancer research in which the properties of an individual patient and their tumor are more explicitly taken into account in the development of new diagnostic and therapeutic modalities.
The euroCAT project encompasses four major tasks: building a data extraction system, setting up a clinical trial infrastructure, creation of predictive models in lung cancer and simulation of new treatments via in-silico trials. The data extraction system will extract locally available medical data from all cancer patients from multiple centers. Effective and efficient IT tools will be developed to extract, browse and query all the relevant data from heterogeneous databases. A multi-centric clinical trial infrastructure will be built in which access to the data and predictive models will be provided through a unified interface in a privacy-preserving matter. The data extraction system will be used to develop and validate machine learning based predictive models for lung cancer using data from multiple institutes. Finally, new treatments will be simulated and useful selection criteria for a clinical trial will be identified.
The euroCAT project consortium consists of the University Hospital Aachen, the Catharina Hospital Eindhoven, Interreg, Centre Hospitalier Universitaire de Liège, Limburgs Oncologisch Centrum, Maastricht Radiation Oncology, Siemens and Maastricht University.
The duCAT project (funded by the Dutch funding agency "Technologiestichting STW"[198]) is closely aligned with euroCAT. The objective of this project is, similarly to euroCAT, to develop a new IT framework, for clinical cancer research in which the properties of an individual patient are more explicitly taken into account in the development and translation to clinical trials of new diagnostic and therapeutic modalities.
The project focuses more on the development of machine-learned predictive disease models and will work on developing predictive modality models (simulating the effects of the new modality, and identifying useful patient selection criteria for enrolment into clinical trials). The consortium consists of MAASTRO clinic, Academisch Ziekenhuis Maastricht, Siemens AG, VU Medisch Centrum and Universitair Medisch Centrum Groningen.

EURECA – euroCAT & duCAT link
EURECA, euroCAT and duCAT all aim to provide on ontological layer around EHRs and clinical research information systems. By choosing widely accepted standards and ontologies such as NCI, ICD, SNOMED CT, LOINC etc. and deciding on a common core dataset for each tumor type and/or application domain all three solutions should be able to achieve a high level of semantic interoperability. This makes it possible to create semantic aware query applications that are decoupled from the actual data. Fully integrating the euroCAT and duCAT solutions on the EURECA architecture, would further even allow accessing the three networks as one.
Similar to duCAT and euroCAT, EURECA considers offering services for commercial clinical trials an important exploitation possibility and road to sustainability. The targeted market consists of pharmaceutical companies that require high quality and fast access to data, before and during drug development and subsequent clinical trials. By linking the euroCAT and duCAT networks with EURECA, a joint business model can be developed that requires only a single commercial vehicle providing access to all three networks at once.

The ENCCA project
European Network for Cancer research in Children and Adolescents
ENCCA is funded under the FP7-HEALTH programme (Action line: Structuring clinical research in paediatric and adolescent oncology in Europe). The Consortium includes most of the European Research Centers active on research in the domain of childhood oncology (34 full partners and more than 40 as associate partners). The project is coordinated by SIOP Europe. SIOP Europe (SIOPE) -with a membership of over 800 individuals- is a European organisation promoting research and optimal standards of care for children and young people with cancer.
ENCCA aims to establish a durable, European Virtual Institute clinical and translational research in childhood and adolescent cancers that will define and implement an integrated research strategy and will facilitate the necessary investigator-driven clinical trials to introduce the new generation of biologically targeted drugs into

[197] http://www.eurocat.info/
[198] http://www.stw.nl/

standard of care for children and adolescents with cancer. It is a very ambitious effort, with many challenges at various levels.
Amongst its objectives is the definition of the optimal technical infrastructure supporting such a pan-European virtual organization. This relates to the challenge of ensuring a sustained use of well-established ICT components with a simultaneous consideration of a seamless integration into a collaborative research network based on standardised interfaces.
This technical infrastructure should – amongst others:

- Increase the efficiency of implementation of pan-European multinational clinical research, in particular investigator driven trials (ICDT) in paediatric and adolescent oncology, by agreeing and adopting standardized procedures to ease the bureaucratic burden under the EUCTD, through the support the feasibility, exploration, design and execution of clinical studies and long-term surveillance of patient populations;
- Provide harmonized access to multiple heterogeneous and distributed clinical (EHR) systems and integration with existing clinical trials infrastructure products (e.g. EDC systems);
- Facilitate improvements of data quality to enable routine clinical data to contribute to clinical trials, and importantly vice versa, thereby reducing redundant data capture.

EURECA – ENCCA link

It is apparent from this short description of the objectives of ENCCA that the projects are largely complementary. Although ENCCA addresses a far broader set of challenges, the linking and reuse of clinical data for research purposes is one of the technological challenges that the project is focusing on, which at the same time represents a key objective of EURECA.
The ENCCA partner responsible for the related technical task is also involved in EURECA as a key contributor in most technical activities of the project. This situation provides an excellent opportunity for ensuring that both technical solutions (platform architecture) drafted by respectively ENCCA and EURECA are at least interoperable, and at best based on the same integration framework (allowing service to be interchanged between projects).

The TRANSFoRm Project

TRANSFoRm (Translational Medicine and Patient Safety in Europe) will develop rigorous, generic methods for the integration of primary care clinical and research activities, to support patient safety and clinical research via:

- Rich capture of clinical data, including symptoms and signs rather than just a clinical diagnosis. A generic, dynamic interface, with potential to operate with any electronic health record (EHR), will facilitate both diagnostic decision support and identification of patients eligible for research, thus enhancing patient safety;
- Distributed interoperability of EHR data and other data sources that maintains provenance, confidentiality and security. This will enable large-scale phenotype-genotype association studies and follow-up of trials;
- Software tools and services to enable use of controlled vocabulary and standardised data elements in clinical research. This will enable integration and reuse of clinical data.

The scientific aims of this project are to advance the understanding of ICT system interoperability relevant to healthcare and clinical research and to develop an EU-wide system capable of integrating primary care record systems and research systems.
TRANSFoRm focuses primarily on primary care. The project will enable real-time, within-healthrecord flagging of subjects eligible for recruitment or follow up, distributed anonymised-linked searches and extraction of data for analysis and a standards-based, open-source clinical trial data management system (CTDMS).

TRANSFoRm – EURECA link

One of the objectives of TRANSFoRm is to facilitate the use of existing care data in the clinical research domain. TRANSFoRm however focuses on primary care data, contrary to EURECA which aims to provide solutions for re-use of larger datasets from hospital and medical centre EHRs.
The TRANSFoRm project is interesting to EURECA for different reasons:

- To a large extend, the same issues associated with semantic integration of medical data are to be tackled in both projects. However the scope of information (secondary vs. primary care data) to be dealt with within EURECA is much broader and more specialist. The information modelling (and especially model management) will thus be more complex. Nevertheless, EURECA could benefit from the expertise on semantic integration gathered during the exploratory phase of TRANSFoRm in order to make the correct technology and standards choices (the project started roughly 2 years earlier than EURECA).
- Primary care records are a rich source of routine healthcare data as primary care is responsible for first contact (mandatory or not, depending on the country), continuing, and generalist care of the entire population from birth to death. The most complete (electronic) view on a patient's health status (past and present) is thus formed by combining the primary care data with hospital EHR data (and PHR). Primary care data is out of scope of EURECA, but the project will investigate how (and what effort would be required for that) to interface with the TRANSFoRm platform to unlock it as an additional data source for EURECA. This way EURECA EHR data could be contextualised (with general health information), completed (e.g. eligibility criteria for which data is typically recorded by GPs) and the functionality of EURECA extended (e.g. broader hypothesis testing).

Interaction with the TRANSFoRm project will not be restricted to concertation events. Several partners within EURECA cooperate with key partners of the TRANSFoRm project in different projects dealing with the same topic of semantic health data integration (e.g. P-Medicine and EHR4CR). This will facilitate direct, informal interaction between the researchers involved in the different initiatives, enabling in depth scientific discussion.

1.2.25. **Performance indicators of the EURECA project**

Section B1.1.3 Quantification of Results provides measurable criteria and approaches that we will use to measure the progress, results and impact of the project. In this section we link those criteria to the EURECA objectives. The progress in the major areas of innovation as described in the previous sections will also be evaluated.

Enabling semantic interoperability among EHR and clinical trial systems
This objective will be addressed by developing the EURECA semantic interoperability layer that will link research and clinical infrastructures in the EURECA environment. A measure of success will be the ability of the interoperability layer to support users of the EURECA environment, in well defined scenarios, to access relevant data out of existing research and clinical systems. Additionally, the interoperability layer needs to enable the EURECA services in the defined user scenarios, and will also be validated through the validation of the implemented services. The adherence to widely adopted standards (HL7, IHE) and the use of standardized terminologies will also be a subject of evaluation.

Enabling secondary use of care data for research
This objective will be demonstrated by enabling software services that reuse clinical care data to support clinical research, such as long term follow up and trial matching. This objective is closely linked to achieving semantic interoperability (previous objective). The success of the current objective will be measured by evaluating the services demonstrating secondary use of care data in the context of the concrete scenarios and requirements developed based on input from a wide user community. Of utmost importance for the success of this objective is compliance to all legal, ethical and privacy requirements when providing access to care data for research.

Enabling efficient recruitment by matching relevant patient data with eligibility criteria from clinical trials

This objective will be achieved by improving the current manual recruitment process by providing efficient matching algorithms. In a concrete case study for a completed trial, we will compare the performance of the EURECA matching service with the standard process. The success of this objective also relies on the quality of the EURECA semantic interoperability solution.

Enabling long term follow up of patients beyond the end of a clinical trial, for better research and improved safety
This objective will be demonstrated in concrete scenarios showing how semantic interoperability among care and research systems enables the follow up of patients beyond the end of the trial and supports data mining applications for the generation of new research hypotheses and for the identification of potential adverse events that occur outside the scope of a clinical trial. These data mining applications will be evaluated and validated with our user community in relevant clinical settings.

Improving efficiency by reducing the need for multiple data entry in the information exchange between clinical research and clinical care
This objective will be validated in the context of the previous objectives supporting recruitment and long term follow up. There are several steps in the execution of a clinical trial (e.g. eligible patient identification, entry of Clinical Report Forms, entry of lab results, adverse events reporting, etc.) in which the same data needs to be entered twice, in the EHR system at the clinical side and in the clinical trial system at the research site. We will study the current clinical workflow in clinical trials run by our clinical partners and show how the use of the EURECA services mitigate the need for multiple data entry at several points along the workflow. . As concrete case study, we will perform a “simulation study”, i.e. we will evaluate for concrete clinical trials that have been already completed by our clinical partners how many CRF fields describing information that also resides in the EHR can be automatically filled in.

Exposing a uniform presentation of clinical trial information, validated clinical trial results and other relevant external knowledge and data resources
The level to which we have achieved this objective will also be measured based on the user needs and in concrete clinical scenarios. Our environment will provide access to those external sources of information that are relevant according to the user requirements provided by WP1. Another measure of performance will be the adherence when possible to common standards and the integration through standardized interfaces

Extracting relevant clinical information from the EHR and contextualizing it to the patient case
This objective addresses the needs of clinical care, and will be evaluated and validated together with the clinical users in relevant scenarios defined in WP1. The measure of performance will be given by the ability of the EURECA services to extract out of the patient record those items of information that are relevant for the patient case. Ease of use as perceived by the clinical users will also provide a measure of success.

Building solutions for faster transfer and dissemination of relevant research results and literature to the care
This objective as well addresses the needs of clinical care, and will be evaluated and validated together with the clinical users in relevant scenarios as defined in WP1 based on the input of a wide user community. The measure of performance will be given by the ability of the EURECA services to extract out of external sources of clinical knowledge (trial results, literature, predictive models, risks models) those items of information that are relevant for the patient case under evaluation. Ease of use as perceived by the clinical users will also provide a measure of success.

Prototyping and validation of the EURECA services/modules and tools
All the previous objectives depend in their realization on the development of prototypes and their evaluation and validation with our user community. The development of prototypes and their demonstration at relevant events are also essential for our dissemination and exploitation strategies. We will measure our performance according to this objective by evaluating the integration of our prototypes into the deployment environments of the clinical pilots and by measuring the uptake of our solutions within and beyond our user community (in terms of registered and active users, number of data

downloads, etc.). Additionally, the validation of the EURECA environment and services will be measured according to the evaluation and validation procedures and criteria defined in WP8 Q&A, Evaluation and Validation.

1.3. S/T Methodology and associated work plan

1.3.1. Overall strategy and general description

In accordance to the presented vision of *EURECA* and its conceived technological architecture we have formulated the *EURECA* proposal and structured the scheduled activities, so that a coherent and integrated plan is produced.

The overall project structure is shown in Figure 14. The project consists of a number of interrelated components. The R&D work involves the integration of a number of innovation components. A continuous interaction between the research and verification components of the project is also planned for in the project plan. *EURECA* will actively seek to re-adjust its research activities and objectives based on evidence, as it becomes available from the verification component of the project. In more detail, the project components are:

I. ***User requirement Analysis and Specifications.*** In this component user needs, state of the art and key technological opportunities, issues of trust and security, as well as legal and ethical environment that determine the complex of user requirements will be identified and assessed. The initial output of this work package will be a comprehensive set of user requirements specifications covering all major research and development topics of the project. Users' requirements identification is a prior condition for the development of a demand-led service: *EURECA* will be designed and validated according to market conditions and target users needs and it will fully take into account ethical-legal features and regulatory requirements. A user-centred approach and iterative development process with participation of users in all phases will be implemented to test alternative scenarios. This project component is further divided into a number of activities (WPs), as identified by the analysis already done. Namely, WP1: User needs and WP7: Ethics, legislation, privacy & security.

II. ***The EURECA Technologies and Services.*** The aim of this major project component is to provide the necessary technology to support the project vision and objectives. This component will be continuously fed from the Portfolio Management and Evaluation with the objective of identifying additional generic tools and services to be developed as part of the *EURECA* technological framework. This project component, responsible for developing the generic *EURECA* platform and its services, is further divided into a number of activities (WPs), as identified by the analysis already done. These are: (a) WP2: Architecture, standards and integration, (b) WP3: Information extraction and data access, (c) WP4: Semantic interoperability, (d) WP5: Data mining and knowledge discovery, (e) WP6:Applications, semantic reasoning and decision support, and (f) WP9: Models, deployment & clinical pilots.

   The generic services identified are integrated according to the integration guidelines (developed by a corresponding Activity in the WP2: Architecture, standards and integration to a functional end-user services framework.

III. ***Trust and security.*** The aim of this component will be to provide the required Pan-European trust and security infrastructure for *EURECA* services. For security specification, input is also taken from the Tasks on Legal and Ethical Issues, performed under WP7. This component is responsible for specifying and implementing the overall trust and security infrastructure, and for integrating it into the *EURECA* technological framework.

IV. ***Models, deployment and clinical pilots*** This component also includes all technical activities related to evaluation. The corresponding WP will provide all the processes and expertise through which each *EURECA* pilot will be evaluated on an ongoing basis, and provided with the guidance and supervision necessary in order to optimise the impact and ultimate success of the pilot.

V. ***Horizontal Activities.*** These span the lifetime of the project and provide support to all other project Components. These include a) Exploitation and dissemination, and b) Project management. The Project management component will provide the effective and professional management services required, so that all other project components run unobstructed for delivering their objectives and hence making sure that *EURECA* as a whole does meet its stated objectives.

The Project Components themselves comprise a number of work-packages (WPs). Each WP in turn includes a number of tasks detailing the work to be carried out in *EURECA* in order to achieve its objectives. These

activities cover the aspects of the project from management through to dissemination and exploitation of the results. Each WP is adequately self-consistent and hence managed as an individual activity within the overall *EURECA* project structure. The project components are presented and discussed in more detail in the following sections.

1.3.1.1. The EURECA methodology

The *EURECA* project will cover technological, medical and legal research, service development, large-scale system integration, testing, and uptake activities in a complex interaction. Therefore a clear and well-structured methodology for the project is essential. Figure 21 depicts the main tasks across all work packages for the entire duration of the project and the relevant associated milestones.

The project has the following four phases:

Phase one: Definition

The definition phase marks the beginning of the project. There are three fundamental project objectives, which are of the utmost importance to the overall success of the project, to be addressed in the definition phase: 1) User acceptance, 2) availability of technology and 3) exploitability of the *EURECA* platform.

The requirements engineering process is a structured set of activities which lead to the production of a requirements document. Inputs to the requirements engineering process are information about existing systems, stakeholder needs, organizational standards, regulations and domain information.

The definition phase involves the following multidisciplinary and cross-sector tasks:

Creating scenarios of user behaviour and interaction with platform functionality is an extremely useful instrument for identifying key technological, security, socio-economic and business drivers for future user requirements of new work methods and collaborative work environments. The scenarios will be deduced from the domain settings defined by users and provide the framework for all subsequent user requirements specifications.

Defining **functional user requirements specifications** based on the user scenarios involves addressing seamless interoperability, openness, and specific user requirements for the selected user cases must also be integrated in the functional requirement specification.

Defining **trust and security user requirements specifications** involves identifying not only trust and security issues, but also legal and ethical issues, which in turn translate to new trust and security requirements.

Defining **societal user requirements specifications** will be done by correlating ethical, regulatory and policy issues with the deployment and wide spread use of the *EURECA* environment and services. Aspects of e.g. social acceptance, regulatory frameworks for surveillance and control of private citizens, privacy of data, governmental provisions, etc. will be addressed and integrated with the functional and trust and security user requirement to round off the package of complete and cohesive user requirements specifications for *EURECA*.

Volatility and Evolution of Requirements

Requirements change. During the time it takes to develop a system the users' needs may mature because of increased knowledge brought on by the development activities, or they may shift to a new set of needs because of unforeseen organizational or environmental pressures.

Hence, the requirements engineering process of elicit, specify, and validate will not be executed only once during system development, but rather will be returned to so that the requirements can reflect the new knowledge gained during specification, validation, and subsequent activities. The adopted requirements engineering methodology will be iterative in nature, "so that solutions can be reworked in the light of increased knowledge"[199].

[199] Sudhakar M., Managing the Impact of Requirements Volatility, Master Thesis, Department of Computing Science, Umeå University, Sweden, 2005, http://www.cs.umu.se/education/examina/Rapporter/MundlamuriSudhakar.pdf

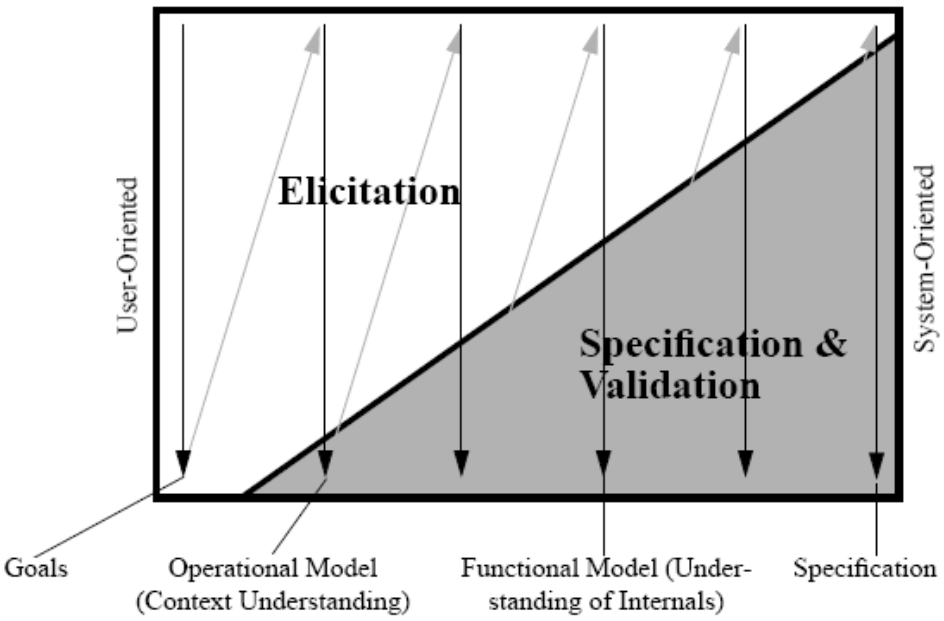

Figure 13 Requirements Engineering as an Iterative Process

An iterative process of requirements development can address the problems of volatility[200]:

> The traditional notion of the software development life-cycle with requirements capture being completed before the design stage is no longer satisfactory. Requirements capture and design are now seen to be symbiotic. The initial set of requirements needed to start off the design process is gradually refined into a systematic and coherent statement of requirements hand in hand with the refinement of design.

A user-centred approach and iterative development process (see Figure 13) with participation of users in all phases will be implemented to test alternative scenarios. Validation will show the success in terms of user quality, user preferences and acceptance.

Phase two: Research and development

The complexity of the domain which is addressed by the *EURECA* project necessitates that a spiral process of requirements analysis, elicitation, technological implementations, integration and validation is adopted. Specific techniques have also been selected for the elicitation, negotiation and agreement of requirements as well as their validation. These techniques are ***scenarios*** – during the requirements specification phase - and ***prototyping*** – during the R&D phase.

The research and development phase takes off from the initial user requirements specifications and comprise a range of integrated, multidisciplinary research and technology tasks:

Identifying technology requirements from user requirements specifications. The Consortium knows large parts of the technological requirements of the *EURECA* platform. However, the scope of the platform goes well beyond state-of-the-art, particularly in the areas of *service oriented infrastructures, distributed and privacy preserving data mining, semantic interoperability* and it will be important to ensure that the technologies behind these services are properly specified before the platform development begins. In particular the integration work in building the platform needs to be specified, including agreeing on interfaces, standards, and establishing security levels.

Identifying current offerings encompass research of existing technologies available from consortium partners, from commercial off-the shelf vendors and from other EU, international and/or national research projects. Identifying gaps and specifying research needs involves a comparison of needed and available

[200] Sutcliffe A., Scenario-Based Requirements Engineering, p. 320, 11th IEEE International Requirements Engineering Conference (RE'03), 2003.

technologies leading to a set of specifications for new research and development demands for the *EURECA* platform.

Designing a prototype platform involves constructing a UML description of the software architecture. It also involves prototyping new tools, and services to be used in the user scenarios.

Designing a business environment, which can provide the framework for adoption and exploitation, including identifying actor roles, value creation, value nets as well as analysing user acceptance through the conduction of open Awareness Workshops.

Phase three: System integration and Testing

With the successful completion of all tasks in the research and development phase, the project has reached the stage, where realization of the platform and installation in user environments are possible, with the following activities:

System integration and testing involves building, installing and testing a suitable infrastructure with appropriate communication links to locations and users for interoperability of services. The system architecture must include the appropriate measures for trust and security and identified societal needs, as described in the societal user requirements specifications. The integration must also provide concise guidelines for developers so that tools and components are able to integrate seamlessly to the architectural framework. The prototype will be installed at the pilot sites for initial validation. In agreement with our requirements engineering process the integration and evaluation process will not be executed only once during system development, but rather will be returned to so that the requirements, new prototypes and their evaluation do reflect the new knowledge gained during specification, validation, and subsequent activities.

Phase four: Demonstration, evaluation, validation, and uptake

The final phase of the project will be concerned with validation of the prototype *EURECA* platform in the various user cases as well as overall project evaluation and preparation for take-up activities and exploitation.

Demonstration. The EURECA tools and services will be installed at the pilot sites by the users and exposed to the cases described in the user scenarios.

Validation of the prototype platform will involve user testing of the prototypes (letting the users execute the user scenarios on location). Also the socio-economic and security aspects will be tested, by involving focus groups and questionnaires. Validation will be performed jointly by the users and the developers and documented in a report for each user scenario. Evaluation of the entire project and its achievements compared to the project objectives will be performed jointly by all partners and documented in the final project report.

Uptake activities involve, in the first instance, a comprehensive demonstration program, including the possibility to provide demonstration services to various potential users (as consumers or providers of EURECA services). During this process, the partners will work on implementing their exploitation strategies for the ultimate goal, a Europe wide deployment and exploitation of the EURECA environment and software services.

1.3.2. Timing of work packages and their components

The work breakdown is based on the envisioned research activities, each work package carrying out a set of coherent, related and manageable tasks. The overall project structure is depicted in Figure 14. In this section we also provide a brief description of each of the project's work packages.

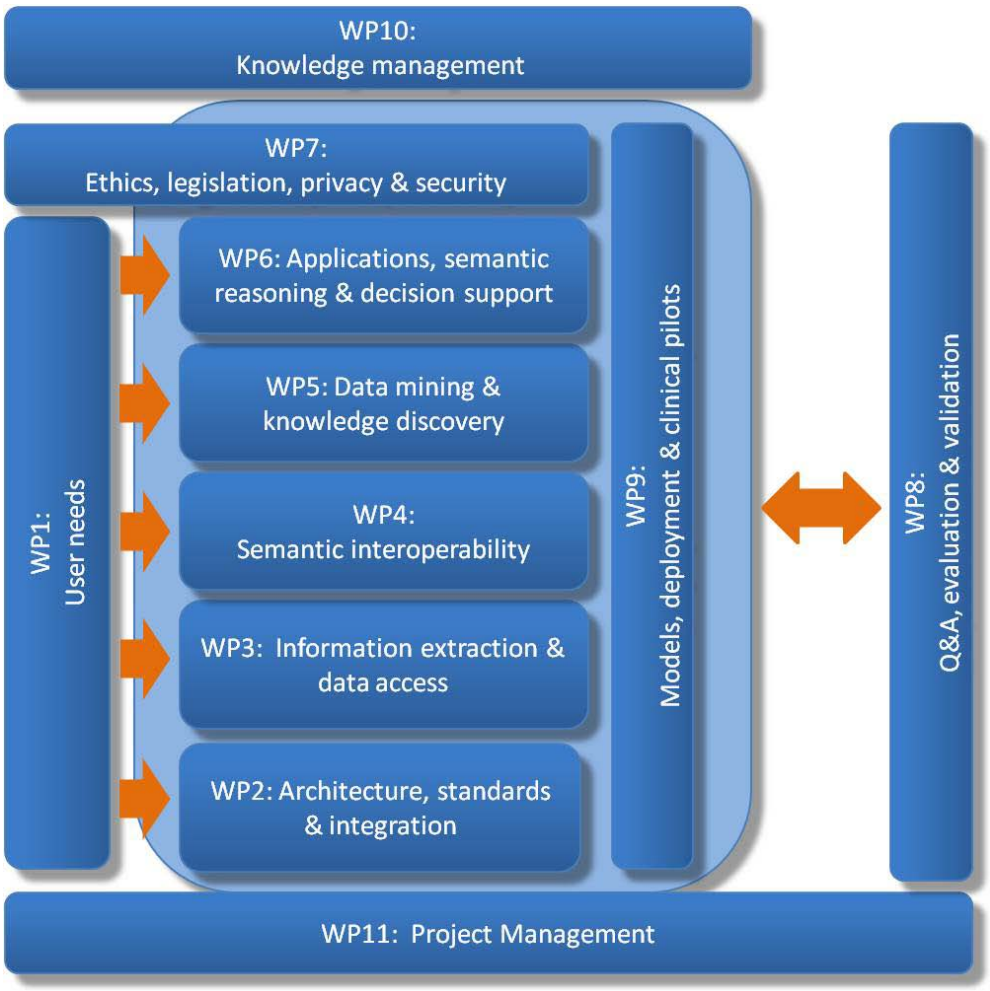

Figure 14 Preliminary work breakdown

WP1 – User needs

WP1 of EURECA will elaborate the user needs and the various requirements that are necessary for the development of the proposed environment. This WP will address the user requirements (in conjunction with all other WPs) from the technological, the clinical application, and the legal and ethical standpoints. Based on the analysis of the requirements gathered, this work package will build comprehensive scenarios and use cases that will be used in the design and iterative development of the INTEGRATE solutions. An important task of this work package will be to provide end-user input, through the contribution of a wide community of users in the chosen disease domains, to the design, development, evaluation and validation of our environment.

This WP will address the needs for developing a seamless, secure and consistent linking of clinical care information in electronic health records (EHR) with information in clinical research information systems, such as clinical trial systems. This WP will address the technological requirements (in conjunction with all other WPs) from the clinical application standpoint.

Daily clinical care

Today only 5% of adult cancer patients are enrolled in prospective clinical trials. Diagnosis and treatment of most cancer patients within Europe cannot be evaluated easily, because data of patients treated outside of clinical trials are only stored in clinical information systems which are heterogeneous among different hospitals and which are not allowed to be exported for analysis by the hospital due to security issues. This makes the detection of severe adverse events (SAEs) and suspected unexpected severe adverse reactions (SUSARs) nearly impossible if a patient is not treated within a clinical trial and the SAE or the SUSAR is not announced to regulatory bodies.

Clinical trials

Even if patients are treated within prospective clinical trials it is not possible to efficiently analyze data across trials on a wide scale. There is also a insufficient information regarding existing clinical trials in Europe. This

WP will analyze the needs and requirements to assist in automatic identification of patients for clinical trials and tools to bring state-of-art research knowledge in the decision process of treating physicians.

The pharmaceutical industry

The pharmaceutical industry is an important user group for our solutions and therefore needs to be closely involved in the requirements gathering and analysis phases and in the elaboration of the EURECA use cases.

The project has opted to establish a strong link with the industry as a whole, rather than to include a single pharmaceutical company as project-partner. The latter risks that the project would focus on the particular needs of one company and miss out on the more general industry issues.

The project has made the connection with the pharmaceutical industry through the establishment of a Pharma Advisory Board, which includes representatives of several large pharmaceutical companies.

Current members include (cf. Letters of commitment to the board):

- Amgen, through Andrew Roddam, International Head Center for Observatinal Research
- Janssen Pharmaceutica (a Johnson&Johnson company), through Maik Stumpf, Director Health Information Sciences
- Sanofi-Aventis, through Pierre-Yves Lastic, Senior Director, Data Privacy & Healthcare Interoperability Standards

The three listed industry representatives are actively working on the issue of EHR data re-use in clinical research. They are as such also personally involved in the EHR4CR IMI project. This board will be further extended in the first months of the project.

The involvement of this board in EURECA will not be restricted to WP1, it will also play an important role in validation and in the exploitation of the EURECA solutions.

Patient empowerment

The role of patients in clinical trials, besides to be a patient, is not defined yet. For all the tools and software that will be developed we will also analyze the needs of patients. This will increase the empowerment of patients in clinical care and will also address safety issues. Both aspects will lead to better, safer and more personalized medicine. Additionally, our safety scenario will also involve access to patient-managed data out of PHRs. It is well acknowledged that serious side effects of drugs and treatments may be missed if they manifest themselves outside the hospital setting and relying on the patients' contribution to recognize and report them could significantly improve safety.

User Scenarios

In this work package we will develop relevant clinical scenarios and use cases to support the development of the EURECA environment. The following scenarios will be analyzed and refined, and additional scenarios emerging from the user needs analysis will also be defined.

- Improve patient recruitment and trial search by automatic identification of potential candidates for a clinical trial: It is currently often the case that research and care systems are disconnected and finding patients for a clinical trial requires a lot of manual search in the EHR system and subsequently entering overlapping data in the Clinical Report Forms and in the patient file. Achieving semantic interoperability among the two types of systems (EHR and CTMS) helps to have more efficient patient recruitment and to avoid double data entry.
    - Find patients for clinical trials: This scenario will demonstrate how matching data in the Clinical Report Forms with data extracted from the EHR can help identify eligible patients for clinical trials.
    - Find relevant trials for patients: Once patient data extracted from the EHR system can be matched against the eligibility criteria of running clinical trials, this information can also be used in clinical practice to find relevant trials from the perspective of a patient case.
    - In this scenario we will also demonstrate automatic storing of clinical trial eligibility criteria into the EHR and use of the EHR for automatic EDC (pre-filling).
- Enable long term follow up of patients, beyond the end of a clinical trial: Access to EHR data would enable researchers to follow the patients in their trials beyond the end of the trial and to better assess the outcome and the effects of the therapies and to build new hypotheses and follow-up research. Currently the consistent flow of information between the clinical researchers running the trial and the clinician treating the patient is stopped at the end of the trial. Further follow up is ad-hoc and inefficient, for example by sending postcards to the treating physician to ask for information, and the data collected is insufficient to allow for proper evaluation and analysis.
- Data mining and association studies on EHR data to detect safety issues that could not be found within the scope of a clinical trial: A semantically interoperable linkage among Clinical Trial Systems and EHRs enables the detection of side effects to treatment or medication that could not be identifiable within the trial itself but that can be found by analyzing the longitudinal records of patients in the EHR, that may span

many decades after the trial ended. Building a feedback loop from clinical practice to the clinical trials would support the detection of relevant side effects that are currently missed and the continuous evaluation and improvement of treatment. This would also allow to generate and test research hypotheses, and markers or models that help predicting late effects. This scenario will also evaluate standards-based semantic linkage to a PHR system in order to use patient-provided data to detect relevant safety issues.

  - Late effects of radiotherapy: The side effects of radiotherapy are currently only monitored during the trial, but some late effects become manifest long after the closure of the trial. Providing access to EHR data would allow to monitor the patients on a longer term and to adapt the dosage of the treatment to obtain the best outcome with the least harmful side effects.
  - Treatment induced secondary tumors: In rare cases radiotherapy can induce secondary tumors. Large scale data mining of EHR data can identify those rare cases.
  - Side effects of treatment that become manifest after a period of time larger than the duration of the clinical trial: Side effects of drugs can be monitored beyond the trial to detect health risks with late onset and to generate new research hypotheses
  - Severe side effects of treatment that are rare and cannot be detected within the limited population of a clinical trial: There are several known cases when serious but rare side effects could not be detected within a clinical trial because of the limited sample. Having access to EHR data after the end of the trial, when the drug have entered large scale practice, may help identify rare side effects and build hypotheses for identifying biomarkers that could predict those effects. The same way, rare or unexpected drug-drug interactions may be detected.

- Generate research hypotheses and retrospective validation for rare diseases based on EHR data: For rare diseases the access to EHR longitudinal records of many patients (across institutions, at national or even international level) can provide the needed data to systematically learn about the disease and progress research.
- Consistent and efficient detection and reporting of adverse events (SUSARs, SAEs): The reporting of adverse events, although obligatory in clinical trials, is currently inefficient, inconsistent (terminology, timing, etc.) and error-prone involving multiple actors that need to enter the same data. These issues can be addressed by a consistent linking of EHRs and Clinical Trial Systems and suggesting a process that would eliminate the need for multiple data entry.
  - This scenario will also evaluate standards-based semantic linkage to a PHR system in order to use patient-provided data to detect relevant safety issues.
- Contextualization of information in clinical care: As the amount of data collected in clinical practice for a patient and the relevant published research results, treatment options and guidelines increase significantly the physicians are often overloaded with the amount of information that they need to digest for a patient case, especially given the often severely limited time available. This may lead to suboptimal treatment and medical errors. By linking EHRs with external repositories (public data, validated research results, literature) and based on the set of relevant concepts selected to model the data in the EHR (part of achieving semantic interoperability) the relevant data for the patient case can be efficiently extracted and presented.
  - Knowledge transfer into care. Improving patient care by providing relevant validated research results: This scenario shows how the relevant therapy information for a patient case can be extracted and presented using external resources of information and a repository of validated clinical trial results.
  - Extract EHR information relevant for the patient case under investigation: In this scenario the relevant data for a patient case is selected out of the large longitudinal record of the patient by intelligent reasoning based on the model and of the set of concepts relevant for the disease.
- Propose a uniform presentation of validated clinical trial results: The view on clinical trials run in Europe and on their validated results is currently incomplete and inconsistent. To support future research and prevent wasting money with duplicating effort of other institutes or exploring research dead-ends, all validated results of clinical trials, including negative results that are currently seldom reported, should be made publicly available. The trial description including the results should be detailed enough and uniform to enable its easy use. This data could also be used in future clinical decision support to bring research results faster to the front-line clinician and improve patient outcome. Taking into account existing initiatives, standards and solutions we aim to propose a model for data that should be reported out of a clinical trial

and demonstrate their management and use. We also aim to provide uniform access to external repositories of clinical trial results.

WP2 – Architecture, standards and integration

Approach

The objective of this Work Package is to provide an architecture defining a coherent, consistent, interoperable, extensible and scalable Service Oriented Architecture (SOA) platform that can support the EURECA services. The WP will define an architecture specification emphasizing interoperability and interfacing for integrating the EURECA services. Subsequently this WP will provide a reference implementation to demonstrate these services.

This work package will base the architectural design on the consolidation of system requirements and technical workflows (derived from usage scenarios) defined in WP2. The architecture will surely address the following topics:

- The architecture will include a security framework based on widely accepted security standards, allowing the integration of different security solutions into the EURECA platform. The security architecture will support integration of: Identity Management (IM), Authentication and Authorisation, establishing trust between service providers, audit and provenance, Privacy Enhancing Technology.
- The architecture will enable a high degree of scalability (maintaining acceptable performance) and incorporate mechanisms for providing highly available services.
- The architecture can be extended to new services and capabilities, which can be added without impacting the existing, deployed system. The EURECA standards-based service framework should be able to support current and future scenarios with respect to re-use of EHR data (i.e. be future-proof).
- The architecture will account for high degrees of heterogeneity at all levels of the system, ranging from platforms and programming languages to the semantics of the data used by EURECA. This applies equally to the EHR systems in the healthcare domain and the clinical data management systems in the research domain.
- Integration of existing systems (EHR, CDMS) into the architecture should be possible with a limited amount of effort (modification of the interfacing capabilities), and ideally without requiring major changes to those systems.
- (Technical) Criteria will be defined that must be met by the existing system, in order to ascertain the usability of the system's data for clinical research purposes.
- The architecture will support the "local ownership, local control" principle, meaning that data providers can define and manage their access policies to be consistent with local and national statues and regulations.

The user scenarios and user requirements from WP1 will be analyzed and translated into technical requirements, technical workflows (composition of services) and system specifications; these will validate the functional capabilities of the system.

The Technical Board (lead by the WP2 leader) will govern the major decisions to be taken in the architectural design process. For example determine how semantic interoperability services (WP4), security and privacy services (WP 7) and the service workflows ("end-user application") will interact architecturally.

Additionally, this WP will define the approach for extending the platform through the development of new services and the integration of legacy services into EURECA. Integration profiles will provide a framework for extending EURECA in the future as new services are developed. These new services will need to comply with the integration profiles which will guarantee interoperability with existing EURECA services (through the interoperability framework). It is expected that multiple integration profiles will be developed gradually over time, enabling tighter levels of integration of services in EURECA where necessary. The integration profiles will also provide a mechanism for verification and validation of new services and components.

This WP will create a reference implementation of the EURECA architecture that integrates the different components built in the other Work Packages into a functional platform that will be validated in several clinical pilots (WP8, WP9). As the system is progressively (iteratively) developed and deployed, it will be periodically evaluated for usability with target user groups.

The adoption of international standards is critical in the creation of this platform. Therefore, a specific task in this WP is dedicated to the identification, evaluation and selection of appropriate standards. EURECA envisages a pragmatic approach, using existing solutions where possible (identified through the state-of-the-art analyses of WP1), thus avoiding the risk of performing "double work".

Architecture

The EURECA architectural design will be based on a loosely coupled SOA interconnecting independent services. Where needed for meeting the requirements, central SOA core components will be provided. Examples include central registries of data sources (cataloguing data sources contents), service registries (e.g. storing access policies and service availability) and intermediate brokers that coordinate and orchestrate interactions between data providers and service access platform.

The adoption of the SOA paradigm, coupled with the use of established international standards, will result in a vendor neutral solution, making EURECA an open model. This should encourage wide adoption and open up different routes (commercial, open source, mixed) towards long-term sustainability of the EURECA architecture.

EURECA will define an architecture in which services can be interconnected based on integration profiles, without enforcing a “monolithic” solution. Different service implementations should be interchangeable in a EURECA compliant environment. Ultimately, the goal of EURECA is to provide an open architecture specification that can be used by collaborative initiatives to efficiently set up medical data integration projects (i.e. the possible application scope transcends EHR data re-use). In this way, such collaborations can easily interconnect with EURECA compliant data sources, re-use EURECA compliant services, etc. Whether their particular platform instance is based on the EURECA reference implementation or on another implementation of the EURECA architecture (commercial or open source) is of little importance.

Additionally, the SOA approach allows for a gradual introduction of EURECA services, allowing data providers to gain trust in the platform with less intrusive applications (e.g. physicians looking for trials which could be useful for their patients), before exposing their complete EHR to the more advanced services.

EURECA aims to address the needs of a very wide community and offer a solution that can be extended across various European regions and establish an attractive and integrated pan-European environment for re-use of EHR data. The modular approach allows EURECA to remain flexible towards current unknowns such as data provider or service provider specific preferences, changing regulations, future services to be integrated and technological advancements. As previously mentioned, this loosely coupled approach allows any given service to be replaced by an equivalent alternative (compliant with the EURECA interoperability framework). This implies that for many services (consent services, terminology services, identification services, etc.) a decision can be made to rely on a common shared service or on a local implementation (e.g. for use of some proprietary components).

Finally, it is worth noting that the described approach allows for parallel development of components. This decreases the risk of running into development bottlenecks (component development dependence is decreased).

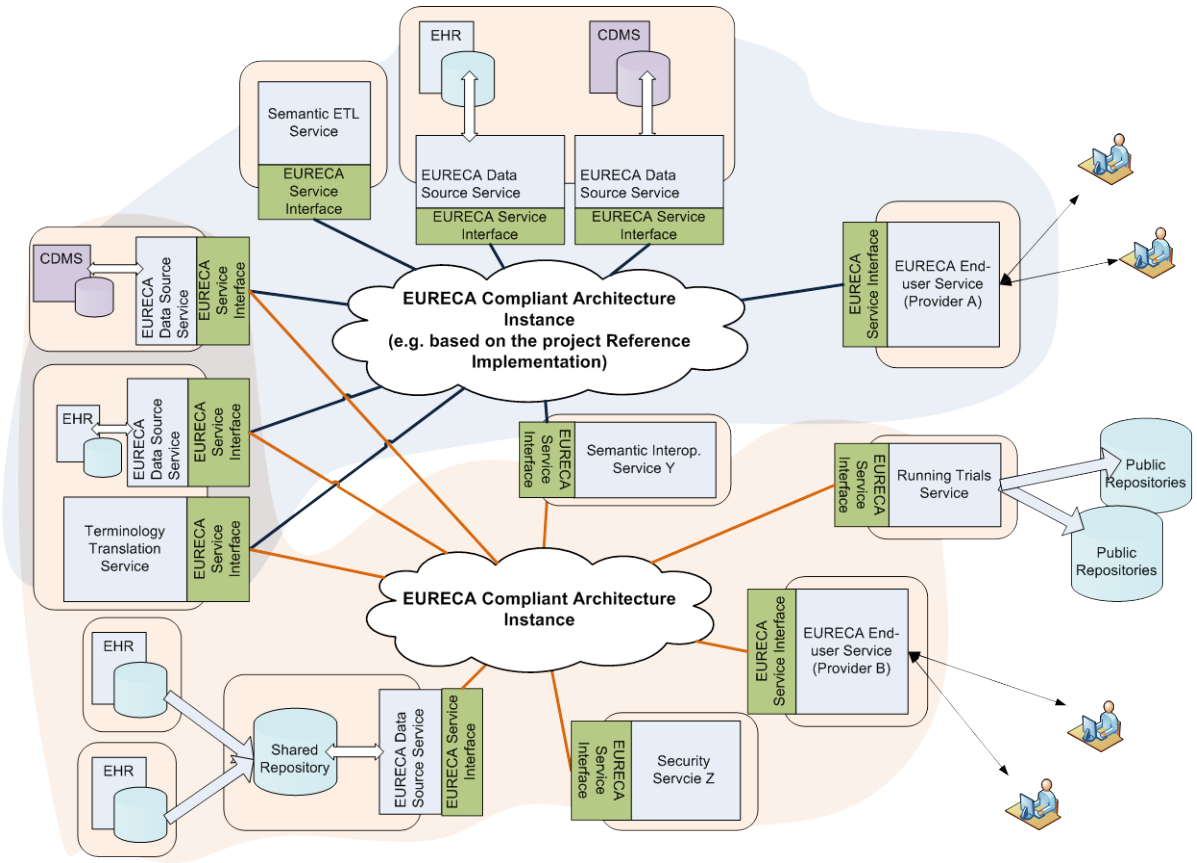

Figure 15 High-level EURECA concept overview showing two independent instantiations of the EURECA architecture (e.g. set-up by different organizations, covering different regions, etc) , sharing different services based on EURECA interoperability.

End-user functionality will be provided by a composition of services that will be deployed on the EURECA architecture and will consist of both distributed and central components. One example of this is a patient recruitment application for clinical researchers. This application requires a central management component that allows semantic queries to be distributed over the different data sources for execution, implying the need for a query module in the “EURECA data source endpoints”.
Designing the EURECA interfaces for the latter end-points is one of the most challenging (joint development) tasks in this project. These EURECA service interfaces will need to provide an execution environment for local components as explained by the example. It will also be crucial in hiding the heterogeneity of the external EHR, CDW and CDMS data sources through the “External Systems Interface” (cf. Figure 15).

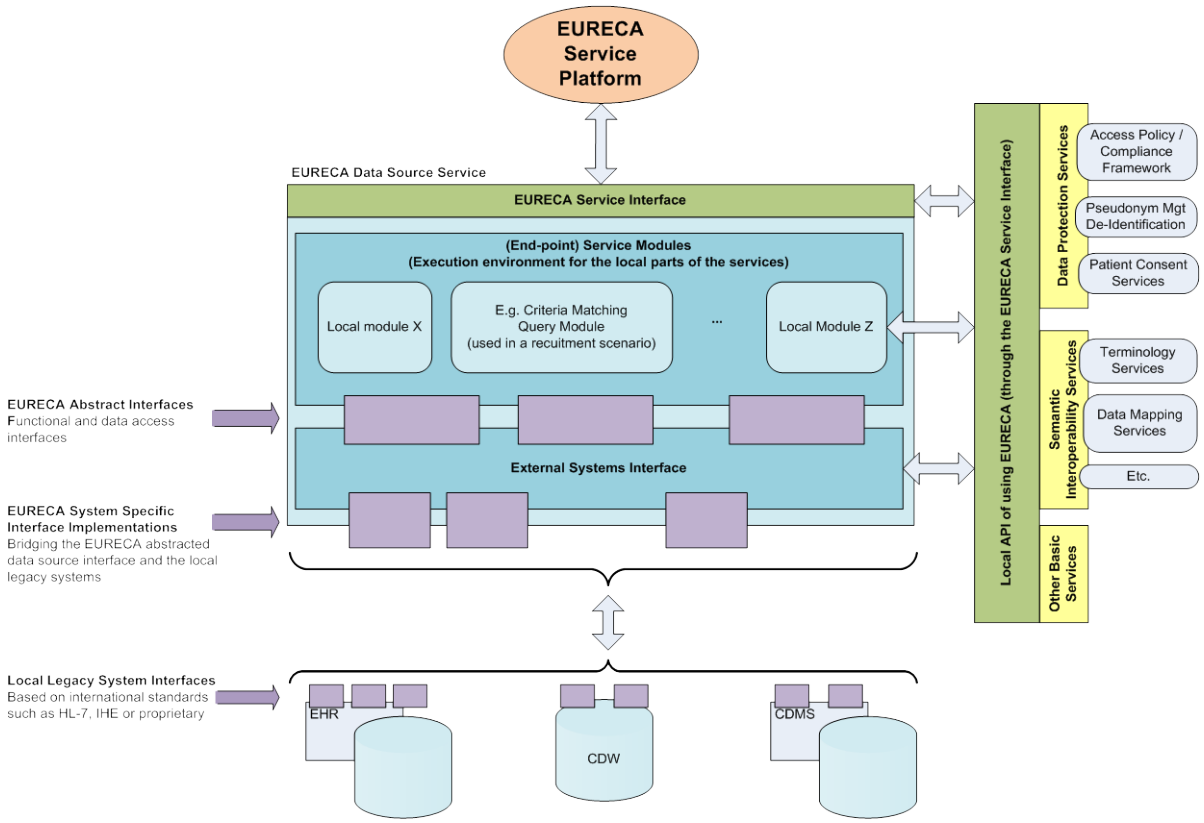

Figure 16 EURECA data source end-point can be deployed on an EHR, CDW or CDMS and will rely on different local and remote services. The picture illustrates the interface abstraction.

This layer will be defined as a set of abstract interfaces that provide uniform data access towards the external data sources and expose the specific functionalities of the source systems (e.g. querying). For each different external source (e.g. EHR from a specific vendor), a separate implementation of these abstracted interfaces will need to be provided. As is the case with other services, those end-points will be able to rely on locally and/or centrally deployed services, e.g. for semantic interoperability (mapping towards a common information model and common knowledge model, terminology service, etc.). Figure 16 illustrates the conceptual (high level) architecture of these data source end-points.
Different technical details may impact the EURECA architecture. These include for example “where” and “how” semantic transformation of a distributed query will be performed, if source “shadow” data bases (e.g. when a direct connection with an EHR or CDMS is not allowed) are required as part of the interface layer, etc. These questions need to be answered at an early stage of the EURECA project. In order to mitigate the risk of running into severe integration issues, the architecture will provide flexibility to accommodate for different technical approaches (cf. focus on interface definition and interconnection).

WP3 – Information extraction and data access

Of course, it is not realistic to expect that all available EHR and CT information will be structured, coded according to standardized terminologies and stored according to standard representations in the local sources. To be able to build appropriate and comprehensive information models of the sources, link the local information and data to defined core datasets for the relevant domains (extracted out of widely adopted standardized terminologies/ontologies), and create a reliable semantic linkage among the various domains, concept tagging and information extraction approaches will also be required.
To identify, correctly classify and extract relevant concepts out of free text reports in the EHR or out of free text clinical trial eligibility criteria in the CT protocols, the project will also focus on natural language processing and tagging/text mining of relevant data. Whenever possible, existing techniques and tools will be reused and the most pragmatic approaches will be adopted, in order to ensure data quality. Our main goal is to use these techniques to support the effort of the domain experts and IT experts to build comprehensive, standards-based information models of the sources, including the information currently only available in free text reports. This extraction process will be gradual and based on the defined core dataset, i.e. search with a clear goal in mind, not a random search in an infinite space. The idea is to gradually understand the data

available in the free text report, find, classify and annotate the relevant concepts and also find the relations when possible. To ensure quality, the quality of the results will always be supervised by an expert (e.g. annotator, domain expert, etc.).
We will provide tools to perform syntactic and semantic analysis of textual data available in the EHR and Clinical Trial systems available in the EURECA environment. Information Extraction in this context will go far beyond simple pattern matching techniques. Pattern matching techniques are not enough if we want to make a step towards knowledge discovery and reasoning. Information Extraction (IE) usually consists of two main tasks: concept identification and event identification (i.e. relations between concepts). However, in order to obtain accurate results in both concept and event identification, complex linguistic phenomena have to be considered. We distinguish the following steps for IE:

1) Recognising in text the concepts of interest.
2) Negated and hedged information processing using NLP techniques
3) Temporal information processing
4) Links detection between concepts

1) Concept recognition
Recognizing concepts of interest present in free text is a common task in the processing of medical textual data. Concept recognition is very dependent on the chosen terminologies used in the project as the concepts labels that are retained. Concept recognition is thus strongly linked with decisions taken in WP5: Semantic interoperability.
But simple term extraction without taking into account contextual information remains unsatisfactory for accurate results. An important subtask in IE consists in detecting the status of the information conveyed by the textual data, more specifically detect if information is negated or speculative.
2) Negated and hedged information processing
Negated information is often present in medical text, especially in EHR information where a negated fact or negated exam results can be extremely informative regarding the state of the patient. Negation is a well-known and difficult problem to handle in NLP. In the medical domain, previous work on this topic has been performed, with interesting results that we believe can still be improved. In a similar way, hedged information can also be extremely misleading in an information extraction task. Hedged information consists of assertions that are not factual. It is estimated that MEDLINE contains about 11% of sentences carrying speculative information. Obviously this information has to be treated in a different way than definite assertion. We intend to use refined NLP techniques based on deep linguistic knowledge for both negated and hedged information processing.
3) Temporal information processing
Temporal information is another element that has to be considered for refined information extraction tasks in the medical domain. Xerox already developed a temporal processing module which is embedded within the general purpose parser. This module detects temporal blocks in patient discharge summaries coming from different units of French hospitals. In this specific application, temporal coordinates are associated to infection occurrences in order to help to determine if these infections are acquired in hospital or not. We intend to enrich the temporal processing module to other kinds of medical documents available in the EURECA project and to support the EURECA clinical scenarios. Once again the temporal dimension in IE is a necessary first step for knowledge discovery and reasoning.
4) Links between concepts
Once concepts are detected and annotated with the correct background information (negation, modality, time), an important aspect is to be able to see how these concepts are related and the kind of relations holding between them. Concepts and relations build complex events. An event is thus a frame in which participants consist of terms of interest which are linked according to given types of relations. These links are supported by syntactic dependencies holding between concepts. These dependencies are extracted using NLP techniques. We propose to adapt the Xerox parser to the EURECA domain. The parser computes syntactic and semantic dependencies between the linguistic units (words or terms) present in texts. As the approach adapted is a dependency-based approach, long distance relations between textual fragments can also be calculated. In the EURECA context, events of interest have first to be defined (which kind of relations, which kind of entities participate to the relations). The parser will then be adapted accordingly in order to match occurrences of the defined events in texts.

Language heterogeneity and multiple data resources
In EURECA we will also need to deal with language heterogeneity, as we aim to bring together Clinical Information Systems containing data in several European languages. It is known that the performance of NLP techniques is highly dependent on the data itself and training is required for new resources. Working with different languages adds additional complexity.

We believe that instead of a one-fits-all solution, a variety of methods and tools (concept annotators, chunkers, entity recognizers, etc.) need to be combined for each available system to obtain best performance. High quality data extraction can help us build accurate information models of the sources, improve the quality of the data, and ensure high quality semantic linkage among the various systems available at the clinical sites. The IE task is not meant to replace the experts, but to support them in identifying concepts, annotating medical text, and building the information models of the sources.
The data extraction output will also be used in WP5 and WP6, to support advanced data mining and semantic reasoning algorithms.

Access to external heterogeneous data resources

Another important task of WP3 is to provide uniform and standards-based access to relevant external sources of data and knowledge to be used by the EURECA tools and services. EURECA will provide uniform interfaces to access relevant external repositories of clinical trial information and results, clinical guidelines, literature, data warehouses, etc.

An important use case will include standards-based access to a Personal Health Record system, enabling the use of patient-authored data in a scenario aiming to improve safety by early and efficient detection of serious side effects of drugs and treatments.

WP4 – Semantic interoperability

Semantic Core Dataset
The investigations undertaken by SemanticHEALTH[201] suggest that full semantic interoperability (Level 3) is required in order to take full advantage of computerized medical records. It is however also recognized that due to steep investments needed, the highest level of semantic interoperability should only be sought in specific areas with the high potential for improvements.
The semantic interoperability layer provided by WP4 is the underpinning of the EURECA semantic services. Semantic interoperability requires agreement on meaning of concepts that capture the domain of discourse as well as the labels for those concepts. Such concepts are often captured in coding systems and ontologies such as SNOMED CT[202]. The foundation of the semantic interoperability layer will be the semantic core dataset consisting of well defined set of relevant domain concepts that sufficiently describe the semantics of the chosen clinical domain. A semantic core dataset will be directly extracted from, or soundly mapped to concepts from relevant existing standardized terminologies e.g. well established and widely used ontologies such as SNOMED CT to facilitate semantic interoperability beyond the scope of EURECA. Identifying sound semantic subsets of SNOMED covering a certain clinical domain is also on the roadmap of SemanticHEALTH[201].
The EURECA semantic data set(s) will be validated in concrete use cases, for the different EHR and clinical trial systems available at the clinical and clinical trial sites. The semantic core dataset is an essential prerequisite to semantically-aware access to both EHR and Clinical trial data in a machine processable manner. Concepts in the dataset will have their unique identifiers, well understood meaning as well as a set of synonyms they can be referred as.
Considering the problem of language heterogeneity, in EURECA we plan to address this issue by offering a gradual approach, semi-automatically translating only those parts of the clinical ontology identified as the core semantic dataset. The core semantic dataset is essential for capturing the necessary semantics of a particular clinical context and is expected to be considerably smaller in number of concepts than for instance the entire SNOMED CT. The main rationale here is that only a confined subset of relevant concepts from the clinical ontology will be needed for data extraction and reasoning in a given clinical context/domain while most of the remaining concepts would never be used by reasoning algorithms. Hence, translating (only) the selected semantic core dataset and not the entire clinical coding system enables a modular and scalable approach where the initial translation effort is limited in scope and delivers immediate benefits in increased semantic interoperability. Addressing the anticipated language heterogeneities, when no translation of the relevant standard terminologies exist in that language, we will work out together with the clinical experts a translation of the core dataset to the languages that are used for the primary data capture.
We will also conduct a study demonstrating possible extensions of the semantic core dataset approach to other clinical domains.

[201] http://www.semantichealth.org/
[202] http://www.ihtsdo.org/SNOMED CT/

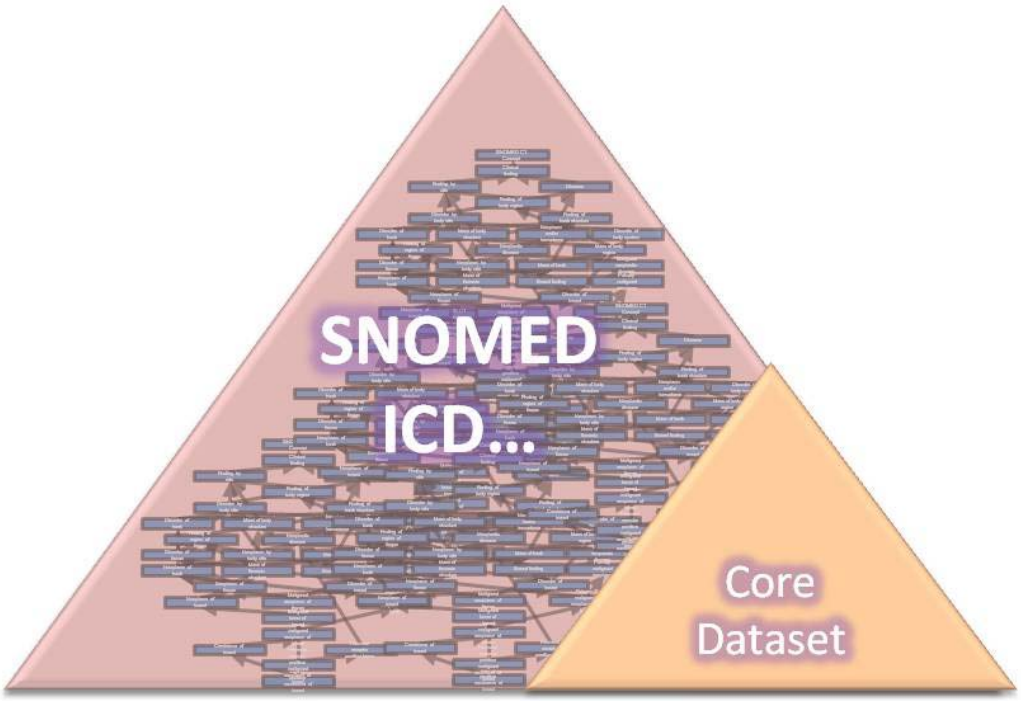

Figure 17 Core dataset covering a chosen clinical domain and a mapping formalism

Mapping formalism, mappings, and semantic interoperability platform

The canonical information models of EHR and CT systems identified in WP9 will be mapped to the semantic core dataset in order to guarantee a well defined meaning of various data elements across the entire EURECA platform.

Moreover, the close relation of the semantic core dataset and the clinical ontologies such as SNOMED CT will also guarantee that the semantics of these data elements can also be easily understood by external systems.

We will identify together with WP9 the requirements for mappings that bridge the semantic core data set with the information models representing the EHR systems and the clinical trial systems. These information models provide a canonical view, reflecting the content and the structure of the respective information management system. The proposed mapping formalism should be able to mitigate the foreseen structural and contextual differences between the semantic core dataset and the information models. We will use this formalism to instantiate the necessary schema-level mappings that will be executed by the semantic interoperability platform during the data extraction process.

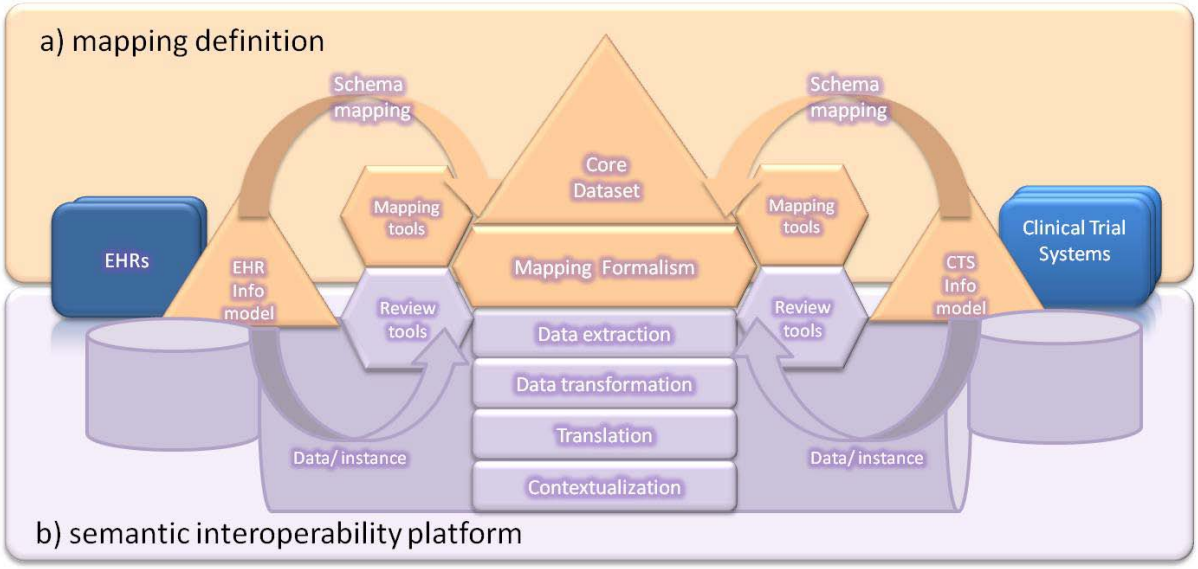

Figure 18 Schema-level mapping definitions a) and the underlying semantic interoperability platform that handles the necessary data transformations b)

In order to facilitate the necessary data transformation among various information systems which need to interoperate, we will build a prototype of the EURECA semantic interoperability platform. This platform will utilize the semantic core dataset as well as the schema-level mappings that link to the EHR and CT information models. The platform will be able to execute these mappings during the data extraction phase, instantiating thus the semantic concepts with patient data and/or clinical trial data (see Figure 18). The semantic interoperability platform will be an essential software engine behind the EURECA services, enabling linkage between the patient data in the EHR and the clinical trial systems.
The process of discovering the underlying information models of the sources and their representation in a desired syntax in our canonical information models will be supported by the information extraction tasks carried out in WP3.

WP5 – Data mining and knowledge discovery

WP5 focuses on the development of intelligent tools for knowledge discovery in integrated EHR and research data. The EURECA project will approach the discovery and use of clinically relevant knowledge and data in several areas, such as:

1. Enabling the use of care data for research and setting up a framework for knowledge discovery with the goal to allow efficient data analysis.
2. Based on the EURECA core set of concepts, specific multi-level knowledge discovery tools will be developed that address issues such as patient selection for maximal benefit and minimal risk, hypothesis generation and early detection of risks, meta-analysis, and long-term monitoring of patients.
3. Intelligent tools for faster transfer of research results to clinical care and contextualization of information.

The EURECA approach will supply medical researchers with fast access to significantly larger sets of data, including data from standard care, than currently possible. Part of this work package, a generic knowledge discovery framework based on the EURECA architecture will be developed. The KD framework will provide basic services and building blocks to guarantee the inter-operability of services, scalability and re-usability. This will prevent one-off solutions and promote the re-usability and adaptability of components.
One particular challenging task in EURECA will be the efficient long-term follow up of patients and data mining of EHR data. Mining of patient care data will allow to detect survival, recurrence, and serious side effects long after the end of the trial. Mining of clinical data may enable the early detection of patient safety issues, but also the generation of new research hypotheses. Existing solutions are usually not very efficient, with the effect that only very basic statistical procedures can be used. The field of data stream mining[203]

[203] C. Aggarwal (Ed), Data Streams: Models and Algorithms, 2007

investigates more efficient solutions, but often only for specific analytic problems. WP5 will develop approaches based on recent developments in distributed data mining[204]. This will allow to process more patient data, while looking for more risk patterns, which may result in the better understanding of diseases and higher patient safety standards in clinical care.
A basic technology that will facilitate the re-use of existing solutions and existing information, is similarity learning, i.e. the automatic construction of appropriate measures of similarity. This approach would enable the automatic identification of similar trials for meta analysis, the identification of similar patient cases for comparison of treatments, or the identification of similar data sets in a clinical-trial data repository in order to re-use existing solutions and prevent the repetition of dead ends in research[205].
Finally, to improve the transfer of clinical results to care, specific knowledge discovery tools will be developed. Based on approaches from the domain of information retrieval, users of the EURECA system both clinicians and researchers will receive appropriate, personalized information, recommendations and alerts, which will flexibly adapt to their preferences. Building on techniques developed for search engine optimization[206], user feedback will be exploited to assess the relevance of a specific piece of information. The particular new challenge will be to adapt existing approaches, which are targeted to unstructured text data, to the various types of data, with different degrees of structuring, available in the context of EURECA.

WP6 – Applications, Semantic Reasoning and Decision Support

We will demonstrate the software services and tools of the EURECA environment and the achieved semantic interoperability among EHR and clinical trial systems by implementing several clinical scenarios. It is the role of this work package to implement the advanced reasoning functionality required the end-user applications/services that are described in those scenarios:

- Improve patient recruitment and trial search by automatic identification of potential candidates for a clinical trial.
- Consistent and efficient reporting of adverse events (SUSARs, SAEs) and early detection of safety risks by making use of data out of a variety of sources, including external knowledge repositories and data sources.
- Contextualization of information in clinical care:
    - Knowledge transfer into care. Improving patient care by providing relevant validated research results
    - Extract EHR information relevant for the patient case under investigation
    - Extract relevant information out of external knowledge sources (guidelines, literature, risk models, etc.) and use it for a specific patient case.

Next to the already mentioned benefits of increased semantic interoperability with respect to clinical research and patient safety (e.g. improved patient recruitment, long term follow-up, minimizing multiple data entry), the improvements in semantic interoperability will have clear benefits to clinical practice as well. The increased semantic interoperability provided by the EURECA environment will also enable new applications in the clinical decision support domain where computerized guidelines or validated research results can be directly applied to the patient data, helping the clinical practice staff to make the right decision with respect to patient safety and outcomes. Our clinical scenarios will demonstrate these benefits.
EURECA aims to research the contextualization of validated research data and literature to a patient case, to be used in standard care. Relevant external data, such as references to articles, applicable literature, guidelines, protocols and standardized descriptions of validated trials, will be selected based on the current patient case under investigation. In this process we will use the information model of the source EHR, the EURECA core of concepts and the external sources of information to which EURECA provides uniform access, such as computerized clinical guidelines, external research public databases and repositories of clinical trial information and validated results.
A proper semantic link between EHR and clinical trial systems will enable, among others, an efficient generation and adaptation of computerized clinical guidelines such that they reflect the latest validated research results. As a result, standard clinical care can use ICT tools to quickly adopt and personalize these guidelines, mapping them to the actual patient records, identifying optimal treatment plans based on the extracted relevant information from the EHR. These technological and conceptual advancements will lay the foundations for a new generation of clinical decision support systems that are both semantically aware and up-to-date with the latest validated research results.

[204] J. Sharfman et al., A Geometric Approach to Monitoring Threshold Functions over Distributed Data Streams, ACM Transactions on Database Systems 32(4), 2007
[205] N. Punko et al., Facilitating Clinico-Genomic Knowledge Discovery by Automatic Selection of KDD Processes. ICML 2008
[206] F. Radlinski et al, How does clickthrough data reflect retrieval quality?, ACM Conference on Information and Knowledge Management (CIKM) 2008.

For example, extending the EURECA platform in such a way that the state-of-the-art computerized clinical guidelines can draw the necessary patient data directly from the EHR will help the clinicians by delivering personalized, patient-centric guideline knowledge at the point of care, further helping to avoid unnecessary medical errors and improving the standard of care. The semantic linkage between the EHR systems and computerized guidelines will lay the foundation for the next generation of commercial decision support systems which are patient-centric and context aware. The clinicians will no longer have to (re-)enter the patient data into the clinical decision support systems, the clinical guidelines will be directly contextualized with respect to the patient case. An essential prerequisite for intelligent processing of clinical data is to achieve an adequate semantic linkage between the patient records and the clinical guidelines.

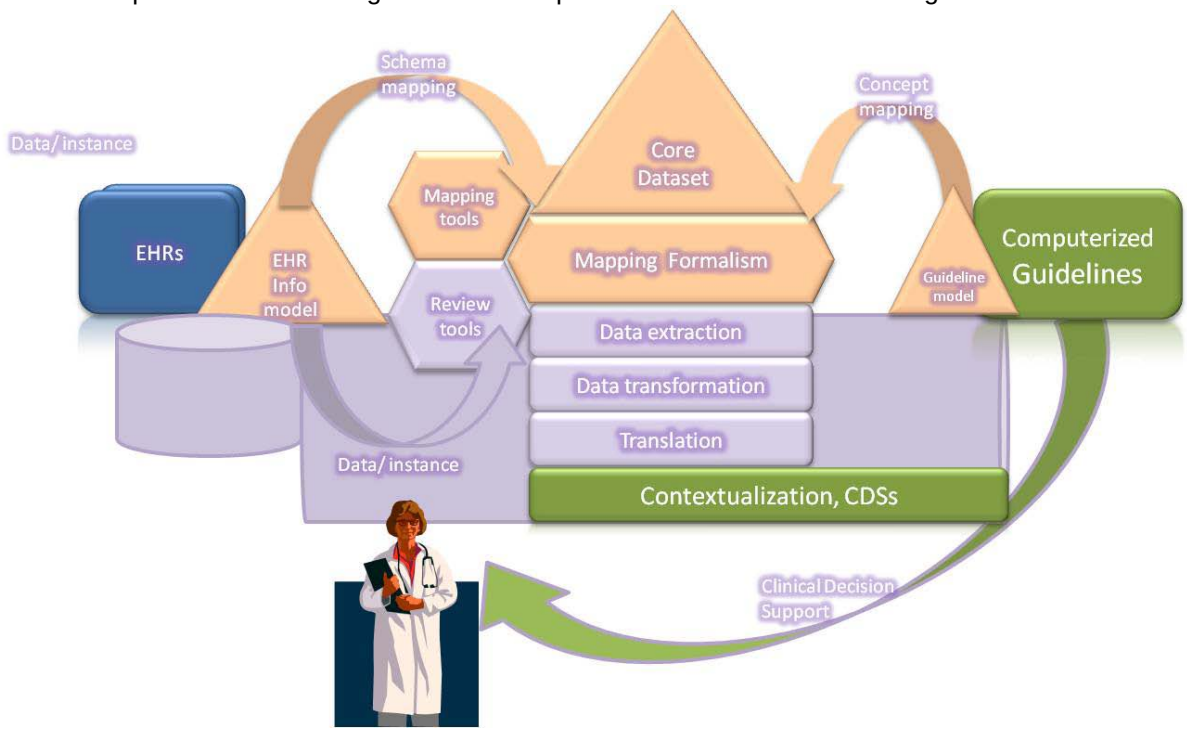

Figure 19 The EURECA semantic interoperability platform enabling knowledge inference (clinical decision support)

WP7 – Ethics, Legislation, Privacy and Security

EURECA aims to open up large sets of medical data for wider use (re-use of EHR data). It is well known that health information is of particularly sensitive nature and from an ethical and legal point of view requires rigid protection. WP7 is dedicated to ensuring that the solutions proposed and demonstrated by EURECA are in line with governing legislation and best practice[207]. WP7 combines ethical, legal (IT Law) and technical (e-security, e-privacy) expertise in order to define an operational framework and provide a set of tools and services which will aid EURECA applications to achieve regulatory compliance with minimal effort. Besides this, WP7 will accompany other WPs with regard to legal and ethical information and give guidance in Intellectual Property Issues. An important role of this workpackage will be to ensure that the project addresses and manages all the requirements coming from the ethics reviews, and to report to the European Commission on compliance with the Ethics requirements of the relevant National bodies.
Fundamental to this work is the design of a Data Protection Framework for EURECA (eDPF – EURECA Data Protection Framework) which incorporates the rules set by the relevant National and EU legislation and sector best practice policies (ethics). The initial WP7 work will focus on defining the requirements regarding

[207] Such as the CHARTER of fundamental rights of the European Union; DIRECTIVE 95/46/EC on the protection of individuals with regard to the processing of personal data and on the free movement of such data; CONVENTION No. 108 for the protection of individuals with regard to automatic processing of personal data; CONVENTION No. 05 for the protection of human rights and fundamental freedoms; DIRECTIVE 2001/20/EC on the approximation of the laws, regulations and administrative provisions of the Member States relating to the implementation of good clinical practice in the conduct of clinical trials on medicinal products for human use; DIRECTIVE 2005/28/EC laying down principles and detailed guidelines for good clinical practice as regards investigational medicinal products for human use, as well as the requirements for authorisation of the manufacturing or importation of such products.

patients' data for the EURECA applications, as to a large extent, the technical (e-security, e-privacy) framework is depending on this analysis. Special attention will need to be given to the (regulatory concept of) "purpose of use" of health data, as EURECA aims to enable re-use of electronic healthcare record data for purposes outside of the care domain (clinical research). In general a change of the purpose of use is forbidden according to Art.6 Nr. 1 lit b of the Data Protection Directive 95/46EC. The linkage of data coming from EHRs with data coming from clinical trials will therefore need a legal justification. The latter is dependent on National enactment of an exception foreseen by the Directive for Member States wanting to enact it and thus makes projects like EURECA face a multitude of different national regulations. A second important aspect is the identification, from a data protection point of view, of the responsible data controllers[208] in the complex data flows.

The framework to be set up will outline the boundaries within which the EURECA services have to operate. It not only includes rules defining "Who has access to what data for which purpose, and under what conditions", but also suggests (novel) solutions which enable (compliant) access to otherwise unavailable data (e.g. through a TTP supported de-identification scheme).

WP7 will be able to build upon knowledge gathered during the ACGT project (Advancing Clinico Genomic Trials – FP6-2004-IST-4)[209]. In this FP6 project, an innovative Data Protection Framework[210] capable of dealing with regulatory issues of large scale sharing of medical and biological data in the clinical trial context was designed (by the same legal and security partners involved in EURECA). EURECA has a different scope as it targets the integration of clinical care information in EHRs with information stored in clinical trial systems (in the context of EHR data re-use) and especially has to deal with data protection issues associated with the cross-domain interaction between the care and research domain.

WP7 is responsible for the development of the security services implementing this eDPF. Solutions will need to be provided for dealing with: Identity Management, Authentication and Authorisation, trust establishment, audit and provenance. Where possible, existing solutions (that are compliant with industry standards) will be integrated into the EURECA security architecture (e.g. Identity Provider solutions like Shibboleth or ZXID).

The implementation of the EURECA Data Protection Framework (eDPF) will rely on policy based authorization services to translate the complex legal rule sets into authorisation decisions for "access to" or "processing of" highly sensitive data over distributed resources. This approach ensures flexibility towards changing legislation and policies. This research will start by examining the suitability of existing policy based authorisation mechanisms, i.e. policies and associated decision engines such as defined by XACML[211], PERMIS[212], GAS[213], etc. To meet the very specific eDPF requirements, extensions will need to be defined. Support will be needed for things like the "notion of datasets" and data bound concepts such as "purpose of use" and "conditions on use" (e.g. by introducing sticky policies[214] associated with datasets, or other types of privacy-metadata). For this work, EURECA will be able to rely on results of the TAS3[215] and INTEGRATE[216] projects, in which EURECA partners are involved in work on policy based access control.

Patient consent (with respect to data protection) is unmistakably connected to the topic of re-use of personal data beyond its originally intended use (see earlier). WP7 will analyse how this concept can be incorporated into the framework through "Consent Management Services", which are considered to be specialised policy based authorization services (consent rules form a policy). Such services need to enforce integrity of consent directives and correctly combine them to avoid conflicting preferences. Consent services will serve as information source for other services which may evaluate policies containing references to "consent directives".

Next to implementing preventive measure for data protection, the eDPF will also rely on audit trails for detection of security breaches and proper incident handling. The majority of auditing mechanisms log individual events per application or computer system. In order to reconstruct a logical chain of events for proper audit, these different logs need to be combined (which is in general not possible). Few standards and solutions are available providing manageable uniform audit trails in distributed systems. In order to be useful

[208] A possible solution is to integrate a central data controller within the project, responsible for all data processing and central contact point for patients' requests. A legal body providing such a service would be for instance the "Center for Data Protection", a non-profit organization under Belgian law, see: http://www.privacypeople.org .

[209] http://www.eu-acgt.org/

[210] A data protection framework for trans-European genetic research projects. Claerhout B, Forgó N, Krügel T, Arning M, De Moor G. 2008, Stud Health Technol Inform, pp. 67-72.

[211] XACML - eXtensible Access Control Markup Language

[212] PERMIS - PrivilEge and Role Management Infrastructure Standards

[213] GAS - Gridge Authorisation Service

[214] Chadwick D, Lievens S. Enforcing "Sticky" Security Policies throughout a Distributed Application. http://www.cs.kuleuven.be/conference/MidSec2008/sticky.pdf.

[215] TAS3 - Trusted Architecture for Securely Shared Services - http://www.tas3.eu/

[216] INTEGRATE - Driving Excellence in Integrative Cancer Research through Innovative Biomedical Infrastructures

for checking compliance of a system with data protection legislation, audit trails need to include a lot of extended contextual information which they rarely do (e.g. type of data accessed, identity of the person listed in the medical record accessed, etc.). Moreover they need to be readily accessible in a user-centric and data-centric way (i.e. provide all audit trails about one person, or about one dataset). Reconstruction of such user-centric or data-centric audit trails based on standard logs using query interfaces is not workable in practice. Difficulties encountered in "live systems" include: the volume of audit trail data is too large to efficiently query, identity of data subjects is usually not recorded, identities listed in audit trails across applications can no longer be linked after a certain period of time, etc.
WP7 will address this issue, while taking standards efforts such as ISO standards 13606 Part 4 (Health informatics -- Electronic health record communication -- Part 4: Security), ISO 27789 (Audit trails for electronic health records, draft) and IHE ATNA (Audit Trail and Node Authentication) into account.
In order to correctly assess the compliance of data flows with regulations, the provenance of received information and stored data must be recorded. The audit trail described above will need to be extended with "how" the data was transformed and where it originated from. Equally, with respect to data privacy, knowing the provenance of a data set can inform a user or system about the applicable data privacy policies (cf. consent).
Furthermore, WP7 will deal with technology specifically designed for protecting data privacy. In order to guarantee regulatory compliance, business processes need to be evaluated as a whole and subsequently "enhanced for privacy" (e.g. by ensuring anonymity of patients as much as possible during a process). How this will be done exactly is dependent on the process itself (defined by the user requirements), the technical implementation thereof (architectural and technical constraints), the legal and ethical boundaries. A set of "building blocks" which can be used to set-up compliant data flows will be researched and developed by WP7. Some of these components will be made available as true services, deployed by a Trusted Third Party which governs the data protection related aspects of the process. Others will rather be stand-alone tools. Building blocks will be specified as the need arises, however they will most certainly include:

- De-identification (anonymisation & pseudonymisation) tools
    - EURECA can build on previous work such as CAT, the Custodix Anonymisation Tool, (developed within the ACGT project) which takes a generic approach towards de-identification of various data sources.
- Pseudonym management & patient matching
    - E.g. for longitudinal and multicenter follow-up of patients.

Finally, WP7 will address Intellectual Property Issues resulting from the EURECA architecture in a European and International context[217]. Issues to be analysed include property of patient data, licensing of the software developed, licensing of external software tools, legal protection of inventions made within EURECA and copyright issues. As the WP will create and provide software licence agreements for software developed within the project and contracts on data exploitation, it is likely that the policy based data protection enforcement can also serve as tool in technically governing IPR agreements. Legal advice in IP issues will furthermore be provided regarding the exploitation and dissemination of EURECA results, e.g. access rights, rights of use, etc.

WP8 – Q&A, Evaluation and Validation

Considering the user needs as described in WP1 and the corresponding intended pilots as described in WP 9, WP8 will identify specific application objectives to be tested, define the evaluation criteria and devise monitoring procedures to be executed by the involved stakeholder groups. Special care will be taken to involve the biomedical research and clinical end-user community, and the pharmaceutical industry as early as possible in the evaluation and validation effort. Technical validation will be conducted in tight collaboration with WP2-WP7 and WP9, and the procedures for the assessment of the adequacy of treatment of personal data will be established jointly with WP7. Adequate personal-data treatment is of special importance in the EURECA project as foreseen pilots will involve real clinical data.
This WP is crucial in the sense that it will continuously assess the prototype and applications within the EURECA environment and will report back to developers or users in order to ensure that the original goals will be reached. A set of guidelines and checklists to be used by testers in writing the validation and feedback reports will be proposed in the first 12 months of the project. The reports will suggest possible improvements, modifications and additional functionalities.

[217] Such as Agreement on trade-related aspects of intellectual property rights" (TRIPS); The European Patent Convention (EPC); DIRECTIVE 91/250/EEC on the legal protection of computer programs; DIRECTIVE 96/9/EC on the legal protection of databases; DIRECTIVE 2001/29/EC on the harmonisation of certain aspects of copyright and related rights in the information society.

Validation will include:

- Information input and output and respect of data-exchange standards,
- Semantic interoperability of EURECA components,
- Execution of realistic data mining and knowledge discovery scenarios,
- Integration and access to clinical trial registries and data,
- Appropriate coding and protection of access to personal data.

Attention will be given to:

- Usability and user-friendliness,
- Clarity and standardisation of definitions and vocabulary,
- Speed and robustness to user errors.

Evaluation Workshops will be held to discuss the reports between the groups which implemented the tools being tested and the testing groups. Developers and testers will assess the outcome of workshops and suggest refinements of the usability and evaluation criteria.

WP8 will coordinate with the evaluation group managers and will report to management and steering groups on progresses achieved and remaining problems to be addressed.

Through the implementation of key testing scenarios and sound reporting procedures, it is expected that design and implementation issues will be trapped soon in the development process. This work package thus establishes some of the important means to ensure that EURECA will provide a high-quality, secure and user-friendly software environment in the time frame of the project.

In addition to the scenarios which will be derived from the user needs analysis conducted in the context of WP1, nine clinical scenarios are potentially foreseen in view of the validation of the EURECA platform, described in Section 1.2.22.

For the effective execution of the worked planned, the WP will also focus on delivering the required training to end users, through structured user workshops to be held at each clinical site.

WP9 – Models, Deployment and Clinical Pilots

WP9 represents a critical activity in the project implementation plan as it is the place where all evaluation and validation activities of the technologies delivered by the project will take place, in the context of elaborate clinical pilots, under the coordination of WP8.

Another important task of WP9 is to build the development environment allowing the EURECA semantic interoperability layer and the EURECA software services to be designed and built. This includes providing sufficient schema-level data describing the structure and content of both the EHR and CT systems, based on the systems and clinical domains at the clinical sites participating in EURECA, to enable the development of canonical information models for the EHR and CT systems. Examples of such schema-level data are CRF fields from clinical trials, and archetypes of EHR data. Additionally, sufficient (and well-matched) patient data from both care and research systems need to be provided to enable the testing and validation of our interoperability framework and our software services. This data needs to be prepared according to the legal, privacy and security requirements. All this data will be stored in “surrogate” information systems that will be used during development, in order not to disrupt the clinical environment. In this WP the clinical sites will carry out the necessary work concerning structuring of data and proper interfacing with the EURECA environment.

In WP9 the canonical information models for EHR and CT systems will be built. We will describe the models making use of suitable standardized formalisms, such as CEN13606 (RM and archetypes), and HL7 v3.

In WP9 we will also evaluate existing mapping formalisms and tools that can enable seamless data linkage between the canonical model and the underlying implementation of the sources. Next to this, a solution will be provided to enable access to the various implementations of the sources via the uniform interface of the canonical model. Finally, WP9 will collaborate with WP4 to develop the appropriate mappings between the canonical models and the EURECA core data set.

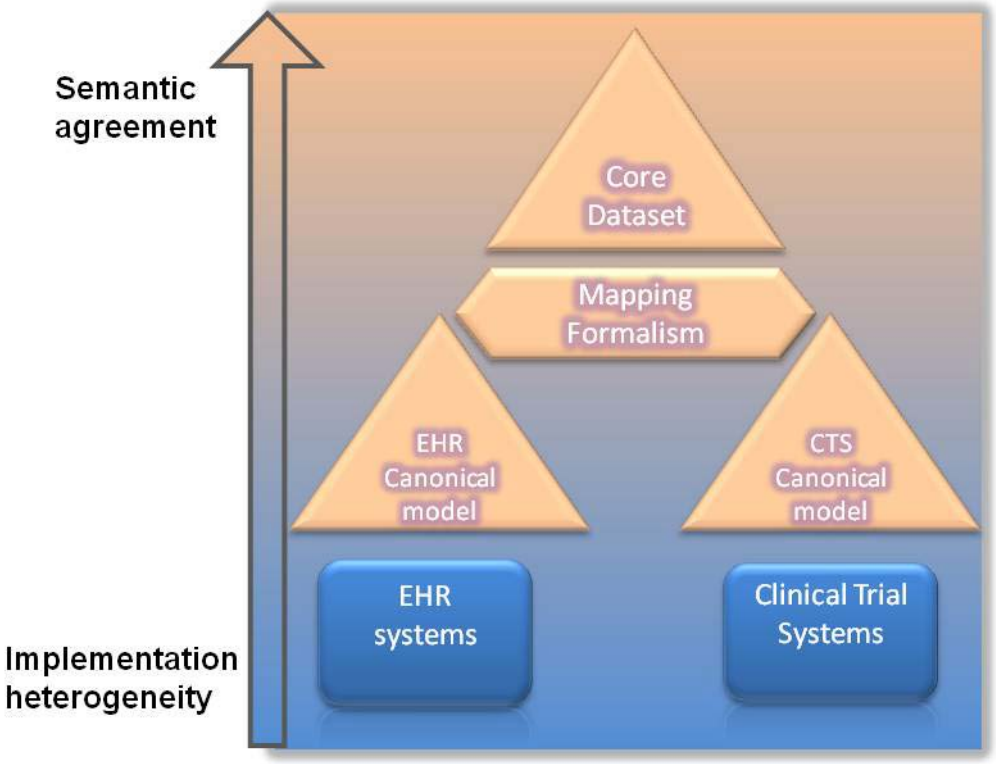

Figure 20 From implementation heterogeneity to semantic agreement

It is therefore also the responsibility of this WP to:

1. Interact with all the clinical organizations and clinical trials organization in the project with the objective that good canonical information models representing the EHR systems and the clinical trial systems of the pilot sites are developed. These information models are a required input to WP4 so that it can proceed with the mappings that bridge the semantic core data set with the information models representing the EHR systems and the clinical trial systems.
2. Prepare the technical and procedural infrastructure – in compliance with the defined legal and security framework of the project – for the installation of EURECA technologies and tools for their extensive evaluation and validation.
3. Organize the effective evaluation and validation activities of the EURECA technologies – once these are ready and delivered by the other project WPs – in accordance with the detailed evaluation and validation methodological framework developed by WP8.

All clinical partners have committed to participate as deployment and demonstration sites for the EURECA's semantic interoperability environment. Within the semantic interoperability environment, the EURECA software services/tools will be deployed, evaluated and validated. In each case the evaluation and validation will be carried out in the context of one or more of the EURECA scenarios that are most important for that organization. In the pilots we will also assess the qualitative and quantitative improvements brought by the EURECA solutions compared to the current state of practice.

Finally, it is the role of this workpackage to coordinate, based on contribution from all technical workpackages and on the lessons learnt during development, deployment and validation, the elaboration of a set of guidelines including minimum requirements and steps to be followed by an external clinical organization in order to become EURECA-compliant and to be able to join the EURECA interoperability environment and to make use of our advanced services.

WP10 – Knowledge Management

This work package has a multiple role: It has the responsibility to investigate and support the exploitability of the EURECA results and to propose realistic exploitation models, to disseminate the project's results to the user community, to contribute to standards whenever applicable and to protect the IPR of the project partners. The uptake of the project results by user communities is of key importance. These aspects are elaborated in Section 3.2.
Important exploitation goals of INTEGRATE are the sustainability of our solutions beyond the duration of the project and the use of the project outcomes by stakeholders outside of the consortium.
We believe that a successful dissemination of our results is key to the wide-scale adoption of our solutions. All project partners (especially academic partners and research centers) will present findings at high-level conferences and workshops, as well as publish research results in premium peer-reviewed journals. Industrial partners will complement dissemination of results through the realization of proof-of-concept prototypes, field tests, and demonstrations at conferences and professional exhibitions.

Exploitation

The loosely-coupled service-oriented interoperable EURECA environment offers the possibility to combine different models of exploitation. On the one hand, the tools and services could be provided as FOSS[218] to the community (which can contribute in turn). On the other, this does not exclude the potential of commercial exploitation for specific services and tools (offering maintenance, support, minimum service level, customisation, etc.).
The main exploitation goal of EURECA is the use of project outcomes by stakeholders outside of the consortium. One of the shortcomings of many exploitation efforts is that they fail to provide compelling usage scenarios and at the same time involve a steep learning curve for third parties often with disproportionally small expected benefits. To avoid this, already in the proposal phase, we have linked to several large user groups (BIG, GBG, EORTC, EuroSarc, METOXIA, SIOP[219]) that expressed interest to support us with clinical scenarios and in the requirements collection phase, and to test and use the EURECA software services and tools. We will also address these user groups to provide a prioritization of the EURECA outcomes according to the expected benefit. This prioritization is crucial for assessing the exploitability of the various tools and services.
We will focus on validating EURECA through measurable results in concrete clinical scenarios (e.g. percentage of automatically extracted CRF items from an EHR, number of patients that could be automatically recruited for a given trial, etc.). During the project we will examine together with our users the various workflows which the EURECA services can facilitate and disseminate that information to the user community. Consortium partners who are end users themselves will work from the beginning of the project to recruit other groups (or individuals) on the merits of the proposed outcomes.
EURECA refrains from developing a large tightly integrated infrastructure. Not only does such a tightly integrated infrastructure needs to be maintained and exploited by a single entity (restricting business models), it offers little or no flexibility for adapting to existing workflows, requiring most legacy to be replaced. The latter implies large investments, which seriously undermines sustainability.
EURECA envisages a different approach and therefore focuses on added value (as opposed to basic infrastructure functionality) by developing loosely coupled components deployed within a framework architecture. The tools and services provided will fit the different architectural frameworks at the user sites, allowing us to provide solutions that can be integrated into the existing workflow of various users with minimal effort.
Our services and tools need to be able to work with different clinical information systems at the clinical sites, and should not be dependent on a single EHR or clinical trial system solution. We will work to identify early on in the project existing procedures that represent high value to end users and the concomitant IT systems currently in use. A major goal of the exploitation work will then be to ensure seamless integration of our tools and services with those systems so as to ensure transparent use by the end users.
The above is particularly true for use of EURECA results within the commercial pharmaceutical domain, which can be important for the continuation of EURECA beyond the project end-date. Pharmaceutical companies generally do not share clinical trial related data. In order to address this user segment, which works in closed environments, it is important that the EURECA services can be deployed in the architecture existing at the user site, with little additional development effort.

[218] Free and open source software
[219] http://www.siop.nl/

Because of their importance to future exploitation, EURECA will make sure to meet all needs of the commercial pharmaceutical domain and has therefore established a strong link with industry through the EURECA Pharma Advisory Board and the relation to the EHR4CR project.
Enabling hassle-free integration of service components is key to the success of the EURECA interoperability framework. We will setup a certification programme checking services for “EURECA compliance” (similar to caBIG). This programme will be coordinated by the EuroRec Institute of whom certification and quality labeling is the core business. The “EURECA compliant” label will guarantee system integrators that a service can be deployed in a EURECA compliant environment without needing additional development.
Within this work package, we will explore different possibilities for exploiting the results of EURECA beyond the end of the project. Collaborations with the bio-pharmaceutical industry may lead us to propose exploitation models allowing us to create a sustainable eco-system around the EURECA architecture while still providing solutions for the non-profit community.

Dissemination

WP10 will set up a unique website, placed within ecancer.eu for posting of all significant results and notices from the intramural work of each WP. This involves interacting and understanding the nature of the obstacles and their ultimate solutions as they evolve. This requires close collaboration with all other WPs in EURECA. Through eCancer the project outcomes will be advertised and evaluated using established cohorts of clinicians, pharma and patient representatives.
The rules of engagement on communication to stakeholders, internal and external will be agreed and implemented from the beginning of the project. eCancer will advertise the existence, aims and progress of the project, agreed with all partners, by publication on ecancer.eu, editorials and advertorials in key newspapers, lay and technical, and promotions at cancer conferences in Europe and the rest of the world. The potential market for the deliverables of this project will not be restricted to Europe and the USA, as it is clear that the Far East and Australasia are fast becoming a growth market for clinical research and trials.
This WP will also address the patient-related community, and aim to describe relevant project results in language understood by patients and their carers. The patient preferences of styles of presentation of clinical information (e.g. clinical trial information to support enrolment) will be evaluated.
Next to eCancer and EuroRec, all clinical partners will contribute to dissemination, by sharing the results of the project within the various collaborative user networks in which they participate. The clinical partners will also be actively involved in the technical work in the project through the contribution of their IT departments. Within the project, they will become show cases demonstrating to the clinical community that standards-based semantic interoperability has a significant role in achieving better data integration and reuse, increasing the efficiency of clinical research, reducing costs of research and avoiding medical errors.

WP11 – Project Management

This work package comprises all activities related to the non-technical management of the project, including overall and work package coordination, preparation of and participation in project meetings (e.g. yearly review meeting), and project administration within the partners organisations. The project management is described in more detail in section 2.1.

1.3.3. Timing

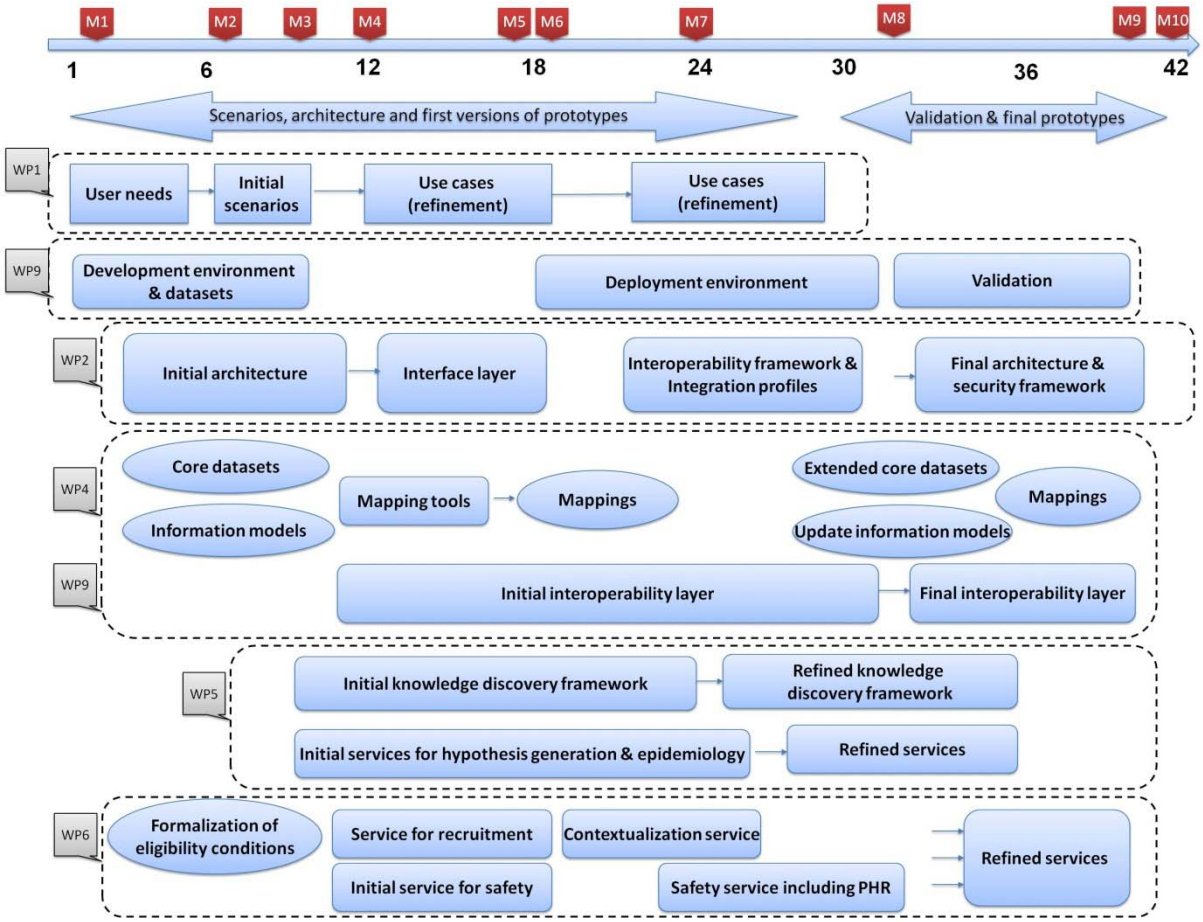

Figure 21 Main EURECA components

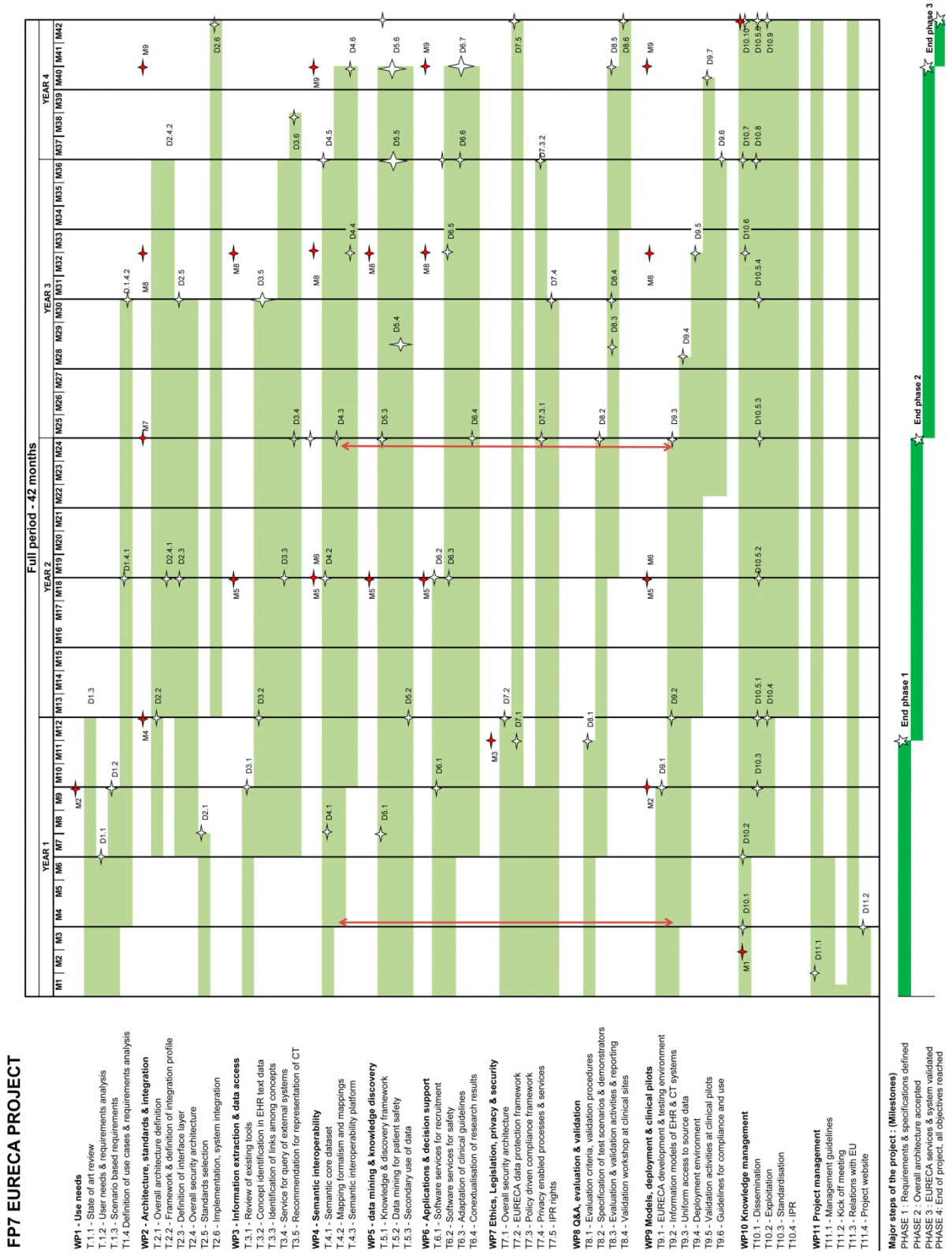

Figure 22 Planning overview

2. Implementation

2.1. Management structure and procedures

2.1.1. Project Management Structure

The project management structure given below is defined in order to successfully accommodate the complexity and scale of an Integrating Project. It is defined this way to clearly identify the responsible members of the various entities of the consortium, as well as to optimise communication between the different partners and the managing bodies of the project – the Management Board and the GA and the Advisory Boards – and the European Commission. The management structure is based on rules and regulations described in the Consortium Agreement.
Next to the management board, we have established a Technical Board of the project, chaired by a Lead Architect (the leader of WP2: Architecture, standards &integration). The role of this board is to discuss and to make decisions concerning the daily technical work in the project and to ensure close collaboration among the project's work packages during integration. All partners contributing to the ICT research and development work will be represented in the Technical Board. Through this Technical Board we will also ensure the collaboration with the semantic interoperability NoE and with other related projects on technological development aspects.
The project's Consortium Agreement will include amongst others, measures for the arrangement of confidentiality, IPR, exploitation rights, decision-making and change procedures, possible addition of participants and negotiation with third parties, as well as possible cooperation after the project's end. The Consortium Agreement will be in accordance with the EC Model Grant Agreement and General Conditions.

Organisation

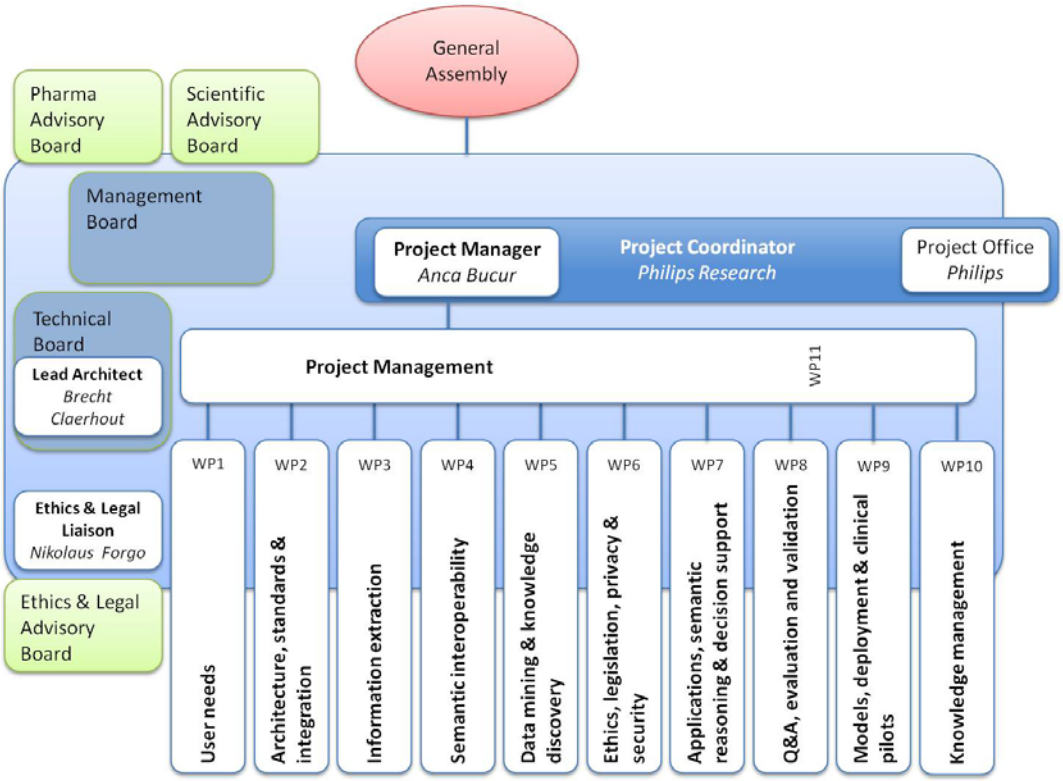

Figure 23 Project management structure

Specialist advisory panels

The management board has the authority to establish panels to advise it and support it in the proper management of and co-ordination of the project. These panels have an advisory role only.

The members of the panels will be senior representatives from the governmental health care authorities in the participating countries, with responsibility for ICT policy making and strategic direction of ICT developments, or recognized experts in areas relevant to the project. Representatives from industry and Standards' Organisations may also become part of the panels. The Panels will provide recommendations and guidance to the consortium and review project outcomes on specific topics. The project board will meet annually with the members of the specialist advisory panels and discuss the direction of the project, the project outcomes, and project strategic decisions. When possible, to save costs, these meetings will be collocated.

We have identified the need for three external advisory panels:

- An Ethics & Legal Advisory Board to include representatives of (a) European and National Patient Organisations, (b) national ethical committees form several EU countries, (c) experts. The committee will primarily provide external advice on ethical and legal issues related to the areas of the project. We have appointed an Ethics and Legal Liaison to make the link to the Ethics & Legal Advisory Board and to manage the regular communication (between the annual meetings with the Project Management Board).
- A Scientific Advisory Board to include representatives of user groups from clinical research and care and ICT research. Already identified members of the Scientific Advisory Board are EORTC, METOXIA, EuroSarc.
- A Pharma Advisory Board to include representatives of the pharmaceutical industry. Several members of the board have been identified. The leader of the WP10: Knowledge management will be in charge of organizing the annual meetings of the Scientific Advisory Board and of the Pharma Advisory Board with the Project Management Board.

Meetings

Collaboration is highly dependent on good communication at all levels. Partners communicate better if they have met each other in person. Therefore, EURECA will organise a **kick-off meeting** at the start of the execution of the project, in the presence of the project members of all partners.

A **General Assembly meeting** will take place at least once a year. This meeting will, if possible, coincide with major project events (e.g. contractual reviews, major milestones or decision moments, modifications to the project having a major impact), conferences or fairs. A representative of the coordinating partner, in general the Project Manager, chairs the GA meeting.

Interactive management by a regular **Board meeting** plays an important role in the communication strategy. The communication strategy aims to keep all the partners fully informed about the project status, the planning and all other issues that are important to the partners to obtain maximum transparency for all involved and to increase the synergy of the cooperation. The Board meets regularly, in person or by telephone conference, to ensure a timely and effective discussion on overall technical progress as well as technical issues. The Project Manager coordinates the Board and chairs its meetings and is responsible for the agenda and meeting minutes.

Work Package meetings will be organised to discuss technical issues and to monitor and track the technical progress of the WP. Each WP is responsible for a number of deliverables and milestones due at certain dates. The Work Package Leader coordinates the WP meetings and is responsible for the agenda. Meeting minutes will be made available to the Project Manager, the Board and to all the project members involved. In addition **internal workshops** (consortium meetings) will be arranged to discuss specific technical issues, with attendees in varying configurations, to synchronise dependent tasks, intra as well as inter WPs. **Telephone conferences** (telco's) will be organised, on a regular basis, in varying constitutions, to deal with the day-to-day coordination of the project.

The Board members will hold **internal review meetings** to review and approve all the deliverables, technical reports, publications and press releases. Sufficient time will be planned to allow for corrective comments before the final release of deliverables or publications.

The **Annual Review meeting** of the project will take place on request of the European Commission. During this meeting, the technical progress of the project is evaluated by the EC Project Officer and external auditors, who are experts in the field. The EC Project Officer together with the Project Manager determines date, agenda, duration and location.

- Period 1: M1 – M12, review expected in M14
- Period 2: M13 – M24, review expected in M26
- Period 3: M25 – M42, review expected in M43

Concertation Meetings are organised by the EC. On invitation of the Scientific Project Officer, a delegation of the EURECA Project will participate actively in these meetings.

Travel

The EURECA project will require collaboration and co-ordination between work packages, between the General Assembly, Management Board and Steering Committees, and with external boards to be set up for providing external expert advice and guidance (including patient organisations), such as the External Advisory Board and the Ethical Advisory Board.
In order to create such collaboration and ensure coordination, there will be a high requirement for travel so participants may interact on a personal level. To avoid multiplying travel expenses, electronic multimedia collaboration will be used as much as possible and physical meetings restricted when appropriate. Also, when possible meetings involving the same participants will be co-located. We also assume that each cycle of meetings will be held in the same week, in the same location, so that costs can be reduced for participants involved in multiple sessions.
The travel needs are estimated on implementation of consistent, cost-effective, fully documented travel policies, procedures and practices. The requirement to travel will not be taken lightly and we will ensure appropriate measures are in place to control expenditures as tightly as possible.
The travel obligations include:

- 1 General Assembly meeting annually attended by all Contractors
- 4 management board meetings annually attended by all members of the Management Board
- 4 technical management committee meetings annually
- 1 meetings of the steering committees and specialist advisory panels annually
- 1 consortium meeting annually
- 2 trips by each participant annually for promotional activities (conferences, workshops, fairs, etc)

Information Interchange

All information such as minutes of meetings, technical contributions to deliverables, and relevant publications will be communicated to the Project Manager. She is responsible for channelling this information to the other partners, where appropriate. It is the strategy of the Consortium to guarantee a fast and complete flow of information.
Communication will be done preferably by e-mail and telephone. Technically, the communication is supported by specific **e-mail reflector lists**. Specific e-mail reflectors can be requested and will be maintained by the Project Office.
Furthermore, at the start of the project and as part of the management activities, a dedicated web-based communication structure will be set up, with **shared workspace environment** properties. This environment offers possibilities for posting and retrieving documents, discussion groups, mailing lists, FAQ, news, calendar, etc. All consortium members will actively use this interactive knowledge-sharing platform for information exchange. This communication structure is consortium confidential with restricted access.
The communication strategy also aims to communicate effectively with parties outside the consortium, such as the European Commission, other European project consortia, other potential users of the technology and standardisation committees. Therefore, the project will launch a **public project website** as part of this strategy. This website will also be used as a portal for dissemination and contains therefore, amongst others, a public project summary, public presentations, a public events calendar and an overview of dissemination activities.

2.1.2. Management Procedures

Decision-making Mechanisms

Regular way-of-working

Procedures for making decisions at a managerial level will be detailed in the Consortium Agreement. At the highest level, the General Assembly (GA) will provide a forum for discussion and will decide on management issues. Decisions of the GA are binding for the project. The GA will decide on major modifications of the overall work plan, budget distribution, possible addition of and calls for new partners, and will prepare requests for amendments to the European Commission. The Project Manager of the coordinating partner Philips will chair the GA. The GA serves as an advisory and overall monitoring forum for EURECA and holds a decision-making role when it would come to a disagreement in the Board.
The **Board** will decide on strategic and operational issues, such as the monitoring of the technical progress of the project, the implementation of procedures and structures, the definition and (re-) allocation of activities, technical roadmaps WP interactions and conflict management.

Day-to-day decisions at the technical level will be made by the Work Package Leaders (WPLs), where needed after consultation with the Project Manager.

Problem handling

A Consortium Agreement will be made for the arrangement of confidentiality, IPR, exploitation rights, decision-making and change procedures, possible addition of participants and negotiation with third parties, as well as possible cooperation after the project's end.
The philosophy of the Consortium regarding problem handling and corrective actions is in the first place based on prevention. In case a problem arises, it will be tackled as soon as possible and at the lowest possible level, meanwhile bringing the problem and the proposed solution to the immediate attention of the Project Manager.
Each partner of the consortium is responsible (liable) for the execution of any part of its share of the project and for fulfilling other EC contractual obligations. In case, however, that a partner performs poorly, due to technical and/or resource reasons, this will be promptly documented in project meeting minutes. The project manager will try to solve the problem immediately by all possible means thus limiting the impact for the project. If necessary, an extra meeting with the Board will be organised. Meanwhile, the Project Manager will, if necessary, send an official letter to this partner (with notification to GA and EC project officer) stating the possible jeopardy of the project.
In case the problem cannot be solved by the Board, it will be escalated to the next level, the GA and an official letter, stating the possible jeopardy of the project, will be sent by the Project Manager to the defaulting partner. A meeting of the GA will then be organised on short notice. The GA will either solve the problem or decide on the final corrective actions which may be banning the partner from the project. All actions will be properly documented. More details on the procedure are described in the Consortium Agreement.

Progress Monitoring

A dedicated Work Package, WP11 – Project Management, is in charge of all project management issues, including overall technical progress monitoring.

Reporting

The technical Work Packages are responsible for a detailed planning for the tasks of their WP. Each of the partners and Work Package Leaders will report on a regular basis to the Project Manager on:

- the progress of the work;
- status of the deliverables;
- compliance to and updates on the planning;
- possible problems and their solution.

With this information, the Project Manager will keep track of the overall project status, in order to control the project, take timely measures and regularly update the overall project Gantt chart and the manpower matrix.
The consortium will provide the contractual obligatory technical reports, as mentioned in the Model Grant Agreement with the EC, with the reporting frequency agreed upon. The Project Office will coordinate the realisation of these reports. The Project Manager will take care of the distribution of the technical reports to all parties concerned.

Project Calendar

The Project Coordinator will maintain a Project Calendar, containing all relevant dates and deadlines regarding technical events, deliverables and milestones of the Project. It is the responsibility of each Partner to keep the calendar up-to-date and to inform the Project Coordinator on changes. External or public events will also be put on the public project website.

Quality Assurance

All partners will execute quality control and assurance procedures according to their internal standards and procedures. Furthermore, quality assurance issues will be on the agenda of the Board meetings, and preventive and corrective actions will be taken where necessary.

2.2. Beneficiaries

Partner 1 – Philips

Organisation Profile

Royal Philips Electronics of the Netherlands is a diversified Health and Well-being company, focused on improving people's lives through timely innovations. As a world leader in healthcare, lifestyle and lighting, Philips integrates technologies and design into people-centric solutions, based on fundamental customer insights and the brand promise of "sense and simplicity".

Headquartered in the Netherlands, Philips employs approximately 116,000 employees in more than 60 countries worldwide. With sales of EUR 23 billion in 2009, the company is a market leader in cardiac care, acute care and home healthcare, energy efficient lighting solutions and new lighting applications, as well as lifestyle products for personal well-being and pleasure with strong leadership positions in flat TV, male shaving and grooming, portable entertainment and oral healthcare.

Healthcare

We combine our expertise in medical technology with the clinical know-how of our customers to produce innovative solutions that meet not just the needs of individual patients, but which also enable healthcare professionals to work faster, more easily and more cost-effectively. Our medical customers have ranked us number 1 in cardiovascular X-ray, digital X-ray and ultrasound, patient monitoring systems, nuclear medicine, cardiology systems and critical care systems.
Key products: Brilliance CT scanners, Panorama MR, iU22.

Technology

Technology is the driving force of our Healthcare and Lifestyle products. And in today's world, in which increasingly more aspects of our lives are being enabled by technology, it is the continuing Philips tradition of innovation that will enable us to provide the solutions that realize the full potential of fast-evolving digital technologies.

Healthcare Information Management

The department Healthcare Information Management of Philips Research Europe focuses on Information Management in Clinical Applications and in Home Care environments and on Clinical Decision Support Systems. The two main clinical domains currently addressed are oncology and cardiology, next to generic solutions regarding clinical information systems, information integration, domain modelling, medical imaging, standardization and interoperability. Beside carrying out industrial research projects for Philips Healthcare, the members of the group have participated in and led numerous collaborative projects at National (Dutch-government funding) and European level.

Participant's role in the project

In the EURECA project Philips will be responsible for the coordination of the project, carry out research in the areas of semantic interoperability, standards-based modelling, reasoning and decision support, and providing uniform access to external sources of data and knowledge based on standard formalisms. Philips will contribute to the development of the EURECA architecture and tools, and to the integration of the EURECA environment, and lead WP9 Models, deployment and clinical pilots. The partner will also contribute to the elaboration of the exploitation strategy and plan and to the dissemination of the project results.
Philips will be involved in all work packages of the project.

Staff Members Profile

Dr. Anca Bucur holds a Master in Compilers and Distributed Algorithms from the Technical University of Bucharest, Romania, and a PhD in Distributed Systems from Delft University of Technology, the Netherlands. She has been with Philips Research Europe since 2003, as senior scientist, and has lead and participated in several industrial research projects in the healthcare domain. She has carried out research projects in Clinical Information Systems, medical imaging, and computational genomics, having as main client Philips Healthcare. She has also led Philips' contribution to the Dutch-funded collaborative VL-e project and to the European FP6 ACGT project, where she was also responsible for the work package "Distributed data access, tools and application". Her research interests and expertise include clinical information systems, clinical decision support, parallel and distributed algorithms and applications, and computational genomics, and is

the author of many research papers and patents in those areas. She is currently the coordinator of the FP7 INTEGRATE project.

Dr. Richard Vdovjak graduated informatics at the University of Zilina, Slovakia. He obtained a professional doctorate in software engineering and a PhD in computer science focusing on model-driven distributed ontology-based information systems, at the Technical University of Eindhoven, the Netherlands.
In 2005 he joined the Healthcare Systems Architecture Group in Philips Research Europe as senior scientist. He has led several research projects dealing with issues related to distributed picture archiving and communication systems, as well as issues related to semantic interoperability in the context of clinical information systems in the domain of breast cancer. He was also involved in research on model-driven software development. He has published numerous research papers and patents in the above areas.

Dr. Hans Jonkers studied Mathematics and Computer Science at the Technical University of Eindhoven, the Netherlands. From 1978 to 1982 he worked for the Centre for Mathematics and Computer Science in Amsterdam where he was involved in the Algol 68 project. After receiving his Ph.D. in Computer Science at the Technical University of Eindhoven he joined Philips Research in 1982. He has since then been involved in research and technology transfer in the areas of software development methodology in general, and formal methods and model-driven development in particular. He has been involved in the development of the COLD specification language, the SPRINT software development method, the ISpec interface specification approach and the Vampire model-driven development environment that have been or are being applied in product development in Philips. His current interests are in meta-modeling, model-driven development, computer interpretable guidelines and HL7 Version 3. He is a principal scientist in the Healthcare Information Management group, and has extensive expertise in industrial research.

Partner 2 – FORTH

Organisation Profile

The Foundation for Research and Technology – Hellas (FORTH) is one of the largest research centres of Greece with well - organised facilities and a highly qualified staff. The research and technological focus of the foundation is centred on selected areas of great scientific, social, and economic interest. Two Institutes of FORTH, with their relevant laboratories and groups, are to be involved and contribute to the current project.
The Institute of Computer Science (ICS), since its establishment in 1983, is a pioneering contributor towards the deployment and adoption of Information Society Technologies in Greece and plays a leading role in worldwide efforts towards the development of an Information Society accessible and acceptable by all citizens. The FORTH-ICS group involved in the current proposal is the **Biomedical Informatics Laboratory (BMI Lab)**. The BMI lab at FORTH-ICS has established a tradition of internationally acknowledged excellence in conducting high-level R&D work and in developing innovative systems and services. Its research activities focus on the development of innovative computer methods and tools in the area of medical and biomedical informatics, ehealth, m-Health, medical imaging and bioinformatics. Recently the lab is also focusing its R&D activities on biomedical modelling and simulation in the wider VPH research context. The laboratory is currently involved in several projects and initiatives related to the topics of the project. In specific the group is coordinating the **ACGT project**- an IST Integrating project aiming at the development of the European Biomedical Informatics Grid for cancer research, the **ContraCancrum project** – an FP7 STREP project focusing on multi-level cancer modelling, and the **TUMOR project** – an FP7 STREP project focusing on model interoperability for cancer research (international collaboration with MGH).

Participant's role in the project
FORTH will be involved in all work packages of the project and lead WP8: Q&A, evaluation and validation, being in charge of organizing the evaluation and validation of the project. FORTH will also have significant contribution to the definition of the architecture, the development of the semantic interoperability environment and of the EURECA software services.

Staff Members Profile
Dr Manolis Tsiknakis received a B.Eng. degree in electronic engineering, a M.Sc. in microprocessor engineering, and a Ph.D. in control systems engineering from the University of Bradford, U.K. In 1992, he joined FORTH's Institute of Computer Science (FORTH/ICS) where he is currently a Principal Researcher, coordinating the activities of the Center for eHealth Technologies of the Biomedical Informatics Laboratory. He has been the Principal Investigator in many collaborative R&D projects. He is the initiator and co-chair of the ERCIM Biomedical Informatics Working Group. He is a member of the Programme Committee of a number of relevant conferences and on the editorial board of relevant scientific journals. His current research

interests are in the areas of biomedical informatics, service oriented architectures, semantic information integration, and service platforms for pervasive eHealth and mHealth service. He is a member of IEEE and ACM. Dr. Tsiknakis is the scientific coordinator of the ACGT Integrating project.

Dr. Kostas Marias holds a Principal Researcher position in the Institute of Computer Science (ICS-FORTH). During 2001-2003, he worked as a Researcher at the University of Oxford on a medical image analysis project related to the early detection of breast cancer in women that use Hormone Replacement Therapy. He completed his PhD in the field of Medical Image Analysis/ Medical Physics in 2001 (UCL London, Royal Free Hospital) working in the Medical Vision Lab, University of Oxford. His PhD research focused on the development of software for the non-rigid registration of mammograms which he further developed as a part of a commercially available CAD system. During 2000-2002, he was a senior consulting scientist with the diagnostic software company Mirada Solutions Ltd. (UK), a spin-off from the University of Oxford. He also holds an MSc degree from Imperial college of Science, Technology and Medicine in Physical Science and Engineering in Medicine as well as a Electrical Engineering Diploma from the National Technical University of Athens (N.T.U.A). Currently he is the coordinator of 2 EC projects on cancer modelling (ContraCancrum and TUMOR) and is actively involved in providing open access image analysis/modelling tools in the clinical setting for the promotion of predictive oncology. He has published more than 60 papers in international journals and conference proceedings in the above fields.

Dr V. Sakkalis holds an Associate Researcher position in the Institute of Computer Science (ICS-FORTH). He holds a PhD in Electronic and Computer Engineering and is currently a member of the Institute of Computer Science – Foundation for Research and Technology (ICS - FORTH). Previously he completed his Masters degree at Imperial College of Science, Technology and Medicine, UK. His background falls in Biomedical Engineering, Atomic-Molecular Physics, Optoelectronics and Laser. His research interests include biosignal and image analysis, visualization, classification algorithms, biostatistics and biomedical informatics and modeling. He is currently the Technical Coordinator of TUMOR (ICT EC project concerning caner modeling interoperability). He has published more than 60 papers in scientific archival journals, proceedings of international conferences & workshops and scientific newsletters, related to his fields of expertise.

Mr Stelios Sfakianakis received his BSc in Computer Science in 1995 and his MSc with highest distinction in Advanced Information Systems in 1998 from the University of Athens. In January 2000 he joined the ICS-FORTH's Biomedical Informatics Laboratory (BMI). His interests include the semantic integration and composition of services in state of the art computational environments such as the Grid and the Semantic Web. In the past he has worked in the design and implementation of a service oriented architecture for the realization of the Integrated Electronic Health Record by the means of CORBA and Web Services middleware technologies. On the technical side his experience spans the application design and development using the Unified Modeling Language (UML), the development of distributed systems using CORBA, Web, REST, and Grid Services, and the design of OWL/RDF-S ontologies and their employment in the semantics-based description of services. Additionally he in interested in the design of programming languages and their various paradigms, such as functional, object and data oriented, and concurrent. He was actively involved in the ACGT integrating project.

Partner 3 - IJB

Organisation Profile

Located in Brussels, the **Institute Jules Bordet (IJB**) is an autonomous comprehensive cancer centre devoted entirely to the fight against cancer. The institution is connected with two important hospitals networks: it is the anticancer centre of the Free University of Brussels and is part of IRIS, the network of public hospitals in Brussels and it coordinates the “Programme de Soins Oncologiques” for these two networks. It is also an active member of the Organisation of European Cancer Institutes[220]. Its mission is triple: to provide optimal care to patients, to organize high level teaching activities and to conduct fundamental, translational and clinical research.
Every day more than 115 physicians and 600 other individuals join efforts to screen and treat men and women from all social strata suffering from cancer, to perform research, and to disseminate their expertise around the world through educational programmes, lectures and publications. More than 2000 patients with an incident cancer are yearly taken in charge and the data related to the patients or to the tumours are recorded in the hospital registry. Leading laboratory research is developed at the Institute and around 70 innovative clinical trials are initiated each year thanks to the efforts of more than 100 dynamic researchers.

[220] http://www.bordet.be/en/presentation/organ/oeci_e.html

The Institute Jules Bordet has an international reputation in various fields, especially in breast cancer research. It provides multidisciplinary care to more than 500 new breast patients each year and plays a leading role in translational /clinical research for this disease, both through the enrolment of women with breast cancer into clinical trials and through leadership in international collaborative research groups (EORTC/BIG). In particular, the **department of Medical Oncology** is actively participating into more than 200 clinical trials (35 of Phase I, 80 of Phase II and 78 of Phase III as well as many academic trials). Most of the active trials are conducted at European level and few of them worldwide.

The **development of IJB information systems** is mainly focused on aggregating and consolidating medical information on an harmonized platform. We aim at storing and exchanging information according to international norms (mainly DICOM and HL7), geared towards semantic interoperability. Many reports (discharge documents, consultation visits) are already stored and exchanged using the Clinical Document Architecture (CDA) format of HL7, and are undergoing a process of normalization according to HL7 RIM Information Model.

The Breast European Adjuvant Studies Team (BrEAST) is a specialized clinical trials unit (data centre) in conducting large, international Phase III studies in breast cancer aiming to register new drugs. The unit is responsible for setting-up, coordinating and managing the data collected in these trials. The BrEAST's team consists of over 40 professionals, including physicians, project managers, data managers, computer scientists, clinical monitors, quality assurance managers, safety specialists, data entry staff, and administrative assistants.

Participant's role in the project

In **EURECA** the IJB ICT department participates in WP4 and WP9 by providing models for EHR data, collected at various key points in the patient's clinical process, that were coded and normalized using international coding systems (as LOINC, SNOMED CT, ICD9 and ICDO) for each domain relevant to screening for inclusion a in a clinical trial and adverse event reporting. An export of clinical trial databases will be provided for many studies handled by IJB data centres. Moreover, a database of all clinical trials approved by IJB Ethics Committee is also maintained including a description of their eligibility criteria, and will serve as a starting point for providing models for clinical trial description. IJB also contributes to user requirements and to the development and validation of the scenarios concerning improved recruitment, hypothesis generation and protocol feasibility.

Staff members profile

Ph. Hennebert, B. Sc, earned a degree in Mathematics from the Free University of Brussels, Belgium in 1993. After working as a mathematician for the Pharmacokinetics Laboratory and the Microarray Unit of IJB, he joined the team of the Medical Oncology Department, where he set up FDA-compliant Clinical Trial Databases. In 2001, he set up ICT workflows, database and software for the local Cancer Registry at IJB. Since 2002, he has been designing the Electronic Health Record software for IJB, focusing on semantic interoperability norms, in particular HL7 & Clinical Document Architecture (CDA) formats, and is now acting Director of IJB IT Department.

Pascal Gil, MD received his Medical Doctor degree from the University of Paris in 1994 and his degree in healthcare law in 1991. He has an extensive experience in pharmacovigilance, clinical trials set up and quality control. He entered the French Health Authorities in 1987, was in charge of implementation of the new European Regulation and then extended his skills through various experiences as an expert (laboratory, CRO). He acts as the clinical trials coordinator at the IJB

M. Paesmans, M. Sc, received a degree in mathematical sciences (1985) and in actuarial sciences (1988), from the Free University of Brussels. She started working at the Institut Jules Bordet as ELCWP statistician and joined progressively several other collaborative research groups. Responsible for the IJB's Biostatics Unit since 1990, she is head of the Data Centre since 2001 that developed and maintained the hospital cancer registry and underwent many research activities.

Stella Dolci was responsible for the setting up of the BrEAST data centre. Today her main responsibilities cover project management from study set up to final analysis, including team and workflow management, liaison with study statisticians for data transfer, assistance with clarifications on database set-up and provision of descriptive statistics as required.

Partner 4 - Custodix

Organisation Profile

Custodix is a private limited company established in 2000 specialized in data protection solutions for eHealth. Today Custodix is recognised as one of the most advanced and reliable trusted service providers (TSPs) in the healthcare sector providing both consultancy and technical privacy protection solutions. Its data protection services are supported by a combination of high-level technical and legal expertise in privacy protection and e-security and considerable experience with a number of advanced e-security solutions such as PKI, VPN, Web Services security frameworks, time-stamping and more.
Custodix helps businesses deal with all security and privacy related aspects of modern data management (cf. European Data Protection Directive 95/46/EC, 2001/20/EC, 2005/28/EC, the WHO GCP, etc) by offering end-to-end solutions for data protection in e-clinical trials, disease management and other longitudinal studies. Custodix has an international customer base, deploying services in Europe, Asia and Australia. Customers include commercial companies and government research organisations.
In its short existence, Custodix has already been involved in several IST projects (FP5, FP6, FP7) as e-privacy and e-security and distributed architecture experts. These include the EU flagship project: ACGT (Advanced Clinico-Genomic Trials on Cancer, developing a bio-medical GRID infrastructure for sharing clinical and genomic expertise, 2006-2010) and the TAS3 (Trusted Architecture for Securely Shared Services) project developing an architecture with trusted services to manage and process distributed personal information, 2008-2011.
Furthermore, Custodix leads the technical workgroup and system architecture design for the EHR4CR (Electronic Health Records for Clinical Research) IMI (Innovative Medicines Initiative), in which 11 pharmaceutical companies together with 22 public partners aim to develop a scalable and cost-effective approach to interoperability between Electronic Health Record systems (EHRs) and Clinical Research (for use in pharmaceutical clinical research).

Participant's role in the project
With its background as commercial IT service provider, Custodix will coordinate the work on defining and implementing the EURECA architecture. Custodix is also the main responsible for IT-security and privacy related technical tasks.

More specifically, the contribution will include:

- Coordinating the design and implementation of the EURECA architecture and reference infrastructure.
- Design and implementation of the main security infrastructure and services (Identity Management and Provision, Authentication, Authorization services, etc.), which includes design of a framework for making data protection policies (derived from legislation and regulations) manageable, enforceable and auditable over different, heterogeneous data sources (in light of regulatory compliance).
- Design and implementation of specialized tools and services necessary for ensuring privacy protection of patients, such as de-identification and patient identity management services (pseudonym management & record linkage).
- Contributions to the EURECA exploitation strategy.

Custodix will be involved the following work packages of EURECA: WP1, WP2, WP4, WP6, WP7, WP8, WP9, WP10, WP11. Custodix will also lead the Technical Board of the project and coordinate the contribution to the common info-structure set up by the NoE on semantic interoperability.

Staff Members Profile
Brecht Claerhout holds a master degree in electronics engineering. He has previously been active in FOSS development as author of a major network security tool (Sniffit). He has worked at the IMEC (Interuniversity Microelectronics Center) and RAMIT (Research in Advanced Medical Informatics and Telematics) research groups. Brecht is currently employed by Custodix, one of the first Trust Service Providers in the world focusing on data protection, and has been actively involved in a large number of European research projects as company scientific lead, the two most recent being ACGT (Advanced Clinico-Genomic Trials on Cancer, 2006-2010) and TAS3 (Trusted Architecture for Securely Shared Services, 2008-2011).

Brecht has previously advised the former Belgian Telematics Commission (Ministry of Health) on several security subjects; and is acting as a guest speaker at Ghent University. Brecht has published several conference and journal papers on the subject of security and privacy protection.

David Voets holds a master degree in software development. He has participated in several local (Flemish IWT) and European (IST) research projects (e.g. INFOBIOMED NoE, ACGT IP), aimed at enforcing European (Bio)medical Informatics as an integrative discipline with a view on supporting individualised healthcare. In these projects, he has worked on privacy protection solutions and generic policy driven privacy protection services. He has experience with building service oriented architectures for healthcare applications, such as data collection platforms, clinical trial and EHR software.

Partner 5 - UdS

Organisation Profile

The Saarland University was founded in 1948 in co-operation with France. Today the University counts 15.500 students of whom 7 percent are foreign students. The Saarland University has 8 faculties and provides the broad spectrum of disciplines typical of a classical universitas litterarum. At the Faculty of Medicine (University Hospital), located in Homburg / Saarland more than 1800 people are studying medicine. There are 36 hospitals or institutions treating more than 54.000 inpatients and nearly 190.000 outpatients each year. Participants from the Saarland University is the department of Paediatric Oncology and Haematology, that is responsible for the care of patients in the Saarland and the surrounding area. The focus in research of the Department of Paediatric Oncology and Haematology is nephroblastoma (clinical study and trial and basic research in cooperation with different institutes) and brain tumour.
The **clinical expertise** is most relevant for the project. Prof. Graf is a board member of different study groups in the German Paediatric Oncology and Haematology Society (GPOH). This includes trials for Osteosarcoma (COSS 86, COSS 91, COSS 96), Non-Hodgkin-Lymphoma (NHL-BFM 90, NHL-BFM 95), Brain tumour (HIT 91, HIT-SKK 92, HIT-Rez 97, HIT-LGG), Nephroblastoma (SIOP 93-01/GPOH and SIOP 2001/GPOH) (chaiman), Acute Myelogenous Leukemia (AML BFM 98). He is a board member of the international study Committees on SIOP Brain Tumor and SIOP Nephroblastoma. He is the chairman of the International SIOP Renal Tumour Study group (SIOP-RTSG). Prof. Graf is a board member of the Informatics Section of GPOH. Furthermore UdS has participated in the FP7 project ContraCancrum (ICT-2007-…..) as leader of WP9, and in the BMBF funded KPOH project, a Competence net for Paediatric Oncology.

Participant's role in the project

An important of the Saarland University is being the link between clinicians and IT-people to bridge the gap between research and clinical daily practice. UdS will be leader of WP2, dealing with the user requirements for the EURECA tools and services. UdS will also actively participate in the development of the semantic interoperability environment and in the development, deployment and evaluation of several relevant EURECA services.

Staff Members Profile

Prof. Dr. Norbert M. Graf is Professor of Paediatrics and Director of the Clinic for Pediatric Oncology and Hematology, a member of the Faculty of Medicine of the University of the Saarland in Germany. Prof. Graf is a member of the German Society of Paediatrics, the Austrian Society of Paediatrics, the German Society of Paediatric Oncology and Hematology (GPOH), the German Cancer Society, the Cancer Society of the Saarland, Germany, the German TNM Committee (representative for the German Paediatricians), the European Bone Marrow Transplantation, the International Society of Pediatric Oncology (SIOP), the Pediatric Society of Bone Marrow and Stem Cell Transplantation, the European Haematology Association and is an Associate Member of COG (Children's Oncology Group, USA). He is a member in the Editorial Boards of the Cochrane Childhood Cancer Review Group and Paediatric Blood and Cancer. He is currently the coordinator of the EU FP7 P-Medicine project.
Holger Stenzhorn studied computational linguistics at Saarland University in Saarbrücken, Germany and is currently a research associate at the Department of Paediatric Oncology and Haematology of the Saarland University Hospital in Homburg, Germany. Before, he had positions as a researcher at the Institute for Medical Biometry and Medical Informatics of the Freiburg University Medical Center, Germany and the Institute of Formal Ontology and Medical Information Science in Saarbrücken, Germany, and he was visiting researcher at the Digital Enterprise Research Institute in Galway, Ireland, before he worked as a software engineer at XtraMind Technologies in Saarbrücken, Germany. His work focuses on the representation and management of information and data, ontologies and Semantic Web technologies, biomedical informatics, natural language processing, user interfaces and software design and development. In the past he participated in the development of multilingual document retrieval, information extraction, and natural

language generation systems, both in industry and academia. He has been involved in several ontology engineering and application tasks: an ontology for clinical trials on nephroblastoma and breast cancer (EU-funded ACGT project), an ontology for the research on cerebral aneurysms (EU-funded @neurIST project) as well as the BioTop top-domain ontology. His main work at the moment focuses on developing a software system (ObTiMA) for the improved management of clinical trials. Further, he is a member of the Healthcare and Life Sciences Interest Group of the World Wide Web Consortium.

Partner 6 - UOXF

Organisation Profile

Molecular Oncology Laboratories at the Weatherall Institute of Molecular Medicine, the latter is also a University Institute and houses over 300 Scientists working in many different areas of human biology and medicine. Adrian Harris group comprises several tenure-track or tenured Scientists working on different aspects of cancer biology, particularly hypoxia and tumour angiogenesis applied to breast cancer biology. The Clinical Unit has 8 dedicated beds for clinical research in Phase I Trials and on-site facilities for MRI, PET scanning, antibody scans, tumour biopsies and pharmacodynamic monitoring of Phase I Trials. The Unit is a designated Phase I Centre by Cancer Research UK. National Translational Research Network funded by the National Health Service for Phase I / II Trials.
Cancer Research UK Oxford Cancer Centre: CRUK have funded a Translational Research Unit in Oxford, via the University since 1989. This comprises the CRUK Medical Oncology Unit at the Churchill Hospital, adjacent to the Clinical Oncology and Radiotherapy Department, and the CRUK Molecular Oncology Laboratories at the Weatherall Institute of Molecular Medicine. The quinquennial review for the CRUK Cancer Centre gave it alpha star ranking. It is also a CRC Phase I centre. CRUK bring in expertise in discovery and analysis of new genes in the hypoxia transcriptome and in the analysis of human cancers, with focus on breast cancer, to discover pathways related to hypoxia and angiogenesis; and development of therapeutic approaches.
EuroSarc: Translational research and clinical trials in rare bone sarcoma (this is under evaluation, Oxford is co-applicant and active component).
Eurobonet: translational research in rare bone sarcoma (ccomponent based in Oxford University, lead based in Leiden). This project will end in July 2011.
These two components connects Oxford University with European work programmes on the basic biology of rare bone sarcomas, tumours that frequently occur in the younger age group and therefore have treatment late effects. The translational element includes whole genome expression array analysis of primary tumour material that requires linking with patient outcome data and health records.
Oxford is also an active component of Metoxia and of the P-Medicine integrating project.
UOXF bring to EURECA expertise in the areas of breast oncology, bone tumour and soft tissue sarcomas, data-mining and bioinformatics. The organization will provide the project with clinical user requirements, clinical knowledge, clinical trial and care data, and all needed information on clinical systems.

Participant's role in the project

UOXF is the leader of WP5: Data mining and knowledge discovery. UOXF is a pilot site for several scenarios concerning semantic interoperability, hypothesis generation, and identification of potential safety risks based on large scale data mining of EHR data. UOXF will also bring to EURECA their vast expertise in clinical and translational research and link to large external networks of clinical organizations (Metoxia, EuroBoNet, EuroSarc)

Staff Members Profile

Dr. Francesca Buffa is a Senior Research Fellow in Bioinformatics, at the Weatherall Institute of Molecular Medicine, University of Oxford. Previously, Dr. Buffa has held the following posts: Research Fellow, Human Cancer Biology and Informatics, Gray Cancer Institute (2003-2006); Research Fellow, Institute of Cancer Research, London (2001- 2002) ; PhD Fellow, Institute of Cancer Research, London (1997-2000). Dr. Buffa specialises in Bioinformatics and Biostatistics. Her Areas of Interest are: computational genomic, prognostic and predictive factors and markers, data-mining in molecular medicine and translational research, and she was involved in the ACGT EU FP6 project and is a co-applicant of the recently awarded P-Medicine FP7 project.

Prof. Adrian Harris is the Director of Molecular Oncology Laboratory and Professor of Medical Oncology, at the Weatherall Institute of Molecular Medicine, University of Oxford, since 1988. His current appointments: Cancer Research UK Professor of Medical Oncology, Oxford University; Director of Molecular Oncology Laboratories Weatherall Institute of Molecular Medicine; Consultant Medical Oncologist, Oxford Radcliffe Trust and a Professorial Fellow of St Hugh's College, University of Oxford. Professor of Clinical Oncology,

University of Newcastle, Newcastle upon Tyne. Professor Harris specialises in Breast Cancer and Melanoma. His Areas of Interest are: Tumour Angiogenesis and Immunotherapy, Signal Transduction Therapy. Prognostic and predictive factors and markers

Professor Bassim Hassan is Professor of Medical Oncology at The Weatherall Institute of Molecular Medicine, University of Oxford since 2006. His current appointments: Professor of Medical Oncology (Visiting), Cancer Research UK Clinical Scientist, Consultant Medical Oncologist, Weatherall Institute for Molecular Medicine, University of Oxford and Oxford Cancer Centre, Honorary Consultant Medical Oncologist, Oxford Sarcoma Service, Nuffield Orthopaedic Centre, Oxford. His Areas of Interest are: sarcoma, regulation of tumour growth, mouse models, new cancer treatments that aim to block the action of tumour growth signals. He has been the WP5 leader for EuroBoNeT. The project will end in July 2011; however Prof. Hassan is co-applicant on EuroSarc, a European trial network under evaluation in the FP7 Rare Cancer. If EuroSarc is awarded, Prof. Hassan will guarantee direct collaboration which is already reflect in the scenarios suggested in this proposal; furthermore, he will bring requirements from the EuroSarc into EURECA and will disseminate results from EURECA to the EuroSarc network.

Partner 7 - FhG

Organisation Profile

The Knowledge Discovery Team at the **Fraunhofer Institute for Intelligent Analysis and Information Systems (IAIS)** in Sankt Augustin, Germany, is a research group located in the field of Data Mining and Machine Learning. Professional experience and expertise in this group include Data Mining, Statistics, Distributed Systems, Geographic Information Systems, and Data Warehousing. The KD group of FhG IAIS has a long tradition on basic and applied research in the field of knowledge discovery and in particular **data mining**. This has been proven by numerous publications and a broad variety of industrial and publicly funded projects. In particular, the field of distributed and grid-enabled data mining is a major research focus of Fraunhofer IAIS, which for example participates in the related EU projects ACGT, DataMiningGrid, and Simdat

The **Institut Biomedizinische Technik (IBMT)** applies its potential on subjects like non- or minimal-invasivity, microsystems engineering, implant technology, molecular and cellular biotechnology, nano(bio)technology, cryotechnology, biocompatibility, ultrasound technology, sensor manufacturing technology, telemetric data and energy transfer, health telematics, intelligent personal health systems and multilocal sensor systems connected by communication technologies. Important R&D fields are industrial applications of the molecular and cellular biotechnology and the cryotechnology for storage of living samples at low temperatures as well as isolation, cultivation and differentiation of stem cells for the regenerative medicine.

FhG's role in the project will be to

- Develop and implement tools for hospital-based knowledge discovery, i.e. KD tools that are designed to support medical practitioners with little or no knowledge of statistics and data mining.
- Develop and implement machine learning approaches for the analysis of clinical data and biomedical research data, i.e. KD tools for the expert user. In particular, novel algorithms for the monitoring of large sets of patient data and the meta-analysis of trial data will be developed.
- Contribute to the design and implementation of the general distributed architecture of the project, including appropriate meta-data representations
- Contribute to the technical implementation and testing of the EURECA system
- Contribute to the testing and usability evaluation tasks, both in general and in particular regarding the effectiveness of the KD tools.
- Contribute to the development of the semantic interoperability environment and of the services supporting the scenarios concerning early detection of safety risk, improved recruitment, and hypotheses generation.

Staff Members Profile

Dr. Stefan Rüping is the leader of the Integrated Data Mining group at Fraunhofer IAIS. He served as a member of the program committee of several international conferences, e.g. the European Conference on Machine Learning (ECML) 2005-2008. His project experience includes leading the machine learning activities in several EU and commercial projects.

PD Dr. Michael Mock received his Doctor degree from the University of Bonn in 1995 and his Habilitation degree from the University of Magdeburg in 2004, being a member of the Department of Computer Science

since then. Dr. Mock is currently working as senior scientist at FHG-IAIS in the department KD (Knowledge Discovery). His research interests include distributed systems, integrated data-mining systems and real-time systems.

Fatima Schera received her degree in physics in Tbilisi State University, Georgia and worked for 12 years in the department Automated Information and Measurement Systems at Tbilisi Scientific Association Analitpribor before she came to Germany in 1999. Since then she is senior software engineer at IBMT in various projects. Her expertise includes software development, software analysis and design, semantic web, web services, e-health systems and their interoperability. Ms. Schera has over 9 years of experience in developing (inter-)national software projects in the domain of e-health.

Partner 8 - VUA

Organisation Profile

The AI Department of the Vrije Universiteit, Amsterdam hosts a world-leading Semantic Web research group. The group initiated and coordinated the EU's first Semantic Web project (On-To-Knowledge) in 1999, and has since participated in key EU-funded projects such as Knowledge Web, SEKT and SWAP. The group played a significant role in the EU contribution to the definition of OWL, the W3C standard Semantic Web representation language.

Currently the LarKC project (http://www.larkc.eu) is running. The project is aiming to develop the Large Knowledge Collider, a platform for very large scale semantic web reasoning.

For the past ten years, the group has also been active in the area of medical guidelines and protocols, with an emphasis on constructing formal and semi-formal models of guidelines (using the Proforma and Asbru languages), and on using such models for validation and verification. Much of this expertise was built up in previous EU-funded FET- open projects (Protocure-I, Protocure-II).

Participant's role in the project

VUA will contribute to the following topics of the EURECA project:

- semantic interoperability between health-care record systems and clinical trial systems
- matching patient records to eligibility criteria of clinical trials
- support for development of eligibility criteria of clinical trials by approximate reasoning
- reuse health-care records for clinical research
- medical logic representation based on experience on representation of ontologies and guidelines.

The partner brings key expertise in the area of Semantic Web technology. Through leadership in research on Semantic Web, it has significant expertise in the area of ontologies, their use in modelling medical knowledge, and their use as key tools in achieving service interoperability. The partner has also acquired significant expertise in constructing formal and semi-formal models of medical guidelines and protocols, and in using these formal and semi-formal models for quality assessment and verification.

VUA will lead WP6 Applications, semantic reasoning and decision support, and contribute to the definition, implementation and validation of the EURECA semantic interoperability environment and of several software services, such as patient recruitment and contextualization of information to a patient case.

Staff Members Profile

Prof. Dr. Frank van Harmelen is a full professor in Knowledge Representation and Reasoning. After studying Mathematics and Computer Science in Amsterdam, he obtained his Ph.D. from the University of Edinburgh (Department of AI) for his research on meta-level reasoning. He is one of the designers of OWL, the W3C standard Web Ontology Language. He is scientific advisor of Aduna, one of the earliest companies in the Semantic Web arena, and developers of the Sesame RDF storage and retrieval engine. He has published over 100 papers, many of them in leading journals and conferences, and many of them highly cited (Hirsch index of 47). One of his five books is the first text book on Semantic Web technology (now deployed in university courses across the world, with translations in Spanish, Japanese, Chinese and Korean). He was the 2002 Programme Chair of the European Conference on Artificial Intelligence, the General Chair of the 2004 International Semantic Web Conference, and Chair of the Semantic Web track of the 2005 World Wide Web conference. He has been keynote speaker in numerous events, among which the 2005 European

Semantic Web Summerschool, the 2006 European Semantic Web Conference, and the 2006 Web Intelligence Conference in Hong Kong. He is currently scientific director the LarKC project (http://www.larkc.eu), aiming to develop the Large Knowledge Collider, a platform for very large scale semantic web reasoning.

Dr. Annette ten Teije is lecturer at the Vrije Universiteit Amsterdam. She gained her PhD (on the automatic configuration of diagnostic knowledge-based systems) from the University of Amsterdam in 1997. She has been a visiting research scientist at the Imperial Cancer Research Fund in London, and was a lecturer at the University of Utrecht before joining the Vrije Universiteit. She has published over 50 research papers, many of them in leading conferences. Relevant work includes verifying safety-properties for knowledge-based systems, and the automated configuration of Web Services (in the EU-funded IBROW project under the FET-O programme). She was VUA team-leader in the Protocure-II project, concerned with formal modelling and verification of medical guidelines and protocols). She was also involved in the WS-DIAMOND FET-Open project concerned with self-healing web-services. She has edited a book volume on the state of the art in research on medical guidelines and protocols, and has co-chaired the primary European workshops in this area over the past years: ECAI-workshop in 2006, Guideline workshop 2007, KR4HC2009, KR4HC2010.

Partner 9 – Breast International Group

Organisation Profile

Founded by leading European opinion leaders in 1996, the BIG network currently unites 44 groups based around the world that conduct more than 30 clinical breast cancer trials together[221]. In line with its objective of integrating breast cancer research, BIG has established a sustainable model of collaboration with the U.S. National Cancer Institute (NCI) and North American collaborative research groups, which has led, amongst other projects, to the publication of international guidelines for the harmonization of data across clinical trials. Links with other European networks (e.g. EU FP6 MetaBre, ACGT and the FP7 COGS projects) are also being built. This approach will be applied to the EURECA project.
Under the BIG umbrella, a network of 29, predominantly European, cancer research centres, hospitals and specialised laboratories have joined together to create NeoBIG, an innovative programme of neo-adjuvant breast cancer trials aimed at rapidly test novel targeted agents and potential predictive biomarkers. The focus of NeoBIG is to build, through the sharing of strategies, expertise, technologies, methodologies and protocols, a durable, multi-dimensional translational research structure that will support these neo-adjuvant trials and to provide a strong foundation for future adjuvant trials in breast cancer and research in other cancer domains.

Participant's role in the project

BIG will lead the Knowledge management workpackage of the project (WP10), will play a key role in providing clinical expertise related to large multicentric clinical trials in breast cancer, and will contribute to the definition of user requirements in the context of EURECA clinical scenarios (WP1: User needs). In addition, BIG, taking advantage of its experience as Network of Excellence leader (FP6 TRANSBIG), will contribute to the dissemination and exploitation activities of the Consortium. Finally, BIG will contribute to the definition of the legal and regulatory frameworks applicable to the project and to the evaluation and validation of the clinical scenarios at the pilot sites. Scenarios concerning improved patient recruitment, and protocol design and feasibility will be implemented in the context of BIG-led completed clinical trials.

Staff Members Profile

Professor **Martine J. Piccart-Gebhart**, MD, PhD, founded BIG in 1996, and currently serves as its chair. Dr. Piccart-Gebhart, who is Professor of Oncology at the Université Libre de Bruxelles and Director of the Medicine Department at Institut Jules Bordet (IJB), currently serves as president-elect of the European Society for Medical Oncology (ESMO). She is the immediate past-president of the European Organization for the Research and Treatment of Cancer (EORTC) and recently served on the American Society of Clinical Oncology (ASCO) Board. **Dr. Kamal Saini**, MD, will be involved in the scientific management of the project assisted by **Dr. Carolyn Straehle**, Managing Director, and an experienced EU project Manager, **Livia Meirsman**, while the NeoBIG-dedicated project manager, **Dr. Amal Arahmani** will ensure the link with the NeoBIG trials. It is foreseen that BIG's role in the "Knowledge Management" aspects of the project (WP10)

[221] http://www.breastinternationalgroup.org/

will be coordinated by **Cecilia Waldvogel,** BIG Communications Manager. BIG's in-house Legal Advisor, **Emmanuelle Ceysens / Diane Van Vyve** (shared position) will also play a key role in addressing any IP, legal or ethical issues arising in the preparation and conduct of the project.

Partner 10 - LUH

Organisation Profile

The Institute of Legal Informatics (IRI) of the Leibniz University of Hannover (www.iri.uni-hannover.de) at the Law School of the University of Hannover is the oldest establishment dedicated to scientific research on legal problems of Information and Communication Technologies at a German University. It was given its grant in 1983.

The subject matter of the Institute's research includes the requirements, applications and consequences of computer usage both in the legal system and in practice. A multitude of structural and specific questions in numerous fields (above all with regard to economic considerations) are investigated empirically and on an interdisciplinary basis. Topics have included information systems in personnel, telecommunication, data protection law, international data traffic, medical data protection law, medical data in the health insurances, electronic registration as well as the development of a comprehensive database for the law regarding public procurement and the creation of an innovative Data Protection Framework for EURECA, that enables researchers to share their data, but protects the privacy of the data subjects as well.

In the course of more than 20 years of its existence, IRI can look back on a long line of successfully completed research projects, which have been undertaken in addition to the further research and publications of the Institute's members. These projects have been supported by such institutions as the German Research Association, the Volkswagen Foundation, the Thyssen Foundation, the European Commission, the German Federal Ministry for Labour, the Austrian Federal Ministry for Education, the Arts and Culture, the Ministry for Science and Culture of the Lower Saxony, the Lower Saxony Commissioner for Data Protection and different private enterprises.
IRI is one of the leading institutions in Europe dealing with data protection, data security and intellectual property issues in the ICT for health area. The Institute and its members advise national and European institutions and projects in all relevant issues of the field.

The objective is to ensure that no legal and ethical barriers are in the way of accomplishing the EURECA goals. LUH's focus will lie on identifying and creating the data protection framework for EURECA. Main attention will be attributed to the implementation of the necessary security infrastructure as well as the ethical implications and requirements of the legal framework. In this respect informed consent forms will be drafted to ensure patients are properly informed about their rights and to secure those rights.

Accordingly, mainly the Directive 95/46/EC on the protection of individuals with regard to the processing of personal data and on the free movement of such data will be revised and considered. As a result, the necessary data protection framework will be established. It will be developed as a useful tool for the Partners in order to conduct the research in compliance with relevant laws and will allow to balance interest of the patients data safety and the development of the project itself. Moreover, the International and European legislation regarding Intellectual Property rights will be briefly described by providing guidance which measures should be taken for the Intellectual Property protection in EURECA. Finally, special emphasis will be given to the best possible equilibrium between the interests of patients giving their data and the researchers producing results out of these data within the project.

Staff Members Profile

Prof. Dr. **Nikolaus Forgó** has studied law, philosophy and linguistics in Vienna and Paris. In 1997 he got his Dr. iur. (Dissertation in legal theory). Between the years 1990-2000 he has been an Assistant Professor at the University of Vienna (Austria). From 2000 he is a full Professor for Legal Informatics and IT-Law at the University of Hanover. He has many publications, and has carried teaching and consulting in all fields of IT-law, legal informatics, civil law, legal history and legal theory.

Dr. **Tina Krügel**, LL.M. has studied law in Hanover and completed her legal clerkship in Hanover/Johannesburg (SA). In 2002 she attended the LL.M.-Programme EULISP in Hanover/Oslo. In 2005 she got her Dr. iur. (e-commerce law). Since 2004 she is an attorney at law and works for the IRI as a

research associate with the main focus on data protection law. She has several publications in the field of e-commerce law and data protection law.

Partner 11 - Xerox

Organisation Profile

Xerox Research Centre Europe's (XRCE) primary activity is research aiming at increasing productivity in the workplace through new document technologies. While drawing on the strength of the Xerox Corporation around the world, XRCE's focus is Europe. XRCE forms partnerships, collaborates with a wide range of European research organizations, and works with the business divisions of Xerox and with customers to understand their strategy and requirements. The Xerox Research Centre Europe is committed to inventing and designing document technologies that business both needs and wants. It pursues a vision of document technology where language, physical location and medium - electronic, paper or other - impose no barrier to effective use. Its specific mission is to become a Centre of Excellence for the understanding of document processes and for the invention of technologies that support them. The research carried out at XRCE covers the following areas: "Content Analysis", "Parsing and Semantics", "Document Structure", "Image Processing" and "Work Practice Technologies".
"Parsing and Semantics" group will be involved in the project. "Parsing and Semantics" developed for many years a competence in the Natural Language Processing in different languages. Lower-level linguistic processing tools (morphological analyzers and part-of-speech taggers) are available for 13 European languages. At a higher level of linguistic processing (deep syntax, word-sense disambiguation, NER and advanced semantic processing) the institution developped tools for French and English. Lately, the group have developped a strong interest in the health-care domain. This interest has been materialized in the participation in a French national project in which Xerox developped a system for the automatic detection of hospital-acquired infections in hospital discharged summaries.

Participant's role in the project
Xerox will lead EURECA's WP3: Information extraction and data access, and focus on the information extraction task enabling the extraction of relevant concepts and their relations out of free text documents in the clinical research and EHR systems.

Staff Members Profile

Caroline Hagege, PhD, is a senior research scientist in the Parsing & Semantics group at XRCE. She is currently in charge of the information extraction task in a French national-funded project in the medical domain. Caroline Hagege holds a PhD in Computational Linguistics and joined XRCE as a researcher in 2001. She developped the current XRCE's English syntactic and semantic general-purpose robust analyzer. She also coordinated the multilingual development of Named Entities Recognition Systems for Spanish and Portuguese. Her recent interests and activities include information extraction in the medical domain and temporal information processing from running texts.
Caroline has published many articles in the NLP domains. She also holds several patents in the field and is regularly member of reviewing committees of international conferences in the field of Natural Language Processing.
Caroline Brun is a research project leader in the Parsing & Semantics group at XRCE. She is in charge of R&D project in the area of natural language processing and information extraction.
She joined XRCE as a Ph.D. student in computational linguistics, working on coordination syntax for French within LFG, a unification-based grammar formalism. She then worked as a postdoc on lexical semantic disambiguation and document authoring. Now, her main interests include shallow and deep parsing techniques, named entity recognition and disambiguation and more recently, NLP techniques applied to opinion mining. Caroline has published many articles in NLP-related journal and conferences. She also holds several patents in the field and is regularly member of reviewing committees of international conferences such as IWP05, FinTal06, WTP'07, ACL-SemEval07, GoTal'08, etc.
Aaron Kaplan is a researcher in the Parsing and Semantics group at Xerox Research Centre Europe. He holds a PhD in computer science from the University of Rochester (USA). His work concerns the application of natural language processing techniques to information retrieval and information extraction tasks, as well as knowledge representation and reasoning for NLP. He has been responsible for the design, development, and integration of NLP components in information retrieval systems in a number of European and French research projects, as well as systems used internally at Xerox.
Frederique Segond is area Manager of the Parsing and Semantics activity at XRCE. She holds a PhD in Applied Mathematics from the "Ecole des Hautes Etudes en Sciences Sociales" in Paris (France) and is managing several projects in Natural Language Processing.

Partner 12 – UPM

Organisation Profile

The Biomedical Informatics Group of UPM was established in 1993. The staff at the group includes one full professor, two associate professors, one assistant professor, ten external professors and researchers collaborating with the lab and around 15 graduate and undergraduate students.
Collaborations include exchanges with foreign institutions, such as Harvard University (where six members of the lab have worked during the past years for mid-term and long-term stays), Georgia Tech and Rutgers University, in the USA, and research agreements with hospitals and health institutions in Spain. Contracts in biomedical informatics have included the Ministry of Defence, Spain, and industry, including funding from computer companies (E.g., HP), and from Spanish software companies. The Biomedical Informatics Group has obtained various awards for excellence in research from the Spanish Society for Health Informatics and other institutions.

Participant's role in the project
UPM will be mainly involved in WP3, WP4 and WP9. The group will contribute to these work packages with its expertise in ontologies and semantic heterogeneities overcoming. Concretely, in WP4 UPM will participate in the construction of the Semantic Core Datasets, and in building the mapping formalism and mappings between the core data sets and EHR and clinical research models. UPM will also collaborate to the task evaluating existing mapping formalisms and tools to link the canonical model with the underlying sources and to WP9: Models, deployment and clinical pilots.
UPM has a significant experience in participating in research and development activities related to Biomedical Informatics. The group has collaborated in many national and international projects related to the area, and is currently coordinating Action-Grid, the First European Initiative in Biomedical Informatics, Grid and Nanoinformatics. In recent years, the Biomedical Informatics Group has produced several publications in the fields of Public Health and clínico-genomic data integration

Staff Members Profile

Prof. Victor Maojo is a Full Professor and Director, Biomedical Informatics Group, Artificial Intelligence Lab, UPM, Madrid. He holds a M.D. degree (medicine) and a Ph.D. in Computer Science. He has been a Visiting Professor and Consultant in Georgia Tech (Atlanta, USA, 1990-91), and a "Research Fellow" in Medial Informatics at Harvard Medical School-MIT (Boston, USA, 1991-93). He has been the Principal Investigator in more than 15 national projects and has been awarded two international grants from Hewlett-Packard. He has been an Expert for the European Commission in the IV, V, VI and VII Framework Programmes. He has participated in the IST study BIOINFOMED, the INFOGENMED and ACGT projects, the INFOBIOMED network of Excellence and as a Project Coordinator for ACTION-Grid. He has been the Chairman of the 2nd International Symposium on Medical Data Analysis (ISMDA 2001). He has published 160 research papers in scientific journals and international conferences. He is in the editorial board of various scientific journals. He has published numerous articles and communications in top-ranked journals (e.g., JAMIA, Journal of Biomedical Informatics, Methods of Information in Medicine, IEEE EMB, IEEE TITB, Computers in Biology and Medicine, NATURE, JASIST, Pattern Recognition and others). Prof. Maojo has given a large number of invited and Keynote speeches and has been in the Program Committee of around 25 international conferences. He will give the "Reed Gardner Lecture" at the University of Utah, USA, on April, 2009.

Prof. Jose Crespo is a telecommunication engineer (1989). At the Georgia Institute of Technology, USA, he received a Master of Science in electrical engineering in 1990, a Master of Science in management in 1993, and a Ph.D. in electrical and computer engineering in 1993. He has been a post-doc researcher at the Centre de Morphologie Mathematique (E.N.S. des Mines de Paris) and since 1985, he is at Universidad Politécnica de Madrid, where he is currently an associate professor. His research interests are image processing and analysis, mathematical morphology, medical imaging, medical informatics, algorithms, videoconference and multimedia. He has been author and co-author of research papers in journals such as Signal Processing, Pattern Recognition, Optical Engineering and the Journal of Mathematical Imaging and Vision, Methods of Information in Medicine and others. He has been a member of IEEE and ACM.

Partner 13 - MAASTRO

Organisation Profile

MAASTRO clinic (Maastricht Radiation Oncology clinic see www.maastro.nl), has a strategic agreement with the GROW research institute (growth, and development - www.grow-um.nl), the Faculty of Health Medicine and Life Sciences at Maastricht University and the University Hospital of Maastricht. Our aim is to develop a patient database of all the known and potentially predictive factors that are associated with the prognosis of cancer. The profiles that can then be determined provide a basis for identifying subgroups or individuals for whom specific treatments are most suitable, with the least chance of side effects. This should lead to a Decision Support System.

Three of the strategic highlights of MAASTRO are: Lung Cancer, Tumour Hypoxia and Molecular Imaging. MAASTRO Clinic offers state-of-the-art radiotherapy to more than 3500 cancer patients each year from the Mid and South Limburg area in the Netherlands. MAASTRO Clinic currently has 200 employees, 7 modern linear accelerators all equipped with EPID (electronic portal imaging) and some with Megavolt Cone Beam CT. It was the first centre with a CT-PET scanner dedicated for radiotherapy, which now includes the possibility to acquire dynamic PET scans and 4D-CT-PET in lung cancer patients. Both a small and large bore multi-slice CT with 4D-CT option is available. At the MAASTRO Clinic stereotactic and intensity-modulated radiation therapy (IMRT), image guided radiotherapy (IGRT) and in-vivo dosimetry using both EPIDs (dose guided radiotherapy) and MOSFETs are used in daily practice. Within our department approximately 25% of our patients are included in clinical trials that have been approved by an ethical committee.

In 2006, we underwent an international site visit (Prof. Martin Brown, Stanford University, USA; Prof. Jan Lagendijk, University Medical Centre, Utrecht, the Netherlands) which resulted in a very positive report and approval of our research plan, which focuses on lung cancer and hypoxia. An integrated IT system with three modules (multiparametric interconnected databases, Machine Learning based predictive algorithms and User interface (including eCRF partially automatically populated from the Electronic Health Record) has already been implemented for lung and rectum cancer.

MAASTRO has extensive experience in leading clinical trials in comparisons of treatment planning studies, in large multi-centric in-silico trials such as the ROCOCO trial where treatments with photons, protons and carbon ions are compared. Much experience has been gained on developing methodology to handle large amounts of data, use PET imaging information, develop robust software tools for multifactorial data handling leading to predictive models and Decision Support Models. We are one of the first centre in Europe having eCRF populated from the Electronic Medical File. Dr Verhaegen, who recently joined MAASTRO, has extensive experience with Monte Carlo techniques.

Partner's contribution to the project

MAASTRO will participate to the development of the EURECA interoperability environment and in the implementation and validation of several EURECA scenarios (e.g. patient recruitment, trial execution, etc.). We will contribute data and clinical knowledge and link the EURECA project to the large regional euroCAT and duCAT projects.

Our aim in **EURECA** is to also develop a patient database of all the known and potentially predictive factors that are associated with the prognosis of cancer. The profiles that can then be determined provide a basis for identifying subgroups or individuals for whom specific treatments are most suitable, with the least chance of side effects. This should lead to a Decision Support System. Three of the strategic highlights of MAASTRO are: Lung Cancer, Tumour Hypoxia and Molecular Imaging.

Staff Members Profile

Prof. dr. Philippe Lambin is a clinician, Radiation Oncologist, pioneer in translational research with a focus on hypoxia and imaging. He has a PhD in Radiation Biology. Moreover, Prof. P. Lambin has extensive experience with clinical trials. He is one of the international experts in the Flims workshop "Methods in Clinical Cancer Research organized jointly by the FECS, AACR and ASCO. He is also member of the scientific committee of KWF (The main Dutch funding body in cancer research).

Furthermore he has extensive experience with research into the use of systems biology approaches combined with large databases of biological (geno-proteomic), clinical, imaging and treatment data coupled

with a Machine Learning System. He is involved in several successful EC grants (e.g. Biocare, Euroxy, Metoxia in which he is the cluster leader).

Partner 14 - eCancer

Organisation Profile

ecancermedicalscience is a non-profit organisation which manages the free, online open-access peer-reviewed cancer journal, publishing original science articles (including video clips), reporting on cancer news and webcasting cancer conferences. ecancer is embracing web 2.0 technologies to help create a virtual social network of oncology professionals and provide an online community for those involved in all fields of cancer research and treatment. The journal was founded by Professor Umberto Veronesi and Professor Gordon McVie at the European Institute of Oncology in Milan, in partnership with ECCO - the European CanCer Organisation and is supported by the Fondazione Umberto Veronesi, Fondazione IEO and Swissbridge. *e*cms AG is the not for profit entity based in Switzerland which manages *e*cancermedicalscience.
ecancer aims to improve communications between sub-specialised cancer scientists and clinicians by working interactively and faster – offering a rapid peer review process as well as video, audio, and event listings. ecancer actively encourages the communities of sub-specialised scientists and cancer carers to exchange ideas and research, speeding up the time it takes from discovery, to patient benefit. The journal is published by Cancer Intelligence Ltd in Bristol and is editorially independent of the European Institute of Oncology and ECCO.
ecancer is involved in the FP7 Eurocancercoms project, in WP1, project managing the dissemination of the project over the two year duration. ecancer is also involved in WP6, evaluating the suitability of new dissemination technologies made possible by the internet. It will also act as the web address and reservoir for all papers, reports and publications from the CA. Currently it is visited by 15,000 readers each month from 171 countries. ecancer currently hosts hundreds of oncology videos, thousands of hours of content since the launch of ecancer.tv (the video section of the site) at the beginning of 2009. ecancer's experience will allow the broadcast of videos relating to the project and testing of pilot television materials.

*e*cancer is the journal of the IEO (European Institute of Oncology) in Milan and of the OECI (Organisation of European Cancer Institutes). It is also the science and multimedia partner for ECCO, which represents 50,000 cancer scientists, clinicians and paramedical specialists throughout Europe. It is visited by 25,000 individuals each month, from 171 countries to date. Its international presence is therefore unparalleled. The editorial board consists of world experts in oncology, and is particularly strong in clinical trial expertise. Together with ECCO, ecancer coordinate an FP7 project, Eurocancercoms, which is led by IEO.

Staff Members Profile

Professor Gordon McVie is widely regarded as a leading international authority in the research and treatment of cancer. He currently is responsible for Clinical Research Coordination, Strategy and International Affairs at the IEO. He is part-time Director of Cancer Intelligence, which publishes ecancermedicalscience. Previously Professor McVie was Chief Executive of the Cancer Research Campaign (CRC), then Cancer Research UK the largest grant giving charity in the UK. Throughout the Eighties, he was Clinical Research Director at the National Cancer Institute of the Netherlands, and Consultant in Oncology at the Antonie van Leewenhoek Hospital, Amsterdam. As President of EORTC, he set up the present Drug Development Group in Brussels, and with NCI support, the European New Drug Development Network. In the UK he was one of the architects of the Cancer Trials Networks in Scotland, Wales, and England, and was a founding member of the National Cancer Research Institute. He led CRC into a merger with ICRF to form Cancer Research UK, the world's largest charity supporting cancer research.

Susi Burke is the Communications Director of ecancermedicalscience and board director of ECMS AG. She has communication and marketing experience across multiplatform media with global brands. She is an expert in cancer communications and launched the cancer media centre – a service designed to connect the media and cancer experts across Europe to ensure fair and honest reporting.

Jonathan Birch has completed a Masters in Science, Media and Communication at Cardiff University. He has studied a range of sciences and has experience analysing health related news, working in broadcast media and translating scientific research into clear, concise copy.

Partner 15 - EuroRec

Organisation Profile

The Eurorec Institute (http://www.eurorec.org) is an independent European not-for-profit organisation that promotes the use of high quality Electronic Health Records (EHRs). It overarches a network of national ProRec and other affiliated centres in 23 member states. EuroRec and its centres represent all relevant eHealth stakeholders (a.o. healthcare authorities, EHR system vendors and health professionals). EuroRec has developed a repository of validated EHR quality criteria (+1700, translated in 13 languages) and a series of tools for using these criteria in the context of quality labelling/certification, procurement and product documentation of EHRs. In December 2009, EuroRec released a first profile identifying the functionalities required of an EHR system in order to be considered as a reliable source of data for regulated clinical trials.
EuroRec has participated or still participates in several EU-funded projects:

- RIDE (Partner). A Roadmap for Interoperability of eHealth Systems EC FP6 (2005-2007).
- Q-REC (Coordinator). European Quality Labelling and Certification of Electronic Health Record systems (EHRs). IST-27370-SSA (2006 -2008).
- EHR-Implement (Partner). National policies for EHR Implementation in Europe: social and organisational issues. PHEA/CEC 2006112. (2007 – 2010).
- EHR-QTN (Coordinator). Thematic Network on Quality Labelling and Certification of EHR Systems. ID: 238912. (2009 – 2012).
- HITCH (Partner). Healthcare Interoperability Testing and Conformance Harmonisation. FP7-ICT-2009-4 (2010-2011).
- ARGOS (Coordinator). Transatlantic Methods for Handling Global Challenges in the European Union and United States: Health Related ICT) partim Interoperability and Certification). RELEX C1/2009/PP, SI2.548981 (2010-2011).
- EHR4CR (Managing Entity). Electronic Health Records for Clinical Research. The EHR4CR project is funded by the IMI Programme.

Staff Members Profile

Georges J.E. De Moor (MD, CP, PhD), (President, Senior Researcher) studied Medicine and specialised in Clinical Pathology and Nuclear Medicine at the State University of Ghent (Belgium), where he also obtained in 1994 his PhD in Medical Information Science. He is head of the Department of Medical Informatics and Statistics at the State University of Ghent, Belgium and he is head of the Clinical Pathology Laboratory of the St. Elisabeth Hospital in Zottegem, Belgium. At State University of Ghent, he teaches Health Informatics, Medical Statistics, Decision Theory and Evidence Based Medicine
As president of RAMIT (Research in Medical Informatics and Telematics), he has been involved in both European and International R&D projects (+85) and standardisation activities and was the Founding Chairman of CEN/TC251, the official Technical Committee on standardisation in health informatics in Europe. As a result of the conducted research, Prof. De Moor has been founding or co-founding a number of spin-off companies mainly active in eHealth, including the domain of privacy protection.
He chairs in Belgium and in Europe a number of official Committees related to either ICT in Health or to Laboratory Medicine. He has edited twelve books related to ICT in Health, published over 200 articles in international peer-reviewed scientific journals. In 2005 he was awarded with the international Rory O'Moore Award for Health Informatics.
In 2004, he has been elected President of the European Institute for Health Records EuroRec, (http://www.eurorec.org) which is the de facto body for the harmonisation of certification of Electronic Health Record systems in Europe.
His input in the project will be for free.
Jos Devlies (MD), Medical Director and Senior Researcher
Dr. Jos Devlies, medical director of EuroRec, is a physician and involved since 1986 in the development and the commercialisation of EHR systems for General Practitioners. He was president of several health IT companies and is still chairing a cooperative company of Belgian physicians. He was also involved in the development of a clinical glossary, a medicinal product database and integrated clinical decision support applications. He was also involved in standardisation activities, chairing the EN 12610 on Medicinal Product Identification, and participating either as partner or as coordinator in a large number of nationally and European Union funded projects: SHARE, C-Care, C³, eProLearn, Liverdoc, ReMINE and Q-REC. He joined the team of Professor Georges De Moor in 2006. He is actually the main responsible for the content of the EuroRec Certification Repository and has a long experience in certification of EHR products for General Practitioners. He is coordinating the EHR-QTN Thematic Network project, promoting the use of EHR systems and certification of health care systems in more than 25 European countries. He is also author of several

recommendations for the Belgian National Health Insurance Authorities, especially regarding clinical pathways management. He is responsible for the EuroRec Certification Repository and is experienced in certification of EHR products for general practitioners.

Pascal Coorevits (BICT, MSc, PhD), Researcher

Dr. Pascal Coorevits received his master's degree in motor rehabilitation and physiotherapy from the Ghent University (Belgium) in June 1998 and a bachelor's degree in ICT in June 2003. In February 2007 he obtained his PhD from the Ghent University. From 1998-2005 he worked as a scientific researcher at the Department of Rehabilitation Sciences and Physiotherapy of the Ghent University and moved to the not-for-profit organisation Research in Advanced Medical Informatics and Telematics (RAMIT vzw) in 2006, where he is a senior research associate in both research and management of several national and international R&D projects for medical informatics and statistics. In 2008 he became assistant professor of medical informatics and statistics at the Department of Public Health at Ghent University and in January 2009 a member of the Biostatistics Unit of the Faculty of Medicine and Health Sciences of the Ghent University. He joined EuroRec in January 2010.

Geert Thienpont is a computer scientist and studied at the Industrial School of the State B.M.E. GENT. Since 1992 he has been involved in national and international R&D projects in the field of Medical Informatics & Telematics and e-Learning & e-Testing. Currently, he is project manager of RAMIT for the European R&D projects. Furthermore he was/is involved in national and European associations such as EuroRec (European Institute for Health Records), MS-HUGe (Microsoft Healthcare User Group Europe), MedNet and PROREC-Belgium.

Geert Thienpont was working for two years (1994 -1995) in Luxembourg at the Centre de Recherche Public - Centre Universitaire. He was involved in a LRE project named ANTHEM ("Advanced Natural language Interface for Multilingual Text Generation in Healthcare") as IT researcher.

He is since 2009 also president of WHITe³ (World of Health ICT: Education, Events and Exhibitions).

Partner 16 - SIT

Organisation Profile

For years now, Stoneroos has been active in the field of data processing and extraction of patterns from large sets of information. This activity has become more and more essential, as data repositories keep on growing at a faster than linear rate. Furthermore, this evolution is seen in a multitude of different domains. To do so, we deploy new technologies data mining and Semantic Web, to facilitate services like personalization, data recommendation and data interoperability.

Within the EURECA project, one of the important ingredients involves the semantic interoperability between different systems within the environment. Recommendations and contextualization of information to a specific patient, alignments between different systems like the EHR systems and clinical trial systems, mappings providing new links between external data and these systems, etc. are within the scope of the project. However, such alignments are not always trivial. Often, several data transformations, translations, etc. are required to match different concepts within separate and potentially heterogeneous data repositories. Moreover, an extensive and flexible language or structure is required to describe these alignments and allow automated processes to interpret and exploit them.

Next to the exchange and alignments of data between data repositories, a second important issue is the interplay of users and data within the system. Obviously, users are no machines and need to be served in an adapted and personal way to be able to contribute their part in the greater whole. Moreover, every person is different and might have a different information need. Therefore, systems need to be able to adapt and personalize information to the person who is receiving it.

For both of these general challenges, Stoneroos believes it can contribute valuable knowledge and expertise. Over the past years Stoneroos has specialized itself trying to find an answer to the problem of data overload in the TV domain, where 10's of thousands of TV programs where described by an extensive metadata description and needed to be matched to thousands of user profiles, facilitating an unobtrusive and personal access to the data. Here, we translated data into Semantic Web languages and aligned different data structures and ontologies. After this data integration process, recommendation and personalization algorithms were deployed to see which parts of the data and which presentational forms of the data should be chosen to accommodate each specific user best. Similarly, these techniques would be the basis for data integration and personalization in the EURECA project.

Over the last years we participated in various European projects including the ITEA Passepartout project on Personalized Ambient Media, and the FP7 NoTube project striving to facilitate personalized creation, distribution and consumption of TV content. In spite of the fact that these project were largely targeted at

smart multimedia in the home, Stoneroos amassed a large knowledge of cutting edge technologies and expertise in Semantic Web, personalization and recommendation systems. Although our main domain has been multimedia and entertainment, these technologies are nowadays becoming more and more crucial in a variety of domains, including scientific literature applications, medical treatments and inter-operation, exchange and presentation of various types of data collections, etc. Therefore, we are striving to deploy our gained knowledge in new and different domains, perfecting techniques and gaining new insights.

Participant's role in the project

In EURECA, SIT will contribute to the contextualization and personalization services in WP6. Additionally, we will contribute to the development of the semantic interoperability environment and to the design and development of user interfaces for the EURECA services.

Staff Members Profile

Annelies Kaptein, graduated in Business Administration (ing) and Psychology (Drs). She has build up an extensive experience in TV and Media and worked for several Dutch Public Broadcasters in different areas: as a TV director, as a project leader (NOB / Dutch view) , as consultant and in the board of directors responsible for new media and marketing. In 2002 she started the new media company Stoneroos with a focus on personalized television. Since the start Stoneroos has won several international prizes for iTV enhanced TV applications.

Pieter Bellekens, PhD, is head of research within Stoneroos. He obtained his PhD degree in 2010 at the Technical University of Eindhoven in the Department of Computer Science, on the topic of Context-sensitive and User-adapted Access to Heterogeneous Data Sources, illustrated in the TV domain. During this time a close collaboration was kept with Stoneroos eventually leading to the development of iFanzy, a personalized Electronic Program Guide. Current research focusses on personalization of interfaces and systems next to the generation of recommendations of various objects like TV programs, advertisments, etc.

Jan Van Nunen is software developer at Stoneroos. In 2009 he obtained his Master degree in Computer Science from the Technical University of Eindhoven. During the last year he developed many Web 2.0 and multiplatform applications, broadening his knowledge over a wide variety of technologies.

Erik van Dijk is creative designer. He designed the user-interfaces for Stoneroos' interactive EPG (iFanzy). He also designed all other user-interfaces Stoneroos made in the last years, including the interfaces of 2 of the biggest Dutch (IP)TV distributors.

Partner 17 - GBG

Organisation Profile

The German Breast Group (GBG) Research GmbH is an academic research organization with 90 employees specialized in the implementation of investigator-initiated trials of malignant breast cancer, providing the platform for all clinical studies of the German Breast Group. This academic research institute has recruited up to 4250 patients annually in clinical breast cancer trials. The GBG is currently conducting 18 breast cancer trials, in 10 of those GBG is the legal sponsor of the study. GBG covers the complete portfolio to conduct large scale phase III studies including protocol development, monitoring, data-management and statistics. GBG has developed its own web-based data capturing system Medcodes® which is used for all trials started after 2009.
The clinical affiliation of the Managing Director, Prof. G. von Minckwitz, is the Luisenkrankenhaus in Düsseldorf (head: Dr. Rezai), which is having a strong link to GBG.
The Luisenkrankenhaus is among the 3 largest breast cancer centres in Germany dedicated only to the treatment of breast cancer. 600 to 800 new breast cancer patients have received surgical and systemic treatment in this institution yearly. A clinical trials unit was set up 5 years ago and recruiting now around 15-20% of these patients into clinical trials.
The trials unit is also linked to the German International Medical Centre (www.g-imc.de) for Research and Education which supports the conduct of clinical trials on the administrative level but also by promoting the trials at educational events and conferences organized by this institution.
The work concerning semantic interoperability to which GBG will participate in EURECA will be carried out in close collaboration with the Luisenkrankenhaus Düsseldorf as a clinical site.

Participant's role in the project

GBG will provide clinical user requirements, clinical knowledge based on the group's outstanding expertise in running clinical trials, clinical data, and all needed technical information on its clinical information systems in a non-academic but highly specialized centre in Germany (Luisenkrankenhaus Düsseldorf).
GBG will also contribute to the dissemination of the project's results within their large network and to the evaluation and validation of the interoperability framework and several EURECA services in common scenarios with BIG and IJB.

Experience relevant to these tasks

Strong expertise on the surgical and medical treatment of breast cancer. Strong participation in clinical breast cancer trials. Long experience in developing case report forms.

Staff Members Profile

Prof. Dr. Gunter von Minckwitz is Managing Director of the German Breast Group (GBG) Research GmbH, an Alternate Director of Oncologic Senology at the Luisenkrankenhaus in Düsseldorf and associate professor at the University of Frankfurt. He qualified at the University of Heidelberg Medical School, Germany, and completed his fellowship and residency at the University Women's Hospital, Heidelberg and the University Women's Hospital, Frankfurt. Dr. von Minckwitz has a 20 years experience in conducting breast cancer trials and has published approx.. 150 Medline cited publications almost all on breast cancer. He served as principal investigator in multiple national and international multi-centre phase I-III breast cancer trials. As chairman of the AGO Breast Commission he was in charge for developing standard treatment guidelines for primary and metastatic breast cancer. He is member of the St. Gallen Consensus Panel for Primary Breast Cancer.

Partner 18 – NRC

Organization profile

The National Research Council (NRC) is Canada's principal federal government agency for research, development, and technology-based innovation. NRC comprises more than 20 institutes and national programs, spanning a wide variety of disciplines in Life Sciences, Physical Sciences, and Engineering. NRC employs over 4,000 personnel. The NRC Institute for Information Technology (NRC-IIT) conducts research in software and systems technologies, including algorithms, methodologies and software engineering.
With Canada as a French-English bilingual country, special focus is given within NRCIIT to developing language-independent text analysis methods, developing tools for cross-language tasks (parallel text alignment, translation tools), and developing methods for processing multi-lingual documents. This expertise is continuously expanded into languages beyond French and English. Special interest is given to applying research components to Inuktitut, the languages of the Northern first nations (Inuit). In several of its research groups, including the Interactive Information group, NRC-IIT focuses its research on informatics for medical applications, including medical language processing in English and French.

Participant's role in the project

NRC will contribute to the Information Extraction research in WP3.

Staff Members Profile

Dr. Berry de Bruijn has been a researcher at the Interactive Information group at NRCIIT for over eleven years. He earned his PhD in Medical Informatics from the Faculty of Medicine, University of Maastricht (The Netherlands) in 1997. After his successful thesis defence, he accepted a two-year postdoctoral position at the University of Ottawa, School for Information Technology and Engineering. Here, he studied methods for interactive information retrieval and text summarization. He joined NRC-IIT in December 1999, starting research in literature mining for molecular biology. He has since conducted and led research on text analysis methods applied to biomedical text, including text mining from scientific abstracts and full-text, and classification and analysis of textual patient records. He has published 30 articles which, together, were cited close to 500 times.
Berry de Bruijn was one of the first researchers to combine text mining with bioinformatics. He contributed the text mining toolkit named Textomy to the BIND (Biomolecular Interaction Network Database) initiative and the PreBIND repository was created with data automatically extracted from literature by Textomy. His 2003 paper was one of the first in the field and has been cited over two hundred times. His work has contributed to bringing text mining from concept to maturity; today, text mining begins to be recognized as an established set of methods in the biomedical domain. Textomy developed into LitMiner (2002-2007), a comprehensive testbed and application environment for literature mining in the biomedical field. Not only did it innovate the

field of text mining, interactive information management, visualization, and data integration, it was also an access point to rich information for biology researchers. It enabled researchers at NRC-IBS to summarize the literature on biomolecules and physiological processes, and chart gaps and missing linkages in the published research. Participation in the 2003 NIST Text Retrieval Competition (TREC) gave Dr. de Bruijn an opportunity to measure the performance of the methods developed for LitMiner against a gold standard data set. It also allowed a comparison with other participating teams from around the globe. The two sets of output as produced by the LitMiner system ranked second and fourth among the 50 submissions, only bested by the submissions from the US National Library of Medicine.

2.3. Consortium as a whole

EURECA targets a specific research objective in a sharply focused approach while at the same time it includes a coherent and integrated set of activities dealing with multiple related issues and provides state-of-the-art responses to the challenges identified in the call, i.e. integration of clinical research and care, semantic interoperability, re-use of care data for large scale epidemiology and cohort studies, and more efficient (faster recruitment, avoid multiple data entry) and safer research. Special emphasis is given in making sure that the new tools, services and applications to be developed in EURECA will also be evaluated on their effectiveness and validated in use cases with a high potential for improving patient safety in research and epidemiology.
In order to build, verify and demonstrate the proposed solutions, *EURECA* has carefully selected 18 leading organisations from 8 European countries.

The collective expertise, commitment and prior delivery record of these organisations, which were specifically selected for their diverse experience and essential competencies as well as for their complementarities, guarantee the successful outcome of the proposed project.

The idea for the *EURECA* project has evolved over a significant period of time, as a result of experiences and R&D results from previous National and EU projects, but also as a result of urgent needs for technologies assisting the optimization of the process for new discoveries. An indication of this need is the fact that several EURECA partners (e.g. Institute Jules Bordet) have internally begun to locally address the problems of how best to “link their clinical patient data with their clinical research (CT) data, in an attempt to optimise both the care and new discoveries processes.” However, achieving semantic interoperability at a scale that has impact requires EU-wide consolidated effort and collaboration.

Subsequently, a suitable consortium has been built for starting up the project, which contains relevant partners and stakeholders from ICT organizations, clinical organisations, academic and research institutions, as well as two specialised SMEs. All these diverse organisations are needed from the beginning, in order to build and verify the envisaged computational and service delivering environment of the EURECA project.

Successful completion of the EURECA workplan and realization of its objectives requires the concurrent presence (and obviously successful collaborative work) of a diverse set of expertise. Specifically:

1. ***HEALTHCARE ICT:*** In the Healthcare ICT section, the project requires knowledge and expertise in healthcare information technology, and healthcare related standards. A number of such organisations have been selected with a proven track record of involvement in such R&D activities. Philips, FORTH, FhG, Custodix, UPM, Xerox, IJB, MAASTRO, EuroRec have a proven record of healthcare IT research and development. They are also strongly linked and contributors to key standardisation activities, such as CEN, HL7, IHE. Also, our partner VUA has significant expertise on Semantic Web Technologies and Semantic Interoperability in e-Health.

   On the other hand several partners are also active in research and development of clinical research systems (i.e. clinical trial systems). Saarland University, FhG, and FORTH are the key developers of the ObTiMA system, an open source Clinical Trial design and Management System in compliance with the CDISC standard[222]. The IT department of Institute Jules Bordet has also for a long period been involved in the implementation of systems and services for clinical research (i.e. trial management systems, adverse effects reporting modules, etc). They have also identified the need for standards-based interoperability and are extending their EHR system making use of HL7 CDA and SNOMED coding.

[222] The Clinical Data Interchange Standards Consortium (CDISC) is an international, non-profit organization that develops and supports global data standards for medical research.

Next to the experience brought in by the partners described above, the Consortium felt that it needed additional expertise in the domain of "archetypes" with respect to specific tooling. Clinical archetypes are specifications of the knowledge data and their interrelationships that play an important role in determining how clinical information is represented and organised inside EHRs when they are interoperably communicated between systems. They form an absolutely essential part of the semantic interoperability underpinning of e-Health. Much of the work already done in this area is now embodied within international standards for EHR architecture requirements (ISO 18308), EHR interoperability (ISO/EN 13606) and through a major open source initiative: the openEHR Foundation. In responding to this need the consortium invited EuroRec to join the project. EuroRec have long experience in modelling clinical content and standards for health record architecture and communication. Their contributions in EURECA will be very important.

2. ***SOA:*** The second main area in which the project needs to move beyond current state-of-the-art is the domain of "service oriented science". Taking into consideration that our ambition is NOT to define and propose new solutions for the healthcare IT systems (EHR systems) or new clinical research systems but to develop bridges that enable the semantic interoperability of these two worlds, and modules/software services that provide added-value to these worlds by making use of existing data collected in either of these domains, the obvious architectural choice is a "service oriented approach". A number of EURECA partners have demonstrable experience in developing state-of-the art SOA compliant solutions in healthcare. Philips, FORTH[223], Custodix, and FhG are such partners. More specifically FhG-IBMT is the coordinating partner in the smartHEALTH[224] integrating project of the EC focusing on the implementation of semantic web services.
3. ***CLINICAL RESEARCH:*** The third key area of expertise for EURECA is that of clinical research. Again, the expertise of our clinical partners, BIG, GBG, UOXF, IJB, UdS and MAASTRO, is unparalleled. They all represent top research institutions in the oncology area but also large hospitals treating significant numbers of patients both in and outside clinical trials, a fact that will ensure the full compliance of our solutions with best practices.

   In selecting the appropriate clinical partners the Consortium was also aware of the fact that the success of projects in the biomedical area is often endangered by two closely connected issues: Lack of access to relevant patient and research data and inability to deal with the legal, ethical, privacy and security needs related to patient data.

   The EURECA project will not be faced with these issues, as by the start of the project we will already have access to significant amounts of anonymized/pseudonymized patient data from clinical trials, provided by our clinical partners in full compliance with the legal, ethical and security requirements. These data sets, described in section 1.2.23, will provide a quick start and high value in EURECA for building, testing and validating our prototypes. In addition to having access to clinical research data we recognise that access to large longitudinal EHR data sets is also required for the successful implementation of the EURECA workplan. Our clinical partners are fully committed to also provide access to EHR data, following a legislation-compliant process similar to the one used in ACGT. The legal research concerning handling patient data from clinical trials that we have carried out in ACGT will also provide the EURECA project a fast start in setting up the appropriate legally-compliant framework for access to patient specific clinical data. Additionally, EURECA will establish an Ethics and Legal Advisory Board that will include top European experts.
4. ***INVOLVEMENT OF PATIENTS AND PATIENT ORGANIZATIONS:*** We consider patients and patient organizations essential to the EURECA project, as we aim to support patient empowerment by providing technology solutions, removing legal barriers and promoting a culture change. **WP1: User needs** specifies the important role of the patients and of patient organizations as users of EURECA, and plans to involve them in the user requirements, in the evaluation and in the validation phases of the project. Through our clinical partners we involve several patient organizations to identify patient needs and define user scenarios that are relevant from their perspective. Such organizations are: Tenovus, ECPC, Europa Donna, Favo, ICCCPO. Patient organizations will also be represented in EURECA Ethics & Legal Advisory Board. Additionally, eCancer will leverage their extensive collaboration with patient organizations and their outreach to patients to describe relevant project results in language understood by patients and their carers, and to evaluate the patient preferences of styles of presentation of clinical information (e.g. clinical trial information to support

[223] D.G. Katehakis, S. Sfakianakis, G. Kavlentakis, D. Anthoulakis, M. Tsiknakis, (2007), Delivering a Lifelong Integrated Electronic Health Record based on a Service Oriented Architecture, IEEE Transactions on Information Technology in Biomedicine, Vol 11, No 6, pp. 639-650.

[224] http://www.smarthealthip.com/

enrolment). In EURECA, eCancer will publish relevant content on eCancerpatient, the patient-focused area of eCancer. The EURECA scenario that evaluates feasibility of use of PHR data for improved safety will be developed and implemented taking into account the patient perspective: It relies on patient-recorded and managed data, requires patient commitment to create content (recognize serious side effects and report them), and needs the patient consent to use their entered PHR data.

5. ***ETHICS & LEGAL:*** As already mentioned, we expect that legal issues, currently unsolved, with respect to the compliance of our proposed solutions in this area with the existing legal regulations and ethical framework of Europe, in particular with respect to "seamless access and analysis" of patient clinical data, need to be addressed by EURECA. In responding to this need, we have brought into the consortium the University of Hannover. They are undisputable experts in scientific research on legal problems of Information and Communication Technologies with application in Healthcare. They will be supported in this task by the Breast International Group, with their vast expertise in running clinical trials in an international setting, and dealing with all regulatory, legal and ethical requirements. On ensuring compliance with all legal and ethical requirements, we will work closely with the EURECA Ethics & Legal Advisory Board that will include independent top experts recognized at European level.
6. ***EXPLOITATION:*** Realizing the huge exploitation potential that lies behind the EURECA scientific and technological objectives has also been at the center of our concerns. In maximizing this potential we have selected specific industrial and commercial partners who are committed to innovation and exploitation. Apart from the project Coordinator, Philips – who sees the EURECA project as a strategic initiative in the Healthcare Domain and has concrete exploitation plans – two SMEs are also included. Custodix and SIT are innovative SMEs focusing their activities on the medical and clinical research areas and respectively in ICT research in the context of semantic reasoning for personalization and contextualization, with a proven record of results-exploitation. It is also the strong interest of our clinical partners in achieving efficient, state-of-the art data integration and achieving semantic interoperability to speed-up their research and improve patient care that guarantees the future exploitation of the tools and technologies to be delivered and validated in EURECA. In the project, Xerox will extend their advanced information extraction tools to new types of documents and domains.
7. ***DISSEMINATION:*** One last item that the Consortium had to respond to was the question of dissemination and linking with the stakeholder groups. We have seriously thought about *what is the most appropriate organisation for leading such an effort.* The conclusion was made that we need to move away from "possibly traditional approaches" to dissemination, and decided that the most appropriate such partner is eCancer.

   eCancermedicalscience is a non-profit organisation which manages a free, online open-access peer-reviewed cancer journal, publishing original science articles (including video clips), reporting on cancer news and webcasting cancer conferences. ecancer is embracing web 2.0 technologies to help create a virtual social network of oncology professionals and provides an online community for those involved in all fields of cancer research and treatment. ecancer is the journal of the IEO (European Institute of Oncology) in Milan and of the OECI (Organisation of European Cancer Institutes). It is also the science and multimedia partner for ECCO, which represents 50,000 cancer scientists, clinicians and paramedical specialists throughout Europe. It is visited by 25,000 individuals each month, from 171 countries to date. Its international presence is therefore unparalleled. The editorial board consists of world experts in oncology, and is particularly strong in clinical trial expertise.

   They will be supported by the Breast International Group, the leader of the Knowledge management workpackage. BIG through their specialized personnel, have an outstanding expertise in dissemination, publication of materials and event organizing. Additionally, the organization brings to the project the link to their very large, world-wide network of breast cancer research organizations. BIG comprises over 40 collaborative academic breast cancer research groups, research partnerships and clinical trial units from Europe, Canada, Latin America, and Asia-Pacific, and collaborating closely with the U.S. National Cancer Institute (NCI) and North American Breast Cancer Group (NABCG).

   Additionally, each individual partner will contribute by working through their usual dissemination channels in the commercial (e.g. customer contacts and trade fairs) and academic world (e.g. conferences and publications) which go well beyond the cancer domain.

   Especially the Eurorec Institute, with their permanent network of National ProRec centres, has access to an European wide dissemination channel towards the healthcare industry (developers and vendors), healthcare providers (buyers), policy makers and patients. They will use this broad network

to disseminate the EURECA results and encourage adoption of the EURECA architecture (also through their certification programme).

Finally, the pharmaceutical industry is well represented in the EURECA project through the Pharmaceutical Advisory Board (cf. letters of commitment of various companies). This group of people is involved in data integration and possibilities of re-use of data on a day to day basis. They are committed to disseminating useful result out of the EURECA to their respective communities.

8. ***IMPACT & SCALE:*** Several members of the consortium (BIG, GBG/Luisenkrankenhaus Düsseldorf, IJB, UOXF) are top players in the global breast cancer research and treatment community. They will constitute a show case of EURECA, enabling us to prove the applicability of our solutions in a real multi-centric, multi-language (English, French, German) environment, also heterogeneous with respect to legal aspects, country-specific regulations, technology and ICT infrastructures. In this complex context we will validate the architectural framework, the services, and especially the viability of the core data set-based approach in an EU-wide setting. The success of this scenario will be a strong motivator for other large clinical networks (EORTC, METOXIA, SIOP, EuroSarc, etc.) to adopt the EURECA solutions at a wide scale and convince our collaborators from the pharmaceutical industry of the significant added benefits of EURECA (cf. also the involvement of the EURECA Pharmaceutical Board).

   Additionally, in order to gain momentum and community support, we are committed to provide open architectural solutions, disclose code as open source, and to ensure low entry barrier into EURECA for open source systems and tools.

Obviously, one could argue that there are other organisations and research groups which could contribute towards the achievement of the EURECA objectives. Although this may be true, we recognize that an additional but very crucial factor for the success of the project is the actual size of the Consortium - too many partners can create serious management problems. As a result we have purposely chosen the minimum number of participants which possess the required skills and expertise for realizing the vision of the project. These partners, as well as their role, skills and experience are described in detail in Table 2.3, which follows below.

Additionally, we have made by design provisions for developing global partnerships and linking to other pools of expertise and with major stakeholder groups. For this purpose we have setup a Scientific Advisory Board, which will consist of recognised experts. Already several such individuals and/or stakeholders have agreed to participate. Next to the Breast International Group and the German Breast Group who are partners in EURECA, other large consortia of clinical organizations who expressed interested in the results of the EURECA project are EuroSarc, Metoxia, EORTC and SIOP. These networks are also represented within the consortium by clinical partners who are as well members of those networks and will take a leading role in requirements gathering and dissemination of results in the context of those large networks.

Of particular importance for the further exploitation of our results is the collaboration with the pharmaceutical industry. Therefore, we have set up a Pharma Advisory Board, including experts representing several large pharmaceutical companies. The participation of these experts and their active engagement in our Advisory Board will enable us to align EURECA's approach with those of the pharmaceutical industry with respect to standards-based harmonization of efforts.

A special mention must be made to the EURECA's Consortium Prime Contractor. The contribution of Philips to the *EURECA* project started from its earliest stages. In close collaboration with a core group of partners, Philips has been the driving force in gathering partners possessing the required skills and expertise, for the successful completion of the project. Thus, Philips will bring its experience of management of multi-disciplinary and multi-cultural research projects. Within *EURECA*, Philips will act as the project coordinator and carry out the administrative and financial coordination of the project, while ensuring all exchanges with the representatives of the European Commission. Philips will have close interaction with all the *EURECA* partners and provide all the necessary tools and framework for all work packages to carry out their work and collaborate extensively among themselves.

Finally, we are aware that *interoperability is by definition a global issue which cannot be tackled in isolation, requiring both critical mass and openness*. Therefore, it is our goal to join forces with other initiatives and contribute to common solutions in Europe and beyond. All EURECA partners who participate in other relevant projects with focus on semantic interoperability in healthcare (as described in section 1.2.24) will actively work to establish working collaborations with those initiatives.

2.3.1. Expertise of the participants

Table 2.3 – Consortium Overview

Participant	Participant Short name	Type[1]	Country	Expertise	Role in the project
1	Philips	IND	NL	Clinical technology, clinical information systems, information integration, domain modelling, medical imaging, standardization and interoperability.	Project management, semantic interoperability, HL7 standards, mapping formalisms, core data set based on standardized terminologies, uniform access to trial information, pilots, validation, software services for recruitment and safety, semantic reasoning.
2	FORTH	RES	GR	Post-genomic research and innovative computer methods and tools in the area of medical informatics, ehealth, m-Health, medical imaging and bioinformatics.	Architecture, standards and integration. Models, deployment and clinical pilots. Evaluation and validation. Software services.
3	IJB	PB	BE	Clinical care, clinical trials (breast cancer), clinical and translational research. EHR and CT systems, HL7 CDA, SNOMED CT, CDISC, etc.	User needs. Models, deployment and clinical pilots. HL7 standards, CDISC, CT systems, SNOMED CT terminology. Semantic interoperability environment. Pilot site. Patient recruitment, hypotheses generation, feasibility.
4	Custodix	SME	BE	Privacy protection and e-security. Architecture, SoA, CT and EHR systems, standardization.	Ethics, legislation, privacy and security. Architecture. Semantic interoperability environment. Evaluation and validation. Software services for recruitment and protocol feasibility.
5	UdS	HE	DE	Clinical care, clinical trials and clinical research in paediatric oncology. Clinical trial management systems, EHR systems.	User needs and clinical research scenarios. Provide datasets and information on EHR and clinical trial systems. Semantic interoperability environment. Pilot site. Validation. Scenarios in patient safety, avoidance of multiple data entry, hypotheses generation, recruitment.
6	UOXF	HE	GB	Clinical care, clinical trials (breast cancer, bone sarcoma, etc), clinical and translational research. Bioinformatics and data mining; Diagnostic, prognostic, predicitive markers.	User needs and clinical/clinical research scenarios. Provide datasets and information on EHR and clinical trial systems. Bioinformatics in clinical research, data mining and knowledge discovery. Pilot site. Detection of safety risks, cohort studies, epidemiology.
7	FhG	RES	DE	Data Mining, Statistics, Distributed Systems, Data Warehousing, healthcare and research standards, development of CT systems.	Data mining and knowledge discovery. Semantic interoperability, implementation of tools for recruitment and safety, Healthcare and research standards.

Participant	Participant Short name	Type[1]	Country	Expertise	Role in the project
8	VUA	HE	NL	Semantic reasoning, semantic web technologies, logic representations, ontologies, medical guidelines. Data models.	Healthcare standards, semantic interoperability, semantic reasoning, support for development of eligibility criteria of clinical trials by approximate reasoning. Reuse health-care records for clinical research, medical logic representation, ontologies and guidelines.
9	BIG	PB	BE	Clinical trials in breast cancer, clinical and translational research, predictive and prognostic models, ethical, legal and regulatory frameworks	User needs and scenarios. Models, deployment and clinical pilots. Evaluation and validation. Ethical, legal and regulatory requirements. Clinical research standards.
10	LUH	HE	DE	Legal and ethical problems of ICT in the healthcare domain.	Ethics, legislation, privacy and security.
11	Xerox	IND	FR	NLP, information extraction, ontologies, semantic reasoning	Information extraction. Evaluation and validation. Development of the semantic interoperability environment.
12	UPM	HE	ES	Ontologies, biomedical informatics, clínico-genomic data integration, semantic mediation tools.	Semantic interoperability, mapping tools, ontologies and standardized terminologies, data integration. Evaluation and validation. Implementation of services, deployment at pilots
13	MAASTRO	PB	NL	Clinical site specialized in radiation oncology, clinical trials. Automatic eCRF completion with EHR data	User needs and clinical scenarios with focus on radiotherapy. Evaluation and validation site of clinical pilots. Interoperability, EHR/eCRFs.
14	eCancer	OTH	CH	Dissemination of oncology knowledge.	Dissemination
15	EuroRec	PB	FR	Healthcare standards (HL7 v3, CEN 13606, archetypes), semantic interoperability, clinical trial systems, standardization	EHR systems, standards (HL7 v3, CEN 13606, archetypes), clinical research standards, standardization, dissemination.
16	SIT	SME	NL	Semantic web technologies, semantic reasoning, personalization, contextualization, recommendation systems, software design and development, user interfaces, web design.	Semantic reasoning, contextualization of information for a patient case, software design and development, user interfaces.
17	GBG	PB	DE	Clinical trials in breast cancer, protocol development, patient recruitment, surgery, clinical and translational research, predictive and prognostic models, ethical, legal and regulatory , clinical guidelines.	User needs and scenarios, clinical knowledge. Models, clinical pilots. Evaluation and validation. Clinical guidelines. CDISC, CT systems, SNOMED CT terminology. Semantic interoperability environment. Pilot site. Patient recruitment, hypotheses generation, feasibility.

Participant	Participant Short name	Type[1]	Country	Expertise	Role in the project
18	NRC	PB	CA	Language-independent text analysis methods, processing of multi-lingual documents,	Information extraction of concepts and data out of EHR and clinical trial systems in a multi-lingual context

[1] Type: **IND**: Industry, **HE**: Secondary and Higher Education Establishment, **RES**: Research Organisation, **SME**: Small or Medium sized Enterprise, **PB**: Non-profit Public Body, **OTH**: All other organisations

2.3.2. Complementarity of the participants

Each of the partners of the *EURECA* project has been selected in a way that ensures that the full spectrum of skills and expertise required for carrying out the proposed project, are present in the *EURECA* consortium.

It must be emphasised once again that the partners were selected to be complementary in terms of their skills and knowledge, as well as for the role they will play within *EURECA*. Each partner has an impressive track record in knowledge creation, innovation and commercialisation in their respective domains of expertise. As a result, the partners that have been included in the consortium were selected, based on their ability to ***add value to the project***, through their ***commitment to joint innovation*** at a Pan-European level, their ***specific knowledge***, and their capacity for dissemination and exploitation. This plan also incorporates the extensive experience and knowledge of the other members of the consortium in participating in previous EU framework programmes.

These points become more evident from the short description of roles, expertise and experience which are presented in Table 2.3a. Moreover, Table 2.3a clearly highlights the roles and the functions (responsibilities and involvement) of each participant in the *EURECA* consortium.The more extensive profiles of the partners of the *EURECA* project, as well as the short CVs of key personnel from all project participants were presented in Section 2.2.The figure below graphically depicts the areas of expertise relevant to the EURECA workplan and attempts a mapping of the various partners, based on their prior expertise, in these areas.

Figure 24 graphically depicts the areas of expertise relevant to the EURECA workplan and attempts a mapping of the various partners, based on their prior expertise in these areas.

As already mentioned in the section above, we have taken special care in selecting our clinical partners. The most important criteria for their selection were: (a) possession of relevant data, (b) willingness to share this data (conditionally or unconditionally), (c) innovation in methods for knowledge discovery and (d) adherence to and compliance with legal and ethical issues. Their adherence to these “quality attributes”, i.e. openness and collaborative innovation, represents an additional complementarity dimension of project partners.

Finally, it is our experience that an additional attribute that in many cases influences the success of a project is the ability of the various partners to function together as a coordinated and coherent group and perform high-level collaborative research. The partners in the EURECA Consortium have a proven ability to working together. Most of them have successfully collaborated in a number of flag-ship research projects in their domains of expertise.

The strong leadership of the project will also assist in further developing the “collaborative innovation culture” of the consortium. The project manager, with the support of the very experienced Philips personnel, will focus on efficient, effective accomplishment of planned tasks, including proper handling of the consortium agreement, intellectual property rights, etc.

Concluding we may reemphasize our belief that the current consortium:

- Is of the highest possible quality, as individual partners
- Possesses all required skills and expertise relevant to the project
- Has a very impressive track record in innovative work while working together
- Is highly motivated

It is our belief, therefore, that the current consortium will be successful in responding to the various challenges described in our workplan and in delivering a truly intelligent suite of models, tools, services and guidelines for linking clinical research with clinical practise, and in setting up appropriate structures for exploiting – either individually or collectively – project results.

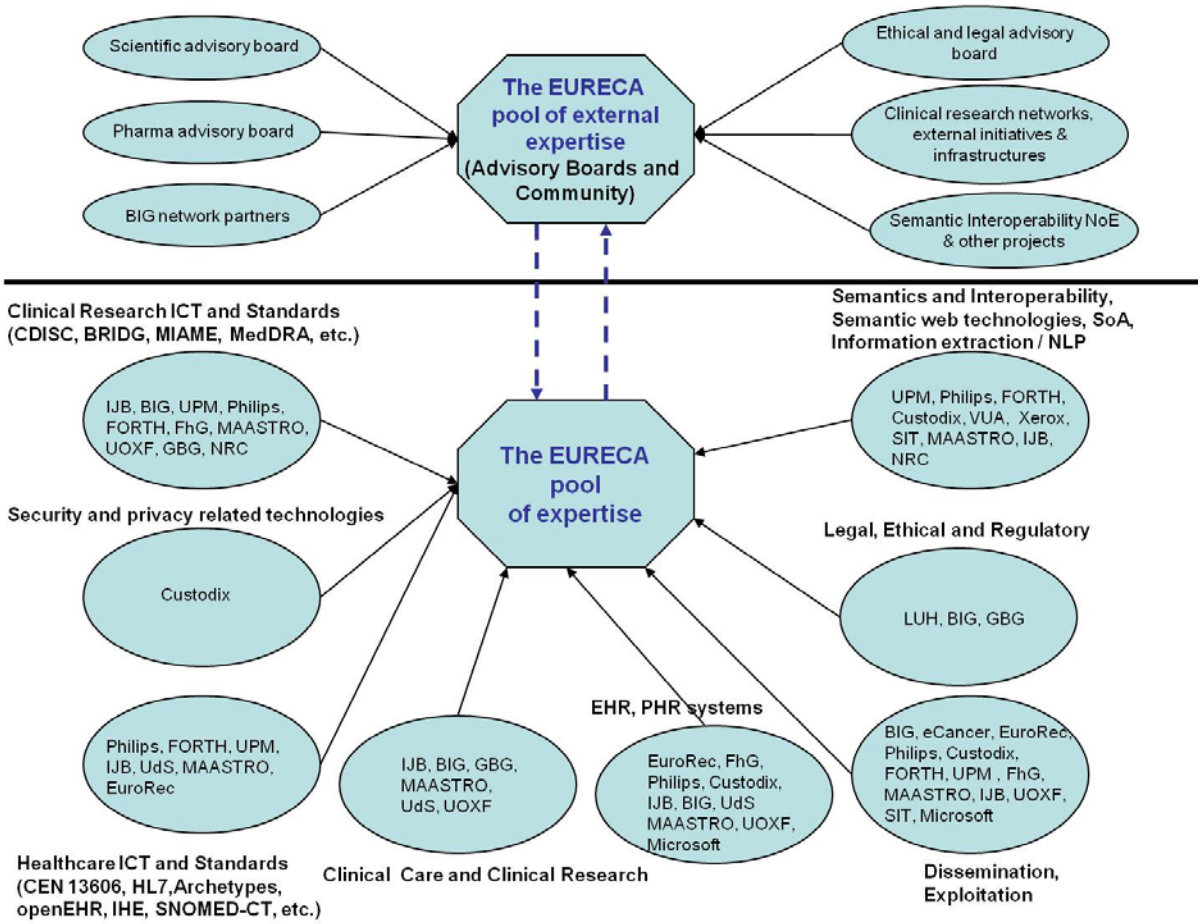

Figure 24 Expertise of the EURECA partners

2.3.3. The EURECA Clinical Partners

As already stated, the EURECA clinical partners represent top research institutions in the oncology area but also large hospitals treating significant numbers of patients both in and outside clinical trials.

1. ***The Breast International Group:*** The participation and endorsement of the BIG clinical research network ensures access to valuable data across all spatial scales relevant in the context of oncology (molecular to organism) and to top clinical research knowledge, and brings to our consortium the opportunity to demonstrate a large impact and reach a large clinical and biomedical research community willing to make use of our solutions. Founded by leading European opinion leaders in breast cancer research and treatment, BIG is a pioneer in integrating breast cancer research, both in Europe and further around the globe. While working independently from but fruitfully collaborating with the pharmaceutical industry, BIG has established a successful model for conducting clinical and translational breast cancer research. BIG will also provide their legal expertise to the consortium to ensure that our solutions are fully compliant with all legal requirements concerning the management and sharing of patient data. Next to that, requirements will be extracted from regulatory frameworks (such as the EU Data Protection Directive, FDA 21 CFR part 11) to which compliance is required.

   While BIG has a global reach, the activities in the EURECA project will primarily focus on the European part of the BIG network, especially for solutions regarding interoperability. However, within the legal and regulatory requirements evaluation task of EURECA, we will also investigate alternatives for future extension of our solutions beyond the European context. A unique quality of the project is the full commitment of the Breast International Group, as a partner in the project, to contribute large amounts of patient data and the extensive basic, translational and clinical research expertise of their network to build solutions based on challenging but realistic use cases.

   More specifically, BIG will lead WP10: Knowledge management, contribute from the user side to the development of the interoperability solutions and of relevant scenarios focusing on improved recruitment and better safety. Next to user requirements, BIG will also provide input and requirements to WP7, concerning legal and ethical issues and regulatory frameworks with which the

project needs to comply. Finally, BIG will participate to the evaluation and validation of the EURECA results.

2. ***The University of Oxford:*** The University of Oxford – through its Molecular Oncology Laboratory of the Weatherall Institute of Molecular Medicine, which is a designated Phase I Centre by Cancer Research UK, the Cancer Research UK Oxford Cancer Centre (CRUK) Translational Research Unit in Oxford, and the EuroBoNet component in translational research in rare bone sarcoma (Component based in Oxford University, lead based in Leiden) will bring into EURECA expertise in the areas of breast oncology, bone tumour and soft tissue sarcomas and bioinformatics.

   The quinquennial review for the CRUK Cancer Centre gave it alpha star ranking. It is also a CRC Phase I centre. CRUK bring in expertise in discovery and analysis of new genes in the hypoxia transcriptome and in the analysis of human cancers, with focus on breast cancer, to discover pathways related to hypoxia and angiogenesis; and development of therapeutic approaches.

   In ***EURECA*** the University of Oxford will provide clinical user requirements, clinical knowledge, clinical trial and care data, and all needed information on clinical systems. The University of Oxford will lead WP5: Data mining and knowledge discovery, contributing to the scenarios addressing re-use of EHR data for detection of safety risks, hypothesis generation, and protocol feasibility. Through the involvement of their IT department next to the clinical experts, University of Oxford will also participate in the tasks around the definition and implementation of the semantic interoperability environment.

3. **Saarland University:** The focus in research of the Department of Paediatric Oncology and Haematology of the University of Saarland is nephroblastoma (clinical study and trial and basic research in cooperation with different institutes) and brain tumour.
   Prof. Graf is a board member of different study groups in the German Paediatric Oncology and Haematology Society (GPOH). This includes trials for Osteosarcoma (COSS 86, COSS 91, COSS 96), Non-Hodgkin-Lymphoma (NHL-BFM 90, NHL-BFM 95), Brain tumour (HIT 91, HIT-SKK 92, HIT-Rez 97, HIT-LGG), Nephroblastoma (SIOP 93-01/GPOH and SIOP 2001/GPOH) (chaiman), Acute Myelogenous Leukemia (AML BFM 98).
   In ***EURECA*** UdS will be leader of WP1, dealing with the user requirements for the EURECA tools and services. In addition it will provide clinical trial and care data, clinical scenarios and technical information on all its clinical and research information systems. UdS will also be a pilot site for some EURECA services. Through the involvement of their IT department next to the clinical experts, Saarland University will also participate in the tasks around the definition and implementation of the semantic interoperability environment between the hospital EHR and their CT systems, and to scenarios demonstrating improved safety, elimination of multiple data entry, and more efficient patient recruitment.

4. **Institute Jules Bordet:** The Institut Jules Bordet (IJB) is an autonomous comprehensive cancer centre devoted entirely to the fight against cancer. The Institute Jules Bordet has an international reputation in various fields, especially in breast cancer research. It provides multidisciplinary care to more than 500 new breast patients each year and plays a leading role in translational /clinical research for this disease, both through the enrolment of women with breast cancer into clinical trials and through leadership in international collaborative research groups (EORTC/BIG). In particular, the department of Medical Oncology is actively participating into more than 200 clinical trials (35 of Phase I, 80 of Phase II and 78 of Phase III as well as many academic trials). Most of the active trials are conducted at European level and few of them worldwide.

   In ***EURECA*** the IJB ICT department will provide models for EHR data, collected at various key points in the patient's clinical process, that would be coded and normalized using international coding systems (as LOINC, SNOMED CT, ICD9 and ICDO) for each domain relevant to screening for inclusion in a clinical trial and adverse event reporting. Also, an export of clinical trial databases will be provided for many studies handled by IJB data centres. Moreover, a database of all clinical trials approved by IJB Ethics Committee is also maintained including a description of their eligibility criteria, and will serve as a starting point for providing models for clinical trial description. Additionally, IJB will provide clinical expertise and constitute a pilot site for EURECA.

   IJB will participate to the implementation and validation of the EURECA environment and tools in the context of the scenarios demonstrating efficient recruitment, long term follow up, single data entry and improved safety. The scenarios will be validated in the context of previous multicentric trials led by BIG.

5. **Maastricht University:** The Maastricht Radiation Oncology clinic focuses its research on Lung Cancer, Tumour Hypoxia and Molecular Imaging. MAASTRO has extensive experience in leading clinical trials in comparisons of treatment planning studies, in large multi-centric in-silico trials such as the ROCOCO trial where treatments with photons, protons and carbon ions are compared. MAASTRO is one of the first centres in Europe having eCRF populated from the Electronic Medical File.
   In ***EURECA*** MAASTRO will provide access to both research and clinical data and will also focus on the development of a patient database with all the known and potentially predictive factors that are associated with the prognosis of cancer. The profiles that can then be determined provide a basis for identifying subgroups or individuals for whom specific treatments are most suitable, with the least chance of side effects. In addition a clinical-research oriented scenario proposed by the Maastricht Radiation Oncology clinic aims at exploiting external clinical data in order to validate a predictive model of patient safety issues related to radiation therapy (survival, oesophagitis and radiation-induced lung damage). They will also provide scenarios, clinical expertise and contribute to the development and validation of the EURECA framework and services.
6. **The German Breast Group Research GmbH** is an academic research organisation with 90 employees specialized in the implementation of investigator-initiated trials of malignant breast cancer, providing the platform for all clinical studies of the German Breast Group (GBG). This academic research institute has recruited up to 4250 patients annually in clinical breast cancer trials. The GBG is currently conducting 18 breast cancer trials, in 10 of those GBG is the legal sponsor of the study. GBG covers the complete portfolio to conduct large scale phase III studies including protocol development, monitoring, data-management and statistics. GBG has developed its own web-based data capturing system Medcodes® which is used for all trials started after 2009. **The Luisenkrankenhaus Dusseldorf** is among the 3 largest breast cancer centres in Germany dedicated only to the treatment of breast cancer. 600 to 800 new breast cancer patients have received surgical and systemic treatment in this institution yearly. A clinical trials unit was set up 5 years ago and recruiting now around 15-20% of these patients into clinical trials.

   In ***EURECA*** the GBG will work closely with the Luisenkrankenhaus Dusseldorf to contribute to the EURECA interoperability environment. GBG will provide expertise concerning all aspects of clinical trials, and constitute a pilot site for the validation of the EURECA solutions.

2.3.4. Involvement of SMEs

Custodix

This Belgian SME has extensive **expertise in data protection**, with a strong emphasis on Privacy Enhancing Technology for medical data collection. Custodix is one of the first and few Trusted Service Providers (TSPs) in the Healthcare sector. Custodix is involved in a number of commercial projects which are very relevant to EURECA (customer names cannot be disclosed):

- Custodix provides secure information exchange, de-identification and audit functionality in a project where EHR data is extracted from a large number of General Practitioners in Belgium and used for medical research.
- Custodix has provided security modules for Electronic Data Capture (EDC) systems used in several Clinical Trials.
- Custodix has its own Electronic Data Capture system used in disease management.

Custodix can offer **pharmaceutical companies** possible cost reducing measures in the clinical trials, allowing them to achieve cost savings in the following areas:

- Secondary-use of medical data: the effort associated with collection of (new) data (entry, validation, etc.) is huge.
- Automated patient recruitment: for new trials, patient recruitment is currently a painstakingly manual process and thus costly (with people actually visiting physicians to ask them for patients who match the trial inclusion criteria).

Both topics are at the core of the EURECA project and both can only be solved in practice when data privacy is appropriately addressed. This is right within the **core business of Custodix**: data protection within eHealth and more specifically disease management and clinical trials. What is more, Custodix is actually already involved in a (less technological advanced) small commercial pilot initiative which tries to address those two needs. Custodix sees a clear market need and thus aims to use the results of EURECA to (jointly with other partners) exploit secure compliant services to facilitate re-use and patient recruitment.

SIT
SIT has been active in the field of data processing and extraction of patterns from large sets of information. This activity has become more and more essential, as data repositories keep on growing at a faster than linear rate. Furthermore, this evolution is seen in a multitude of different domains. To do so, they deploy new technologies data mining and Semantic Web, to facilitate services like personalization, data recommendation, data interoperability, and context-sensitive and user-adapted access to heterogeneous data sources
Over the past years SIT has specialized itself trying to find an answer to the problem of data overload in the TV domain, where tens of thousands of TV programs where described by an extensive metadata description and needed to be matched to thousands of user profiles, facilitating an unobtrusive and personal access to the data. Here, we translated data into Semantic Web languages and aligned different data structures and ontologies. After this data integration process, recommendation and personalization algorithms were deployed to see which parts of the data and which presentational forms of the data should be chosen to accommodate each specific user best.
SIT amassed a large knowledge of cutting edge technologies and expertise in Semantic Web, personalization and recommendation systems. Although our main domain has been multimedia and entertainment, these technologies are nowadays becoming more and more crucial in a variety of domains, including scientific literature applications, medical treatments and inter-operation, exchange and presentation of various types of data collections, etc.
We are confident that the above mentioned technical knowledge and the expertise that we have achieved through the years is readily transferrable to the health ICT domain. In EURECA, these techniques would be the basis for personalization and contextualization of information.

2.3.5. Sub-contracting

No subcontracting.

2.3.6. Third Parties

No third party.

2.3.7. Other Countries

We have invited the **National Research Council of Canada** to join our consortium, enabling us to benefit of their top expertise in language-independent NLP and multi-lingual (e.g. English, French, Dutch, etc.) and information extraction (the quality and performance of the tools developed by the NRC have been proven by their winning of international challenges in medical NLP), and of their work in building clinical trial results repositories (the TrialBank, together with UCSF). Both these aspects are highly important in EURECA, as the success of an EU-wide semantic interoperability solution also relies on our ability to address language heterogeneity (a significant barrier to scalability of current solutions in EU), and the need for improved reporting of clinical trials results is a recognized issue in the clinical research community. Finally, applying a pragmatic solution to extracting concepts and structure out of free text reports in EHR is essential in order to build a comprehensive information model of the EHR system.
The coordinator's budget includes budget for this new partner (105k Euro to cover for 24PM research work). If EURECA enters negotiations, we will present our argument (during the negotiations with the EC) concerning the uniqueness of this partner and the very high value they bring to the consortium.
The NRC will bring to the consortium the following advanced expertise and tools:

Language-independent medical NLP:
Most if not all of the text processing methods that we use are statistical, and therefore language independent. We have extensive experience in applying machine learning algorithms (for instance Support Vector Machine) to NLP tasks. One concrete example of a bilingual environment for text processing is a multi-year project on Syndromic Surveillance. Here, emergency room visit patterns are monitored for outbreak indications by analysis of the chief complaints. This system is applied in a bilingual area therefore reports can be in French and in English (and occasionally, in a mix of the two); the system is designed to take either language as input and provides language-independent aggregates.
A current project that we work on is ICD coding for day surgery reports. A portion of these are in French, and while retraining needs to be done to deal with French, the algorithms themselves remain unchanged.

Standardizing reporting of clinical trial results
This work concentrated on extraction of trial results from full-text scientific publications.
The extracted information is subsequently mapped to formalized data structures to allow for cross-collection analysis or for information retrieval purposes. This study furthered the state of the art in that it was the first study where full-text articles were used rather than only the abstracts (some of the information can only be found in the full text of the article), combined with a broad set of information elements that was extracted (rather than just a handful or less). We extracted about two dozen diverse information elements, using the same algorithm throughout. This study is currently expanded into using other document types than journal publications, specifically research/ethics board approval requests, grant applications, consent forms, etc. Adding these documents allows for earlier capture of studies, as well as for surveying trial progress, tracking negative results, and accounting for publication bias.

Building a repository of clinical trials
The work on clinical trial information extraction was done to feed the information into a comprehensive structured repository of clinical trial information. This repository is hosted by the University of California San Francisco under the name HSDB (Human Studies DataBase), and it uses the Ontology of Clinical Research (OCRe), and is built under the Human Studyome Project. The NRC-IIT ExaCT system combines information extraction with a user interface for curating the extracted information against the input document, thus guaranteeing high quality data but collected at a reduced cost.
For the Bind and PreBind project, we built a tool for extracting protein-protein interactions from literature, to be included into a large biomolecular interaction database.
For the LitMiner project, we constructed our own repository of biomolecular entities which included links to external databases but also lists of synonyms, “not-to-be-confused-with terms” (for disambiguation), and a publication index (which had ambiguities removed).

Expertise with bilingual NLP
NRC-IIT has broad expertise with processing bilingual text corpora. This expertise includes translation tools (especially targeted to solve restricted-domain discourse particularities – i.e., finding the specialized jargon that is hard to find through a dictionary), collection and analysis of parallel texts, cross-language information retrieval, dealing with accentuation (or missing accentuation), dealing with mixed-language documents, and semantic, syntactic, and morphological analysis of text.

Collaboration
We have been working in good collaboration with partners in many projects, where each of the partners brought their part of the overall expertise to the table. In many cases, these partners were from the health care or clinical research community. Examples are: Samuel Lunenfeld Research Institute at Mount Sinai Hospital Toronto, the NRC Institute for Biological Sciences, the Ottawa Hospital Research Institutes, the Ottawa Heart Institute, Ottawa Public Health, and the University of California San Francisco. A number of these projects were multi-year multi-partner projects.

2.4. Resources to be committed

The table at the next page shows a good balance between RTD over the partner’s involvement for such an industry-driven research project.

Proposal Acronym		EURECA	Duration in months		42
Effort Information in Person months					
Nr.	Partner	RTD	MNGT	OTHER	**Total pm**
1	**PHILIPS**	**106**	**42**		**148**
2	**FORTH**	**84**	**2**		**86**
3	**IJB**	**70**			**70**
4	**Custodix**	**106**	**2**		**108**
5	**UdS**	**71**	**2**		**73**
6	**UOXF**	**82**	**2**		**84**
7	**FhG**	**95**			**95**
8	**VUA**	**72**	**2**		**74**
9	**BIG**	**40**	**2**		**42**
10	**LUH**	**40**			**40**
11	**XEROX**	**50**			**50**
12	**UPM**	**91**	**2**		**93**
13	**Maastro**	**55**			**55**
14	**Ecancer**	**40**			**40**
15	**EuroRec**	**15**			**15**
16	**Stoneroos**	**40**			**40**
17	**GBG**	**30**			**30**
18	**NRC**	**23**			**23**
Total		**1110**	**56**	**0**	**1166**

2.4.1. Repartition between partner categories

The EURECA project partners plan to mobilize an amount of 1166 person-months for the realization of the overall project. The picture below shows the overall distribution of the efforts (in %) among the partners.

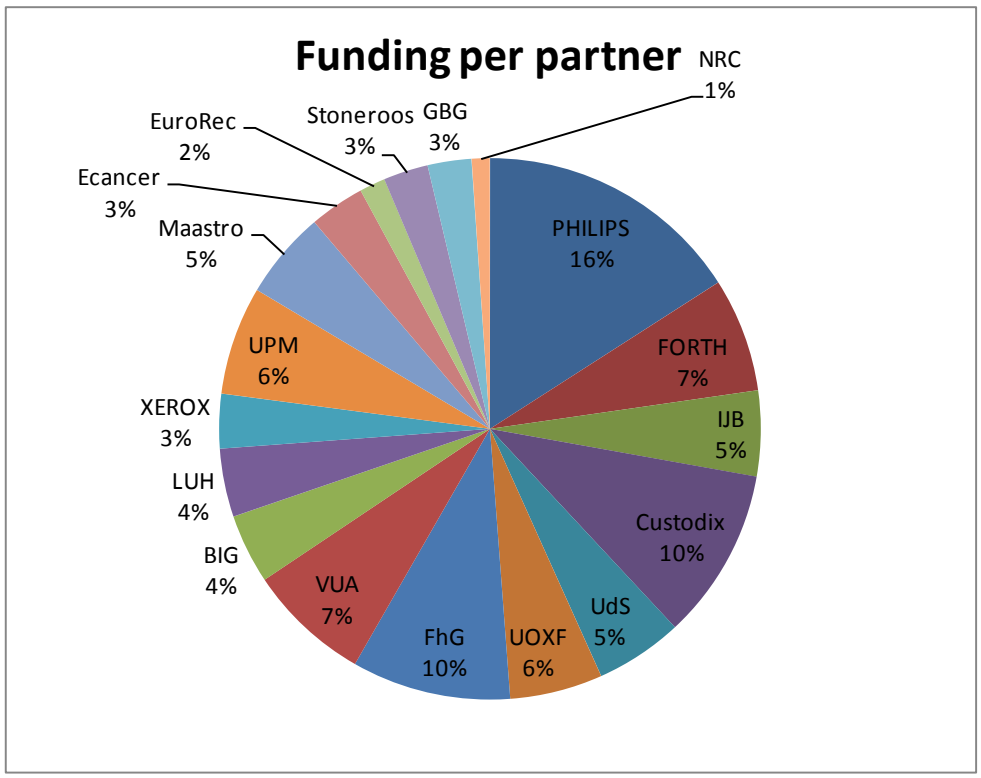

Figure 25: Effort distribution by partner.

All budgets listed in the A3 forms reflect the estimated costs expected to be incurred in carrying out the project and were calculated according to the accounting principles of the partners, which are subject to formal annual financial audits. Moreover, it is agreed that for the "EURECA" project Philips is the formal coordinator and that BIG will coordinates the dissemination, exploitation and standardisation activities.
The compulsory cost certificates, if needed, are registered under the management activity, and are to be seen as subcontracting costs.

Proposal Acronym:	EURECA					Duration in months:		42			
Financial Information											
						RTD		MNGT		**TOTAL**	
nr	Partner	Effort pm	Cntry	RTD funding	Indirect costs	Costs	Funding (50% or 75%)	Costs	Funding (100%)	Costs	Funding
1	PHILIPS	148	NL	50%	real	1,918,634	959,317	577,094	577,094	2,495,728	1,536,411
2	FORTH	86	GR	75%	real	834,336	625,752	31,200	31,200	865,536	656,952
3	IJB	70	BE	75%	60%	652,000	489,000	3,000	3,000	655,000	492,000
4	Custodix	108	BE	75%	real	1,207,750	905,812	81,700	81,700	1,289,450	987,512
5	UdS	73	DE	75%	60%	624,000	468,000	35,600	35,600	659,600	503,600
6	UOXF	84	UK	75%	60%	686,128	514,596	18,172	18,172	704,300	532,768
7	FhG	95	DE	75%	real	1,212,712	909,534	7,500	7,500	1,220,212	917,034
8	VUA	74	NL	75%	real	896,000	672,000	33,125	33,125	929,125	705,125
9	BIG	42	BE	75%	20%	410,316	307,737	92,280	92,280	502,596	400,017
10	LUH	40	DE	75%	60%	462,987	347,240	48,000	48,000	510,987	395,240
11	XEROX	50	FR	50%	real	625,500	312,750		-	625,500	312,750
12	UPM	93	ES	75%	real	804,298	603,224	21,067	21,067	825,365	624,291
13	Maastro	55	NL	75%	60%	670,912	503,184	3,000	3,000	673,912	506,184
14	Ecancer	40	CH	75%	60%	418,400	313,800	3,000	3,000	421,400	316,800
15	EuroRec	15	FR	75%	20%	199,350	149,512		-	199,350	149,512
16	Stoneroos	40	NL	75%	60%	344,000	258,000		-	344,000	258,000
17	GBG	30	DE	75%	60%	336,720	252,540		-	336,720	252,540
18	NRC	23	CA	75%	sm	139,020	104,265		-	139,020	104,265
Total		1166				12,443,063	8,696,262	954,738	954,738	13,397,801	9,651,000

The total management costs include the reserved budgets for the External Advisory Boards and for active cooperation with the NoE on semantic interoperability and European infrastructure.

- Personnel costs 55%
- Other direct costs 6%
- Indirect costs 39%

2.4.2. Repartition between RTD and Management activities

The total person month allocation shows a significant RTD work share for research organizations of 452 PM or 41% of the total work, which demonstrates the weight of breakthrough research activities in the project with a good balance between large industry (14%), Universities (HES 32%), two SME's will have the opportunity to use the generated knowledge to extend their portfolio (13%).

	1110		All
Distribution of RTD "PM" per organisation type	156	14%	IND
	146	13%	SME
	452	41%	RES
	356	32%	HES

Definition of State of the art, user needs, scenarios and use cases (WP1) mobilizes 78 PM of the total workforce. This work will be the input that will be carried out in the technical work packages 2 – 9.
The technical work packages Architecture (WP2), Information extraction (WP3), Semantics (WP4), Data mining and knowledge discovery (WP5) and Applications – decision support (WP6) will use 630 PM of the total work force. In parallel with these WP's, the ethics, privacy and security will be investigated in WP7 (68 MM). Finally, in WP8 the evaluation criteria and validation procedures will be set up (56 MM), and in WP9 the clinical pilots take place (197 MM).

In WP10 (81 MM) the non-technical items are covered like dissemination, exploitation and dissemination and WP11 (56 PM) covers all project management aspects - both technical and administrative) as described in detail in section B2.1. Management and coordination activities account for 4.9 % of workforce.

2.4.3. Repartition between budget categories

costs per cost category										
nr.	Partner	Personnel	Travel	Dur. Eq.	Consum.	Other	Subcontr.	Certificates	Indirect	Total
1	PHILIPS	1,370,397	27,963					9,000	1,088,368	2,495,728
2	FORTH	413,200	42,000					5,400	404,936	865,536
3	IJB	399,500	8,000					3,000	244,500	655,000
4	Custodix	821,000	30,000	25,000		40,000		4,000	369,450	1,289,450
5	UdS	365,000	26,000		15,000	5,000		2,000	246,600	659,600
6	OXFU	417,336	7,143		14,286			2,276	263,259	704,300
7	FhG	623,229	20,000	10,000	1,000	1,000		7,500	557,483	1,220,212
8	VU A'dam	393,340	30,000					5,000	500,785	929,125
9	BIG	309,400	15,000	500	31,430	60,000		3,000	83,266	502,596
10	LUH	259,160	25,207	5,000		30,000			191,620	510,987
11	XEROX	375,000	12,000						238,500	625,500
12	UPM	431,055	30,000	9,000		10,000		4,000	341,310	825,365
13	Maastro	364,320	40,000	15,000				3,000	251,592	673,912
14	Ecancer	200,000	7,000	1,500		53,000		3,000	156,900	421,400
15	EuroRec	148,125	18,000						33,225	199,350
16	Stoneroos	200,000	15,000						129,000	344,000
17	GBG	191,550	18,900						126,270	336,720
18	NRC	95,850	20,000						23,170	139,020
	totaal	7,377,462	392,213	66,000	61,716	199,000	0	51,176	5,250,234	13,397,801

The costs for ***travel*** are planned for all project related activities: for meetings of the project management boards, for technical working meetings when those cannot be combined with management meetings, for dissemination at international conferences, for other dissemination events (seminars, workshops, ICT events, etc.), for technical and management reviews (including reviews by the EC). The partners will carefully plan and harmonise necessary travel to consider possibilities to timely combine meetings of different nature at one place when possible.

Equipment:

The realisation of the ambitious business and technological objectives incorporates several risks as described in section 1.3.6. Therefore, the EURECA consortium carefully planned the distribution of efforts to the different WPs over the project life-cycle. As such the work plan, and especially the first 11 months (phase 1) has been set up according to the potential risks. This will enable to refine the project's technical direction and provide sufficient resources to the second half of the project, if a major redistribution of efforts is necessary (i.e. further detailed contingency plans are elaborated along the project life-cycle, based on the work and the continuous risk assessment).

3. Impact

3.1. Strategic impact

..Integrating EHRs with clinical trials has major potential to increase recruitment rate, as could be demonstrated in first studies…

Prof. Dr. Ohmann, Kuchinke:
Meeting the challenges of patient recruitment.
A role for electronic health records.
Int J Pharm Med 2007; 21: 263-270

The *EURECA* proposal focuses its research and development agenda on several pressing needs in the healthcare and in the bio-pharmaceutical industry. Their successful resolution will have significant and far-reaching implications.
The relevant impact areas listed in the work programme are:

- Common platform for a wide range of ICT-based healthcare services.
- Improve sustainability of healthcare services by enabling better use of resources.
- Increased international competitiveness of European Healthcare Information Services and Software industry.
- Guidance on healthcare information systems issues in “green field” member states.
- Accelerated establishment of interoperability standards and of secure, seamless communication of health data between all involved partners, including patients.
- Wide-scale epidemiology based on Europe-wide healthcare information system.
- Faster medication innovation and lower costs through a more efficient research process.

The EURECA project also supports many of the EU policy and social objectives. The potential impact of a successful implementation of the current project is briefly presented in the sections to follow.

3.1.1. Contribution to Expected Impact

3.1.1.1. Common platform for a wide range of ICT-based healthcare services

The EURECA contribution

EURECA aims to build a seamless, consistent, standards-based semantic interoperability framework linking care-specific information systems with clinical research information systems, to enable efficient and large scale reuse of care data for research. On top of the interoperability platform, we will build a set of software services for more efficient research and better care. The long term vision is to enable easy development and deployment into the environment of other ICT-based healthcare services.
EURECA envisages an open, scalable Service Oriented Architecture platform and will therefore focus on interoperability and interfacing. The main objective is to define an interoperability framework, based on a reference architecture and of sound definition of service interfaces and integration profiles, in which the tools and services created by the technical WPs can be fit and subsequently provide a reference implementation of this framework. This approach reuses existing systems and services whenever possible (speeding up initial deployment), ensures extensibility on the long run, allows for integration of (future) third party solutions. It will enable interfacing with other platforms from similar initiatives in order to rely on each other for scale and critical mass.
We will provide support for early adopters who would like to develop based on our reference architecture, join our environment, make use of our interoperability layer and services, and provide services on top of EURECA. The goal of EURECA is to provide an open architecture specification (interoperability framework) that can be used by collaborative initiatives to set up easily their data integration projects, in a way that such collaborations can interconnect with EURECA compliant data sources, re-use EURECA compliant services, etc. We will provide detailed guidelines concerning the minimum requirements and the steps to follow by an organization that would like to join or deploy a EURECA-compliant environment, make use of the EURECA services, or build new services on top of the existing EURECA environment.
Several ICT-based services relevant to the clinical research and care communities will be implemented in the project, such as: efficient automatic recruitment, protocol feasibility evaluation and support for protocol design; services enabling data mining of large volumes of data to generate new hypotheses, identify safety risks, and carry out epidemiology studies; integrated reporting of serious side effects; use of PHR data for reporting of side effects; and contextualization of information. All these services were identified as highly relevant by our communities of users. They are supported by the scalable and consistent semantic interoperability platform among EHRs and clinical trial systems, and by uniform access within the EURECA environment to a variety of external knowledge and data repositories.

Lowering the complexity threshold for the integration of service components to the minimum is key to the success of the EURECA interoperability framework. We will setup a certification programme checking services for "EURECA compliance" (similar to caBIG). This programme will be coordinated by the EuroRec Institute of whom certification and quality labeling is the core business. The "EURECA compliant" label will guarantee system integrators that a service can be deployed in a EURECA compliant environment without needing additional development.

3.1.1.2. Improve sustainability of healthcare services by enabling better use of resources

Significant effort and financial investments in the biomedical research and in the healthcare industry have resulted into a wealth of data and information with the potential to bring along large qualitative improvements in patient outcome. However, the heterogeneity of data, the low adoption of shared standards, the fragmentation with respect to methodology, infrastructure and tools, the duplication of efforts, and the insufficient collaboration across disciplines, organizations and industries limit the impact of these investments.

Paradoxically, despite the huge volumes of data generated in clinical practice, very little of it is used for pushing the frontiers of medical science forward. Lack of data still hampers basic research, the secondary use of the large amounts of patient data collected in clinical practice is low, and the link between clinical research and clinical practice is often broken, creating a widening knowledge gap between research and care. As the cost of healthcare in Europe becomes almost unaffordable, reducing the expenses while both accelerating research and significantly increasing the quality, safety and efficiency of care is a necessity.

The EURECA contribution

The goal of the EURECA project is to enable seamless, secure, scalable and consistent linkage of healthcare information residing in EHR systems with information in clinical research information systems to support the two currently separated worlds of clinical research and clinical practice to connect and benefit from each other. Next to benefits concerning patient outcomes and safety, this integration has an important potential to bring along significant cost reduction.
On top of the achieved semantic interoperability we build software services and tools to support more efficient research, better care and improved patient safety. The reuse of EHR data for research can reduce the costs and duration of clinical trials and solve the current shortage of data in clinical and biomedical research. Access to large amounts of high quality data from clinical practice will support large scale data mining for epidemiology, better hypotheses generation, better study design and execution.
Besides saving money on the research side through more efficient and better designed clinical trials, we also save the time of the healthcare professionals and help reduce the number of errors by eliminating the need for multiple data entry in several scenarios: trial enrolment, trial execution, reporting of serious side effects. Early detection and reporting of safety issues and better assessment of the benefits of drugs versus risks (including through the use of patient-managed data out of PHRs) can save the patients a lot of suffering (and sometimes even save lives), can save in hospitalization costs, and can reduce the costs of trials incurred by the pharmaceutical industry.

3.1.1.3. Increased international competitiveness of European Healthcare Information Services and Software Industry

Improving the Competiveness of European Healthcare Information Services

There are several European-specific factors that need to be addressed to improve the competitiveness of the European healthcare information services. These are a very high fragmentation and heterogeneity with respect to approach to integration and interoperability, implementation of standards and use of standard terminologies, legislation and regulations, language, and EHR vendor market (often country specific). Operating at EU level in this complex heterogeneous context requires significant additional costs for healthcare organizations and for information systems vendors.

EURECA will significantly facilitate the CT design and implementation phases by adding unique functionalities that will seamlessly link patient EHRs to CT management systems. The mid–term benefit will be a dramatic decrease in the time from drug discovery to marketing and clinical adaptation while in the longer-term is expected to lead to an increase in the number of CT and to the faster advancement of cancer research that is currently hampered by the rigid regulations and the interoperability gaps between the clinical records and CT information systems.

The EURECA contribution

The workplan of EURECA provides remedies to all these issues. At the core of the EURECA project is achieving semantic interoperability among EHR and clinical trial systems, consistent with existing standards (HL7 v3, CEN 13606, CDISC, etc.), while allowing

for/managing the various sources of heterogeneity: technology, medical vocabulary, language, etc. To identify, correctly classify and extract relevant concepts out of free text reports in the EHR, the project will also focus on natural language processing and text mining of relevant data. When possible, existing techniques and tools will be reused.

The EURECA's legal experts (LUH), also seeking advice when necessary from EURECA's Ethics & Legal Advisory Board, will ensure that no legal barriers are in the way of accomplishing the EURECA vision. The focus will lie on identifying and creating the data protection framework for EURECA. Main attention will be attributed to the implementation of the necessary security infrastructure as well as the ethical implications and requirements of the legal framework. In this respect informed consent forms will be drafted to ensure patients are properly informed about their rights and to secure those rights. Being aware of the legal complexities, we will develop solutions that work EU-wide and are not limited to a few member states.

Improving the Competiveness of European Software Industry

ICT is at the very core of the knowledge-based society. Activities will continue to strengthen Europe's scientific and technology base and ensure its global leadership in ICT, help drive and stimulate product, service and process innovation and creativity through ICT use and value creation in Europe, and ensure that ICT progress is rapidly transformed into benefits for Europe's citizens, businesses, industry and governments.

Improving the competitiveness of European industry and enabling Europe to master and shape future developments in ICT so that the demands of its society and economy are met is an important policy objective of the EU.
Industrial partners have a central role in EURECA. Also, small and medium-sized enterprises (SMEs) active in the software industry are important partners with significant responsibilities in the EURECA work plan. This is a result of our belief that their role in promoting innovation is vital and that SMEs play an essential role in the development and nurturing of new visions in IST and transforming them into business assets.

The EURECA contribution

The EURECA project will provide software platforms supporting highly innovative networked businesses on top of an Internet of Services. Our vision is to enable increased flexibility of the resources managed by virtual organisations and facilitate dynamic outsourcing with third parties capability to aggregate services, act as intermediaries for delivery, and provide innovative new channels for interaction.

Advancing Service Oriented Science

EURECA identifies the adoption of *Service Oriented Business Models* as a fundamental shift that is necessary to change the European economy into *"the world's most dynamic and competitive knowledgebased society"*[225] This paradigm shift manifests itself by the *"evolution of business models from the sale of products to the provision of electronic services"*, where services are seen as utilities that can be used but that are not owned by users. Enabling this paradigm shift, it is clear that a new generation of ubiquitous and converged network and service infrastructures is required that support construction and deployment of highly scalable, flexible, manageable, context-aware, dependable and secure services. These network infrastructures need to support an Internet of dynamically combined services with worldwide service delivery platforms. The 'third party generated service' is emerging as a trend supporting the move towards user-centric services, as shown by the advances in Service-Oriented-Architectures and in service front-ends as the interface to users and communities. Virtualisation of resources remains an important research driver enabling the delivery of networked services independently from the underlying platform.

The EURECA contribution

Key research challenges in the area of Service Oriented Infrastructure will be addressed by the *EURECA* project, which include:

- **Service abstraction and virtualisation**: Architectures and tools that enable provision of services as utilities supporting different levels of abstraction and virtualisation of resources, dynamic monitoring of SLAs that are internal to service containers and the dynamic reallocation of services; reliable, multi-protocol dynamic binding, with endpoint virtualisation;

[225] *NESSI Strategic Research Agenda. Vol. 1. Framing the future of the Service Oriented Economy. Public Draft 1 – Version 2006213Revision 3.1*

- **Inherently Stable and Safe Architectures**: Research, identify, and document principles and architectures that confer a high degree of stability, predictability, and trust to infrastructures and related deployed systems when considered from an end-to-end perspective.

3.1.1.4. Guidance on healthcare information systems issues in "green field" member states

The EURECA contribution

The absence of significant (digital) legacy systems can be an advantage for a healthcare organization that aims to deploy an efficient and cost-effective healthcare information system. Taking into account the goal towards an integrated Europe-wide healthcare information system, adopting from the start a standards-based approach can provide huge benefits to the healthcare organization in the future.
The fragmentation in available EHR solutions and the lack of use of standards by vendors are major issues, hampering research and generating unnecessary costs. At the same time, efficient collaboration with the pharmaceutical industry and attracting research funding should be a priority for "green field" member states. In this context, the best solution is the implementation of standards-based, integrated healthcare information systems, and this is where EURECA can help.
Little digital legacy also makes the information extraction task simpler. For the paper based legacy data it needs to be considered whether the effort to digitize the information pays for the benefits and only digitize or store in a structured way data that is likely to be used again.
Additionally, when starting fresh with the development of a new information system it is crucial to introduce proper information models and core data sets based on standardized terminologies, enabling to describe the information in a structured, meaningful way. Not needing to map to prior information systems will make this structuring effort easier and less resource intensive. The creator of the systems can be inspired by current standards and show cases demonstrating best practices. EURECA can provide a bootstrapping solution by delivering core datasets, an open reference architecture and comprehensive guidelines.
EURECA envisages an open, scalable SOA. This approach reuses existing systems and services when possible (speeding up initial deployment), ensures extensibility on the long run, allows for integration of (future) third party solutions and will enable interfacing with other platforms.
We will provide support to users who would like to develop based on our reference architecture, join our environment, make use of our interoperability layer and services, and provide services on top of EURECA. We will provide detailed guidelines concerning the minimum requirements and the steps to follow by an organization that would like to join or deploy a EURECA-compliant environment, make use of the EURECA services, or build new services on top of the EURECA environment.
Additionally we will follow open source developments concerning EHR and CT systems with the goal to provide easy integration of such solutions. There are at this moment already two open source CT systems to which interoperability is envisioned: OpenClinica and ObTiMA. The success of the large effort at VA to build an open source EHR system and the wide adoption of its reference architecture suggests that an open EHR solution may gain momentum in the future. Enabling hassle-free integration of service components is key to the success of the EURECA interoperability framework. We will setup a certification programme checking services for "EURECA compliance" (similar to caBIG). The "EURECA compliant" label will guarantee system integrators that a service can be deployed in a EURECA-compliant environment without needing additional development.
The large clinical research networks that have joined EURECA have a Europe-wide and International reach. This enables us to promote our solutions and approaches in all European countries and far beyond.

3.1.1.5. Accelerated establishment of interoperability standards and of secure, seamless communication of health data between all involved partners, including patients

The EURECA contribution

At the core of the EURECA project is achieving semantic interoperability among EHR and clinical trial systems, consistent with existing standards (HL7 v3, CEN 13606, CDISC, etc.), while managing the various sources of heterogeneity: technology, medical vocabulary, language, etc. This requires building sound information models describing the EHR and the clinical trial systems, and capturing the semantics of the clinical terms by standard terminology systems such as SNOMED CT, ICD, LOINC. The scalability of the solution will be achieved by modularization, identifying core data sets covering the chosen clinical domains. Achieved semantic interoperability among EHR and clinical trial systems opens the way for services and tools that support the health professionals in avoiding double data entry, enable automatic identification of patients for new clinical trials and, on the clinical care side, automatic identification of potentially relevant trials for patients, improve patient safety, and support epidemiology and cohort studies, and large scale collaboration.

EURECA will research and develop solutions to fulfil the data protection and security needs and the legal, ethical and regulatory requirements related to linking research and EHR data, enabling users to exchange data and collaborate in a trusted, secure environment.

We will also provide uniform access to PHR information to support early identification of safety risks and reporting of serious side effects. Patient involvement/reporting has the potential to significantly contribute to improved safety of drugs and treatments and to the early detection and evaluation of safety risks and issues, within clinical trials and beyond their scope.

To reach the patients, EURECA will make use of eCancerpatient, the patient focused area of eCancer containing relevant patient content. eCancer patient disseminates relevant video, news items and hosts patient forums The site is being developed with input from cancer patients and support groups. The site will be an outlet for disseminating the results of the EURECA project that are relevant for patients.

EURECA focuses on open scalable standards-based solutions addressing all security, privacy and regulatory requirements. The wide adoption of our standards-based open framework will contribute to accelerate the implementation of Europe-wide interoperability and of fully secure information exchange.

3.1.1.6. Wide-scale epidemiology based on Europe-wide healthcare information system

There is a strong need in healthcare, especially in the case of complex heterogeneous diseases such as cancer, to integrate the available data and knowledge in comprehensive models, infrastructures and tools, to standardize methodologies, and to promote data sharing and reuse, and multidisciplinary collaboration through the adoption of common standards. Without improvements in interoperability in the biomedical field, the current fragmentation at the level of methodologies, data and knowledge representation, preservation and exchange, terminologies, tools, services, interfaces, etc. will persist and continue to drain significant research efforts.

Despite identified benefits, the secondary use of care data for research, quality assurance and patient safety is still rarely supported. Main barriers to enabling secondary use of data are the lack of interoperability, common standards and terminologies.

The EURECA contribution

EURECA aims to support more effective and efficient execution of clinical research by providing access – in a legally compliant and secure manner – to the large amounts of patient data collected in the EHR systems to be used for new hypotheses building and testing (e.g. to benefit rare diseases), cohort studies, wide-scale epidemiology, as well as protocol feasibility.

Association studies on large volumes of EHR data can also reveal serious side effects to drugs and therapies and link those with patient characteristics (generate new research hypotheses for biomarkers identification). Data mining of the EHR data can be carried out based directly on the information model of the source or based on the EURECA core set of concepts when it is assumed that the end user does not need detailed knowledge of the source. EURECA will demonstrate this secondary use in concrete scenarios and provide uniform access interfaces to the data and tools/services developed to support the execution of the scenarios.

Our services will also provide uniform access to external repositories of data and knowledge relevant in the discovery scenarios (clinical trial databases, biomedical resources, drug databases, literature, etc.).

Access to this wealth of data and knowledge has the potential to enable discoveries currently not possible due to limited availability of data. Additionally, the expected wide adoption of our open framework within Europe will support large scale epidemiology studies that will allow for the early detection of health risks and a fast and efficient reaction when problems occur. Integrated access to health data Europe-wide can help the early detection of outbreaks of disease and the identification of their causes and context, can identify environmental issues (e.g., data mining of a particular disease occurrence versus environment may suggest new hypotheses), and can support early detection of potentially harmful effects of new drugs based on larger populations.

The trend towards EU-wide interoperability based on standardized terminologies will also support the mobility of EU citizens with respect to healthcare delivery by lowering the effects of the complexities stepping from the language- and local information systems heterogeneities.

3.1.1.7. Faster medication innovation and lower costs through a more efficient research process

EURECA will develop solutions to improve the efficiency of the research process taking into account all major stakeholders: independent investigators, clinical researchers, the pharmaceutical industry, patients and healthcare professionals. In our project we create a bridge between large networks of clinical research organizations and the pharmaceutical industry (through the close link to the IMI EHR4CR project and the

participation of representatives of several large pharmaceutical companies in our advisory board). Gathering support for the large scale adoption of our solutions from both communities will help us achieve EURECA's full potential.

Improving the Competiveness of European Biopharmaceutical Industry

The Pharmaceutical Industry is a very important pillar of European Industry. Competition within the pharmaceutical industry continues to intensify, and competing for clinical investigators and patients is no exception.

While pharmaceutical companies invest heavily in marketing approved drugs, they often do not employ that same market research and marketing expertise when it comes to targeting, positioning and communicating the value of clinical trials to study sites and patients.

..European legislation governing clinical trials is slowing research and may even be costing lives, leading scientists have warned….

D. Cressey
European clinical trials rules under fire
Nature News, Published online 13 March 2009,
Nature, doi:10.1038/news.2009.163

The research data underscores the high stakes and urgent need for pharmaceutical companies to improve the clinical trial process. Effective communications can result in better-selected study sites and patients who will remain with a study until it is completed, saving time and money in clinical trials[226].

In Europe a number of new regulations, directives and corresponding guidelines have been issued over the last decade. In particular, the EU Directive 2001/20/EC on Implementing Good Clinical Practice in the conduct of clinical trials on medicinal products for human use has dramatically affected the regulation, oversight and practice of clinical research in Europe. The intention of the Directive was to protect patients and make the European pharmaceutical industry more competitive by ensuring that all Member States had the same rules. The weight of the directive at least doubled the costs of clinical trials[227] and increased the administrative burden so much that many academic researchers could no longer perform such trials.

Also, transposition of the Directive into the legislature of Member States has led States to make additional changes, with often tougher requirements on trial designs. The variety of interpretations of the Directive by Member States has thus ended up in many countries having different rules, hence increasing rather than decreasing disharmonies between EU Member States [228]As a consequence, Sweden, for instance, has seen a 25% decrease in academic clinical trials, Ireland 60% and Poland a staggering 90% reduction[229]. This is also an issue that needs to be quickly addressed.

The EURECA contribution

The pharmaceutical industry has identified the need to integrate clinical trial data with healthcare data to improve efficiency of the clinical trials. By providing standard-based semantic interoperability on domains of concepts (in specific clinical areas) between the two types of systems EURECA supports enhanced clinical trial conduct through improved efficiency and accuracy. The sponsor of the trial will benefit from the achieved interoperability by having access to a larger population, being able to compare safety data from a clinical trial to a much larger baseline, and accessing larger volumes of data for analysis.

The EURECA service that aims to identify potential candidates for clinical trials based on EHR data has the potential to increase the patient accrual reducing the trial duration and the time-to-market of the pharmaceutical products.

Access to longitudinal patient data beyond the end of a clinical trial can allow monitoring of the newly released drugs and support the early detection of serious side effects of treatments and drugs, not only reducing the negative effects on patients, but also reducing the risk of damaging the reputation of the pharmaceutical company and of decreased public trust.

Eliminating the need for multiple data entry along the workflow of a clinical trial, the EURECA environment will improve the ease and efficiency of study execution, reduce the transcription errors and the necessary

[226] Communicating the Value of Clinical Studies: Recruiting and Retaining Investigators & Patients. http://www.benchmarkingreports.com/ businessoperations/op85_clinical_communications.asp
[227] J. Hear, R. Sullivan. The impact of the 'Clinical Trials' directive on the cost and conduct of non-commercial cancer trials in the UK. *European Journal of Cancer*, vol. 43, pp. 8-13, 2007
[228] Editorial. Striking the right balance between privacy and public good. Lancet, vol. 367, pp. 275, 2006.
[229] R. Hoey. The EU clinical trials directive: 3 years on. Lancet, vol. 369, pp. 1777-8, 2007

verification steps, and eliminate redundant processes. All these have the potential to save time and lower the costs of trials, additionally freeing resources and enabling an increase in the number of trials.

Better access to increased patient populations, by accessing EHR data and by achieving more efficient enrolment in trials, may enable research in rare diseases or addressing population sub-groups that are often excluded (e.g. elderly patients).

The EURECA adverse events reporting support aims to eliminate independent data entries in the reporting and description of adverse events in clinical trials, improving the efficiency of the reporting and reducing the number of inconsistencies (e.g. due to different terminology) and errors. The sponsor of the trial can access pertinent information faster and assess sooner the causes and the outcome.

Improving the Capability of Healthcare professionals and Patients

The fact that many physicians, surgeons, and other members of the healthcare team do not encourage their patients to consider participation in clinical trials also contributes to low enrolment. Results from one study showed that a recommendation by their physician was the primary factor influencing patients' decisions to enrol in a trial.

The National Cancer Institute succinctly details the following common perceived barriers that the medical community claims in regards to clinical trial participation.

- **Lack of awareness of appropriate clinical trials.** Physicians are not always aware of available clinical trials. Some may not be aware of the locally available trials, or some may assume that none would be appropriate for their patients.
- **Unwillingness to "lose control" of a person's care.** Most doctors feel that the relationship they have with their patients is very important. They want what is best for the patient, and if the person must be referred elsewhere to participate in a trial, doctors fear they may lose control of the person's care.
- **Belief that standard therapy is best.** Many health care providers may not adequately understand how clinical trials are conducted or their importance. Some believe that the treatment in clinical trials is not as good as the standard treatment. [Investigators] might be uncomfortable admitting that there is uncertainty about which treatment is best in a phase III clinical trial. Debora Paterniti is the author of a study that monitored cancer patients as they considered participation in Phase I and Phase III trials[230]. Paterniti found that "*…a third of patients who were considered for clinical trial participation declined to participate, many of them out of a mistaken belief that investigational treatments are not as effective as standard treatment.*" In fact, many investigational treatments as at least as effective as conventional therapy, and cancer patients who participate in clinical trials frequently have higher survival rates than those who receive standard care.
- **Belief that referring and/or participating in a clinical trial adds an administrative burden.** The length and details of most research protocols may deter providers from participating in clinical trials. The possibility of incurring additional costs and expenses that might be inadequately reimbursed is a deterrent for many.

  According to a study survey conducted in 2000 by the American Society of Clinical Oncology (ASCO), the most significant barriers to patient enrolment included the intensity of paperwork collection and filing, and the extra time needed to train staff in the completion of enrolment and data collection forms. "Recruitment and adherence are very closely linked since those recruited must be followed to study completion as the inception cohort. …There is a clear impact on recruitment in terms of cost and both screening and staff burden"[231].

The EURECA contribution

In responding to this challenge the EURECA project will deliver:

[230] Understanding Cancer Patients' Needs, Concerns, is Key to Improving Clinical Trial Participation. http://news.ucdavis.edu/Capatients_needs.html

[231] Wright JR, Crooks D, Ellis PM, Mings D, Whelan TJ; Factors that influence the recruitment of patients to Phase III studies in oncology. Cancer. 2002; 95(7):1584-1591

- Tools to assist clinicians in understanding their clinical data. Thus they should be quickly supported in identifying appropriate patient cohort for participation in a specific study and identifying relevant trials for a specific patient.
- Simplified and more efficient data entry and improved data retrieval. Tools and services that will ease the administrative burden of participating in CTs, by avoiding duplication of data entry, automating the reporting process, etc. A one-time data entry will replace the current multiple entries of the same data in different systems.
- Improved data quality and consistency, fewer errors and fewer verification steps due to single data entry
- Tool to simplify and increase the efficiency and reliability of serious adverse events reporting and management.
- Increased efficiency in trial management, allowing the contribution to more trials with the same resources
- Efficient presentation of a patient's relevant history, including data from clinical trial participation, contextualized to the current health episode.
- A semantically rich model for describing CTs for publication into CT repositories and a prototype of such a repository.

More physicians could become involved in clinical research due to lower costs and resource requirements, increased efficiency and simplified process of including patients in trials and of data collection.

Improving the Capability of Independent Investigators

Investigator-driven clinical trials are clinical trials that are instigated by academic researchers and are aimed at acquiring scientific knowledge and evidence to improve patient care. Such studies deal with potential diagnostic and therapeutic innovations that do not attract or could even be against commercial interest. Typical examples are proof of concept studies, studies on orphan diseases, comparison of diagnostic or therapeutic interventions, surgical therapies or novel indications for registered drugs. They have a much broader scope and potential impact than industry-driven clinical trials, form a key part of patient-oriented clinical research, and create the basis for continually improving patient care[232].

Independent investigators are faced with all the main issues encountered in pharma-led trials, such as high costs, low recruitment, duplication of trials and complex and inefficient execution, however with a higher impact due to their access to very limited resources. Lack of funding, difficulties with collecting and managing sufficient data and unavailability of appropriate infrastructure have been identified as important obstacles[232].

The EURECA contribution

In responding to this challenge the EURECA project will deliver:

- Services to automatically identify candidates for trials, improving enrolment
- Semantically interoperable linkage between EHR and CT systems to avoid duplicate tasks that currently increase the costs of clinical research
- Semantically-aware uniform access to extensive amounts of data from clinical care to be used for research (new hypotheses generation, research in rare diseases, cohort and association studies, etc.)
- Possibility to seamlessly follow up patients beyond the end of a clinical trial
- Services to assist in the analysis of their clinical data and recommendation services for appropriate patients for participation in specific CTs.
- Dissemination and educational material promoting the benefits of participating in CT, both for the sake of research but also in order to receive better care.

3.1.1.8. Solving Societal Problems

Allowing for discoveries in the laboratory to be quickly transferred to the clinical management and treatment of patients and obtaining societal benefits.

[232] Investigator-Driven Clinical Trials, Forward Look, European Science Foundation, 2009

A standardized representation of a trial would promote the ability to determine the applicability of a trial result in the treatment of an individual patient.

The ultimate purpose of a clinical trial is to discover a method to improve the health or quality of life for an individual patient. One of the major challenges that repeatedly arises relates to the determination of whether or not the outcome of a particular trial or set of trials is applicable to a given patient. The decision often involves a review and evaluation of the structure of the trial with particular attention to the eligibility criteria, treatment regimen, and observed results.

The lack of a standard trial description makes this task more difficult both in terms of simply locating the necessary information within a published trial description, but also in the interpretation and understanding of that information[233]. Additionally, the lack of access to information on trials with negative results may mean that instead of spending money efficiently on new research directions, research dead-ends are repeated with negative consequences for patients.

An IOM report concluded that "it is simply not acceptable for patients to be harmed by the same health care system that is supposed to offer healing and support"[234]. This statement refers to the avoidable harm brought to patients both by errors of omission and errors of commission and by undetected serious side effects of treatment and drugs.

Critical elements for providing safe care to patients include assembling a comprehensive picture of the patient case, applying this knowledge appropriately and effectively, monitoring the effects of disease and therapies on the patient over time and detecting and preventing errors that could harm the patient[234].

Improved patient enrolment in clinical trials does not only hold the potential of improving the treatment for future patients, but it may also provide better outcome for the participating patients as statistics show better outcomes of patients treated in clinical trials.

Other important aspects of patient safety are the integration of all the data that is known about a patient from all previous care encounters and from all systems that hold data about that particular patient, and the enhancement of the adherence to clinical protocols and guidelines[235].

The National Cancer Institute again details the following common barriers that potential volunteers consider in regards to clinical trial participation[236].

- ***Lack of awareness of clinical trials.*** Research has consistently shown that most people with cancer are not aware of the option to participate in clinical trials.
- ***Lack of access to trials.*** The reality or the perception that there are no trials nearby deters many potential participants. In addition, seeking care at a distant trial site presents time and travel barriers.

The EURECA contribution

By achieving semantic interoperability between the EHR and clinical trial systems and by building software services/modules to:

- automatically match potentially-eligible patients with relevant clinical trials,
- provide access and mine longitudinal care data to detect potential safety risks,
- carry out data mining and build research hypotheses,
- contextualize patient and clinical information to the patient case,
- improve efficiency and communication by avoiding double data entry

The EURECA environment supports improvement in patient recruitment, targeted selection of patient case information and clinical guidelines for better outcome, higher data quality and better safety. Faster recruitment also leads to improved research and time-to-market of new drugs.

Improved clinical trial accrual and better patient outcome is also supported by providing uniform access to repositories of running clinical trials information, enabling the physician and the patient to find suitable clinical trials.

[233] Zarin DA, Tse T. Medicine: moving toward transparency of clinical trials. Science 2008;319:1340 –2

[234] LT. Kohn et al., To Err is Human: Building a Safer Health System. Washington DC: National City Press, 2000

[235] PM. Kilbridge et al., The Informatics Oportunities at the Intersection of Patient Safety and Clinical Informatics, JAMIA 15 (4), 2008

[236] Cancer Clinical Trials: The In-Depth Program. http://www.nci.nih.gov/ clinicaltrials/resources/in-depth-program/page7. Viewed 16 April 2003

3.1.2. European Added Value

It is widely recognised today that further progress in life sciences depends on our ability to develop common representations (ontologies, integrated vocabularies, etc.) to model and describe heterogeneous information, as a stepping stone in the process of achieving semantic interoperability in well defined clinical sub-domains.

The complexity of such a task coupled with the fact that solutions developed need to be promoted and supported for wide adoption across the whole of Europe, necessitate that such efforts do take place at a European level. Other reasons for carrying the work at a European level include:

- The required expertise and centres of excellence are distributed across Europe and not concentrated in a single country
- The adoption process becomes easier when the work is performed at a European scale
- Taking into consideration the diversity of requirements and contexts in a multitude of European locations, carrying out the required research at a European level enables the development of the "most appropriate" solutions for all.
- Due to need for collaboration and sharing expertise, increased stratification of diseases and difficulties in enrolling sufficient patients fast, many clinical trials are multi-centric with a European or International scope.
- In order to achieve a successful interoperability solution critical mass of adopters is a requirement. Therefore the approach needs to be EU-wide as opposed to local solutions.
- Bio-pharmaceutical companies are global organizations who need solutions that can be deployed in many countries and across borders.
- Achieving interoperability among EHR systems is a European-level issue, being addressed by many initiatives and determining the adoption of standards such as CEN 13606.
- Large industrial companies involved in building clinical information systems and clinical decision support systems are global players, aiming at solutions that can be deployed at European and International level.

3.1.3. Other national or international research activities and relevant organizations

In order to achieve and deliver real-life results the EURECA project has already linked with and learnt from some key national and international research activities. These national and international research activities include:

- **semanticHEALTH** – The *SemanticHEALTH* Specific Support Action developed a European and global roadmap for deployment and research in health-ICT, focusing on semantic interoperability issues of e-Health systems and infrastructures. The roadmap is based on consensus of the research community, and validated by stakeholders, industry and Member State health authorities. The findings of SemanticHEALTH regarding semantic interoperability and the provisions of the defined roadmap have influenced significantly the underlying thinking and work plan of EURECA. EURECA partner Custodix have been active contributors in the SemanticHEALTH (SSA).
- ***openEHR*** – openEHR is an open, detailed, and tested specification for a comprehensive interoperable health information computing platform for the EHR and other major services such as terminology. It is based on 15 years of research, focussed engineering design and real-world implementation experience, rather than being created as a formal consensus standard. However, over the past five years it has had a significant influence on the development of EHR standards by the three main international eHealth standards development organisations, CEN (European Committee for Standardization), HL7 (Health Level 7), and ISO (International Organization for Standardization). CEN EN13606 is a subset of the full openEHR specification.
- ***EORTC*** – The aims of the European Organisation for Research and Treatment of Cancer are to develop, conduct, coordinate, and stimulate laboratory and clinical research in Europe to improve the management of cancer and related problems by increasing survival but also patients' quality of life. Extensive and comprehensive research in this wide field is often beyond the means of individual European laboratories and hospitals, and can best be accomplished through the multidisciplinary, multinational efforts of basic research scientists and clinicians from the European continent. During the last few years, numerous innovative agents have been discovered as a result of tremendous development in the understanding of the molecular basis of cancer. Further clinical progress in cancer treatment will be accomplished mainly through the conduct of translational research projects, efficient drug development and the execution of

large, prospective, randomized, multicenter cancer clinical trials. EORTC promotes multidisciplinary cancer research in Europe and is linked to other leading biomedical research organizations around the world. EORTC has the aim to facilitate the passage of experimental discoveries into state-of-the-art treatment by keeping to a minimum the time lapse between the discovery of new anti-cancer agents and the implementation of their therapeutic benefit for patients with cancer.

- ***EuroBoNeT*** – EuroBoNeT has been a European project aiming at achieving integration by staff exchange and website-based discussion forums to increase and disseminate knowledge of primary bone tumours at the molecular level for development of new tools for patient care and cure and technology. With this integration exchange of material (virtual BioBank), Standard Operating Protocols and the use of technology platforms will enable EuroBoNeT to obtain statistical significant datasets, otherwise not achievable due to the rareness and large number of sub entities. A joint programme will contribute in obtaining molecular portraits of tumours, separated in 4 research lines. In each research line the biology of the separate group (cartilaginous tumours; osteogenic tumours and related sarcomas; osteoclastogenesis and Giant cell tumours of bone; and Ewing family of tumours) will be examined by genome wide expression and genomic aberration studies. More specific hypothesis driven approaches will be investigated by RNA/protein expression and mutation analysis. Knowledge on normal growth and differentiation will be gathered through in vitro studies. This should lead to further understanding and identification of markers for malignant transformation and/or progression, as well as identification of therapeutic targets.
- **The NeoBIG** clinical program was developed specifically to enhance and accelerate biomarker discovery and validation within the arena of early drug development in breast cancer. In this context, BIG aims to build on its already well established research collaborations to introduce a new model, the NeoBIG program, for the conduct of early breast cancer trials in the era of molecular oncology. This approach – to be launched in a series of international trials (each with 300-600 patients) – will be based on the evaluation of new promising targeted therapies in different molecular breast cancer subgroups. NeoBIG is a network of 29, predominantly European, cancer research centres, hospitals and specialised laboratories that have joined together – under BIG leadership – to carry out an innovative programme of prospective neo-adjuvant breast cancer trials with novel targeted therapies. The advantage of the neo-adjuvant setting is the ability to have an immediate and early surrogate of treatment efficacy as well as the opportunity to take serial biopsies to molecularly characterize the tumour and its response to treatment at multiple data levels. This collaboration between clinical, translational and basic research units ultimate aim is to perform biomarker research and validation in the early clinical setting to accelerate their implementation in the clinical setting.
- ***METOXIA*** – METOXIA is a European project focusing on a clinical application which was born within Euroxy, where the clinical relevance and implications of tumour hypoxia are tested. Duration February 2009-February 2014, an FP7 EU-Collaborative Project/large-scale project with 22 partners. Acronym: *METOXIA*. Project title: “Metastatic tumours facilitated by hypoxic tumour micro-environments”.
- ***Euroxy*** – Euroxy is a European R&D project with the subtitle "Targeting newly discovered oxygen-sensing cascades for novel cancer treatments. The project is supported by EU's sixth framework programme. Euroxy has 25 partners from 12 countries, among these Industrial Partners, Universities, and Research Centres. Human solid cancer most often has hypoxic areas where the effect of radiotherapy and most cytostatics is poor. Different degrees of hypoxia elicit different adaptive cell responses and thereby expose different potential targets for therapy. Euroxy is taking advantage of this insight and in a concerted action develops the technology platform needed for determining and controlling oxygen tension in vitro and in vivo. The cell response to controlled hypoxia is determined, in vitro and in vivo models are being developed and novel putative anticancer drugs are being synthesized. The ambition is to develop compounds of interest for further development by the EU pharmacology industry.
- **EuroSarc** – (Translational research in rare bone sarcoma including clinical trials) connects Oxford University with European work programmes on the genomics and translational trials of rare bone sarcomas. These tumours that frequently occur in the younger teenage and young adult age group and consequently treatment late effects are of major importance. The translational element includes whole genome expression array analysis and next generation sequencing of primary tumour material that now requires linking with patient outcome data and health records. The focus is to recruit patients with rare cancers to clinical trials specific for their cancer. This requires an information network for patient and physician access, including data handling and outcomes for the trials. This also includes biobanks of tissue linked to patient outcomes. The main trials in the workpages are;
    - Phase I/II trials in Ewing sarcoma, including combination studies based in amajor centres for rare sarcoma in Oxford, Leiden, Bologna and Munster.

- Phase II trial in chondrosarcoma, based in the above centres and extedning to EORTC centres.
- Phase IIa trials of MePACT in metastatic osteosarcoma.
- Using patient health records to evaluate diagnostic delays, treatment plans and late outcomes for patients with rare bone sarcomas, e.g. effects of referral pathways, chemotherapy, radiotherapy, radiological surveillance and trial therapy.
- Using patient health records to correlate treatment options and clinical trials for patients with rare bone sarcomas over the age of 40.
- Generate tools to identify patients for health care records in order to disseminate best practice derived from clinical trials for bone sarcoma treatment across Europe.
- Enhance the application of biobank and genomic repositories to health care records and outcomes of patients with rare bone sarcomas.

- ***caBIG*** – Cancer Biomedical Informatics Grid. It is a cancer-based biomedical informatics network developed by NCICB. caBIG aims to connect cancer related data sources, tools, individuals, and organizations, and to help redefine how research is conducted, care is provided, and patients and participants interact with the biomedical research enterprise. *EURECA* has already developed strong collaborations with caBIG.
- ***ACGT*** – An FP6 funded Integrating Project whose ultimate objective has been the provision of a unified technological infrastructure which will facilitate seamless and secure access and analysis of multi-level clinico-genomic data enriched with high-performing knowledge discovery operations and services, in the context of post-genomic clinical trials on cancer. Amongst other things, the project has delivered
  1. A Master Ontology on Cancer[237]. The intention of the ACGT Master Ontology (MO) is to represent the domain of cancer research and management in a computationally tractable manner. The ACGT MO is being maintained, using the Protégé-OWL free open-source ontology editor, Version 4. It is written in OWL-DL and presented as an .owl file. The ACGT MO is re-using Basic Formal Ontology (BFO)[238] as upper level and the OBO Relation Ontology[239]. It is currently been prepared for submission in the OBO foundry.
  2. An ontology based Clinical Trial Design and Data Management System, called ObTiMA, which is compliant with CDISC, as an open source product.
  3. A meta-data model for developing, publishing and discovering semantic services for invocation and execution in eScience Bioinformatic Workflows.
- **P-medicine** – An FP7 project, P-medicine brings together internationally recognised leaders in their respective fields with the aim to create an innovative computational, service-oriented infrastructure that will facilitate this gradual translation from current medical practice to personalized medicine. In achieving this objective p-medicine has formulated a coherent, workplan for the design, development, integration and validation of all technologically challenging areas of work. Our emphasis is on formulating an open, modular architectural framework for the tools and services to be developed, so that adoption of p-medicine services will not be an all-or-nothing decision; on efficient secure sharing and handling of large personalized data sets —including policies, privacy, modelling, cloud storage, etc.; on enabling demanding Virtual Physiological Human (VPH) multiscale simulations, for which standardization and semantic data integration and interoperability are major issues to be addressed; on building standards-compliant tools and models for VPH research, such as the VPH Toolkit , by defining a formalism to make the knowledge that is implicitly encoded in these tools explicit and thus improving the re-use of tools and solutions; on providing tools for large-scale, privacy-preserving data and literature mining, a key component of VPH research. We also focus on ensuring that privacy, non-discrimination, and access policies are aligned to maximize both protection and benefit to patients. The project is clinically driven and promotes the principle of open source and open standards. It targets the urgent needs of the cancer research community, a key area of societal importance, with the intention of strengthening the integration of the European Research Area. To sustain a self-supporting infrastructure realistic use cases will be built that will demonstrate tangible results for clinicians in their daily practice.

[237] http://www.ifomis.org/wiki/ACGT_MO
[238] http://www.ifomis.org/bfo
[239] http://www.obofoundry.org/ro/

UdS (Prof. Graf) is the coordinator of P-medicine, and several EURECA partners including Philips, FORTH and UPM are partners in P-medicine.

- The goal of the FP7 INTEGRATE project (funded under the ICT for Health programme) is to develop flexible infrastructure components and tools for data and knowledge sharing and large scale collaboration in biomedical research, to bring together heterogeneous multi-scale biomedical data generated through standard and novel technologies within post-genomic clinical trials and seamlessly link to existing research and clinical infrastructures. INTEGRATE brings together a wide multi-disciplinary community of biomedical and clinical researchers committed to work together, to establish common methodologies and clinical protocols, to collaboratively build predictive models, carry out research and select the most suitable integrative workflows. This lasting translational research infrastructure will provide a robust, secure and flexible IT solution that will facilitate collaboration between cancer research centres and other relevant stakeholders such as basic research organizations and pharmaceutical companies, and will be validated in the context of the NeoBIG research programme of BIG. The project will provide support for building and linking comprehensive datasets, for format conversion and annotation, and build repositories for the various types of data available from post-genomic clinical trials, for the predictive models and for the annotation and versioning of data and models. To support ease of use and efficiency for clinical research and reduce the distance between research and clinical practice, INTEGRATE also aims to build semantic interoperability with existing research and clinical infrastructures relevant in the context of the NeoBIG program.

  EURECA's coordinator Philips (Dr. Bucur) is also the coordinator of INTEGRATE. Several EURECA partners including BIG, IJB and FORTH are also partners in INTEGRATE.

- ***Eurocancercomms*** – An EU funded Coordination and Support project through the Science in Society Programme. One of the main objectives of Eurocancercomms is to establish an integrated model for a Europe-wide comprehensive cancer information and policy exchange portal. The cancer domain is a good model as its science is far advanced, subspecialisation within basic and clinical research has splintered the area into hundreds of boxes, and its impact on the health of Europeans is a significant issue. Such a model could subsequently be applied to other areas of healthcare, such as cardiology, neurobiology etc. This exchange portal will be hosted by eCancer[240].

- ***smartHEALTH*** – SmartHEALTH[241] has been an FP6 funded integrating project that focused on developing a new generation of intelligent lab-on-chip bio-diagnostic devices for point-of-care that incorporate advanced communication and software capabilities for context awareness, ubiquitous computing in the ambient health information environment, and for new point-of-care driven medical software applications and the provision of e-health services. The project focused on interoperability based on semantic Semantic web services standards. The experiences, knowhow and technologies developed in the project will provide significant reuse in EURECA.

- ***ARTEMIS*** – Another key European activity of relevance to EURECA has been the project ARTEMIS(A Semantic Web Service-based P2P Infrastructure for the Interoperability of Medical Information Systems, IST-1-002103-STP). The Artemis project provides the interoperability of medical information systems through semantically enriched Web services. Artemis project provides the interoperability of medical information systems through semantically enriched Web services, i.e. by defining the service action semantics through a "Service Functionality Ontology". In this way, semantic discovery of Web services are facilitated. In addition, the project examined the use of archetypes in providing semantic interoperability among healthcare systems. In Artemis project, the ebXML Registry semantic constructs were used for annotating, storing, discovering and retrieving archetypes. The work on these areas is complementary to the work planned in EURECA.

- PONTE aims at providing a platform following a Service Oriented Architecture (SOA) approach that will offer *automatic intelligent identification of patients eligible to participate within well-specified clinical trials for drug repositioning* with specific focus on mitigating patient safety risks, reducing clinical trial costs and improving clinical trial efficacy. Work towards this direction will involve the development of *advanced decision support mechanisms* based on *risk assessment* and *techno-economical* models which will be fed with information intelligently retrieved from an innovative semantic search engine operating on top of a novel data representation with excessive descriptive power linking data within drug and disease knowledge databases, clinical care and clinical research information systems by following and extending existing standards. While both EURECA and PONTE have a strong clinical trial recruitment component,

[240] http://www.ecancermedicalscience.com
[241] http://www.smarthealthip.com

the approach taken is quite different. PONTE focuses on the design of clinical trials. PONTE develops Clinical Decision Support technology for providing well balanced, representative lists of patients eligible to participate in clinical trials aiming at mitigating patient safety risks (with the current practice described as "…*poor trial design*…","…*a resulting non-representative patient sample recruited*…"). Therefore, the approach suggested actually changes the clinical trial design and eligibility criteria (and possibly needs more patients enrolled to statistically take co-morbidities into account). EURECA, on the other hand, is not aiming at changing the clinical trial design (and the associated eligibility criteria), but focuses on faster identification of patient eligibility and subsequent enrolment. Subsequently, EURECA's focus on patient safety goes beyond the scope of clinical trials (as serious adverse events may be missed during trials, due to the limited duration and sample, and become apparent in clinical care): Longitudinal EHR data is mined for the early detection of patient safety issues. In addition, in EURECA Clinical Decision Support is rooted in clinical practice. It enables healthcare professionals to extract, in the context of each patient's case, the relevant data out of the overwhelmingly large amounts of heterogeneous patient data and treatment information; providing the healthcare professional *all* the relevant information (but also *only* the relevant information) to make an informed clinical decision for the current patient case. PONT uses Clinical Decision Support only in the clinical research environment, optimizing the clinical trial protocol itself along with optimizing the group of patients enrolled in a clinical trial.

- The vision of the **GRANATUM** project is to bridge the information, knowledge and the collaboration gap among biomedical researchers in Europe ensuring that the biomedical scientific community has homogenized, integrated access to the globally available information and data resources needed to perform complex cancer chemoprevention experiments and conduct studies on large-scale datasets. GRANATUM has as main focus a social collaborative working space for the design and execution of in-silico models and experiments in cancer chemoprevention. It allows researchers to socially share, dynamically discover and fruitfully combine biomedical data, knowledge and experiences. Semantic interlinking is managed through a common ontological reference model. The focus of GRANATUM differs from EURECA. GRANATUM aims to make data accessible in a homogenized, integrated manner, where EURECA aims at providing semantic interoperability among EHR and clinical trials sytems (consistent with existing standards like HL7v3, CEN13606 and CDISC). In addition, EURECA will construct core datasets covering the specific clinical domains under investigation (where modularization ensures scalability). The two projects have different application domains. EURECA focuses on improving clinical care and research by increasing the synergy between clinical trial systems and clinical care systems (such as EHRs). GRANATUM focuses on improving biomedical research activities by providing a collaborative environment in which researcher can jointly design (and execute) models.

SemanticHealthNet (Network of Excellence in Semantic Interoperability) – EURECA will closely collaborate with this project. Through the EURECA Technical Board we will ensure the collaboration with the semantic interoperability NoE on technological development aspects and contribute to the common info-structure set up by this NoE; A budget of 40k Euro has been reserved for this contribution.

- **ALADIN-DTH** (ANR-08-TECS-001) –This project is a 3 years French funded project (Jan 2009-December 2011). It aims at the automatic detection of Hospital-acquired Infection episodes appearing in Hospital patient records. Xerox coordinates the project and is the partner responsible of the Natural Language Processing tasks of Information Extraction.

3.1.4. Assumptions and External Factors

In setting up the current proposal we have made everything possible to minimize risks from external factors and we have not built the EURECA Workprogramme on assumptions, such as that we will find clinical data once the project started.

The Consortium:

- Is aware of all major relevant initiatives in Europe and Worldwide and has already developed links with most of them
- Includes clinical sites that are reference centres in post-genomic clinical research
- Has made sure, at the proposal preparation level, that access to large clinical datasets – necessary amongst others for the modelling activities of the project – are available
- Has brought on board key experts from all required domains – SOAs, legal, security, etc
- The Consortium partners have worked – either bilaterally or collectively – together in the past with significant success. The harmonious cooperation of project partners, often a demanding task in large European project and an influencing success factor, is guaranteed.

- Has set up appropriate management structures and risk-mitigation processes, which will be further elaborated in the Consortium Agreement.
- Has access to significant amounts of clinical trial patient data to be used for prototype development.
- Is aware and has researched in previous work the legal, ethical, privacy and security requirements associated with clinical trials and handling of patient data.
- Has expertise with the relevant standards and clinical information systems
- Includes top clinical research and care sites with expertise in the clinical domains EURECA aims to model.
- Has set up during the project preparation links to external user groups in clinical research and in the bio-pharmaceutical industry, who have expressed interest to closely follow the progress of the project.

As a result we believe that there are no assumptions and we can see no external factors which could prevent the successful implementation of the project.

3.2. Plan for the use and dissemination of foreground

Knowledge Management will be accounted for within WP7 and WP10 of the project. It will include the management of knowledge going into the project, created within the project, and going out of the project, including Intellectual Property Rights (IPR) protection.
To manage knowledge in a structured way, a Knowledge Management Plan is prepared for handling the results of *EURECA* that are generated during the lifetime of the project. The Knowledge Management Plan focuses on the following related topics:

- **Management of knowledge going into the project.**

Good knowledge on the State-of-the-Art is essential for reaching a significant advance over the State-of-the-Art and therefore the import of results from outside the project into *EURECA* is continuously taken into account. In this way the progression and direction of *EURECA* will also be influenced by knowledge generated and shared from other projects.

- **Management of knowledge within the project.**

Since the interrelation between the WPs is important, the knowledge of all partners must be conveyed to the whole consortium. This knowledge exchange is implemented by the communication strategy of *EURECA* as described in Section B2.1.1.
Another very important aspect of managing knowledge within the project is of course the protection of Intellectual Property (IP). One of the main tasks of the coordinator will be to create an optimal climate in which all partners can generate and share the necessary new knowledge, i.e. Intellectual Property.

- **Management of knowledge going out of the project.**

Information exchange with the outside, expert community is considered to be essential, so we will disseminate the strategic and technical results of *EURECA* through publications in journals, presentations at conferences and fairs, contributions to workshops and so on (see also Section B3.2.1). Equally important is the concept of exploitation. We will create an exploitation plan for the commercial use of the knowledge of *EURECA* in such a way that it will generate new business opportunities (see also Section B3.2.3).
The following table gives an overview of the Knowledge Management Plan.

Table 3.2a – Knowledge Management Plan

Knowledge	**What**	**Responsible**	**Plan**
into EURECA	Investigation of State-of-the-Art	all, Advisory Board, Project Officers, Review auditors	Timely knowledge of the State-of-the-Art, keeping informed on progress in the field
within EURECA	Exchange of knowledge	Coordinator	Handbook Web based communication structure
	Creation and protection of knowledge	Coordinator, WP Knowledge Management – Task Intellectual Property	See Section 3.2.4.
out of EURECA	Dissemination	All, WP Knowledge Management – Task Dissemination	See Section 3.2.1.

Knowledge	What	Responsible	Plan
	Exploitation	WP Knowledge Management – Task Exploitation	See Section 3.2.2.
	Standardisation	WP Knowledge Management – Task Standardisation	See Section 3.2.3.

If circumstances induce so, the GA will decide on the installation of a (Dissemination, Exploitation, Standardisation or Intellectual Property) task force to support the already existing management structure.
In the next Sections the several forms of Knowledge Management will be detailed.

3.2.1. Dissemination

EURECA has strong dissemination plans, which include a serious investment in training. Rather than relying on "traditional" project specific web-site, we have decided to use a more established site for our dissemination activities, i.e. the eCancer comprehensive cancer information and policy exchange portal.
Cancer is a good model as its science is far advanced, subspecialisation within basic and clinical research has splintered the area into hundreds of boxes, and its impact on the health of Europeans is a significant issue. This exchange portal is hosted by eCancer[242] .
Dissemination is also strongly linked to the open process we have defined for the specification of the EURECA reference architecture. A series of workshops are planned to be organised in appropriate European or international events, with the objective of promoting the active involvement of external parties. The inclusion of external parties will allow the project to leverage the best-of-breed architectures and technologies, thereby enhancing the quality and applicability of the overall architecture.
eCancer, the main partner in charge of the dissemination of the EURECA results, is the journal of the IEO (European Institute of Oncology) in Milan. It is also the science and multimedia partner for ECCO, which represents 50,000 cancer scientists, clinicians and paramedical specialists throughout Europe. It is visited by 25,000 individuals each month, from 171 countries to date. Its international presence is therefore unparalleled. The editorial board consists of world experts in oncology, and is particularly strong in clinical trial expertise. ecancer aims to improve communications between sub-specialised cancer scientists and clinicians by working interactively and faster – offering a rapid peer review process as well as video, audio, and event listings. eCancer actively encourages the communities of sub-specialised scientists and cancer carers to exchange ideas and research, speeding up the time it takes from discovery, to patient benefit.
The journal is published by Cancer Intelligence Ltd in Bristol and is editorially independent of the European Institute of Oncology and ECCO. eCancer will act as the web address and reservoir for all papers, reports and publications from the EURECA project. eCancer currently hosts hundreds of oncology videos, thousands of hours of content since the launch of eCancer.tv (the video section of the site) at the beginning of 2009. eCancer's experience will allow the broadcast of videos relating to the project and testing of pilot television materials.
Dissemination of knowledge and results will be achieved by presenting the EURECA demonstrators at trade shows, fairs and clinical conferences, e.g., ECCO, ESMO, EBCC, etc.
In parallel to the above main dissemination channels foreseen, the project web site, printed dissemination material and a biannual electronic Newsletter will assist us in reaching out to all stakeholders, promote the project and its results and begin developing the community for exploitation of project concepts, products and services.

Table 3.2b – Conferences and Publications

	Name/Title	Partners involved
Conferences	International Conference on Machine Learning (ICML)	FhG
	European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML/ PKDD)	FhG
	IEEE International Conference on eScience	FhG
	IEEE Int'l Conference on Engineering Management and Service Sciences	FORTH
	IEEE Information Technology and Applications in Biomedicine (ITAB)	FORTH, Philips
	IEEE International Conference on Data Mining (ICDM)	FhG
	IEEE Engineering in Medicine and Biology Conference (EMBC)	FORTH

[242] http://www.ecancermedicalscience.com/

	Name/Title	Partners involved
	IEEE conferences on Knowledge Discovery in Databases (KDD)	FhG
	American Association of Cancer Research (AACR)	UOXF
	European Society of Medical Oncology (ESMO)	UOXF
	ASCO Conference	BIG, IJB, GBG
	EBCC Conference	BIG, IJB, GBG
	San Antonio Breast Cancer Symposium	BIG, IJB
	Connective Tissue Oncology Society	UOXF
	ESTRO, ASTRO, ICTR, ECCO	Maastro
	IEEE International Conference on Bioinformatics and Biomedicine	Philips
	Radiology Society of North America	Philips
	IEEE Symposium on Security & Privacy	Custodix
	Privacy Enhancing Technologies Symposium (PETS)	Custodix
	International HL7 Interoperability Conference	Philips
	Medical Informatics in Europe (MIE)	UPM, FORTH
	International Semantic Web Conference	VUA, SIT
	International Conference on the World Wide Web	VUA, SIT
	American Medical Informatics Association	Philips, UPM, Xerox
Publications	Journal on Machine Learning Research (JMLR)	FhG
	Machine Learning Journal (MLJ)	FhG
	Journal on Data Mining and Knowledge Discovery	FhG
	Lancet Oncology	UOXF, BIG, IJB
	Journal of Clinical Oncology	UOXF, Maastro
	Journal of ASCO	BIG, IJB, GBG
	Cancer Genomics and Proteomics	BIG, IJB
	European Journal of Cancer	UOXF, Maastro, GBG
	The New England Journal of Medicine	BIG, IJB
	Data Protection and Security Journal	LUH
	The International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC)	FORTH, Philips
	BMC Bioinformatics	FORTH, Philips
	IEEE Security and Privacy	Custodix
	International Journal of Information Security	Custodix
	International Journal of Data Mining and Bioinformatics	Philips
	Journal of the American Medical Informatics Association	Philips, Xerox
	Methods of Information in Medicine	UPM, FORTH
	Journal of Biomedical Informatics	UPM, FORTH, Xerox
	Semantic Web Journal	VUA, SIT
	IEEE Transactions on Information Technology in Biomedicine	Philips, FORTH

3.2.2. Exploitation

The work related to the exploitation of the project results will be carried out in WP10 Knowledge management. All EURECA partners will participate in the exploitation task.
Our main exploitation goal is the use of project outcomes by wide user communities, outside of the project consortium. During the project we will work to extend our user base, through different channels. We will also address these user groups to provide a prioritization of the EURECA infrastructure components, tools and services according to their expected benefits for each user group. An important goal of EURECA is to elaborate a realistic sustainability plan that would enable the EURECA infrastructure and tools to live far beyond the duration of the project and to provide benefits to a large number of users in the biomedical community.
Philips: Philips is committed to deliver healthcare solutions that enable continuous improvement in patient care, delivering better healthcare for all at lower costs. Unlocking the information “hidden” in electronic health records and making it semantically interoperable will enable deployment of state-of-the-art computerized

clinical decision support systems that deliver evidence-based clinical knowledge to the point of care, personalized to the clinical case at hand. In complex diseases such as cancer, access to relevant clinical research information e.g., validated results of clinical trials or appropriate trials open for enrolment, will become essential in future clinical decision support systems. Solutions like EURECA will help to detect and prevent avoidable health risks, improving patient safety as well as overall results of the healthcare delivery process. The improved semantic interoperability and adherence to the established standards in the identified areas is an essential prerequisite for successful commercial deployment of such clinical decision support systems. In the absence of a semantic interoperability solution, any deployment of a clinical decision support application may require significant effort with customization and interfacing to each hospital's EHR. We plan to exploit the EURECA results by transfers to appropriate business units in Philips Healthcare.

Custodix: Pharmaceutical companies are confronted with change. Research pipelines are getting thinner and the blockbuster model seems to be at its end. Hence they are dealing with a deteriorating financial situation and thus need to cut cost. Within the clinical trial area, several possible cost reducing measures are considered to be promising, these include:

- Secondary-use of medical data: the effort associated with collection of (new) data (entry, validation, etc.) is huge.
- Automated patient recruitment: it appears that for new trials, patient recruitment is a painstakingly manual process and thus costly (with people actually visiting physicians to ask them for patients who match the trial inclusion criteria).

Both topics are at the core of the project call (and EURECA proposal) and both can only be solved in practice when data privacy is appropriately addressed. This is right within the core business of Custodix: data protection within eHealth and more specifically disease management and clinical trials. What is more, Custodix is actually already involved in a (less technological advanced) small commercial pilot initiative which tries to address those two needs. Custodix sees a clear market need and thus aims to use the results of EURECA to (jointly with other partners) **exploit secure compliant services to facilitate data re-use and patient recruitment**.

Custodix considers **automated consent management** to be a facilitator for health data sharing, both in care environments and in research (secondary use). It is doubtful that consent management can be marketed as a stand-alone service, however it can be a unique selling point of a larger integrated service (e.g. patient recruitment, exploitation of Personal Health Record data, etc.).

FORTH: FORTH has for a long period focused its R&D activities on addressing the R&D challenges involved in the design of the architecture for the future integrated regional health information networks and the seamless delivery of integrated, novel eHealth services. The Laboratory has invested significant R&D efforts towards the design, development, deployment and support of HYGEIAnet, the Regional Health Information Network of Crete and the required Healthcare Information Infrastructure (HII), as a model for the development of such networks. The innovative and integrated technological solutions developed by the Biomedical Informatics Laboratory received the eHealth 2003 award at the eHealth 2003 Ministerial Conference and Exhibition. In the course of its activities the Laboratory has developed a family of innovative products (the Integrated Care Solutions - ICS) which are today in the process of commercialization and exploitation through our participation in a number of National competitive tenders for the development of Regional Health Information Networks for the Health and Social Care authorities in Greece. Specifically, our ICS solutions form the basis for the implementation of Integrated Health Information Systems for the Regional Health System of Western Greece, the 1st and 3rd Regional Health System of Attica, and the 2nd Regional Health System of the Aegean.
At the same time FORTH is a key technology contributor in the implementation of the open source Trial Design and Management System (ObTiMA) in the context of the ACGT project.
The services to be developed in EURECA will act as "bridges" between these two families of technologies, providing added value to both. As a result FORTH is committed to exploit the EURECA results.

UOXF: Integrative system to help in the analysis of relevant scenarios:

1. Ensuring dissemination of trial results to direct recommendation of optimal treatment of patients throughout Europe.
2. To generate enduring IT infrastructure for retrospective and prospective data mining of health care and biobank data sets.
3. Identifying prognostic biomarkers that result in trial eligibility and treatment planning to improve survival and late effects of treatment of bone sarcomas.
4. Identify breast cancer prognostic and predictive biomarkers in breast cancer that can be used for patient stratification and selection for trials and treatments

5. To automatically identify trials for individual patients with rare tumours throughout Europe.

LUH: The main exploitable result will be to build up a research platform where principles of privacy and security regarding medical data are going to be envisaged as a model for future research structures in this area. The platform will be a general help in building an awareness of the data protection rules among the researchers, which will help to improve patients data safety in *EURECA*. Patient data safety plays an important role in building patients' trust in the project and therefore will affect the overall success of the project.

MAASTRO: Main result is a validated data extraction system that can extract locally available medical data from cancer patients included in trials from multiple centers. Advantages are:
- More patients in trials
- creation of a company running clinical trials in the EU-region
- development of new IT tools that could be potentially commercialized
- development of a comprehensive clinical research platform in Europe which will attract companies to run highly efficient clinical trials

IJB: EURECA will Improve the semantic interoperability of the EHR and achieve a good level of performance in the exploitation of the data for the use in research trials (clinical, translational or imagery trials), such as:
- Help with the selection of suitable patients that can comply with the trial inclusion criteria
- Extract safety data to populate a pharmaco-vigilance database for academic trials sponsored by IJB in compliance with EU regulations, hence avoiding double coding and potential safety biases
- Extract normalized data to complete electronic CRFs automatically.

UPM: The results of EURECA are widely usable in the community where UPM carries out its research. In fact, the Biomedical Informatics Group is currently working in a national initiative, funded under the Avanza call, to develop technologies to improve intelligent treatments and patient care. To guarantee the reuse and interoperability of such type of systems, the standardization and mapping efforts that will be carried out in efforts like EURECA become crucial. We believe that the products of EURECA will act as a seed for future efforts in the way of achieving complete semantic interoperability and seamless data exchange in the clinical domain.

SIT: For years now, SIT has been active in the field of data processing and extraction of patterns from large sets of information. These topics have become more and more important, as data repositories keep on growing at a faster than linear rate. Furthermore, this evolution is seen in a multitude of different domains. To that end we deploy new technologies in the domains of data mining and semantic web, to facilitate services like personalization, data recommendation and data interoperability.
Although SIT's activities are already strongly focused in the direction of data extraction and processing, they always have been in the context of the multimedia domain. Moving our expertise and technologies to the medical domain can therefore be seen as an enrichment and diversification of our activities. We expect that this new domain will allow us to apply our knowledge and expertise to new and challenging problems. For the future, we foresee that several aspects in the medical world will undergo far reaching reforms where more and more systems and networks will integrate to improve clinical research and medical services for the patients.

3.2.3. Standardisation

All EURECA partners contributing to the technical work will participate in this task, which will be led by EuroRec. We will benefit this way of their vast expertise in EHR systems and relevant standards.

It is stated in deliverable D 7.1 - Semantic Interoperability Deployment and Research Roadmap of the *SemanticHEALTH* project[243] that:

"...full semantic interoperability (Level 3, as defined by SemanticHEALTH) is required across heterogeneous EHRs in order to gain the benefits of computerised support for reminders, alerts, decision support, workflow management and evidence-based health care i.e. to improve effectiveness and reduce clinical risk. However, it is recognised that achieving Level 3 across the entirety of healthcare would be a lengthy, expensive and possibly unattainable goal. It is instead recommended that Level 3 quality is sought in specific areas of

[243] http://www.semantichealth.org/deliverables.html

clinical practice that are known to be of high patient safety risk, and in priority areas for which the evidence is strongest for a gap to be bridged between current and best practice..."

In addition, the first *SemanticHEALTH* workshop in Copenhagen on 2006/09/30 identified a number of crucial issues to be resolved in the process towards semantic interoperability. Among these issues is the necessity of defining *specific use cases or scenarios on how to integrate eHealth systems in a particular context*. This means for the standards harmonisation process that it should start with clearly defined use cases.

This is exactly the strategy adopted by the EURECA project, i.e. to achieve level 3 SIOp in "*specific areas of clinical practice*" based on "*specific use cases or scenarios*". In EURECA we have chosen cancer to be the clinical domain of reference and we will develop our core sets of concepts on sub-domains in oncology, relevant to our clinical partners and the external clinical organizations and networks associated with the project (EORTC, EuroBoNeT and Metoxia).

Specifically, in EURECA our developments will include the definition of sound information models describing the clinical trial systems, building on existing research results[244] when possible. Electronic health records too need to be properly modelled; to that end we will adopt the appropriate state-of-the-art representation formalisms such as HL7 CDA, the openEHR Reference Model, ISO/EN 13606 etc.
On the clinical trial side adherence to CDISC will be observed. In our previous project ACGT[244], we have developed a Clinical Trial Management system named 'Ontology based Trial Management System of ACGT' (ObTiMA). In order to support seamless data transfer throughout the clinical trial process and interoperability with other clinical trial (sub)systems, ObTiMA is capable of exporting data in the CDISC ODM format. The CDISC ODM format has been extended with an ontological descriptions of items on the Case Report Forms (CRFs). This prior expertise with CDISC will be beneficial in EURECA.
The semantics of the clinical terms should be captured by standard terminology systems such as SNOMED CT, ICD, LOINC etc. The scalability of the solution needs to be achieved by modularization, e.g. instead of aiming at inclusion of the complete SNOMED terminology (more than 300 thousand concepts) ***we will identify a core subset that covers the chosen clinical domain***. Such core data set ***shall be validated*** both by clinical and knowledge engineering experts to assure proper coverage and soundness.
Utilizing the core data set we will devise a mapping system between the information models of CTS and those of EHRs. These mappings will too be ***verified by clinical experts and validated in properly chosen scenarios/usecases*** focusing on those with a high potential for patient safety improvement.

All the relevant work will be publically available through the project's dissemination channels, and they will also be fed towards the relevant standards development bodies.

3.2.4. Intellectual Property

A Consortium Agreement (CA), which will be based on the IPCA version 2, will be negotiated between all partners, settling among other things the internal organization of the consortium, reflecting what has been described about the project management structure of EURECA.
The CA also covers full rights and responsibilities of participants in respect of the confidentiality of any confidential information disclosed by the partners during the project, as well as the publication and communication of information during the project.
Any result generated before the effective date of the Consortium Agreement i.e. background shall remain with the respective party. Any result generated by a party after said date during and within the scope of the project, i.e. foreground, whether or not it qualifies for IPR protection, shall vest in such party. Any jointly generated foreground will be jointly owned. Throughout the execution of the project, all partners will continuously contribute to the identification of foreground that may qualify for IPR protection.
In case certain IP is identified to be essential for the future business opportunities of the involved partners, the necessary steps are taken to protect that IP. The patenting procedure will proceed along the regulations described in the Consortium Agreement.
The agreement will also provide in additional rules for dissemination to ensure smooth dissemination of the results. Settlements of internal disputes and of course Intellectual Property (IP) arrangements will be part of the Consortium Agreement as well. In principle resolution of disputes shall be obtained at the court having exclusive jurisdiction according to the law applicable. Other institutions for dispute resolution may be consulted provided the law applicable so permits.
The IP terms during the cooperation of EURECA will be based on royalty free terms and conditions. After the project, i.e. during exploitation, access rights to background shall require fair and reasonable compensation.

[244] http://www.eu-acgt.org

For the purpose of exploitation, royalty free conditions may apply to access rights to foreground if such terms may reasonably be expected. If not, fair and reasonable conditions may apply for access rights to foreground All Access Rights needed for the execution of the Project and for Use are granted on a non-exclusive basis, are worldwide and in principle do not contain the right to sub-license but do contain the right to have-made. The agreement will also provide in additional rules on the introduction as well as its notification of background that has been made available under controlled license terms e.g. so-called Open Source licenses.

4. Ethical issues (if applicable)

EURECA aims at achieving semantic interoperability among EHR and clinical trial systems. Such bi-directional linking of clinical care data and research data entails several legal and ethical implications. First, health data in general and particularly genetic data contain a large amount of very sensitive information about the person concerned. A person's genetic data provides information about his descent, ethnic origin, and, with a certain probability, also about future diseases and possibly about their healing chances and much more. Secondly, the core idea of EURECA is access to and linking of the large amount of patient data collected in EHR systems. This data is to be used for generating and testing new hypotheses as well as for cohort studies. The result will be an increase in the efficiency and effectiveness of clinical trials, allow faster patient enrolment of eligible patients in clinical trials, detect patient safety issues related to therapies and drugs that would not become manifest in a clinical trial either due to the limited sample or to the limited trial duration and enabling long term follow up of patients as well as a consistent and efficient reporting of adverse events in clinical trials.
At the same time, this involves the merging of health data collected for therapeutic or diagnostic purposes on the one hand and health data collected for research purposes on the other. *For ethical reasons this has to be addressed in the context of the patients' informed consent, particularly their right of withdrawal.* From a legal point of view, this variation in the purpose of data use is generally prohibited. Therefore, the EURECA Data Protection Framework will be of paramount importance. The Framework is to be set up within WP8 relying primarily on pseudonymisation as well as on the project's internal privacy policies, a security framework, binding contracts and the procurement of appropriate informed consent.
In detail:

Table – Ethical Issues

	YES	Page
Informed Consent		
Does the proposal involve children?	X	66
Does the proposal involve patients or persons not able to give consent?	X	66, 78
Does the proposal involve adult healthy volunteers?	X	66, 78
Does the proposal involve Human Genetic Material?		
Does the proposal involve Human biological samples?		
Does the proposal involve Human data collection?	X	66, 78
Research on Human embryo/foetus		
Does the proposal involve Human Embryos?		
Does the proposal involve Human Foetal Tissue / Cells?		
Does the proposal involve Human Embryonic Stem Cells?		
Privacy		
Does the proposal involve processing of genetic information or personal data (about health, sexual lifestyle, ethnicity, political opinion, religious or philosophical conviction)?	X	66
Does the proposal involve tracking the location or observation of people?		
Research on Animals		
Does the proposal involve research on animals?		
Are those animals transgenic small laboratory animals?		
Are those animals transgenic farm animals?		
Are those animals cloning farm animals?		
Are those animals non-human primates?		
Research Involving Developing Countries		
Use of local resources (genetic, animal, plant etc)		
Benefit to local community (capacity building ie access to healthcare, education etc)		
Dual Use		
Research having direct military application		
Research having the potential for terrorist abuse		
ICT Implants		
Does the proposal involve clinical trials of ICT implants?		
I CONFIRM THAT NONE OF THE ABOVE ISSUES APPLY TO MY PROPOSAL		

4.1. Patient informed consents

The patients' participation in *EURECA* will be voluntary and great care will be taken to procure appropriate and legally valid informed consent to the collection of, access to or linking of their health data. In particular, such research and procedures shall only be carried out with the prior, free, informed and express consent of the person concerned in accordance with all applicable international laws and ethics guidelines related to protection of personal data as well as internationally accepted rules on bioethics and human rights. Informed consent shall be procured for the purposes of this project even in cases where the law does not require such consent; for ethical reasons it is vital that each participant of this project is informed and is able to decide what is done with his or her data according to the principle that *autonomy needs consent*.
All decisions and/or interventions to be made shall be made with respect to the privacy of the persons concerned and the confidentiality of such personal data subject to applicable national and international data protection laws.
Data of patients coming from data sources already existing will be analysed in terms of the validity of existing consent. Such patients will be asked to consent again where appropriate. If refreshing consent is not possible, for reasons of the patient's death, a lack of contact details or otherwise, data will only be used when the existing consent is valid.

4.1.1. Right of withdrawal

The EURECA consortium acknowledges the international debate underpinning the importance of a transparent system for withdrawing consent in biomedical research. For this purpose, the patients and participants, having given their consent to the processing of their data, shall be able to withdraw such consent at any time and for any reason without any disadvantage or penalty on the same basis as proposed by other large-scale research undertakings.
EURECA will provide information to participants about the options for withdrawal such as:

- "No further contact": *EURECA* would no longer contact the participant directly, but would still have their permission to use information and samples provided previously and to obtain further information from their EHR.
- "No further access to EHR": *EURECA* would no longer contact the participant or obtain further information from their EHR in the future, but would still have their permission to use the information provided previously anonymously.
- "No further use": In addition to no longer contacting the participant or obtaining further information about them, any information or samples collected previously would no longer be available to researchers. The internal *EURECA* Data Protection Authority would only hold their information for archival audit purposes. The participant's signed consent and withdrawal would be kept as a record of their wishes. Such a withdrawal would prevent information about them from contributing to further analyses, but it would not be possible to remove their data from analyses that had already been done.

4.1.2. Patients not able to give consent

For patients unable to give consent, due to their age or mental disability, *EURECA* will provide information sheets and obtain consent from their legal representatives. For patients who are minors at the time of obtaining consent, agreements will be provided, informing them about their rights upon reaching majority. The *EURECA* consortium is fully aware of and acknowledges the ethical and legal difficulties of conducting research with patients unable to give consent and will reduce such research as much as possible and adhere to internationally accepted standards in relation to such research. Where in doubt, the *EURECA* consortium undertakes to refrain from including such participants in their research.

4.2. Processing of personal data

The *EURECA* project will deal with healthcare, hence highly sensitive, data. However, personal data processing is subject to numerous regulations. Indeed, such data is particularly sensitive and subsequently requires a higher level of protection. Furthermore, because of the therapeutic or scientific stakes involved, such data processing must be reliable enough to absolutely minimise the potential of medical errors or erroneous scientific results.
Any and all relevant legal sources (legislation, case law, studies, surveys prior to legislation) at Community and at international level shall be reviewed and examined thoroughly to identify the applicable policies and

rules to be adopted as well as existing legal gaps to be filled appropriately. The sources considered for the purposes of this exercise include, but are not limited to:

European level:

(1) Art. 3, 7, 8 of the Charter of Fundamental Rights of the European Union

(2) The Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data

(3) Directive 2001/20/EC of the European Parliament and of the Council of 4 April 2001 on the approximation of the laws, regulations and administrative provisions of the Member States relating to the implementation of good clinical practice in the conduct of clinical trials on medicinal products for human use

(4) Directive 2005/28/EC of 8 April 2005 laying down principles and detailed guidelines for good clinical practice as regards investigational medicinal products for human use, as well as the requirements for authorisation of the manufacturing or importation of such products

(5) Directive 98/79/EC of the European Parliament and of the Council of 27 October 1998 on in vitro diagnostic medical devices

(6) Art. 8 of the Convention of the Council No. 5 for the protection of human rights and fundamental freedoms

(7) Convention No. *1*08 of the Council of Europe for the protection of individuals with regard to automatic processing of personal data

Recommendations:

(1) Council of Europe, Recommendation No. R(97)5 on the protection of medical data adopted of 13 February 1997

(2) Council of Europe, Recommendation on human rights and biomedicine, concerning biomedical research, Strasbourg 25th of January 2005

Relevant International Instruments and Documents:

(1) World Medical Association Declaration of Helsinki

(2) *Convention No. 164 of the Council of Europe for the protection of human rights and dignity of the human being with regard to the application of biology and medicine (Convention on Human Rights and Biomedicine)*Additional Protocol to the Convention on human rights and biomedicine concerning biomedical research

(3) UNESCO Universal Declaration on Human Genome and Human Rights

(4) UNESCO International Declaration of Human Genetic Data

(5) UNESCO Declaration on Bioethics and Human Rights

Article 29 Data Protection Working Party:

(1) Working Document on Genetic Data

(2) Opinion 6/2000 on the Human Genome and Privacy

Other relevant documents:

(1) Opinion of the European Group on Ethics in science and new technologies to the European Commission, No. 11, 21 July 1998

(2) International Guidelines for biomedical research involving human subjects (prepared by the Council for International Organizations of Medical Sciences in collaboration with the World Health Organization)

4.2.1. EURECA Data Protection Authority/ Data Protection Framework

In case a problem arises with new legislation relating to health/genetic data collection, data access or patients rights, the *EURECA Data Protection Authority* (to be set up as part of WP8) will evaluate the situation and take appropriate action.

The legal and ethical issues directly related to the research performed by *EURECA* that will be given attention throughout include the following:

- Patient's prior, free, express and informed consent;
- Evaluation, analysis and renewal where appropriate of the informed consent of already existing patient data that will be made available;
- Procedures of withdrawal in case a patient wishes to quit at any time;
- Design and implementation of legally compliant anonymisation and pseudonymisation tools;

- Lawful process, transfer, transmission and storage of health data codified in the *EURECA* ethical and legal policies, such as the *EURECA* General Terms, contracts and information sheets;
- A feedback procedure to the patient where necessary and agreed on in the informed consent;
- The *EURECA* Data Protection Framework:
    - Anonymisation/double-pseudonymisation of health data;
    - Establishment of an *EURECA Data Protection Authority* as legal body
        - Conclusion of binding contracts with all research units with respect to the internal data protection and data security;
        - Authority to be addressed for necessary feedback to the patient;
        - Central contact point for patients to claim their rights;
    - Implementation of a Trusted Third Party as a Data Custodian.

4.2.2. Feedback procedure

The planned *EURECA* Data Protection Framework will include an internal Data Protection Authority. To simplify procedures for patients, this authority will be the central contact point for patients and participants with respect to rights resulting from the processing of their data within *EURECA*. In addition, in order to make unauthorised re-identification of a patient substantially more difficult, a Trusted Third Party will be integrated in the framework, acting as data custodian holding the key necessary for re-identification.
In case there is a need to give feedback to patients, and the patient concerned agreed to receiving feedback as part of the consent provided, the procedure to be followed will be:

- The *EURECA Data Protection Authority* is to be informed.
- It will initiate the re-identification procedure by sending the specific pseudonymised data set to the Trusted Third Party (TTP) for re-identification. The TTP is custodian of the key and is able to assign the data set to the treating physician or hospital, who is in the best position to assess and divulge the clinical data as well as decide on further action, if necessary.
- The treating physician or hospital thereafter gives feedback to the patient.

4.3. Practical management of EURECA legal and ethical issues / Ethics Advisory Board

The *EURECA* consortium acknowledge that many other ethical issues, that are not foreseeable at this moment, may arise as a result of the innovative design of the *EURECA* project. To ensure that at any point in time throughout the project, all ethical, legal, social and safety issues raised by any of the activities of *EURECA* are evaluated in a timely, accurate and careful fashion from the perspective of all stakeholders involved, and that appropriate solutions are found, the internal *EURECA Data Protection Authority* will be established from the very beginning of the project.
Secondly, the clinical beneficiaries' institutional ethics committees will be contacted and involved to provide the maximum available safety.
Finally, within WP 1 an external Ethics Advisory Board (EAB) will be formed from independent experts, leading academics in the relevant field, providing a consultative function. There will be three fixed EAB members, which will be invited to consortium meetings and be contacted for advice as and when needed.

4.4. Requirements emerging from the ethics review

The project will adhere to the following requirements emerging from the ethics review:

1. Prior to the commencement of each relevant WP, and where applicable, copies of ethical approvals/opinions/notifications by the competent legal local/national Ethics Bodies/administrations must be submitted to the European Commission and reported as a deliverable.
2. Detailed information must be provided on the procedures that will be used for the recruitment of participants (e.g. number of participants, inclusion/exclusion criteria, direct/indirect incentives for participation, the risks and benefits for the participants etc) and the nature of the material that will be collected (e.g. sensitive or personal data, etc). It must be explicitly stated if children or adults unable to give informed consent will be involved and, if so, full justification for their participation must be provided.
3. Prior to the start of the project, detailed information must be provided to the European Commission on the informed consent procedures that will be implemented. Copies of examples/templates of Informed Consent Forms and Information Sheets must be included. Applicants must also confirm that

they will provide material to potential participants in language understandable to them, and that proper translation of consent forms and information sheets will be made available to them.

4. The proposed Ethics Board must include independent, external members with relevant experience in ethics. A report by the Ethics Board must be submitted to the Commission with the Periodic Reports. The first report must include an in depth analysis of the ethical and legal framework of the project which can be used as a basis for an ethical follow up audit.
5. Applicants must provide a detailed description of security measures that will be implemented to prevent improper use, improper data disclosure scenarios and 'mission creep' (i.e.: unforeseen usage of data by any third party). It is required that the potential "unforeseen usage" implications of this project be examined and reported to the European Commission.
6. Applicants must provide detailed information on privacy/confidentiality and the procedures that will be implemented for data collection, storage, access, sharing policies especially when third party countries are concerned, protection, retention and destruction. Confirmation that third party countries will comply with national and EU legislation should also be included. A procedure must be implemented by the applicants in order to tackle all processed data that might fall under Article 8 of the Data Protection Directive.
7. In compliance with Directive 95/46/EC and with article 29 working group 8/2010 opinion, a data controller dedicated to the project must be designated. Applicants need to specify how they intend to manage potential conflicts of interest. More specifically, applicants need to document how they will guarantee autonomy towards the commercial third party behind Custodix/ Centre for Data Protection (CDP) if CDP acts as the data controller in the project. Based on the information in the proposal the 288048_EURECA 5/5 panel is of the view that in order to comply with the Directive, CDP cannot act both as an autonomous third party and the data controller in the project.

5. Gender aspects (optional)

Philips and the partners in the consortium are committed to a work environment in which all individuals are treated with respect and dignity. It is believed that each person has the right to work in a professional atmosphere that promotes equal employment opportunity and prohibits discriminatory practices, including harassment. Equal Employment Opportunity (EEO) and non-discrimination has been – and will continue to be – a fundamental principle within the consortium, where assignments and advancement are based upon personal capabilities and qualifications, without regard to race, colour, sex, language, religion, political or other opinion, national or social origin, property, birth or other status. The consortium recognizes the need to attract and retain talent, and that must encompass doing a better job of recruiting and developing women - traditionally less visible in the technology sector.
In view of the low percentage of women active in technical jobs, it is the consortium's policy to strive for women working in the project. The type of work is equally suited for women and men. In EURECA several women are already active in key positions.