ICT-2011-288048

EURECA

Enabling information re-Use by linking clinical Research and CAre

IP
Contract Nr: 288048

Deliverable: D1.1 User needs and specifications for the EURECA environment and software services

Due date of deliverable: (31-07-2012)
Actual submission date: (MM-DD-YYYY)

Start date of Project: 01 February 2012 Duration: 42 months

Responsible WP: UdS

Revision: <outline, **draft,** proposed, accepted>

Project co-funded by the European Commission within the Seventh Framework Programme (2007-2013)		
Dissemination level		
PU	Public	X
PP	Restricted to other programme participants (including the Commission Service	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (excluding the Commission Services)	

0 DOCUMENT INFO

0.1 Author

Author	Company	E-mail
Norbert Graf	UdS	graf@uks.eu
Danny Burke	ecancer	danny@ecancer.org
Cyril Krykwinski	IJB	cyril.krykwinski@bordet.be
Francesca Buffa	UOXF	Francesca.buffa@imm.ox.ac.uk
Gordon McVie	IEO	gordon.mcvie@ieo.it

0.2 Documents history

Document version #	Date	Change
V0.1	17.05.2012	Starting version, template
V0.2	17.06.2012	Definition of ToC
V0.3	13.07.2012	First complete draft
V0.4	24.07.2012	Integrated version (send to WP members)
V0.5	26.07.2012	Updated version (send PCP)
V0.6	27.07.2012	Updated version (send to project internal reviewers)
Sign off	30.07.2012	Signed off version (for approval to PMT members)
V1.0	31.07.2012	Approved Version to be submitted to EU

0.3 Document data

Keywords	User needs, requirements
Editor Address data	Name: Norbert Graf Partner: UdS Address: Saarland University, Campus Homburg Phone: 0049 6841 1628397 Fax: 0049 6841 1628302 E-mail: graf@uks.eu
Delivery date	

0.4 Distribution list

Date	Issue	E-mailer
13.07.2012	V0.3	al_eureca_all@natlab.research.philips.com
24.07.2012	V0.4	al_eureca_all@natlab.research.philips.com
29.07.2012	V0.6	al_eureca_all@natlab.research.philips.com
31.07.2012	V1.0	al_eureca_all@natlab.research.philips.com

Table of Contents

1 Introduction

EURECA aims to build IT solutions for better patient care so that the ultimate winners from EURECA will be the patients, the public and the healthcare services.

Despite improvements in healthcare IT infrastructures, a gap remains in the ability of these systems to deliver knowledge and insight back to the researchers, clinicians and patients they are intended to support. To close this gap software services needs to be built that will interconnect existing data systems, such as clinical trials and electronic health records (EHR). Semantic interoperability between these systems is of utmost importance.

Oncology has been selected as the focus environment for EURECA because of its incidence, the complexity of data collected, and diverse therapy options.

The solutions developed by EURECA will deliver several benefits for patients, including early detection of patient safety issues and more efficient recruitment of eligible patients to clinical trials. The systems will also enable long-term follow up of patients to establish outcomes such as levels of recurrence or late morbidity. Oncology research will benefit greatly from improved interoperability and the ability to reuse the vast amounts of data collected within care.

This deliverable is dealing with user needs and specifications for the EURECA environment and software services. It is of utmost importance that the EURECA environment and the tools to be developed is based on the needs of the stakeholders and in especially on those of the clinicians. This will guarantee that the environment will be used in daily clinical practice. The following methodology was used to achieve this goal:

- A user needs survey was set up
- Interviews of stakeholder were done
- Clinicians taking part in the project formulated scenarios

2 User Needs and requirements

The advanced, standards-based and scalable semantic integration environment enabling seamless, secure and consistent bi-directional linking of clinical research and clinical care systems has the following objectives as written in the DoW:

1. **Support more effective and efficient execution of clinical research** by:
    a. Allowing faster eligible patient identification and enrolment in clinical trials,
    b. Providing access – in a legally compliant and secure manner – to the large amounts of patient data collected in the EHR systems to be re-used in clinical research, for new hypotheses building and testing (e.g. to benefit rare diseases), study feasibility, as well as for epidemiology studies,
    c. Enabling long term follow up of patients, beyond the end of a clinical trial,
    d. Avoid the current need for multiple data entry in the various clinical care and research systems during the execution of a study.
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, and eliminate duplicate reporting (in care and research) of identified serious side effects,
3. **Allow for faster transfer of new research findings and guidelines** to the clinical setting (from bench-to-bedside),
4. **Enable the 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.

The definition of requirements on basis of scenarios is based on the approach from the European Commission's funded ESPRIT 21903 'CREWS' (Cooperative Requirements Engineering With Scenarios) long-term research project. We use a simplified version of the process[1] in order to extract requirements from scenarios. The requirements engineering process can be decomposed into three activities[2]:

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.

2.1 User needs survey

A survey was initiated immediately after the start of the project to retrieve opinions about the project in general and specifically to inform the choices behind the EURECA environment.

2.1.1 Description of the survey

The complete user needs survey is given in Appendix 1 of this document. The online questionnaire is available at: https://www.surveymonkey.com/s/PX55RKC. Altogether 43 questions were asked.

2.1.2 Summary of the results of questionnaire

The following section gives a summary of the results of the questionnaire. For more detail go to Appendix 2 of this deliverable:

- 33 participants from 7 countries and Somalia
- 100% linking of data is useful; 83% also for own research
- 97% know what an EHR is; 76% what PHR is
- 94% think the re-use of personal data is a good idea
  - 97% would share data if privacy issues are solved
- Translational research is enhanced with access to all data
  - 6/33 have already access to all data
  - 61% combine research and care data already
  - 61% have problems at work with not joining EHR and research data
  - 82% believe that new research will be possible having access to research and care data; 76% believe that access to much larger amounts of data will do so
- The research described is covered by the scenarios of D1.1
- Research data are queried in 31% in patient care
- EHR is queried in 57% in patient care
- 79% are aware of problems with EHR
- Unstructured data; no trust in data; lack of standards; no unique patient ID; legal issues; data security; no semantic interoperability; lack of tools
- Data security is not important for 7%
- 72% need additional data sources that are publicly available
- All tools described are covered by the scenarios
- Medical terminologies are used by 50% of participants
- 54% intend to use medical terminologies
- 22% use medical standard specifications for data or research
- 28% intend to use medical standard specifications
- 40 % use biomaterial from biobanks
- 54 % need to have access to clinical data for this research
- None of available tools is available via VPH toolkit
- Only 1 tool is open source
- Only in 2 cases data are available now

According to the questionnaire the following tools are available:

- - Netcord - Worldwide registry de cordon - BMDW (Bone Marrow Donor Worldwide) - Worldwide registry - NMDP (Normal Marrow Donor Program) - USA registry - MDPB (Marrow Donor Program Belgium)

- - SAGE Bionetworks - TRANSMART - BIOMART - Geoportal - caBIG (NCI) (e.g. caTissue)
- Snomed; FDB; HL7
- Altova Map Force (Extract Transform Load - ETL)
- MOFFITT Cancer Center (http://www.moffitt.org/)
- oracle; myrth
- a) Soarian QM http://tinyurl.com/ce7soc5 , b) eurocat environment (www.eurocat.info), c) openphacts (http://www.openphacts.org/) d)
- Part of IJB's internal EHR allows structuring of documents (including pathology and MDT reports, lab results and drug administration data) . The specialized tool SM2008 deals with Clinical Document under the HL7 CDA format
- Documentation of SNOMED
- - caBIG - Seer DataBase (for collecting clinical data in USA)
- not public tools as such, various research groups develop bespoke systems in specific settings

According to the questionnaire 9 participants can make the following data available:

- Case-notes, HICOM, filemaker, LIMS, EPR (Cerner); ARIA; excel
- Nephroblastoma datasets
- Inclusion of publicly available data, easy link and integration
- Our EHR, PACS, radiotherapy treatment information system, biobank
- All parts of IJB EHR, especially: - lab results - pathology and imaging reports - drug administration reports - chemo drug prescriptions - visit notes - discharge reports
- - MDT relative documents - SNOMED-CT code explorer
- CRFs (Clinical Data)
- Academic clinical trials
- EORTC Sarcoma trials EuroEwing trial of Ewing sarcoma long term data retrieval

2.1.3 Recommendations based on the survey

All the information provided by the questionnaire is used for scenarios that are clinically relevant. There are already some tools available. Only one of these tools is open source and none can be retrieved via the VPH toolkit. Available data sources can be used if legal issues are solved.

Therefore the following three recommendations are given:

1. The described clinical scenarios are important as they cover the needs of the participants of the survey.
2. Tools need to be developed build out of use cases as open source tools and stored in the VPH toolkit.
3. Legal issues need to be solved to share data if the data producer wants to share and not only use EURECA tools to work on his own data. Some of the answer highlighted that data should not leave the centre, thus distributed data analysis/mining should be considered.

2.2 User needs interviews

Structured interviews with clinicians as the most relevant stakeholders were carried out by IEO. Altogether 53 stakeholders were interviewed. They are all from Italy. 38 of them were clinicians, including 1 nurse. Six were basic researchers, two pharmacists, one an IT person, three were patients and three with unknown affiliation.

2.2.1 Structure of the interviews

The outline of the structured interview is given in Appendix 3 of this document.

2.2.2 Results of the interviews

The following section displays the results of the different questions. Some stakeholders did give several answers to a single question. Similar answers of different stakeholders were put together to one topic.

Are you aware of any problems with electronic health records (EHR)?

1 not allowed using
9 never used it
10 no
33 yes

Which problems are you facing in your research in relation to EHR?

13 slow, lot of error messages, redundancy of possibilities
9 crush of the system, bugs
8 lack of support to use, understandability of the EHR
7 privacy of data
7 reliability of data, verification of data, completeness of data
3 interoperability, compatibility, communication with other hospitals
3 costs
2 need of a computer at bedside
1 missing availability in other hospitals
1 shortage of storage space

What are your specific needs in this area?

10 No specific needs
25 EHR needs to be easy and fast to use (usability, efficacy, continuously working)
12 To get the most complete information of a patient with follow-up data, including images, avoid duplication of data
4 Nursery Folders, new folders (e.g. for intra and post surgical management, patient management, for specialists)
3 To have always and everywhere the possibility of access
3 Targeted consultation and security
3 Training, order and organization
1 Automatic checks during input according to standards
1 Uniformity, traceability of the compiler
1 Report of personal comments of the patient
1 Less paper
1 Portable devices

- **Which of them are the most relevant needs?**
  16 Usability (e.g. speed, clarity)

10	To get the most complete information of a patient
4	Update of data in real time everywhere, no restrictions
3	Safety and security
2	Folders for observation of patients during surgery, on ward, etc.
2	Uniformity, traceability of the compiler
2	Communication, interoperability issues
2	Accessibility
1	Order
1	Clear understanding

What are general needs for you in this area?

8	None
15	Usability (e.g. speed), functionality
9	To get the most complete and correct information of a patient
4	Possibility to access the EHR from outside the hospital, consultations from everywhere at any time
2	To have only one link between doctor and nurse, redundancy
2	Update of data in real time everywhere, no restrictions
2	Summary of the patient's history
2	Availability of the system
1	Good management
1	Communication between EHR and other software (e.g. for laboratory and radiology)
1	Automatic tracking of the compiler
1	Clear understanding of content
1	Minimize bureaucratic workload
1	Training

What needs to be changed?

1	Nothing
8	Better server, faster informatics platform
7	Usability, make it simpler, less redundancy
6	Do not know
4	Restricted access only to appropriate personal, controlled by central body (1)
3	To change an electronic folder in a simple way
3	Needs to be used by all stakeholders including patients
2	Management system up to date
1	Support by informaticians
1	Management of nursery documentation
1	Number of passwords
1	Continuous update

Can you describe tools that you would like to work with in relation to EHR?

24	iPad, other tablets, PDAs, smartphones, scanners or new computers (e.g. for surgery with apps)
6	Portals in hospitals and outside (general practitioners) with limited access
4	Wireless LAN availability
2	A more immediate level of information sharing (clinical and biological data)
1	Print program that prints forms filled automatically with the data of a patient
1	Forms for consultations of specialists
1	Pharmaceutical handbook giving information about drug interactions
1	Computerized patient and therapeutic administration to nursing staff

1 Resources management system
1 A centralized body that ensures no loss of data
1 Training of staff
1 Quality evaluation of EHR systems

- **Which requirements are necessary for such tools?**
  16 Usability (e.g. speed), intuitive system, also for patients
  6 To use everywhere for all users, interregional network, Wi-fi
  4 Training
  3 Enough money
  2 Strong computer system up to date
  1 Roles and rights management system
  1 Software up to date
  1 Informed consent from patients
  1 data quality, data standards, guidelines
  1 clinical records

Can you suggest other people we should contact?
5 None
8 Hospital staff including nurses
1 IT people
1 Institute BESTA
1 'Ordine dei Medici di Milano'
1 Parents and associations of patients

2.2.3 Recommendations based on the interviews

Not all of the interviewed persons (20/53) are aware of problems with EHR. Those who are aware of problems report mainly on bad IT infrastructures, that are to slow, have a lot of bugs and do crash from time to time. Usability issues need to be taken very seriously. Access to EHR should be possible by all stakeholders with intuitive devices from everywhere taking security and safety issues into account. Data stored in EHRs needs to be complete and regularly updated in real time. Training for people using EHR is needed.

The interviewed people suggest the following tools:

- Portals in hospitals and outside (general practitioners) with limited access
- Tool for immediate information sharing (clinical and biological data)
- Print program that prints forms filled automatically with the data of a patient
- Building of forms for e.g. consultations
- Pharmaceutical handbook giving information about drug interactions
- Patient administration

All of these tools can serve as granular use cases for scenarios described in chapter 3.

3 Specifications for the EURECA environment

3.1 Scenario based environment

3.1.1 General aspects

All scenarios that are developed within EURECA can be categorized into three groups:

- Knowledge discovery
- Data curation
- Basic research and clinical trials support

The relevant scenarios of each group are listed here:

- **Knowledge discovery**
  - Selection of trials for patient enrolment
  - Trial / protocol feasibility
  - Selection and inclusion of patients into trials
  - Detection and prediction of SAEs / SUSARs
  - Pharmacovigilance – Automatic reporting of SAEs and SUSARs
  - Early detection of cancer / individual risk / prevention of cancer
  - Personal medical information recommender
  - Develop or update of guidelines from clinical trial data and literature mining
  - Data mining of consultations
  - Analyse economic data between different procedures / approaches
- **Data curation**
  - Long term follow-up
  - Patient diary (connection between PHR and data management tools)
- **Basic research and clinical trials support**
  - Supporting design of new trials and hypothesis generation
  - Clinical data reuse
  - Opt-out solution for new research
  - Simulation of datasets to combine
  - Rapid learning

See also section 3.1.3 “Summary table of specific scenarios”. Within these groups similar scenarios are put together. The difference between these scenarios might be:

- Scenarios for different stakeholders
- Scenarios using different data sources
- Scenarios being more detailed or sophisticated
- Scenarios having different endpoints
- Scenarios producing different outputs

In EURECA all these scenarios will be dissected into granular use cases by IT people allowing a modular structure of tools that will be developed for these scenarios. The following figure (fig. 3.1) gives a schematic overview of how a part of scenarios can be described. There are only five categories of data sources available:

- Clinical trial data (CT) / hospital information system (HIS) / EHR / PHR
- Laboratory Research data
- Literature
- Public database
- Questionnaires within scenarios

In all scenarios belonging to the group of knowledge discovery patient data are selected from the first group of databases. These specified data will be used for data mining in public databases and in literature. As a result a table with matched data is created. Results of scenario specific questionnaires, produced by clinicians during the development of the specific tool, as selection criteria can further narrow data in this table. The same is achieved by using quality criteria as a next step. Each scenario can define such quality criteria. The results of such a scenario can be shown in different output formats including visualization methods. Depending on the scenario different databases are selected. For data curation no literature or public database is needed and no questionnaires will be done. This results in simplifying the whole scenario (figure 3.2). Supporting basic research and clinical trials the usage of open source databases, literature and CT / HIS / EHR / PHR data is essential. But Questionnaires are not needed (fig. 3.3).

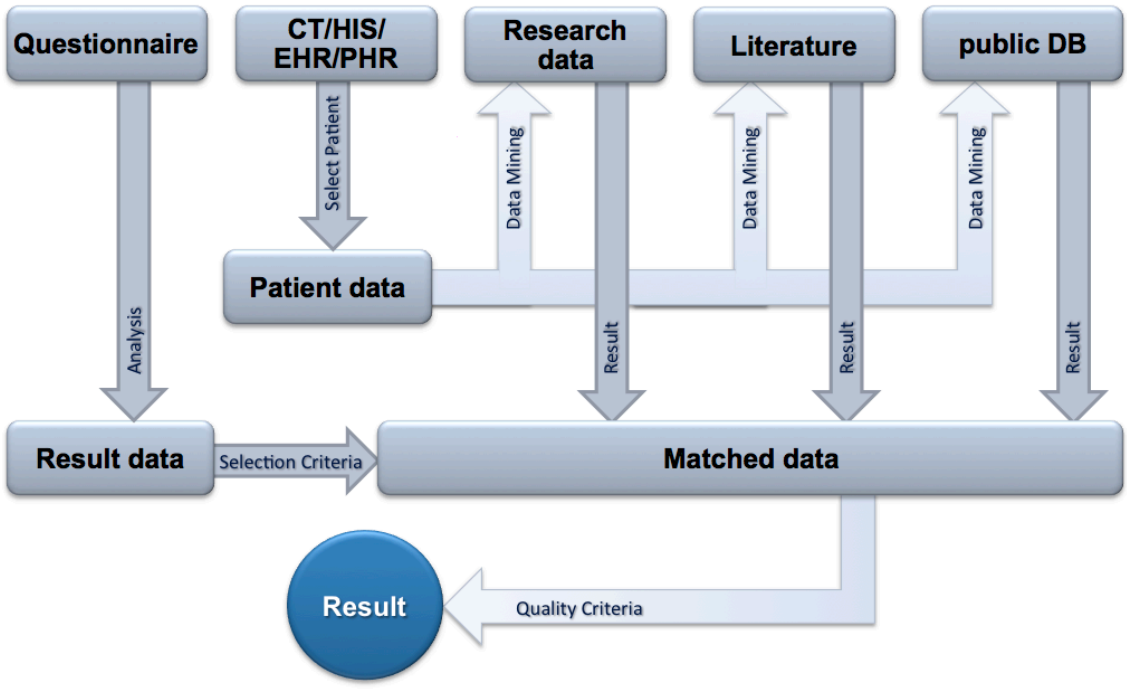


Fig. 3.1: General outline of a part of scenarios for knowledge discovery.

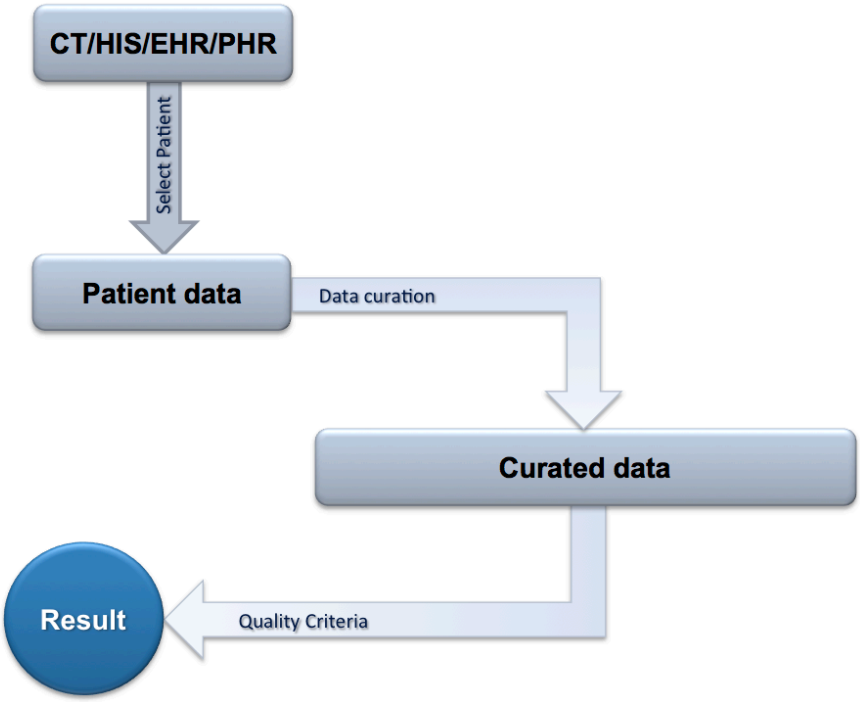


Fig. 3.2: General outline of scenarios for data curation.

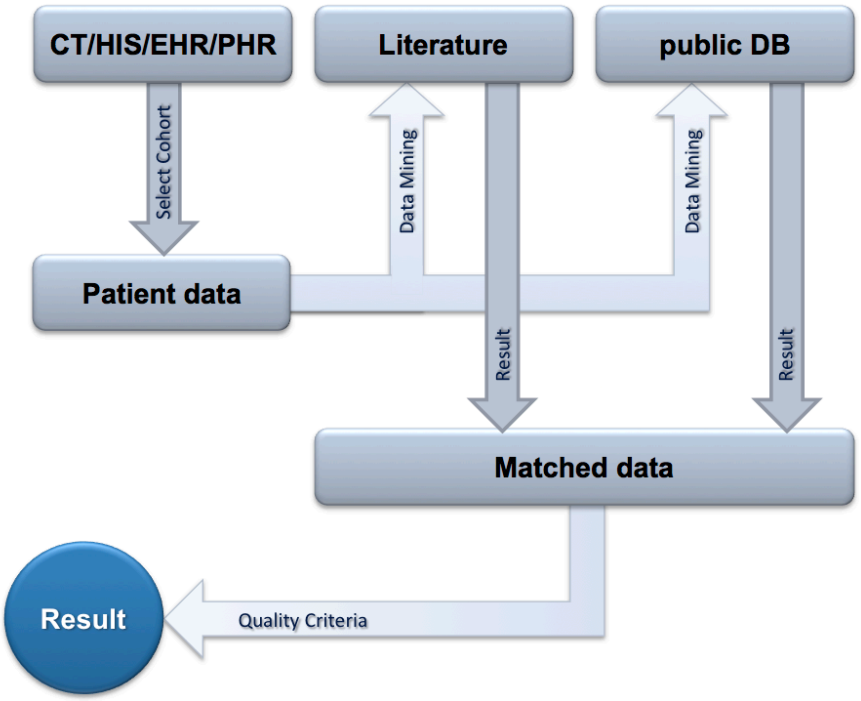


Fig. 3.3: General outline of a part of scenarios for basic research and CT support.

3.1.2 Summary table of specific scenarios

Group	ID	Name	Description
Knowledge discovery	KD1	Suggest clinical trials for a patient	Suggest a list of eligible clinical trials for a patient (prospective), or a list of eligible patients for a trial (retrospective)
	KD2	Clinical trial selection tool	When a patient is first seen at MAASTRO we note features of this patient in our EHR. We also receive images and letters from external hospitals for this patient. We would like to be notified if a new patient is suitable for a certain trial. The notification should be sent a) To our trial nurse / physician assistant / data manager so that they can inform the patient on the possibility of entering a trial and/or b) The physician (perhaps via a notification in the EHR)
	KD3	Ranking clinical trials	Clinicians will be presented with a suggested ranking of available clinical trials suitable for a particular patient based on both eligibility criteria and the characteristics of patients' disease
	KD4	Trial-enrolment advice to clinicians attending the MDT, based on clinical trials protocol eligibility criteria	Clinicians, while consulting the patient EHR during the multi-disciplinary team (MDT) meeting, are alerted about potential trials in which to enrol the patient, based on the matching between clinical trials protocol eligibility criteria and patients EHR data
	KD5	Trial-enrolment advice based on clinical guidelines	A clinician, while using a computerised representation of a guideline to treat a patient, is alerted to potential trials in which to enrol the patient, based on information that is gained from the values of relevant decision points in the guideline.

	KD6	Outcome prediction model tool/Decision Support System	During patient contacts, the physician and patient require an estimate of the outcome for treatment a,b,c,… so that they can reach a decision which treatment fits the patient wishes best. The use case is to provide a tool that uses existing, validated outcome prediction models and allows this kind of decision support. The outcomes that are predicted are cancer specific, but generally local control, distant disease, survival, quality of life, cost, toxicities.
	KD7	Trial / Protocol feasibility	EHR instances are queried with criteria in order to estimate the recruitment potential. For legal reasons, processing can only be done at the hospital sites and only aggregated data can be returned. Protocol feasibility could also be based on private/public data sources such as population information, other protocols, …
	KD8	Select patients for a trial	Find patients for clinical trial. The initiator is the pharmaceutical company. This scenario can also be used as an automatic discovery as a decision support system in the hospital using the recruitment matcher technology.
	KD9	Detection and prediction of SAEs and SUSARs	Data from EHR of a specific patient and data from clinical trials, literature and public databases are used to detect possible SAEs and SUSARs for the specific patients in advance before the drug is given to the patient. This will increase the safety of patients.

	KD10	**"GGO":** Using DNA sequencing in Oncology (for clinical trials & daily health care) to identify patients that have a non-synonymous mutation in a gene that is related to drug or radiation sensitivity	When a young patient with oligo-metastasis (less then 5 metastases, but this just an example) is first seen at MAASTRO we note features of this patient in our Electronic Health Record (EHR). We also receive imaging data and letters from external hospitals for this patient. We would like to maximize this patients' survival and therefore want to address whether a new patient is suitable for certain targeted therapeutic agents (often quite expensive) known to be efficient *only* if certain mutations are present (see table underneath). The drug would be given together with curative stereotactic radiation. If a tumour (but not the normal tissues) has a mutation in a gene that increase radiation sensitivity (e.g. ATM) then an extensive use of radiation at moderate dose but on large fields is also an option. The best way to do so will be through DNA sequencing. This will become more and more rapid and cheaper. We call this GGRT: "Genomic Guided Radiotherapy" or in more general terms GGO "Genomic Guided Oncology"
	KD11	Pharmacovigilance – Reporting SAEs and SUSARs automatically	Data from EHR/PHR/HIS are automatically analysed to find SAEs and SUSARs in patients to report them automatically to regulatory bodies. This includes the identification of episodes of febrile neutropenia and other SAEs from other organ systems.
	KD12	Prevention of medical conditions	Platform where people can obtain support and statistics for conditions they do not have, but for which they might be sensitive. Inclusion of social data can potentially alert users for certain habits (like eating, drinking, smoking, etc.) exacerbating or accelerating those conditions.
	KD13	Early detection of cancer / individual risk / prevention	Data from EHR of a specific patient and data from clinical trials, literature and public databases are used to detect cancer early or to define the individual risk to cancer. This will give a chance of early treatment or even prevention of cancer.
	KD16	Personal medical information recommender	Platform where people can obtain objective information (about trials, treatments, partners in misfortune, etc.) for their specific condition. A PHR can be loaded or provided. The system further allows for the provisioning of objective information and personal help with medical treatments (e.g. pros and cons)

	KD17	Extract Patient data from EHR and PHR	If a patient needs to go to a new physician, he could use this tool to get the relevant information about the patient without going through all his charts. If a patient wants to get a summarize of his data this might be done with this tool as well
	KD18	Guideline development	Data from clinical trials, literature and public databases are used to develop new guidelines for specific diseases or conditions.
	KD19	Rare case literature search tool	Sometimes patient with rare diseases are seen, for which no evidence based guidelines exist. One would like to retrieve from literature information for such a specific case. The use case is to select retrieve a list of relevant literature for a specific case for which we have noted information in the EHR
	KD20	Rare case experience search tool	Sometimes patient with rare diseases are seen, for which no evidence based guidelines exist. One would like to know which centre has experience with treating these patients. The use case is to retrieve a list of colleagues/hospitals that have experience in treating a rare case for which we have noted information in the EHR.
	KD21	Data mining on consultation data	Develop frequently asked questions and to contextualize this info for a particular patient/clinician, identify relevant information for similar patients, etc.
	KD22	Combining data sources for information discovery	A clinician, while looking at a patient record (EHR), gets **pointers** to documents from a wide variety of sources: records of patients with similar conditions, trials in which to enrol the patient or whose outcome might be relevant to the patient, reports of drug side effects, literature, etc. The pointers consist of recommendations to the most relevant documents of each source.

	KD23	Combining data sources for an overview of available information	A clinician, while looking at a patient record (EHR), is presented with an **overview** of the availability of relevant documents from a wide variety of sources. The overview consists of **summaries** of categories of relevant documents. For each category we see **in what way it is related** to the current patient case, and how many documents there are. Examples could be: -"10 records of patients with same disease and age" -"2 trials with matching eligibility criteria" -"50 trials with matching criteria except age" This gives the clinician an overview of what's available. He/she can quickly assess the value of each result category, and select them to see more results.
	KD26	Analyse economic data between different procedures	By joining data from EHR, clinical trials, literature and open databases economic aspects of different procedures (diagnostic and/or therapeutic) can be analysed in respect to outcome and quality of life in an individual patient. This will include data about days to stay in the hospital, expected side effects, costs of diagnostics and therapeutics, etc.
	KD28	Identification of episodes of febrile neutropenia	Detect an episode of febrile neutropenia (chemotherapy treatment side effect) by extracting relevant information
	KD29	Guideline protocol selection	Patients for radiotherapy are usually first seen by our physicians in external hospitals and/or during multi-disciplinary board. Then the patient comes to MAASTRO for the radiotherapy intake. We note features of this patient in our EHR. We also receive images and letters from external hospitals for this patient
	KD30	Clinical Trials finder and patient matcher	Once Sarcoma diagnosis is made, a patient will fall in at least 1/60 subcategories each of which may be the basis of a different treatment and entry into a different clinical trial

Data curation	DC1	Long-term-follow-up of patients from clinical trials by linking PHR data to clinical trial management systems	The service/tool shall collect relevant health data recorded by patients in PHR solutions and link them as long-term follow-up information to corresponding clinical trials in which patients were included. Trial chairman define in one or more specific CRFs which health related information that patients can report in PHR is of interest for his clinical trial. The EURECA service shall then enter the PHR data from the corresponding PHR system into the special "PHR-CRFs" of the clinical trial management system. (The PHR-CRFs will most likely match patient questionnaires.) This mechanism can in particularly be exploited to collect relevant long-term follow-up information from the patient case after the official end of the trial. In this way PHR data can be leveraged for clinical research and long-term follow-up of patients in clinical trials can be improved.
	DC2	Survival follow-up	Extraction of the last follow-up date and patient status (alive or deceased) from Site EHR or the National registry via the site, when the EHR last date of contact is > 2 years old. The date of last contact and the patient's status will be automatically included in the eCRF.
	DC3	Primary and secondary outcome measures	Extraction of primary and key secondary outcome measures from EHR and import in eCRF for a specific protocol. In particular: • Date of first recurrence in breast cancer patients that received an adjuvant treatment and site of recurrence (SNOMED term). • Left ventricular ejection fraction (LVEF) value (%), method used (echo, MUGA or cardiac MRI) and date of assessment
	DC4	Safety reporting of specific adverse reactions after study treatment completion	Extraction of safety data from EHR and import in eCRF to improve safety reporting of adverse reaction with late onset post study treatment completion.

	DC5	Patient diary	Patient will have access to his PHR or to specific eCRFs to enter data by him/her. This scenario is important to curate data beyond the end of clinical trials. There should also be a sharing of data with cancer registries to get dates and reasons of death. Usage of mobile devices might be considered to automatically store such data in ObTiMA or the PHR. This module might also be a module for DC1 (Long-term-follow-up of patients from clinical trials)
	DC6	Access and integrate information from primary care and other clinical databases in patients undergoing clinical trial in Sarcoma	Once patients are involved in clinical trials the ECRF only captures clinical trial related information. Access to other electronic databases or clinical data including nursing functional and emotional data is needed for complete follow-up.
Basic research and clinical trials support	CT1	Supporting design of trial conditions	A trial chair, while designing a new trial, is supported in the design of the eligibility conditions (inclusion and exclusion) based on background information: 1. design of previous trials 2. results of previous trials 3. SUSAR reports from previous trials 4. previous epidemiological (retrospective) studies. 5. published literature
	CT2	Hypothesis generation	Develop new research questions for future clinical trials by analysing clinical trial data and data mining of literature
	CT3	Clinical data re-use	During routine patient care, a lot of information is recorded for patients in local IT systems that also need to be recorded for the trial the patient participates in. The use case is to re-use the clinical data into the trial eCRF systems. This is to avoid double data entry.

	CT4	Identify incident tumours	Identify, in the textual data of the Electronic Medical Record (EMR), all incident tumours for a given period of time
	CT5	Identify first recurrent tumours	Identify, in the textual data of the Electronic Medical Record (EMR), the patients who have a first recurrence on a given period of time
	CT6	Extract and structure textual/structured EMR data	Extract and structure relevant information of a patient tumour from structured/unstructured textual data in the Electronic Medical Record (EMR)
	BR1	Opt-out solution for further research	A researcher having a specific question has the possibility to analyse anonymously EHRs to detect patients that may help to answer his research question with their data. If such a patient is detected he will automatically notified by an email that a researcher wants to use his data for a specific research. The type of research is described in a way the patient understands on a specific website, to which the patient will be linked. The patient can disagree at any time to participate with his data in this research project. Such a scenario is based on the fact that every patients agrees to share his data to any research project and that he can disagree to specific research projects at any time by using the above described website. The same scenario will be possible for research on biomaterial as well.
	BR2	Similarity of datasets to combine	Detection and identification of similar datasets (like Amazon; users who've bought this, also bought...)

	BR3	Rapid learning tool	In rapid learning research we want to learn and validate outcome prediction models from routine patient care data. We need to have access to large amounts (10.000+) of patients pref. with clinical, imaging, biology information
	BR4	Diagnostic sarcoma classifier	Diagnosis of Sarcoma is problematic and prone to misdiagnosis. Diagnosis is made using different data sources that can include genomic, pathology and imaging data depending on the clinical centres and available facilities.

KD: knowledge discovery; DC: data curation; CT: clinical trial; BR: basic research

3.1.3 Specific scenarios

In the following section different specific scenarios are explained

3.1.3.1 Knowledge discovery

Selection of trials for patient enrolment

This scenario describes how a trial can be selected for a specific patient. The goal is to find the optimal trial that fits the needs of the patient the best.

The following steps are needed (see figure 3.4):

1. Select a patient from the EHR, PHR or HIS and retrieve all data that are important for the selection of a trial, like age, gender, diagnosis, histology, stage of disease, tumour volume, primary diagnosis or relapse, comorbidity, and others.
2. Select from these data those that should be used for data mining of trial databases and the literature (Medline, etc.)
3. Select the trial databases (local, national, European, worldwide)
4. Start data mining
5. Rank the found trials according to best fitting
6. Check the so ranked trials according to quality criteria to erase those trials with poor quality. The quality criteria need to be defined (e.g. poor recruitment, question of the trial is already answered by another trial (found by literature mining, …)
7. Let the patient answer a questionnaire online that defines selection criteria from the perspective of the patent and his psychological profile (needs to be developed)
8. Use the results of the questionnaire to further rank the retrieved trials according to the specific needs of the patient.
9. Assign the best trial to the patient
10. Give a list of hospitals that are registered for this trial
11. Print a summary of the best trials with details for the physician to explain the patient what the best treatment would be.
12. Print a summary of the best trials in a language a patient understands for handling such an excerpt to the patient.

The same scenario can be used within guidelines. If a physician is selecting the guideline for the specific disease of the patient, the guideline will forward him to this tool for the selection of the best clinical trial. Or a clinician, while using a computerised representation of a guideline to treat a patient, is alerted to potential trials in which to enrol the patient, based on information that is gained from the values of relevant decision points in the guideline.

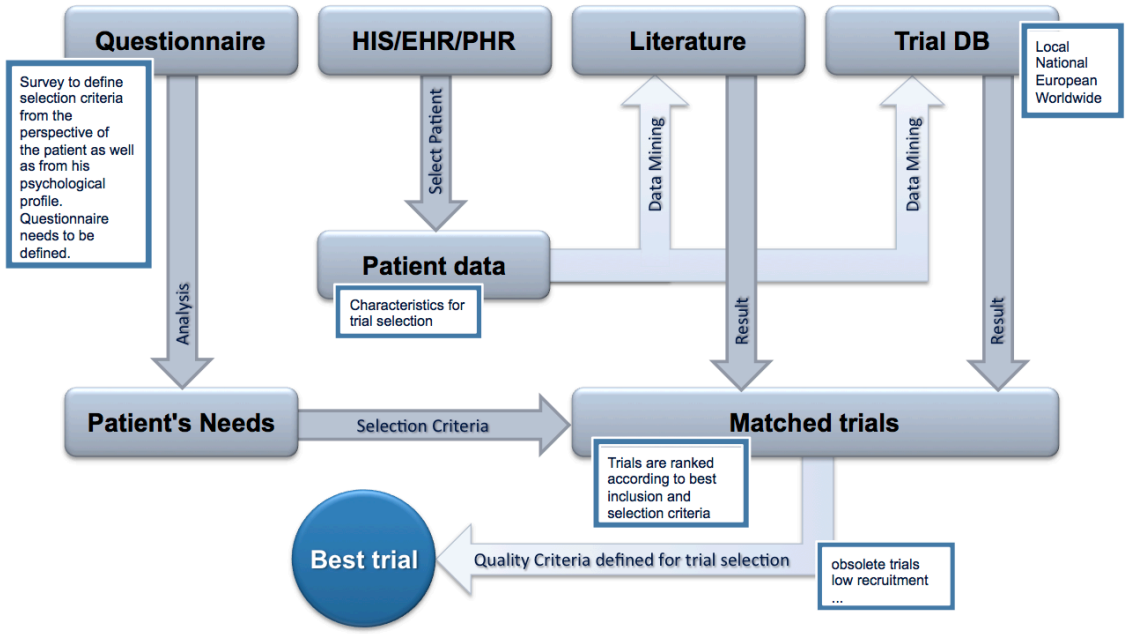


Fig. 3.4: Outline of the scenario for the selection of the best trial for a patient.

In addition to the best trial found the outcome prediction model (KD6) can be used to predict the outcome in respect to local control, distant metastasis, quality of life, acute and late toxicity and cost of treatment.

This is an example how different tools can be easily combined to get additional information and knowledge and underlines the need for building tools in a modular way and taking interoperability issues very serious.

Trial / Protocol feasibility

This scenario describes if a new clinical trial is feasible to start according to the estimation of recruitment potential. Two versions of this scenario are possible:

1. Based on EHR/PHR/HIS data
2. Based on other data sources

For legal reasons the first version can only be done at the hospital site.

This scenario should always be used before the scenario for the section and inclusion of patients into trials. The following steps are needed (see figure 3.5):

1. Patient data from HIS/EHR/PHR are exported and anonymized (data warehouse)
2. Hospital can do data mining to select the cohort of patients that fits recruitment criteria best
3. The number of aggregated data gives the answer if the protocol or the trial is feasible to develop

It might be possible for the hospital to query directly HIS/EHR/PHR instances with criteria to estimate the recruitment potential (not shown in fig. 3.5).

In the second version of this scenario the protocol / trial feasibility can be based on private or public data sets such as population information, other protocols or literature. This is shown in the green part of figure 3.5.

Both versions of the scenario can also be combined if a multicentre trial will be set up.

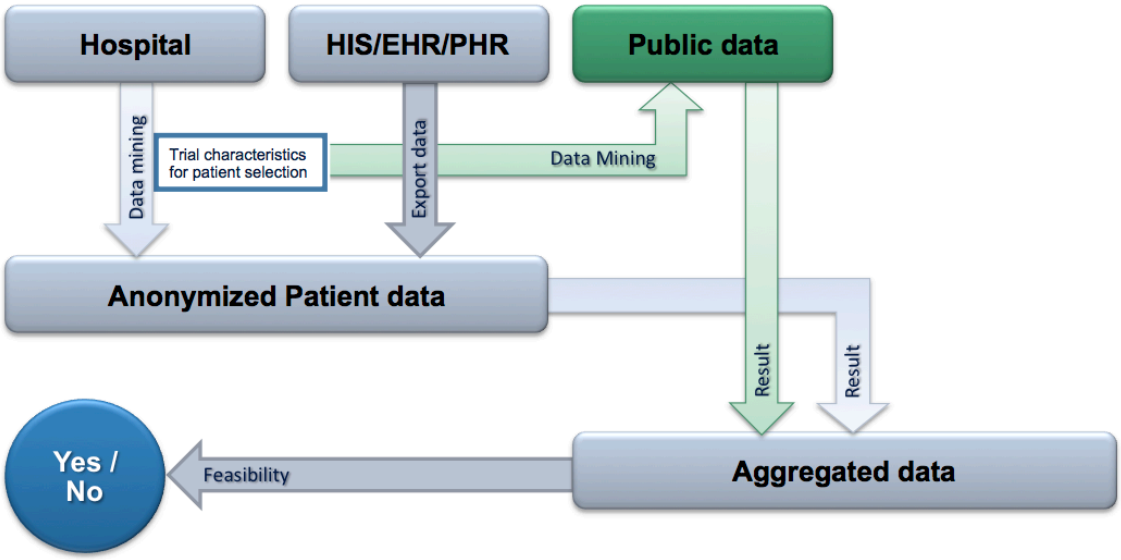


Fig. 3.5: Outline of the scenario for the selection and inclusion of patients into trials.

Selection and inclusion of patients into trials

This scenario describes how patients can be selected for a specific trial. The initiator can be a pharmaceutical company. A pharmaceutical company will never get access to personal data. The search will always be done on aggregated data. There is a relation to the scenario KD16 (Personal medical information recommender).

The following steps are needed (see figure 3.6):

1. Patient data from HIS/EHR/PHR are exported and anonymized (data warehouse)
2. Pharmaceutical company can do data mining only on aggregated data to select the cohort of patients that fits the inclusion criteria of the trial
3. Treating physicians will be announced that specific patients can be enrolled in a trial
4. The physician contacts the patient and explains the trial
5. The patient needs to give informed consent to be enrolled in the trial

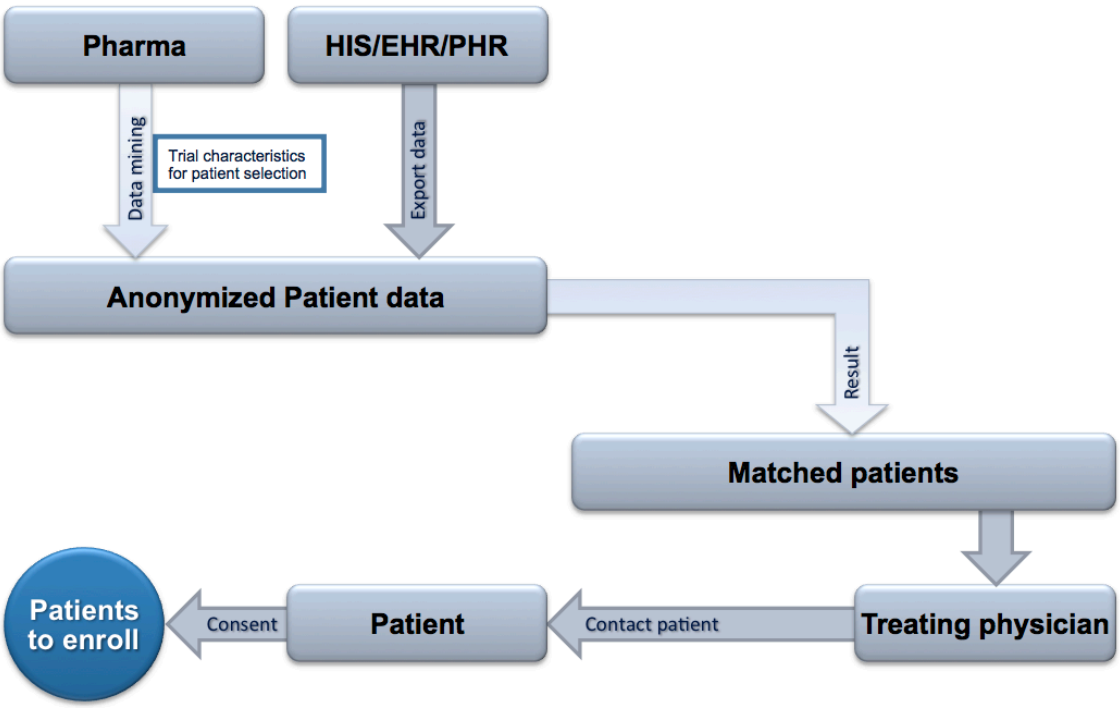


Fig. 3.6: Outline of the scenario for the selection and inclusion of patients into trials.

Two questions need to be solved in this scenario:

1. Is the pharmaceutical company allowed to start the export of anonymized data from HIS/EHR/PHR?
    a. If no, who should start this scenario?
    b. If yes, under which guidelines or contracts?
2. Can the pharmaceutical company contact patients by themselves?

These two questions need to be reviewed according to legal and ethical viewpoints. If the answer of both questioned is yes, then the scenario outline needs adaptations. This scenario can also be used as a decision support service within a hospital to recruit patients for a trial. In this case the role of the pharmaceutical company is the hospital.

Detection and prediction of SAEs and SUSARs

This scenario describes how SAEs and SUSARs can be detected and predicted before a treatment is given to a patient. A database of pharmacogenomics is needed in addition in this scenario.

The following steps are needed (see figure 3.7):

1. Select a patient and relevant clinical data from the HIS/EHR/PHR
2. Select a drug or a treatment that will be given to a patient
3. Do data mining in databases of EMA for SAEs and literature mining
4. Show possible SAEs and list the molecular pharmacogenomics

5. Perform molecular analysis of the pharmacogenomics in the blood of the patient
6. Specify the individual risk of an SAE or SUSAR for the drug / treatment tested in this scenario

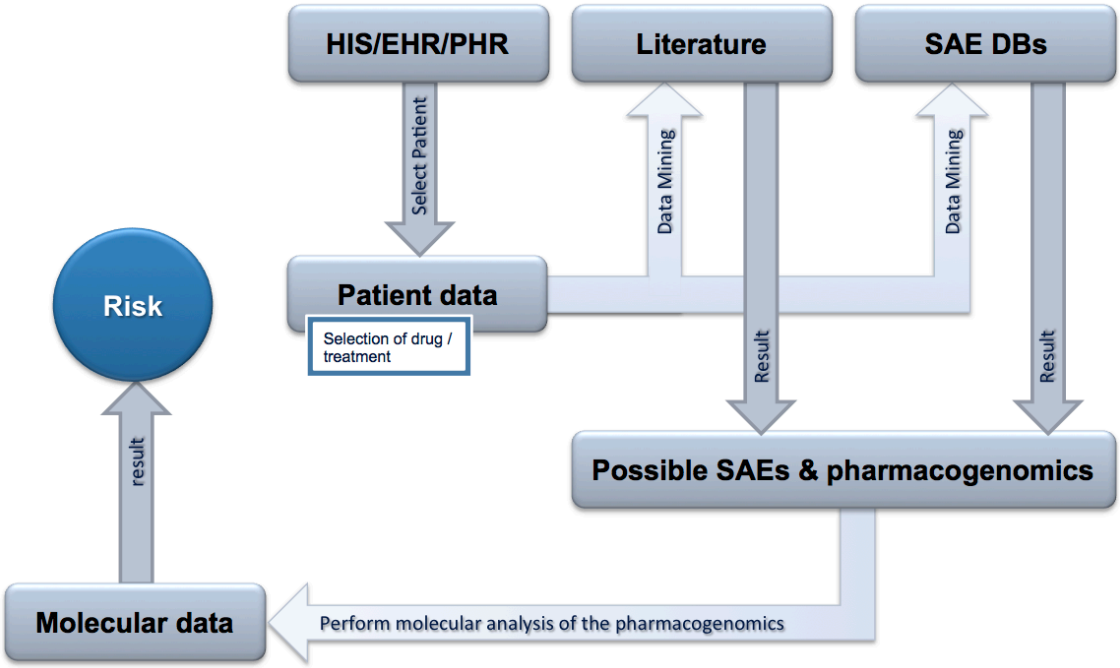


Fig. 3.7: Outline of the scenario for the detection and prediction of SAEs and SUSARs.

"GGO": Using DNA sequencing in Oncology (for clinical trials & daily health care) to identify patients that have a non-synonymous mutation in a gene that is related to drug or radiation sensitivity

When a young patient with oligo-metastasis (less then 5 metastases, but this just an example) is first seen at MAASTRO we note features of this patient in our Electronic Health Record (EHR). We also receive imaging data and letters from external hospitals for this patient. We would like to maximize this patients' survival and therefore want to address whether a new patient is suitable for certain targeted therapeutic agents (often quite expensive) known to be efficient *only* if certain mutations are present (see table underneath). The drug would be given together with curative stereotactic radiation. If a tumour (but not the normal tissues) has a mutation in a gene that increase radiation sensitivity (e.g. ATM) then an extensive use of radiation at moderate dose but on large fields is also an option. The best way to do so will be through DNA sequencing. This will become more and more rapid and cheaper. We call this GGRT: "Genomic Guided Radiotherapy" or in more general terms GGO "Genomic Guided Oncology".

This scenario is a specific example of the scenario for the prediction of SAEs and SUSARs

Pharmacovigilance – Automatic reporting of SAEs and SUSARs

This scenario describes how SAEs and SUSARs are detected in specific patients and will be reported automatically to regulatory bodies.

The following steps are needed (see figure 3.8):

1. SAE and SUSAR definitions are described and used
2. At regular time points that can be fixed (e.g. daily) the HIS/EHR/PHR databases are queried for SAEs and SUSARs.
3. The SAEs and SUSARs are send to a physician
4. He needs to validate the SAEs and SUSARs
5. After validation an automatic report is created according to GCP criteria
6. The SAEs and SUSARS are uploaded to the SAE database at the European Medical Agency (EMA)

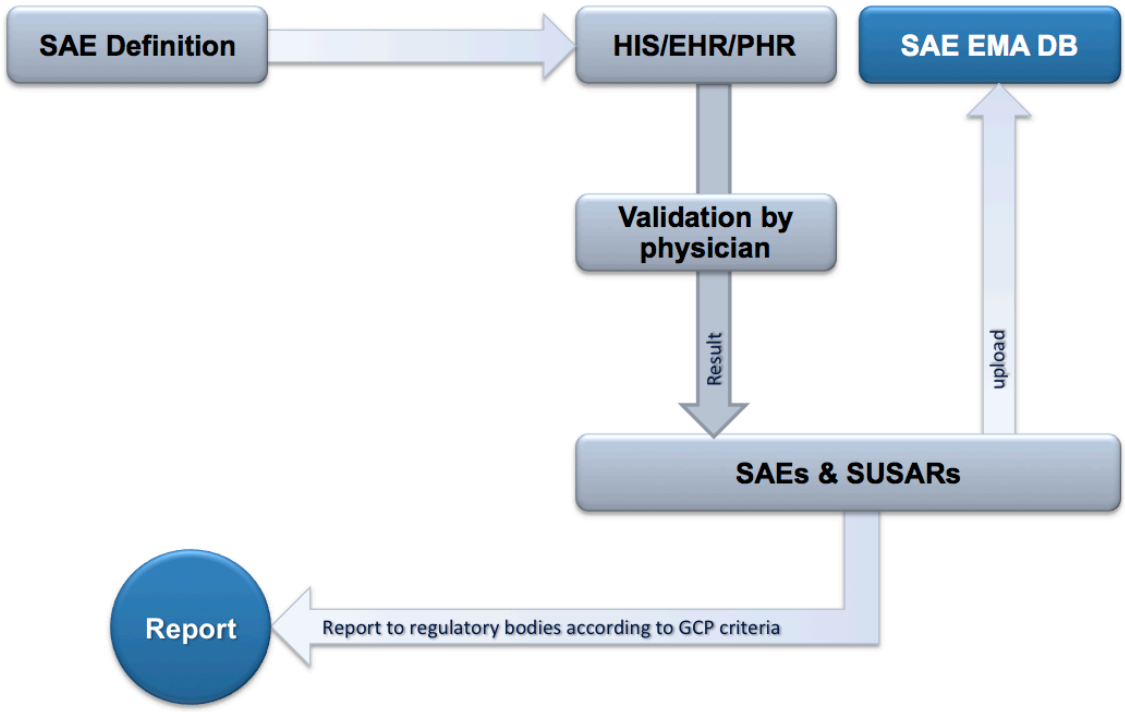


Fig. 3.8: Outline of the scenario for Pharmacovigilance. (SAE EMA DB: database at EMA for SAEs)

Early detection of cancer / individual risk / prevention

According to the patient's personal life style data (social networks), his genetic data and clinical data (EHR, PHR, HIS, etc.) the personal risks for diseases can be listed. This might help to detect cancer earlier by starting a screening program for the patient or advice the patient to change his/her lifestyle to prevent cancer, if such a program exists. The scenario is outlined in figure 3.9.

The following steps are needed (see figure 3.9):

1. Relevant patient data are extracted from the EHR/PHR/HIS
2. Data Mining of literature to find risks for diseases (Cancer)
3. Data mining of social networks to describe the lifestyle of a patient
4. As a result the individual risk can be predicted leading to screening for early detection of cancer or advices to change the lifestyle

The scenario KD16 (Personal medical information recommender) is closely related to this scenario.

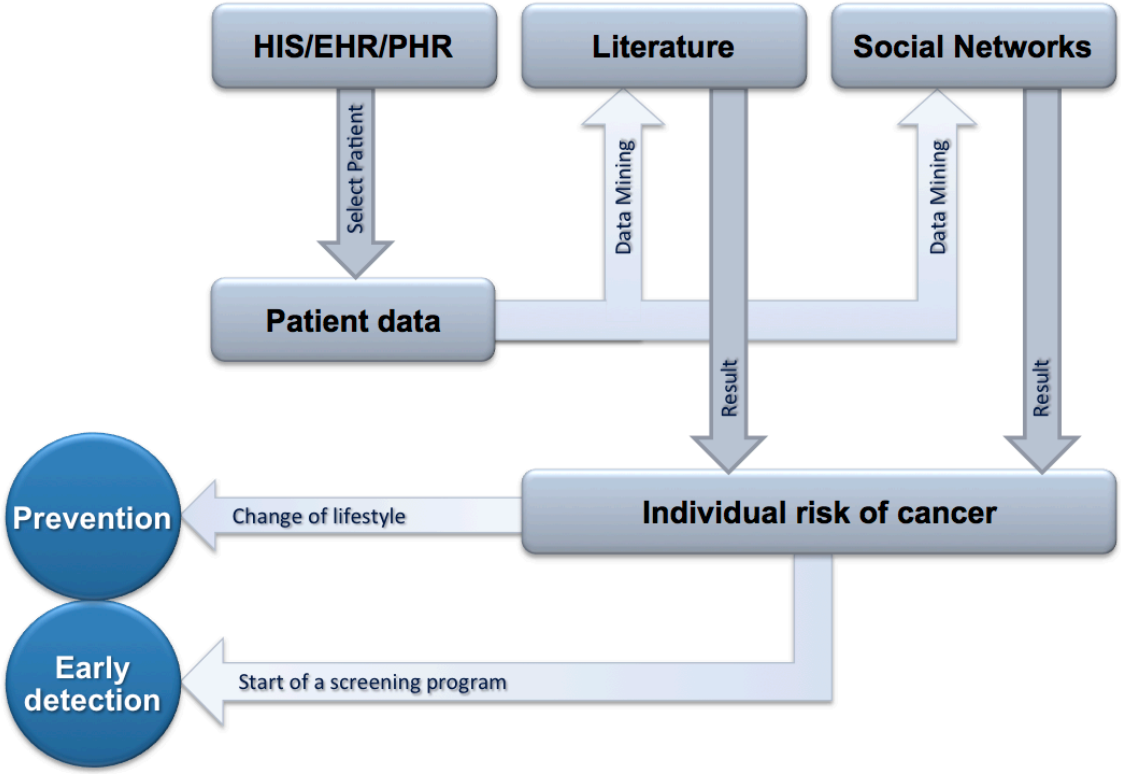


Fig. 3.9: Outline of the scenario for early detection and prevention of cancer.

Personal medical information recommender

This scenario describes how people can obtain objective information about trials, treatments etc. about their specific disease. It defines the condition of a patient and does data mining in all available data sources. The scenario analyses them for writing a report what is the background of the disease, what are the best way to diagnose, what treatment options are available and what is the outcome, including acute and late toxicity as well as life quality. Such a report contains pros and contras, e.g. is surgery better than irradiation in a given cancer of a specific patient? It might also include cost data. A better characterization of a patient according to his/her risk factors will help to predict the outcome of the disease for him/her. This can also be seen as a simulation of the response to different treatments and can be done by selecting patients with the same characteristics from the database and show which treatment results in which outcome. A search for further risk criteria will help to distinguish these patients into more different prognostic groups, to find for an individual patient the optimal treatment.

The following steps are needed (see figure 3.10):

1. Relevant patient data are extracted from the EHR/PHR/HIS
2. Data Mining of literature and trial databases are done, to find relevant information about the disease and possible trials
3. As a result a summary of medical data and objective information about medical knowledge of the disease of the patient is given
4. This information will be written in two different reports
    a. One report for the patients to summarize his condition and explain possible treatments in a language a patient understands.
    b. A detailed report for the physician summarizing the condition of the patient but also pros and contras of possible treatments and procedures. If a patient needs to go to a new physician, the new physician can use this tool to get the relevant information about the patient without going through all charts of the patient. A patient can use the tool as well, if he/she wants to get a summary of his/her data. It should allow the patient also to download his data, including imaging data.

In summary this is a personal medical information recommendation.

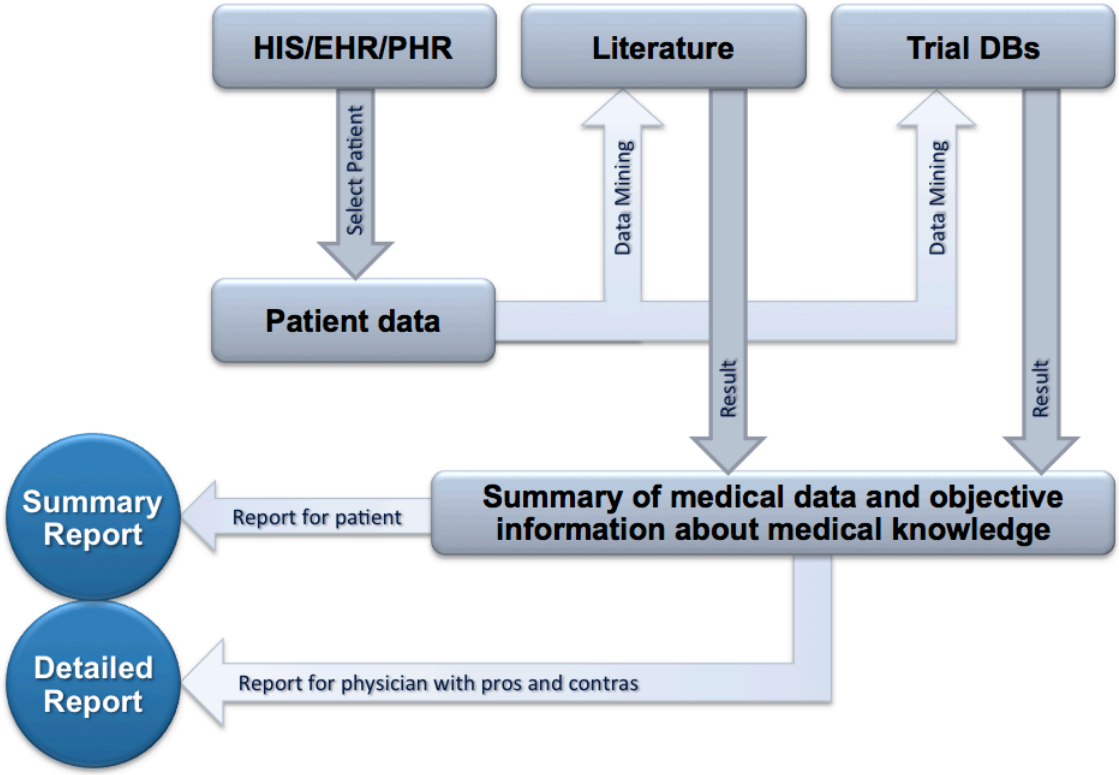


Fig. 3.10: Outline of the scenario for early detection and prevention of cancer.

Develop or update of guidelines from clinical trial data and literature mining

This scenario describes how guidelines can be developed and regularly updated from data mining of clinical trials and literature.

The following steps are needed (see figure 3.11):

1. Select a guideline and items in the guideline that should be updated
2. Use these items for data mining in CT/HIS, Literature and trial databases
3. Search only for data beyond the date of the guideline
4. After data collection do an automatic listing of the updated items
5. The end-user will select the relevant items from these listings
6. These updated items will replace the old items in the guideline
7. The guideline is updated and a new version with the date of update is stored

This scenario covers also the scenarios KD19 and KD20 dealing with rare diseases.

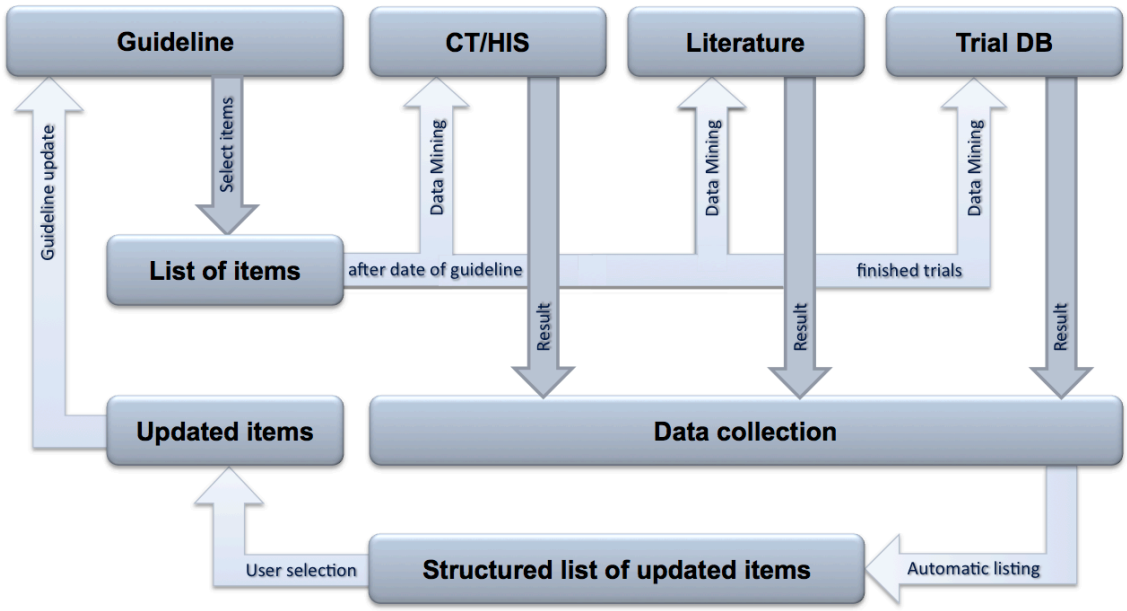


Fig. 3.11: Outline of the scenario for guideline update.

Data mining of consultations

In prospective clinical trials many consultations are performed. A part of the questions of such consultations are repeatedly asked. It would be helpful to generate an automatic answer to questions asked during consultations.

For this scenario the following steps are relevant (see figure 3.12):

1. Select the trial and the documentation of the consultations
2. This documentation can be available in a structured or unstructured way
3. In case of an unstructured way data mining of the text is needed to extract relevant information that will be store in a structured way.

4. The structured data of the trial will be used as a source for data mining for a specific question of a consultation.
5. All answers to the same consultation question will be selected and analysed to create an answer to the consultation question
6. This answer will be validated by literature mining
7. As a result of this validation a final answer will be created

The same scenario can also be used to develop frequently asked questions and to contextualize this information. This will help patients and clinicians to identify relevant information for their needs.

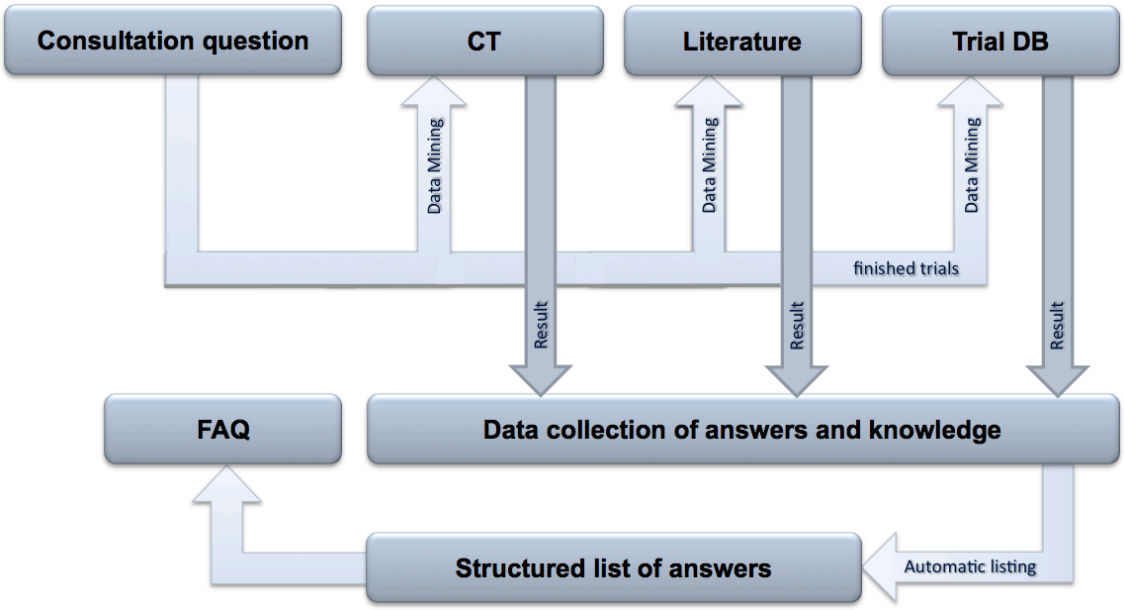


Fig. 3.12: Outline of the scenario for data mining of consultations.

Analyse economic data between different procedures (for funding reasons) compared to outcome and quality of life / data of hospital stays, expected side effects, etc.

By joining data from EHR, clinical trials, literature and open databases economic aspects of different procedures (diagnostic and/or therapeutic) can be analysed in respect to outcome and quality of life in an individual patient. This will include data about days to stay in the hospital, expected side effects, costs of diagnostics and therapeutics, etc.

For this scenario the following steps are relevant (see figure 3.13):

1. Select the relevant data of a patient
2. Search for best diagnostic and treatment procedures for this patient in literature and open source databases for his/her disease
3. List all the procedures

4. Perform data mining of economic databases
5. Select the outcome criteria (quality of life or survival)
6. As a result the best procedures are listed according to economic and outcome criteria

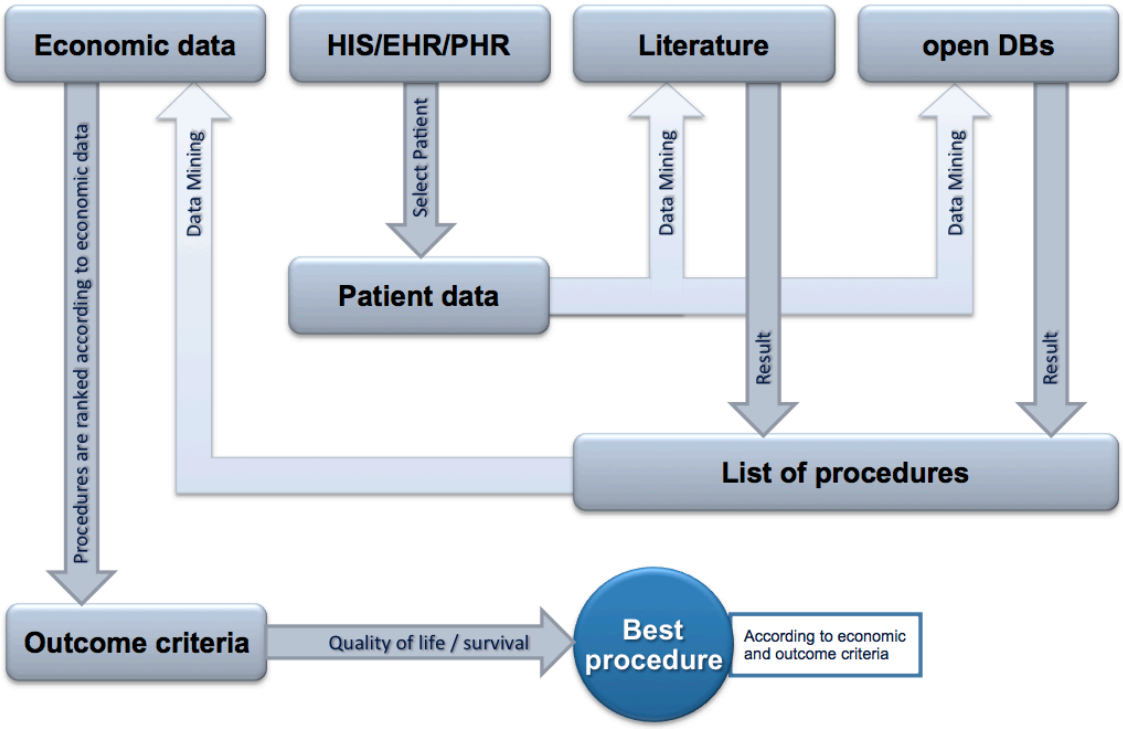


Fig. 3.13: Outline of the scenario for analysing economic data between different procedures.

The following scenarios are part of scenarios described above and need no further explanation.

1. Combining data sources for information discovery
2. Extract patient data from PHR and EHR
3. Extract and structure textual/structured EMR data

3.1.3.2 Data curation

Long-term follow-up

This scenario deals with the curation of data in long-term follow-up. The service or tool for this scenario shall collect relevant health data recorded by patients in PHR solutions and link them as long-term follow-up information to corresponding clinical trials in which patients were included.

The trial chairman defines in one or more specific CRFs which health related information that patients can report in PHR is of interest for his clinical trial. The EURECA service shall then enter the PHR data from the corresponding PHR system into the special "PHR-CRFs" of the clinical trial management system. (The PHR-CRFs will most likely match patient questionnaires.)

This scenario includes also the following scenarios:

Survival follow-up:
This scenario deals with the extraction of the last follow-up date and patient status (alive or deceased) from EHR or the National registry, when the EHR last date of contact is > 2 years old. The date of the last contact and the patient's status will be automatically included in the eCRF of the corresponding trial where the patient was enrolled.

Primary and secondary outcome measures:
This scenario deals with the extraction of primary and key secondary outcome measures from EHR and imports them in eCRF for a specific protocol. In particular in case of breast cancer for example:

1. Date of first recurrence in breast cancer patients that received an adjuvant treatment and site of recurrence (SNOMED term).
2. Left ventricular ejection fraction (LVEF) value (%), method used (echo, MUGA or cardiac MRI) and date of assessment

Safety reporting of specific adverse reactions after study treatment completion:
This scenario deals with the extraction of safety data from EHR and imports them in eCRF to improve safety reporting of adverse reaction with late onset post study treatment completion.

Adverse reaction of interest can be:

1. Congestive heart failure
2. Second primary malignancies (new tumour of different origin than the primary)

Data of interest:

1. Event (SNOMED term)
2. Date of event onset

Other relevant data to complete the safety reporting can be requested by human interaction (grade, concomitant medications, duration, outcome).

For this scenario the following steps are relevant (see figure 3.14):

1. Select the relevant patient from the clinical trial
2. Perform data mining in HIS/EHR/PHR and/or in National registries for death dates

3. Update the data in the clinical trial after checking quality criteria issues

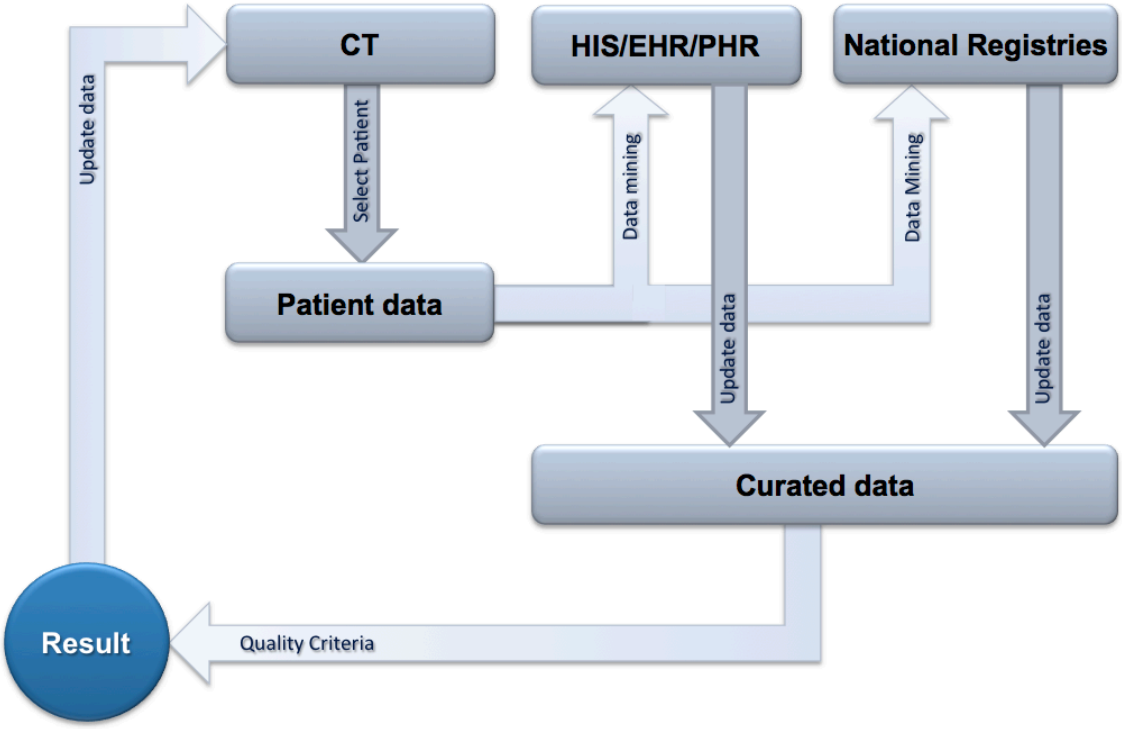


Fig. 3.14: Outline of the scenario for long-term follow-up

Patient Diary

This scenario deals with the possibilities of a patient diary. Such a diary can be used in clinical trials, where there are specific eCRFs for patients. In this eCRFs the trial chairman can define what the patient can be asked. This can include the following items:

- Follow-up of late effects
- Quality of life data
- New surgical data
- Relapse data
- Second malignancy data
- Laboratory values
- Imaging data

The eCRFs can also be filled by data from PHR. ObTiMA, Tolven and OpenClinica can be used as a data management system for such eCRFs. In this case there should be an exchange between the PHR and these data management systems be possible. Out of the clinical trial data and the data from the patient diary a PHR can be build.

For this scenario the following steps are relevant (see figure 3.15):

1. The patient selects the eCRF of the patient diary
2. The patient enters new data
3. From the clinical trial or the EHR corresponding data are compared with those data the patient has provided
4. According to quality criteria these data are matched and an update of the PHR or EHR will be done

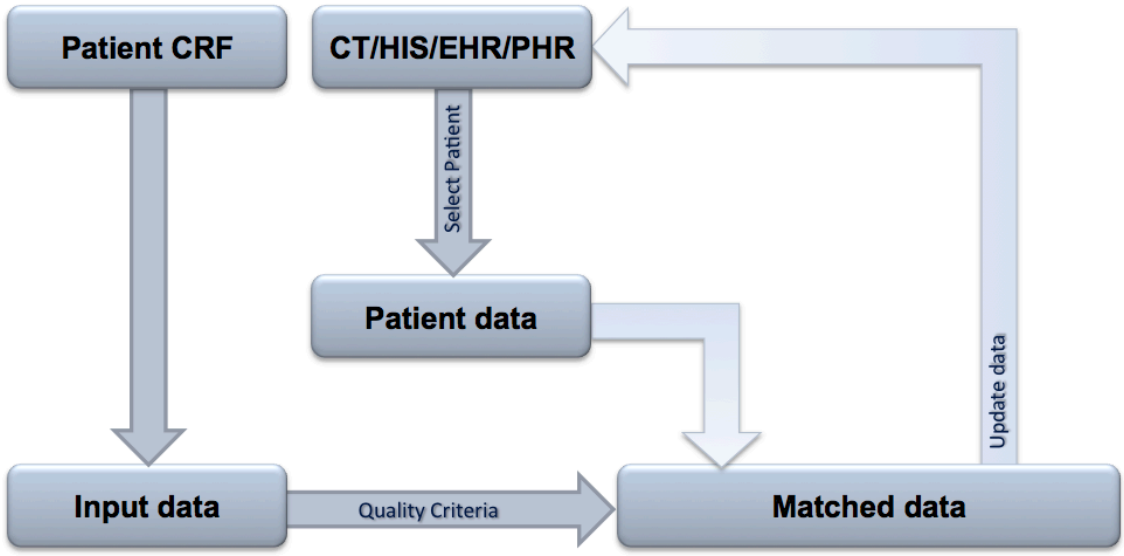


Fig. 3.15: Outline of the scenario of the patient diary

3.1.3.3 Basic research and clinical trials support

Supporting design of new trials and hypothesis generation

Before starting a new clinical trial a new research question is needed. Such a question is of utmost importance and is part of hypothesis generation. Analysing all available data from previous trials, guidelines, literature and others, can support this process. It can also help to find biomarkers that are relevant for the disease suggesting their use in the trial for evaluation or validation purposes.

A trial chair, while designing a new trial, is supported in the design of the eligibility conditions (inclusion and exclusion) based on background information:

6. design of previous trials
7. results of previous trials
8. SUSAR reports from previous trials
9. previous epidemiological (retrospective) studies.
10. published literature

Such a scenario will fasten to write the trial protocol, which is based on scientific grounds.

For this scenario the following steps are relevant (see figure 3.16):

1. The trial chairman starts data mining of clinical trials, literature and public database
2. A hypothesis will be generated to help to ask a scientific based research question

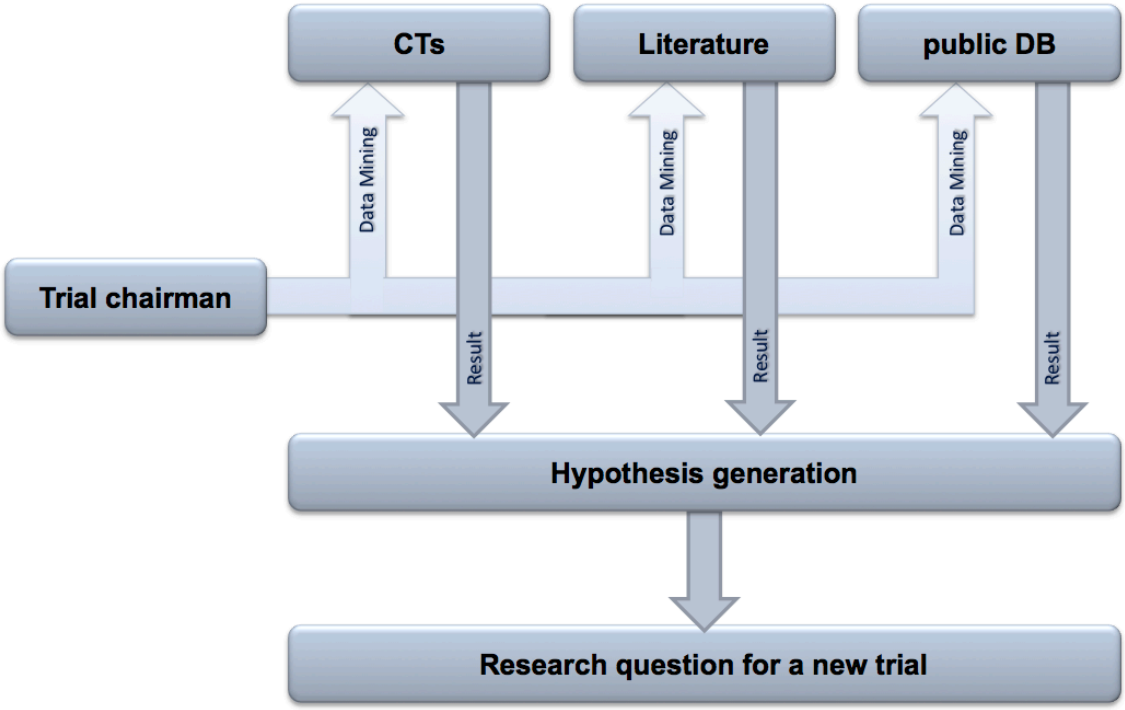


Fig. 3.16: Outline of the scenario for supporting new trials and hypothesis generation

Clinical data reuse

During routine patient care, a lot of information is recorded for patients in local IT systems that also need to be recorded for the trial the patient participates in. The use case is to re-use the clinical data into the trial eCRF systems. This is to avoid double data entry.

Opt-out solution for further research

Provide a platform where patients can select which research they do not like to do with their data or biomaterial.

A researcher having a specific question has the possibility to analyse anonymously EHRs to detect patients that may help to answer his research question with their data. If such a patient is detected he will automatically notified by an email that a researcher

wants to use his data for a specific research. The type of research is described in a way the patient understands on a specific website, to which the patient will be linked. The patient can disagree at any time to participate with his data in this research project. Such a scenario is based on the fact that every patient agrees to share his data to any research project and that he can disagree to specific research projects at any time by using the above described website. The same scenario will be possible for research on biomaterial as well.

For this scenario the following steps are relevant (see figure 3.17):

1. The researcher starts data mining on anonymized patient data that match his research question
2. An email is send to those patients that match and have agreed to participate in future research
3. The patient gets via the email a link to a website, where the research project is described in a language the patient will understand
4. On this website he can actively deny to take part in this research with his data
5. If the patient is not actively going to the website he agrees to the research as he has given a general informed consent for research
6. If the research is finished a new email is send to the patient to be informed about the results of the research that is done with his data.

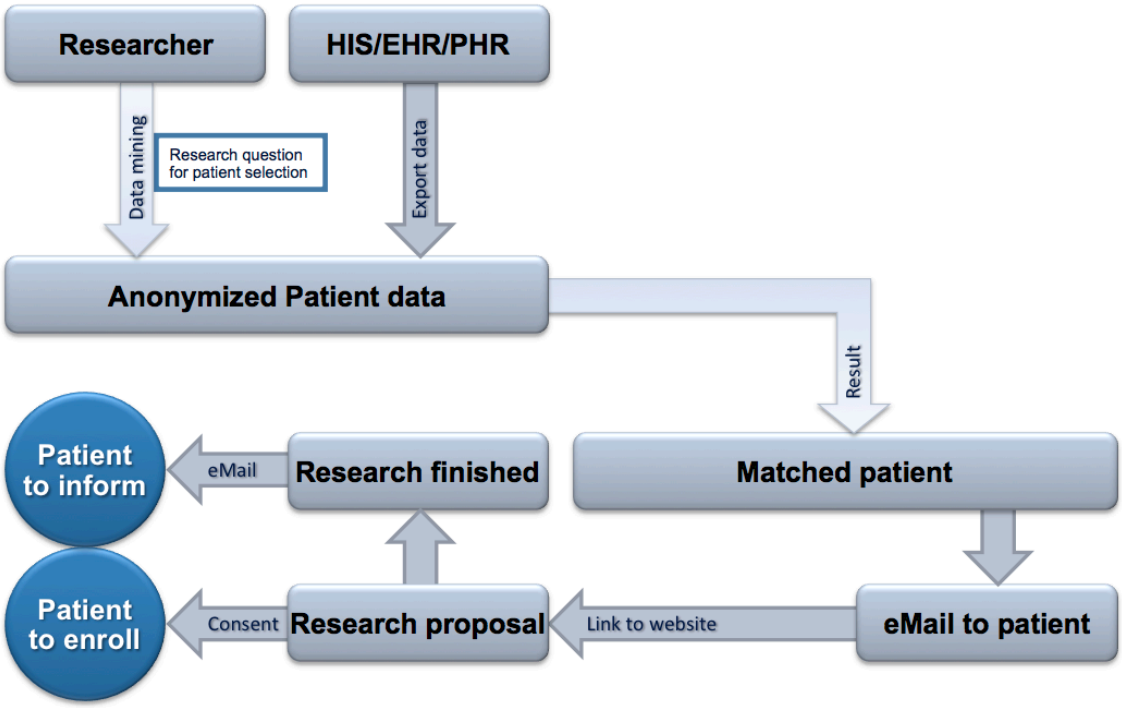


Fig. 3.17: Outline of the scenario for opt-out solution for research

The following two scenarios can be part of above described scenarios and do not need further description.

Similarities of datasets to combine

This scenario describes the detection and identification of similar datasets from different patients. The relevance of such a scenario is to find patients with similar clinical background and compare their treatment, outcome or quality of life. The tool can also be used to find patients for a new clinical trial (KD8) or for a research project (BR1).

Such a scenario can be compared with a scenario used by Amazon: users who've bought this, also bought that.

Rapid learning

In rapid learning research we want to learn and validate outcome prediction models from routine patient care data. We need to have access to large amounts (10.000+) of patient's data preferentially with clinical, imaging, biology information. This scenario can be integrated in the scenario 'Personal medical information recommender' (KD16).

3.1.4 Final Scenarios

This table shows how the above listed scenarios are categorized according to their purpose and similarities. They build the final scenarios that will be further developed and specified (D1.2) in EURECA. It needs to be stated that different scenarios from the scenarios listed in 3.1.3 can be listed in different final scenarios. Dissecting these scenarios into different granular use cases will solve this. Such use case will build the link between different final scenarios. This will guarantee that tools will be modular developed and not from scratch. (see also: http://atlas.ics.forth.gr/EURECA/wiki/index.php/Final_scenarios)

Groups	N°	Scenarios	Scenarios from partners
Knowledge Discovery	1	Selection of trials for patient enrolment	KD1 KD2 KD3 KD4 KD5 KD6 KD29 KD30
	2	Protocol/Trial feasibility	KD7
	3	Selection and inclusion of patients into trials	KD8
	4	Detection and prediction of SAEs and SUSARs	KD9 KD10 KD28
	5	Pharmacovigilance – Automatic reporting of SAEs and SUSARs	KD11
	6	Early detection of cancer / individual risk / prevention of cancer	KD12 KD13
	7	Personal medical information recommender	KD16 KD17
	8	Develop or update of guidelines from clinical trial data and literature mining	KD18 KD19 KD20
	9	Data mining of consultations	KD21
	10	Analyse economic data between different procedures	KD26
Data Curation	11	Long-term follow-up	DC1 DC2 DC3 DC4 DC6
	12	Patient Diary	DC5
Basic Research and CT Support	13	Supporting design of new trials and hypothesis generation	CT1 CT2
	14	Clinical data reuse	CT3 CT4 CT5 CT6
	15	Opt-out solution for further research	BR1
	16	Similarities of datasets to combine	BR2
	17	Rapid learning	BR3 BR4
?			KD22 KD23

3.1.5 Ranking of specific scenarios

The following table shows the ranking of the different scenarios done by clinical partners. The ranking numbers are defined as: 4: most interesting, 3: very interesting, 2: interesting, 1: little interest and 0 no interest. The highlighted average ranking scores are the most relevant to be primarily investigated. These are: Selection of trials for patient enrolment, Protocol/trial feasibility, Long-term follow-up, supporting design of new trials and hypothesis generation, and clinical date reuse.
(see also: http://atlas.ics.forth.gr/bscw/bscw.cgi/d14254/Partners_scenarios_ranking.pdf)

	N°	Scenarios	Clinical partners ranking						AVERAGE RANKING
			IJB	UdS	UOXF	BIG	MAASTRO	GBG	
Knowledge Discovery	1	Selection of trials for patient enrolment	**4**	2	**4**	**4**	**4**	**4**	**3.67**
	2	Protocol/Trial feasibility	**4**	**4**	1	3	**4**	**4**	**3.33**
	3	Selection and inclusion of patients into trials	1	1	**4**	**4**	2	**4**	2.67
	4	Detection and prediction of SAEs and SUSARs	3	**4**	2	2	1	3	2.50
	5	Pharmacovigilance – Automatic reporting of SAEs and SUSARs	3	3	2	2	2	3	2.50
	6	Early detection of cancer / individual risk / prevention of cancer	1	1	3	2	0	2	1.50
	7	Personal medical information recommender	1	1	2	2	1	2	1.50
	8	Develop or update of guidelines from clinical trial data and literature mining	2	2	3	3	3	2	2.50
	9	Data mining of consultations	1	3	2	1	3	2	2.00
	10	Analyse economic data between different procedures	1	1	1	1	1	1	1.00
Data Curation	11	Long-term follow-up	**4**	**4**	**4**	3	3	3	**3.50**
	12	Patient Diary	1	3	3	1	2	2	2.00
Basic Research and CT Support	13	Supporting design of new trials and hypothesis generation	2	3	**4**	3	3	3	**3.00**
	14	Clinical data reuse	**4**	**4**	**4**	**4**	**4**	3	**3.83**
	15	Opt-out solution for further research	1	**4**	2	3	1	2	2.17
	16	Similarities of datasets to combine	1	1	**4**	2	2	2	2.00
	17	Rapid learning	1	1	**4**	1	**4**	1	2.00

4 Tools or service development

To further specify the EURECA environment, it is important to update the needs of stakeholders in an iterative way throughout the run of EURECA. For that purpose at IJB, IEO and USAAR the stakeholder needs are listed according to the three groups defined in chapter 3. This listing is helpful for proposing new scenarios in the future, as it is an evolving process.

→ **Knowledge discovery**
- Meta-analysis
- Statistics (treatment response, prevalence of AEs, demographic data)
- Retrospective studies:
  - v side-effects finding
  - v treatment comparisons
  - v guideline quality and compliance assessment
  - v epidemiological studies (Survival, and disease-free-survival studies) (e.g. for prognosis, drug response prediction)
- AEs automatic detection, monitoring of trial-drug-related AEs
- Drugs toxicity (e.g. long term toxicity)

→ **Data curation**
- Semantic research tool within our EHR
- Improve interoperability between EHR and internal databases (e.g. anatomopathology, biobank, clinical biology and pharmacy labs, internal Cancer Registry)
- To update automatically and quickly the patient database from patient medical anamnesis/history (e.g. Chemotherapy, Radiotherapy, Hormonotherapy)
- To control compliance of treatments with guidelines, to help update treatment guidelines
- Help diagnosis, improving pathologies detection (e.g. silent pathologies)
- Long term follow-up (feedback loop so that the current status of a trial, or a patient in a trial, is known)

→ **Basic research/clinical research**
- Filling eCRFs automatically
- Linking genomic data with clinical outcome

All of the described scenarios need to be translated into use cases to build tools. As not every tool should be build from scratch it is important to dissect the scenarios into use cases of highest granularity. This approach will help to build tools in a modular way and reuse granular tools in different scenarios. This will be described in detail in D1.2 (Definition of relevant user scenarios based on input from the users).

4.1 Operational environment of EURECA

For each of the scenarios it is important to define what are the legal requirements. This work will be done by WP7. From a legal perspective the scenarios can be divided into 3 different settings describing the operational environment. These settings are:

- Trial support and execution
- Research domain
- Care domain

The scenarios that are described in chapter 3 can be linked to these three settings in the following way:

- **Trial support and execution**
  - Selection of best trials for a patient
  - Trial / protocol feasibility
  - Selection and inclusion of patients into trials
  - Pharmacovigilance – Automatic reporting of SAEs and SUSARs
  - Supporting design of new trials and hypothesis generation
  - Simulation of datasets to combine
  - Rapid learning
- **Research domain**
  - Develop or update guidelines for diseases
  - Data mining of consultations
  - Analyse economic data between different procedures / approaches
  - Opt-out solution for new research
- **Care domain**
  - Detection and prediction of SAEs / SUSARs
  - Early detection and prevention of diseases
  - Personal medical information recommender
  - Long term follow-up
  - Patient diary (connection between PHR and data management tools)
  - Clinical data reuse

The distinction between these 3 settings is mainly the question, if anonymized data or personal data are used. In the research domain only anonymized data are needed, whereas in the care domain there is always the need for personal data. In the trial support and execution setting some scenarios do need access to personal data others not. In addition software engineers do need access to patient data during the time they are building the tools. There is no need for patient data later on for IT-people.

4.2 Available databases

For implementing scenarios and developing corresponding tools or services there is a need for usage of concrete databases. From the different clinical partners the following databases will be available, if legal issues are solved.

UdS:

1. Nephroblastoma database of SIOP 2001/GPOH.
   a. This is a clinical trial database for nephroblastoma including tables for consultation
   b. DICOM images of a part of the patients enrolled in the trial
2. Hospital information system data
   a. This is currently under negotiation
3. Cancer registry data
   a. This is currently under negotiation

IJB:

1. Cancer Registry
2. CMO
3. JCMO (future)
4. Oribase - Anapath
5. PACS Telemis
6. Oribase radiology & nuclear medicine data
7. Oribase LAB data
8. Oribase consult & discharge reports (SM2008)
9. SM2008 Day Hospital (oncological day care clinic)
10. CATO chemo prescriptions
11. CRFs

UOXF:

1. PACS (imaging)
2. ARIA (electronic patient record, only used for chemotherapy prescription in Oxford: http://www.varian.com/us/oncology/radiation_oncology/aria/)
   The EPR in use is called EPR and the company supplying it is Cerner Millenium
3. pathology databases
4. different filemaker pro databases
5. next generation sequencing databases (life tech (Oxford) databases)
6. TSB genomic database
7. Biobank
8. MRC CRUK clinical trials unit in Birmingham
9. clinical trials databases:
   a. EORTC
   b. EURO-EWING

MAASTRO:

1. EMD (clinical)
2. PACS (imaging)
3. euroCAT (clinical & imaging)
4. ZyLAB (clinical - OCR scans)

GBG:

1. Closed Trial Databases
2. CRFs

5 SUMMARY

The EURECA environment and the tools to be developed are based on the needs of the stakeholders and in especially on those of the clinicians. This will guarantee that the environment will be used in daily clinical practice. The following methodology was used to achieve this goal:

- A user needs survey was set up
- Interviews of stakeholder were done
- Clinicians taking part in the project formulated scenarios

As a result the following most important scenarios are defined according to ranking done by clinical partners:

- Selection of trials for patient enrolment
- Protocol/trial feasibility
- Long-term follow-up
- Supporting design of new trials and hypothesis generation
- Clinical date reuse

All final scenarios will be translated into use cases to build tools. As not every tool should be build from scratch it is important to dissect the scenarios into use cases of highest granularity. This approach will help to build tools in a modular way and reuse granular tools in different scenarios. This will be described in detail in D1.2 (Definition of relevant user scenarios based on input from the users).

Most important are the data that will be made available within EURECA. To start with as early as possible the legal framework needs to be accepted by all participating partners providing or using data within EURECA. Corresponding contracts are already developed and are sent to the partners for signature. Nevertheless there will also be data used in EURECA that will not be shared but will be used by tools developed in EURECA in the hospital, where these data are coming from.

All tools that will be developed should be modular based avoiding the building of all tools from the scratch.

6 REFERENCES

1. S. Alistair, Scenario-based requirements analysis. Requirements Engineering, 3 (1), pp 48-65, 1998 available online: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.45.6015&rep=rep1&type=pdf.

2. W.E. Rzepka, A Requirements Engineering Testbed: Concept, Status, and First Results. In Bruce D. Shriver (editor), Proceedings of the Twenty-Second Annual Hawaii International Conference on System Sciences, 339-347. IEEE Computer Society, 1989

7 Appendix 1: User needs survey

EURECA (Enabling information re-Use by linking clinical REsearch and CAre) is a collaborative project that is funded by the European Commission under the 7th Framework Programme (Grant agreement no:288048).

The project aims to achieve semantic interoperability between electronic health records (EHR) and clinical trials systems. 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.

We are interested in your opinion about the project in general and specifically we would like to know from you how to improve the EURECA environment. We would also like to ask you to describe what is hampering you in your current work and what applications you would find useful if all the systems were integrated in your organisation and you could seamlessly collaborate with other organisations.

Thank you in advance for taking part in this questionnaire.

1. What is your affiliation? (Please choose all that apply)

- ☐ Clinics
- ☐ Basic research
- ☐ Pharmacy
- ☐ IT
- ☐ Clinical research
- ☐ Translational research
- ☐ Drug/product development

2. Which country do you live in?

3. Do you know what electronic health records (EHR) are?

○ Yes ○ No

4. Do you know what personal health records (PHR) are?

○ Yes ○ No

5. Do you think the re-use of personal data collected in a patient file is a good idea?

○ Yes ○ No

6. Please explain what research you would be able to carry out if you had access to all the data collected in the patient record system of your institution...

7. Can you please give examples of applications that would be improved by combining care data (EHR/PHR) and research data?

8. Do you think linking EHR/PHR data with clinical trial data would be useful in general?

- Yes
- No

9. Would linking EHR/PHR data with clinical trial data be useful for your research?

- Yes
- No

10. Why is data from research useful in your care setting?

11. Do you ever access data in your reseach system and your EHR/PHR system for the same task?

- Yes
- No

If yes, for which tasks

12. Do you ever query the research data when treating a patient?

- Yes
- No

13. Do you ever query the EHR while working on a research question?

- Yes
- No

14. Are you aware of any problems with electronic health records?

- Yes
- No

If yes, please specify:

15. How important do you think data security issues are in this context?

- Very important
- Important
- Less important
- Not important
- I don't know

16. If all security and privacy issues were solved, would you see value in sharing data with other organisations?

- Yes
- No

17. Would any new research be possible if you had access to all the data from both research and care collected in your institution and/or from other institutions?

- Yes
- No

If yes, please describe the possible research

18. Would any new research be possible if you had access to much larger amounts of data then you currently have access to?

- Yes
- No

If yes, please specify the possible research

19. Do you currently ever combine research and care data?

- Yes
- No

If yes, please explain why and when

20. Is your current work hampered by the fact that your EHR systems and research data are independent from each other?

- Yes
- No

If yes, please explain how

21. Are there other other systems and/or data that chould be combined with research and care data to help your work?

- Yes
- No

If yes, please specify which systems or data

22. Are there any additional sources of data or knowledge from outside your organisation that would also help your work? (Publications, databases, ontologies etc)

- ○ Yes
- ○ No

If yes, please specify

The aims of the EURECA project are 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:

-Support more effective and efficient clinical research by:
~Allowing faster identification and registration of eligible patients
~Providing access to large amounts of patient data
~Enabling long term follow up of patients
~Avoiding any need for duplication of data entry

-Allow the mining of longitudinal EHR data for early detection of patient safety issues
-Help the faster transfer of new research findings and guidelines to clinical practice
-Enable healthcare professionals to legally and ethically extract relevant information from the huge amounts of data available on each patient

23. Can you please describe a tool that would help your work within the context of the EURECA project?

24. Are there any specific requirements that the tool mentioned in your previous answer should have?

25. Are you aware of any existing tools that are work in this area?

○ Yes ○ No

If yes, can you please give the tool's name?

26. Which specific tasks is this tool used for?

27. Which groups of professions within your institutions use these tools?

28. Are you aware of any problems with this tool, or how it could be improved?

- Yes
- No

If yes, please specify

29. Which data is needed by this tool to do its job?

30. Is the tool available via the VPH toolkit? (http://toolkit.vph-noe.eu/)

- Yes
- No
- I don't know

31. Is the tool open source?

- Yes
- No
- I'm not sure

If yes, under which license?

32. If you are interested in providing more information about this tool, please provide your email address...
(If you provide your email address someone working on the EURECA project may contact with additional questions)

33. In the context of the EURECA project, can you provide examples of any data sources that would be useful?

- ○ Yes
- ○ No

If yes, please specify

34. In what circumstances would this data be useful?

35. Is this data available now?

- Yes
- No

If no, when will the data be available?

36. If you or your institution is interested in providing data, tools or models to help the EURECA project, please give your email address...
(If you provide your email address someone from the EURECA project may contact you with additional questions)

37. Do you already use medical terminologies (e.g. SNOMED, LOINC, MedDRA) for your data or research?

- Yes
- No

If yes, which ones and for which application?

38. Do you intend to use medical terminologies for your data or research in the future?

- Yes
- No

If yes, which ones and for which applications?

39. Do you use medical standard specifications (e.g. HL7, open EHR) for your data or research?

- Yes
- No

If yes, which ones and for which application?

40. Do you intend to use medical standard specifications (e.g. HL7, open EHR) for your data or research in the future?

- Yes
- No

If yes, which ones and for which application?

41. Do you use biomaterial from biobanks for your research?

- Yes
- No

42. Is it important for your research with biomaterial and biobanks to have access to the anonymised clinical data of patients, which are stored in HIS, PACS, cancer registries, clinical trials or others?

- Yes
- No

If yes, please describe...

43. Can you please suggest email addresses for other people we should suggest for this questionnaire?

Contact 1

Contact 2

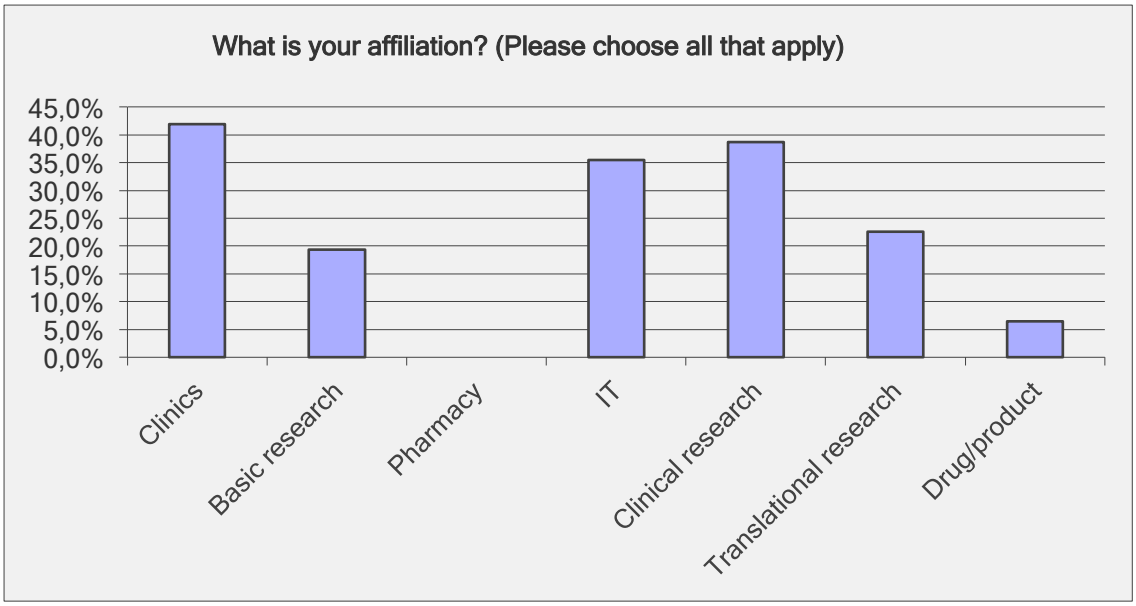
Contact 3

8 Appendix 2: Detailed results of the survey

Altogether 33 people did participate in the survey. Not all participants answered all questions. In the following section the answer to all questions are given:

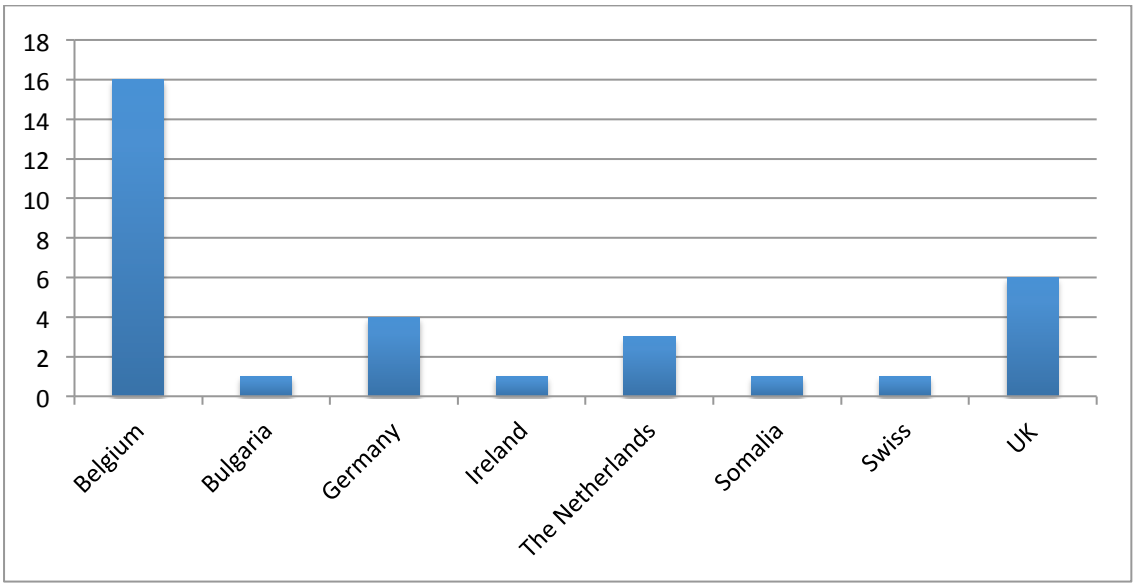
What is your affiliation? (Please choose all that apply)		
Answer Options	**Response Percent**	**Response Count**
Clinics	41,9%	13
Basic research	19,4%	6
Pharmacy	0,0%	0
IT	35,5%	11
Clinical research	38,7%	12
Translational research	22,6%	7
Drug/product development	6,5%	2
Other (please specify)		5
	answered question	**31**
	skipped question	**2**

Number	Response Date	Other (please specify)	Categories
1	Mai 29, 2012 12:11 pm	Cell transplantation	
2	Mai 21, 2012 3:31 pm	CISO (Chief Information Security Officer)	
3	Mai 21, 2012 2:59 pm	Data Manager (Data Centre)	
4	Mai 21, 2012 2:47 pm	Data Warehouse (Statistics)	
5	Mai 21, 2012 2:34 pm	Medical Management	



Which country do you live in?	
Answer Options	**Response Count**
	33
answered question	33
skipped question	0

Number	Response Date	Response Text
1	Mai 29, 2012 12:15 pm	Belgium
2	Mai 29, 2012 12:11 am	Belgium
3	Mai 29, 2012 12:05 pm	Belgium
4	Mai 29, 2012 11:57 am	Belgium
5	Mai 26, 2012 7:02 pm	Somalia
6	Mai 25, 2012 7:16 am	Ireland
7	Mai 23, 2012 10:05 pm	Netherlands
8	Mai 21, 2012 3:31 pm	Belgium
9	Mai 21, 2012 2:59 pm	Belgium
10	Mai 21, 2012 2:47 pm	Belgium
11	Mai 21, 2012 2:34 pm	Belgium
12	Mai 15, 2012 9:44 am	UK
13	Mai 13, 2012 9:02 pm	UK
14	Mai 11, 2012 7:14 am	Germany
15	Mai 10, 2012 12:23 pm	Switzerland
16	Mai 9, 2012 2:14 pm	Netherlands
17	Mai 9, 2012 12:46 pm	Belgium
18	Mai 9, 2012 12:20 pm	Belgium
19	Mai 9, 2012 12:09 pm	Belgium
20	Mai 9, 2012 11:49 am	Belgium
21	Mai 9, 2012 11:41 am	Belgium
22	Mai 9, 2012 11:22 am	Belgium
23	Mai 9, 2012 11:05 am	Belgium
24	Mai 9, 2012 10:01 am	Belgium
25	Apr 25, 2012 9:24 am	UK
26	Apr 25, 2012 8:13 am	UK
27	Apr 24, 2012 11:13 pm	Bulgaria
28	Apr 23, 2012 7:34 am	Netherlands
29	Apr 18, 2012 11:03 am	Germany
30	Apr 18, 2012 8:45 am	Germany
31	Apr 16, 2012 5:13 pm	UK
32	Apr 16, 2012 1:57 pm	Germany
33	Apr 16, 2012 12:27 pm	Oxford UK



Do you know what electronic health records (EHR) are?		
Answer Options	**Response Percent**	**Response Count**
Yes	97,0%	32
No	3,0%	1
	answered question	**33**
	skipped question	**0**

Do you know what personal health records (PHR) are?		
Answer Options	**Response Percent**	**Response Count**
Yes	75,8%	25
No	24,2%	8
	answered question	**33**
	skipped question	**0**

Do you think the re-use of personal data collected in a patient file is a good idea?		
Answer Options	**Response Percent**	**Response Count**
Yes	93,9%	31
No	6,1%	2
	answered question	**33**
	skipped question	**0**

Please explain what research you would be able to carry out if you had access to all the data collected in the patient record system of your institution...	
Answer Options	**Response Count**
	27
answered question	**27**
skipped question	**6**

Response Text	Categories
• Data about geriatric cancer patients: type of tumours, poly-medication, CGA (Comprehensive Geriatric Assessment), comorbidities	
• Efficacy, long term toxicity, genomic-clinical linkage, automated patient enrolment, data mining, Adverse Events automated detection	
• - I have already access - Necessary for translational research	
• The exchange of Tele-health home monitoring PHR data with local hospital EPR's	
• none. I would be able to support clinical genetics research by IT	
• - Controlling (waste, ...) - For patients: monitoring of long term evolution, and detection of silent pathologies	
• I already have access to the EHR of my institution	
• I have already access to the EHR of my institution, but too large amount of data, lack of documentation	
• I have already access to the EHR of my institution	
• phenotype genotype databases; response prediction, stratified medicine	
• Correlations between molecular markers/characteristics of disease and clinical course	
• more modelling of diseases, building of tools for decision support, enhance patient empowerment	
• integration, correlation studies	
• Rapid learning: i.e. building models to predict outcome and use the se prediction models in decision support systems	
• We could gather statistics from retrospective studies	
• First of all, I do already have access to the entire care EHR of my institution. However, much of this data is not structured or usable for research purposes due to lack of quality control. Cross checking and structuring would be needed before use for research. Examples of possible research themes would be : - compliance with guidelines - survival and disease-free-survival studies - epidemiology studies - statistics about use of various treatments - statistics about prevalence of various adverse events	
• Find which patient is eligible for which study to do some statistics (e.g. How many patients HER2/Neu+ do we have?, How many metastatic breast cancer do we have at our institution?)	
• All "historical" information (e.g. Chemotherapy, Radiotherapy, Hormonal therapy treatments) (Patient medical anamnesis) that allow us to update quickly the patient database.	
• - Follow up of patients - Drugs toxicity (phase IV)	
• I already have access to EHR	
• Analysis of association between clinical outcomes etc. with proteomics/ RNA/DNA profiles of tumour tissue research	
• The research would not change but it would take less time - e.g. perform linking of databases	
• for all that necessary	
• data integration and research on clinical decision support systems	
• analyse possible risk factors, analyse treatment options, individualised treatment therapies	

development
- outcome of cancer, new predictors, assessment of interventions on survival
- Outcome data from sarcoma

Can you please give examples of applications that would be improved by combining care data (EHR/PHR) and research data?	
Answer Options	**Response Count**
	27
answered question	**27**
skipped question	**6**

Response Text	Categories
• Patients who participated in clinical trials and toxicities outcome, CGA evaluation, death competing risks	
• - Quicker recruitment to trials - Faster and more comprehensive identification of potential patients for trials - Saving (time, cost, manpower) - Adverse Event automated detection	
• - Biomarker research: predictive, prognosis, surrogate marker (KI67: evaluate a study end point) Circulating biomarker research Extract diagnosis (dans EHR/Anatomopathology), survival data (EHR), treatment response (EHR) - Improve cancer biology comprehension, NGS (Next Generation Sequencing), GEP (Gene Expression Profiling) by structuring data	
• Care-plans for chronically ill (COPD; Heart Failure; Diabetes)	
• genetics, life style analytics, family track back	
• - Epidemiological studies (large amount of data)	
• CRFs: - Fulfil patients' medical history - Find Adverse Events - Fulfil demographic data	
• - Simplifying the filling of CRFs - Avoid duplication of tasks - Minimising of encoding errors - Coherence and validity of data	
• understanding of the clinical relevance and utility of whole genome sequence data	
• Better idea of prognostic factors or factors that predict response to treatment	
• Oncosimulator, decision support tools	
• follow-up, clinical relevance of research	
• 1) Validation of decision support systems based on clinical data in research data. 2) Matching a clinical patient to a past/open trial inclusion criteria to have outcome evidence/ include the patient.	
• Previous statistics may allow us to choose the "best treatment"	
• - Patient eligibility - Follow up of patients	
• - Management and use of biobanks. - filling electronic CRFs - trial eligibility criteria checking - monitoring of trial-drug-related adverse events	
• - Improve enrolment of patients (i.e. Find the right patients for the right studies) - Decrease errors (e.g. in recording laboratory results) - Time saving in data management and monitoring - Improve data recording	
• Limitations inclusion/exclusion criteria on the number of lines of different treatments received to compare with a database that would quickly display this "patient history"	
• Assessment of eligibility	
• To extract relevant information in the EHR to link to research question (biomarkers, eligibility) - (Time consuming)	
• the research process would be very much streamlined and more efficiently carried out,; clinical data from question 6 above would be available with much less effort	
• assigning patients to studies and trials, starting new trials, discover early side-effects	
• there have no possibility to loose the important details. More patients will be included in clinical research.	

- clinical decision support systems
- individualised patient treatment care, advise patients on how to reduce risks of relapse
- analysis of new drugs and their benefit in wider population, adverse drug reporting
- Post clinical trial long term follow-up, morbidity, late effects, fertility effects, and other illness

Do you think linking EHR/PHR data with clinical trial data would be useful in general?

Answer Options	Response Percent	Response Count
Yes	100,0%	30
No	0,0%	0
answered question		30
skipped question		3

Would linking EHR/PHR data with clinical trial data be useful for your research?

Answer Options	Response Percent	Response Count
Yes	82,8%	24
No	17,2%	5
answered question		29
skipped question		4

Why is data from research useful in your care setting?

Answer Options	Response Count
	20
answered question	20
skipped question	13

Response Text	Categories

- Organisation of clinical interdisciplinary care
- Drug development, improved treatments, prognosis, drug response prediction
- Apart from the U.K. Whole System Demonstrator the use of Tele-health data is largely unproven in the improvement of outcomes for those suffering chronic illness
- NA
- - For the follow up of patients - Proactive use of data from patients followed by medical practitioners for the follow-up of another patient (in particular for phase I and phase II)
- response prediction, clinical outcome data from clinical trials is more robust and clean then from routine care
- find new risk factors for stratification of patients with cancer to get better treatments
- direct impact on improving care and possibly saving time and money and effort
- a) Trial inclusion. b) Know if a patient 'fits' the trials on which a guideline/evidence-based-medicine is based. c) Benchmark trial outcome to local quality of treatment (on a population

of patient)).

- For the choice of a treatment
- - Personalising care
- Data from research if useful as a whole, in order to update treatment guidelines; However, this may not - in my opinion - be the result of automated processes. The main direct interaction between research data and the care environment would be, in my opinion, at the patient recruitment stage for trials.
- - To prepare information and education for the patient related to his treatment (reactions, answers to side effects) - "Preventive" anticipation of possible health problems
- - New therapeutic options for treatments - Identify sub groups of patients likely to respond to a treatment - Improving diagnosis
- we are looking for personalised medicine solutions in cancer; these patients may fail on current standard treatments
- NA •
- It is important for individualisation of therapy
- Clinical research provides important data. To ease the way of research into daily clinical practice.
- choosing right drugs for cancer patients
- Sarcoma are rare, patients are young, we cannot currently monitor independently of clinical trials

Do you ever access data in your research system and your EHR/PHR system for the same task?

Answer Options	Response Percent	Response Count
Yes	20,0%	6
No	80,0%	24
If yes, for which tasks		7
answered question		30
skipped question		3

If yes, for which tasks	Categories

- - Biomarker research: predictive, prognosis, surrogate marker (KI67: evaluate a study end point); Circulating biomarker research; Extract diagnosis (dans EHR/Anatomopathology), survival data (EHR), treatment response (EHR); - NGS (Next Generation Sequencing), GEP (Gene Expression Profiling)
- for getting basic laboratory results and clinical outcome data; confirmation of diagnosis
- All kind of patients' data (e.g. medical history, test results)
- Yes, for datasets provided within INTEGRATE (including information from anatomical pathology, clinical biology and pharmacy laboratories)
- Research in EHR is manual because of unstructured data compared to research system that is more structured
- Refer to EHR to capture data items for retrospective studies (ethically approved, with patient consent); this involves reading then re-entering into the research setting, as opposed to an electronic link or download (therefore less efficient than ideal)
- I don't personally data are always prepared for me, but the preparation and curation takes a long time

Do you ever query the research data when treating a patient?		
Answer Options	**Response Percent**	**Response Count**
Yes	30,8%	8
No	69,2%	18
	answered question	**26**
	skipped question	**7**

Do you ever query the EHR while working on a research question?		
Answer Options	**Response Percent**	**Response Count**
Yes	57,1%	16
No	42,9%	12
	answered question	**28**
	skipped question	**5**

Are you aware of any problems with electronic health records?		
Answer Options	**Response Percent**	**Response Count**
Yes	78,6%	22
No	21,4%	6
If yes, please specify:		22
	answered question	**28**
	skipped question	**5**

If yes, please specify:	Categories

- - Rigid structure: not iterative; - Free text narratives; - Lack of standards
- - Unstructured data; - Problems to find information (Biological sample records)
- There is no unique patient ID in Ireland; there is no nationally accepted standards for EHR interoperability; patient data quality is an issue; ownership of data;
- trust in the source of the data
- - Unstructured history (dualité structuré-non structuré); For this purpose it could be useful to incitate directly the physicians to structure the patient record by guiding them in the encoding process (e.g. "NOCTIBUS" project); - Organisation of the EHR (problems grouping oriented OR succession of events oriented)
- - Lack of documentation (Wiki); - Data structure; - Problems of anonymisation for student projects
- - Problem of access to preliminary reports before validation, and use of these preliminary data; - Security of data privacy; - Informed consent may change over time, difficulty of ensuring the patient's consent; - Mix of information in some cases with data from third parties (e.g. for a transplant): problem with an access to data that do not belong to a patient without consent, or problem with automatic extraction of data
- traditional databases are not linked with each other and cannot feed into EHR; Lab IT systems are not web-based, do not allow integration with different departments for on-line sample tracking and reporting

- legal issues
- data security
- a) They are not semantic interoperable, you can't easily exchange data.;
  b) Some vendors protect the syntax (database schema) in such a way that access is difficult
  c) There are many of them, especially between countries
  d) There is no standard being used for the storage of EHR (unlike e.g. DICOM in the imaging world)
  e) They are expensive
  f) They are often not very tuneable to the local processes unless you pay a lot.
- Systematic anonymisation of a result of a query in IJB EHR is problematic, is it necessary inside the Institute?
- The lack of trial-like quality control for routine EHR data, and the lack of structure are the main impediments for statistical / automated use of EHR.
- - No research possible for PDF documents
  - No harmonisation in documents. Everybody should enter data in the same way. We need 1 page for Adverse Events, 1 for concomitant medication, 1 for history, and 1 template for consultation
  - No possibility of sending results for another hospital
- - Corrupted privacy
  - Encoding errors
- - Confidentiality problems, security (access to data, data anonymisation, etc.)
  - Formats problems in data exchange (standards exist but are not easily "computable")
  - Divergence of codification norms of medical concepts (correspondence problems, choice of codes for a specific norm, etc.)
- - Need of more powerful research tool to query the EHR
  - Better interoperability between EHR applications
  - Structuring more the EHR
- Unstructured data make it not easy to extract some relevant information for research and clinical care (the best way would be to do as the DataBase "Tour Mammaire" used in the MDT)
- MAYBE useful data items are available in free text fields. Would be great to have core data in single select format etc.; also consistent definitions of term used in this core items
- it is not ever available and not full at all
- data protection laws
- confidentiality; accuracy

How important do you think data security issues are in this context?

Answer Options	Response Percent	Response Count
Very important	80,0%	24
Important	6,7%	2
Less important	6,7%	2
Not important	6,7%	2
I don't know	0,0%	0
answered question		30
skipped question		3

If all security and privacy issues were solved, would you see value in sharing data with other organisations?		
Answer Options	**Response Percent**	**Response Count**
Yes	96,7%	29
No	3,3%	1
	answered question	**30**
	skipped question	**3**

Would any new research be possible if you had access to all the data from both research and care collected in your institution and/or from other institutions?		
Answer Options	**Response Percent**	**Response Count**
Yes	81,5%	22
No	18,5%	5
If yes, please describe the possible research		20
	answered question	**27**
	skipped question	**6**

If yes, please describe the possible research	Categories

- Already mentioned (e.g. denutrition in order patients participating in clinical trials)
- - Statistics on population
- Linking genetic data with clinical outcomes
- - Accelerate and facilitate with more data
- Critical size of database
- - Meta analysis
- establishing the significance of rare variants or looking at rare patient groups. This is important n the context of personalized medicine
- better modelling of diseases
- more and more meaningful data
- a) We could build prediction models on much more and more diverse patients making them both more accurate and more applicable to a wider population of patient (e.g. ethnic) and treatment options (e.g. advanced and basic treatment such as in developing countries)..b) we could have faster trials as we would be able to include more patients c) the trial data quality would be improved and quicker as all data from the clinical EHR would be available.
- - We could have more statistics on treatment rates (survival or death rates), so that we could improve or adapt treatment (before research publications) - To know the intermediate results of a research, so that we could adapt our treatment without waiting for the conclusions of a research when the partial results make already clear the future conclusions
- - Biobank-related (e.g. genomic) studies - All types of retrospective studies, including side-effects finding, treatment comparisons, guidelines quality and compliance assessment, epidemiological studies (survival, disease-free survival, ...)
- - Better data quality - Time saving
- Sharing information (lab results, Radiotherapy, Scanner) directly with an EDS (Electronic Data System): e.g. eCRF (electronic Case Report Forms) if data are anonymised
- - Safety: utility of having 5-10 years data after the clinical trial for the follow up - Safety with

PHR: for late-toxicity - Retrospective studies - Faster selection of patients
- Best statistics, more robust results
- Data mining for associations that could lead to new hypotheses. Also current ideas could be tested more efficiently , releasing the resource to cover more topics.
- as mentioned in question above (trials side-effects, mining of trials, meta-analyses would be easier, rare diseases could be studied as well))
- organ specific response to therapy, organ specific sites of relapse
- Full appraisal of treatment and patient outcome

Would any new research be possible if you had access to much larger amounts of data then you currently have access to?

Answer Options	Response Percent	Response Count
Yes	76,0%	19
No	24,0%	6
If yes, please specify the possible research		15
	answered question	**25**
	skipped question	**8**

If yes, please specify the possible research	Categories

- Meta analysis
- - Accelerate and facilitate with more data
- identify the statistical significance of rarer variants
- better collaboration between research groups
- extensive validation studies, which will be more convincing
- Identify subset of patients that do worse than expected and for which targeted therapies would be beneficial.
- - Biobank-related (e.g. genomic) studies - All types of retrospective studies, including side-effects finding, treatment comparisons, guidelines quality and compliance assessment, epidemiological studies (survival, disease-free survival, ...)
- Better data quality.
- - Rare mutation patients selection - More significant studies for more powerful analysis
- Best statistics, more robust results
- Expect that a larger sample size would benefit many analyses
- genomic in rare diseases, mining of clinical trials and more powerful meta-analyses which are important for rare events (e.g. some radiotherapy late morbidity)
- Not new but more valid due to the larger amounts of data.
- greater significance of drug response data
- Recruit patients by histology across Europe independent of a referring doctor

Do you currently ever combine research and care data?		
Answer Options	**Response Percent**	**Response Count**
Yes	60,7%	17
No	39,3%	11
If yes, please explain why and when		15
	answered question	**28**
	skipped question	**5**

If yes, please explain why and when Categories

- (idem 11) - Biomarker research: predictive, prognosis, surrogate marker (KI67: evaluate a study end point) Circulating biomarker research Extract diagnosis (dans EHR/Anatomopathology), survival data (EHR), treatment response (EHR) - NGS (Next Generation Sequencing), GEP (Gene Expression Profiling)
- Because I work for a data centre that collect data from care activities from the EHR, and I use these data to fulfil CRFs for clinical trials.
- we combine whole genome data with clinical data, but this is done manually on small numbers of patients
- to analyse clinical trials and find correlations between outcome and molecular biological findings
- We learn prediction models from clinical data and validate them in research data.
- - For the benefit of the patient
- - for biobanks - for the internal cancer registry - as a feedback loop for oncologists in the care environment so that the current status of a trial, or of a patient in a trial, is known
- To increase enrolment of patients by finding eligible patients for adapted studies.
- - Patient file preparation for a future recruitment in a study, - Use data from both research and care to identify patient for a study
- Data Mining from textual data in structured data. Standardisation of norms (in particular HL7v3-CDA).
- - For lab research - Look for prognosis
- Retrospective studies on tissue surplus to diagnostic needs: use associated standard treatment and outcome data (ethically approved and patient consented)
- yes but not personally, they are combined for me
- yes, outcome of cancer, gene array, response to therapy
- Local pathology and clinical database

Is your current work hampered by the fact that your EHR systems and research data are independent from each other?		
Answer Options	**Response Percent**	**Response Count**
Yes	60,7%	17
No	39,3%	11
If yes, please explain how		14
	answered question	**28**
	skipped question	**5**

If yes, please explain how	Categories

- Better collaboration between research and patient data
- Research data is only a subset of a much larger dataset (i.e. EHR)
- - Time consuming - Limiting human mistakes (by automating manual actions)
- - This kind of work needs a manual intervention, where many stakeholders have to intervene according to their rights towards the data - Problems of mapping of data coming from different systems (data correlation from different standards/formats)
- all data has to be entered and re-entered manually leading to clerical errors and inefficient use of time
- We need to do a lot of interoperability work to combine trial data and clinical data. It takes a lot of time.
- It's difficult to know if there is ongoing research into a disease
- - Loss of eligible patients - Waste of time to find patients - Waste of time with patients not eligible - Waste of time in data management - Risk of error
- increasing complexity of procedures for: - security - anonymisation - connectivity
- When we have to come back to Oribase to extract clinical data
- Less efficient, as per previous answers
- mostly in terms of access to new datasets and time it takes to
- It is more complicated to get access to specific records if the patient is not treated at your institution/department etc.
- slows down matching genetics with outcome

Are there other systems and/or data that could be combined with research and care data to help your work?

Answer Options	Response Percent	Response Count
Yes	62,1%	18
No	37,9%	11
If yes, please specify which systems or data		19
answered question		29
skipped question		4

If yes, please specify which systems or data	Categories

- The sharing of more data between Primary Care and Secondary Care
- life style data
- - Medical devices (Groupes / Shadow Groupes)
- CRFs papers, CRFs online
- Would like to have direct access to DB tables
- - Systems to assess toxicity (grades) and tumor responses - System to screen clinical studies in which a patient could enter based on its clinical data (tumour, patients)
- clinical lab data also has to be combined with clinical outcome data and research findings
- matching
- Biobanks. Literature data. Ontologies. Animal research data. Cell or other basic research data. Epidemiology data. Registry data. Financial data (e.g. health insurers reimbursement data).
- The EHR environment could integrate councils of Pubmed publications to keep up-to-date

with new research results
- Planning management tool for the management of patient treatments (e.g. with a planning like in MS Project)
- - Improve our electronic system to find the information more easily - Electronic prescription - Send electronic data (e.g. scanner, labo)
- Automatic and computerised encoding, follow-up, update of different events occurring in patients' life (e.g. side effects)
- A set of integrated medical applications
- Having direct access to the National Registry could help us to keep our research data up-to-date (e.g. survival information, date of death), and thus improve our research results.
- survival data from GP, following discharge from hospital setting
- public databases (trials, genomic)
- imaging, genetic data ,lab data, survival data
- EORTC and EuroEwing

Are there any additional sources of data or knowledge from outside your organisation that would also help your work? (Publications, databases, ontologies etc)

Answer Options	Response Percent	Response Count
Yes	71,4%	20
No	28,6%	8
If yes, please specify		19
answered question		28
skipped question		5

If yes, please specify	Categories

- For elderly cancer problem in other institutions
- - Query the National Registry
- - Data from completed clinical trials - Publicly available annotated genomic datasets
- - Private publications
- - Linking the clinical study and relevant publications that were the basis of the study, or published from the study or similar studies
- publically available genomic variant databases have to be linked with our research data (dbSNP; COSMIC)
- KEGG database, SAE/SUSAR databases, clinical trial databases,
- GUI, workflow
- See previous answer.
- - New guidelines from the "UpToDate" website - New publications from the "PubMed" website
- A cancer registry that can be consulted
- Metadata, especially: - published data about relevant drugs and pathologies - catalogue of open trials
- Knowledge of ontologies
- EUDRAVIGILANCE (European Union Drug Regulating Authorities pharmacoVIGILANCE): European database.
- GP surgeries (survival/relapse data after discharge if patient not re-referred), cancer registries
- databases (genomic/genetic), publications, ontologies (GO, cancer ontology, clinical trials ontology)
- professional organisations and patients organisations

- Publications are important sources of new knowledge. New evidence based knowledge might be used much faster in clinical practice if there were direct access and notification systems
- EORTC

Can you please describe a tool that would help your work within the context of the EURECA project?	
Answer Options	**Response Count**
	19
answered question	**19**
skipped question	**14**

Response Text	Categories

- Any would be interesting I think.
- Tool to improve recruitment to clinical trials
- - Automatic extraction tool to structure data - Patient screening for research studies eligibility - Follow up of patient
- Common international freely available; medical terms; drugs formulary; standards for PHR construction and content
- - Central index to manage all available data within the project
- A generic database for all clinical trials.
- - A tool that could extract chronological information from textual and structured data
- We need a relational research database that can extract data from different NHS clinical and laboratory databases and link it to research results
- annotation tool that allows easy and fast data labelling for search engines, powerful search engine
- a) Tool to convert my EHR structured and unstructured data in such a way that someone else can use it without needing to know my local data specifics. b) Tool to match my individual patient data to an external system to give decision support (e.g. guideline, prediction models, trial inclusion, patientslikeme.com, other hospitals that have experience in treating this particular patient)
- - A tool to access other institutions' data - A global cancer registry available for consultation - A better search tool in the EHR
- NLP extraction of disease staging, treatment and adverse events from the EHR
- Text data extraction module to obtain structured data
- Data mining tools (more global and external) for: - Pharmacovigilance - Laboratory data that we need to standardise
- any tool that would permit Extraction, Manipulation Loading into research databases
- build/lay the foundations of electronic data base
- If you were searching for patients suffering from a specific ICD 10 coded disease with certain inclusion criteria a searching tool that might selectively choose only those patients you need would ease the procedure of patient recruitment
- link molecular variables in lab,to clincal variable in EPR, survival data in national database
- Don’t understand

Are there any specific requirements that the tool mentioned in your previous answer should have?	
Answer Options	**Response Count**
	17
answered question	17
skipped question	16

Response Text	Categories
• CGA (Comprehensive Geriatric Assessment) • - Integrated into current clinical EHR system - Iterative - Suggest suitable clinical trials to doctors treating a particular patient • Should be platform agnostic; available in multiple languages; • Data traceability (with responsibility management) • To have generic functionalities (e.g. same user interface) • - Informed consent of patients - Data security • it should be real time and able to interface with many different databases, it has to be searchable, it has to have an analysis function e.g.: cumulative results, trends, overall response rates • It should run with minimal user interactions. It should be free. It should be real-time. It should run at many other institutions. It should be inside my firewall under my control. • Output should be under standardized, interoperability format (i.e. HL7 + SNOMED) • - User friendly - Easy to use • Advanced text pre-treatment (removal of parasites, substitution of related terms, standardisation of numbers and dates, handling of diacritics, acronyms, abbreviations, etc.) • To identify a same patient in different data sources • Upload every kind of results from research in the system in order that it can be used in the future (eligibility, follow up of patient) • Be aware of issues between research (universities) and care(hospitals) concerns about security confidentiality, include very clear data item to confirm patient consent to use for research (database field with data if consent??) • not sure • Can not imagine • ?	

Are you aware of any existing tools that are used in this area?		
Answer Options	**Response Percent**	**Response Count**
Yes	42,3%	11
No	57,7%	15
If yes, can you please give the tool's name?		11
answered question		26
skipped question		7

If yes, can you please give the tool's name?	Categories

- - Netcord - Worldwide registry de cordon - BMDW (Bone Marrow Donnor Worldwide) - Worldwide registry - NMDP (Normal Marrow Donnor Program) - USA registry - MDPB (Marrow Donnor Program Belgium)
- - SAGE Bionetworks - TRANSMART - BIOMART - Geoportal - caBIG (NCI) (e.g. caTissue)
- Snomed; FDB; HL7
- Altova Map Force (Extract Transform Load - ETL)
- MOFFITT Cancer Center (http://www.moffitt.org/)
- oracle; myrth
- a) Soarian QM http://tinyurl.com/ce7soc5 , b) eurocat environment (www.eurocat.info), c) openphacts (http://www.openphacts.org/) d)
- Part of IJB's internal EHR allows structuring of documents (including pathology and MDT reports, lab results and drug administration data) . The specialized tool SM2008 deals with Clinical Document under the HL7 CDA format
- Documentation of SNOMED
- - caBIG - Seer DataBase (for collecting clinical data in USA)
- not public tools as such, various research groups develop bespoke systems in specific settings

Which specific tasks is this tool used for?

Answer Options	Response Count
	11
answered question	11
skipped question	22

Response Text	Categories

- Web platform for cell registry
- Data Sharing plateforms concerning clinical and genomic data
- Planned interchange of data
- PHR, clinical decision support, analytics, public health
- "Extract Transform Load"
- - Proactive follow up of patients - Sequencing of the tumour
- not yet in use
- Convert local EHR data into a semantic interoperable dataset
- Data transformation and formatting.
- Text pretreatment
- Data base of a large amount of patients, a kind of registry, to do more research based on patients data

Which groups of professions within your institutions use these tools?	
Answer Options	Response Count
	6
answered question	6
skipped question	27

Response Text	Categories
• Transplantation coordinator	
• Researchers, Bioinformaticians, statisticians	
• Health Informatics researchers	
• nobody	
• researchers	
• All medical and care personnel. Medical secretaries and some specialized caregivers use SM2008.	

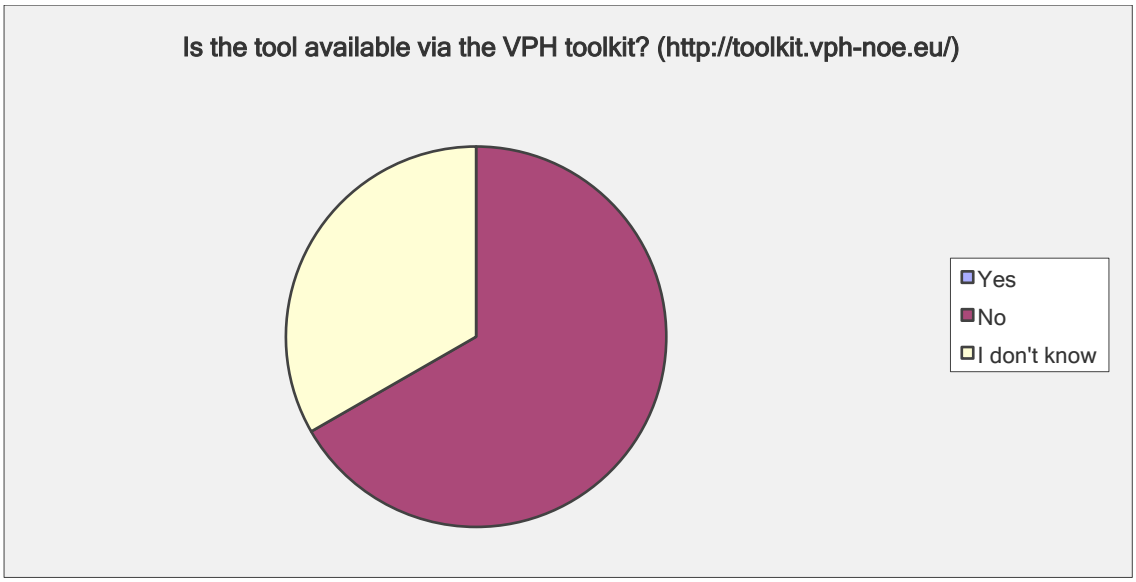
Are you aware of any problems with this tool, or how it could be improved?		
Answer Options	Response Percent	Response Count
Yes	75,0%	3
No	25,0%	1
If yes, please specify		3
answered question		4
skipped question		29

If yes, please specify	Categories
• Cost; complexity and moving standards	
• It is too expensive.	
• Insufficient structuring. Very few items of informations are coded.	

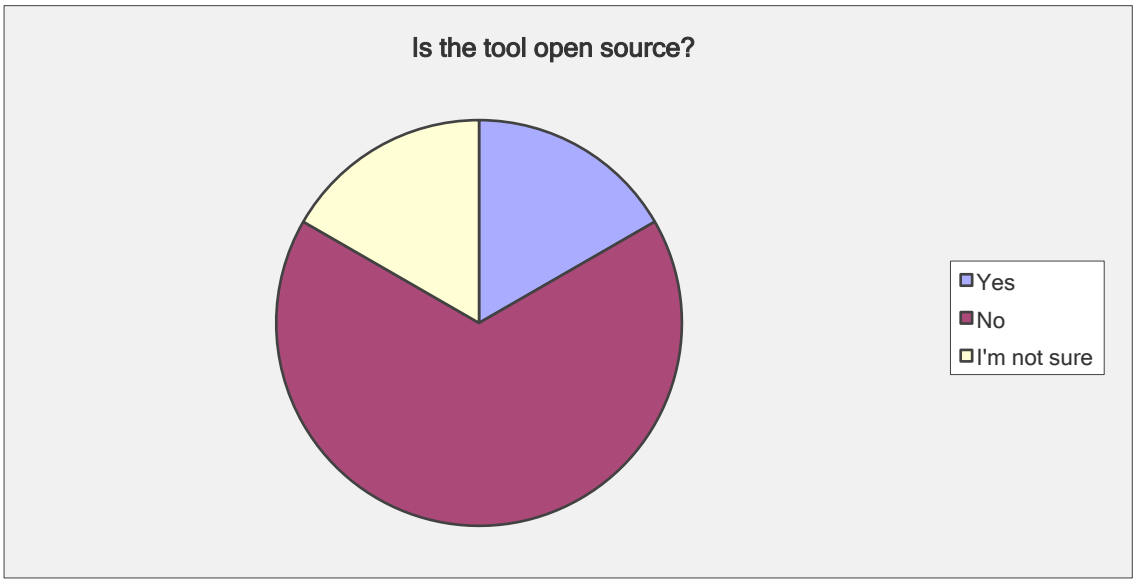
Which data is needed by this tool to do its job?	
Answer Options	Response Count
	5
answered question	5
skipped question	28

Response Text	Categories
Clinical and genomic data	
HL7 bRIM modelled	
see above	
EHR data	
Raw medical information (as text).	

Is the tool available via the VPH toolkit? (http://toolkit.vph-noe.eu/)		
Answer Options	**Response Percent**	**Response Count**
Yes	0,0%	0
No	66,7%	4
I don't know	33,3%	2
	answered question	**6**
	skipped question	**27**



Is the tool open source?		
Answer Options	**Response Percent**	**Response Count**
Yes	16,7%	1
No	66,7%	4
I'm not sure	16,7%	1
If yes, under which license?		0
	answered question	**6**
	skipped question	**27**



If you are interested in providing more information about this tool, please give your email address... (If you provide your email address someone working on the EURECA project may contact you with additional questions)	
Answer Options	**Response Count**
	2
answered question	2
skipped question	31

Response Text	Categories
niemans@pbsolo.nl	
philippe.hennebert@bordet.be	

In the context of the EURECA project, can you provide examples of any data sources that would be useful?		
Answer Options	**Response Percent**	**Response Count**
Yes	42,9%	9
No	57,1%	12
If yes, please specify		11
answered question		21
skipped question		12

If yes, please specify	Categories

- Correspondence tables between medical terminologies/classifications (if it exists)
- National cancer registry (Belgium)
- Case-notes, HICOM, filemaker, LIMS, EPR (Cerner); ARIA; excel

- nephroblastoma datasets
- inclusion of publicly available data, easy link and integration
- Our EHR, PACS, radiotherapy treatment information system, biobank
- All parts of IJB EHR, especially : - lab results - pathology and imaging reports - drug administration reports - chemo drug prescriptions - visit notes - discharge reports
- - MDT relative documents - SNOMED-CT code explorer
- CRFs (Clinical Data)
- Academic clinical trials
- EORTC Sarcoma trials EuorEwing trail of Ewing sarcoma long term data retrieval

In what circumstances would this data be useful?	
Answer Options	**Response Count**
	6
answered question	**6**
skipped question	**27**

Response Text	Categories

- - To complete patients' data - For statistics
- for research and eventually clinical practice
- according to a legal framework
- in all cases
- - Disease typing and staging - Getting the treatment type as coded data - Assessing adverse events
- In defining long term effects of treatments

Is this data available now?		
Answer Options	**Response Percent**	**Response Count**
Yes	28,6%	2
No	71,4%	5
If no, when will the data be available?		1
	answered question	**7**
	skipped question	**26**

If no, when will the data be available?	Categories

- as soon s legal issues are solved

If you or your institution is interested in providing data, tools or models to help the EURECA project, please give your email address... (If you provide your email address someone from the EURECA project may contact you with additional questions)	
Answer Options	**Response Count**
	5
answered question	**5**
skipped question	**28**

Response Text	Categories
anna.schuh@nhs.net	
graf@uks.eu	
andre.dekker@maastro.nl	
philippe.hennebert@bordet.be	
bass.hassan@path.ox.ac.uk	

Do you already use medical terminologies (e.g. SNOMED, LOINC, MedDRA) for your data or research?		
Answer Options	**Response Percent**	**Response Count**
Yes	50,0%	12
No	50,0%	12
If yes, which ones and for which application?		11
answered question		**24**
skipped question		**9**

If yes, which ones and for which application?	Categories

- - SNOMED: for EHR - ICD9: for EHR (used for INAMI) - ICD9-CM (procedure oriented) - Codification for the nursing record (DIRHM: Dossier Infirmier Résumé Hospitalier Minimum) - ALBERT (additional features to LOINC) (data exchange with General Practitioners - Conversion tables between different laboratories)
- - Report of Adverse Events - To anonymise DICOM data
- - Anatomopathology - Structuring data in the Medical Record (allergies, infections, contaminations) - Medical Abstract (RCM, Résumé Clinique Minimum)
- MedDRA, ICD 10, LOINC,
- ICD and CTCAE and NCI Thesaurus
- LOINC is used as a possible transformation target for lab results, but only for data exchange in research projects (e.g. INTEGRATE). LOINC is also used in order to categorize CDA sections. SNOMED is used extensively for pathology reports, and is tentatively being used for MDTs reports.
- - SNOMED-RT - SNOMED-CT - LOINC
- MedDRA, ICD
- - MedDRA: medications for the disease recording, treatment, AEs - CDISC: design CRFs terms to have common language
- SNOMED, TNM cancer staging, RECIST response, CTC toxicities, categories of prognostic factors widely used in literature: receptor positivity
- MedDRA

Do you intend to use medical terminologies for your data or research in the future?		
Answer Options	**Response Percent**	**Response Count**
Yes	54,2%	13
No	45,8%	11
If yes, which ones and for which applications?		11
	answered question	**24**
	skipped question	**9**

If yes, which ones and for which applications?	Categories

- MedDRA
- To extract more relevant data
- don't know
- as above
- SNOMED, Radlex, NCI, ICD
- Non-oncological diagnoses and allergies should be coded in SNOMED in the future. Likewise for surgical procedures.
- CTCAE (Terminology for data recording)
- - TrialDbase (software related to the management of clinical studies) - JCMO (software related to the management of CMO) - CDiAMIC (structured data extraction in anatomical pathology reports)
- For academic trials
- as above
- who did?

Do you use medical standard specifications (e.g. HL7, open EHR) for your data or research?		
Answer Options	**Response Percent**	**Response Count**
Yes	21,7%	5
No	78,3%	18
If yes, which ones and for which application?		6
	answered question	**23**
	skipped question	**10**

If yes, which ones and for which application?	Categories

- - Be careful to respect the published standards (e.g. Z code in HL7v2, the patients' ID is the one of the driving license in the HL7v2 standard definition) - HL7: Standardisation of the process of object creation, but non-standardisation of the objects themselves
- don't know
- HLH 7, CDISC ODM,
- DICOM and HL7
- Discharge letters, endoscopy reports, visit notes, surgery reports, part of the MDT reports are formatted using HL7 v3 CDA. Most medical data (including lab results ...) are being

exchanged between systems using HL7 v2.5 messages.
- HL7v2, HL7v3

Do you intend to use medical standard specifications (e.g. HL7, open EHR) for your data or research in the future?		
Answer Options	**Response Percent**	**Response Count**
Yes	27,3%	6
No	72,7%	16
If yes, which ones and for which application?		7
	answered question	**22**
	skipped question	**11**

If yes, which ones and for which application?	Categories

- - Clinical laboratory data - Identification information data of patients
- don't know
- as aove
- DICOM, HL7, openEHR, AIM
- Most other reports (i.e. imaging, ...) will be converted and/or generated as CDAs.
- HL7v2, HL7v3
- ?

Do you use biomaterial from biobanks for your research?		
Answer Options	**Response Percent**	**Response Count**
Yes	39,1%	9
No	60,9%	14
	answered question	**23**
	skipped question	**10**

Is it important for your research with biomaterial and biobanks to have access to the anonymised clinical data of patients, which are stored in HIS, PACS, cancer registries, clinical trials or others?		
Answer Options	**Response Percent**	**Response Count**
Yes	54,2%	13
No	45,8%	11
If yes, please describe...		6
	answered question	**24**
	skipped question	**9**

If yes, please describe...	Categories

- To analyse results (idem 11) - Biomarker research: predictive, prognosis, surrogate marker (KI67: evaluate a study end point) Circulating biomarker research Extract diagnosis (dans EHR/Anatomopathology), survival data (EHR), treatment response (EHR) - NGS (Next Generation Sequencing), GEP (Gene Expression Profiling)
- - To determine the feasibility of a study - To test new methods in laboratory
- Biological markers improve prediction models for cancer
- Information relevant to translation research may be found in: - the local cancer registry - the national cancer registry
- Otherwise we can't use biomarkers/validate research hypothesis

Can you please suggest email addresses for other people we should suggest for this questionnaire?

Answer Options	Response Percent	Response Count
Contact 1	100,0%	3
Contact 2	100,0%	3
Contact 3	66,7%	2
answered question		3
skipped question		30

Contact 1	Contact 2	Contact 3
jim.davies@cs.ox.ac.uk	john.skinner@ohis.ox.ac.uk	tim.bradford@ouh.nhs.uk
philippe.lambin@maastro.nl	deasyj@mskcc.org	bharat.rao@siemens.com
annemarie.weissenbacher@i-med.ac.at	david.jara@i-med.ac.at	

9 Appendix 3: Structured interview

Interview with Stakeholders

EURECA (**E**nabling information re-**U**se by linking clinical **RE**search and **CA**re) is a collaborative project that is funded by the European Commission under the 7th Framework Programme (Grant agreement no: 288048).

The project 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:

- 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,
  - avoiding the current need for multiple data entry in the various clinical care.
- Allow data mining of longitudinal EHR data for early detection of patient safety issues related to therapies and drugs,
- Allow for faster transfer of new research findings and guidelines to the clinical setting (from bench-to-bedside),
- Enable healthcare professionals within a legal and ethical framework to extract in each patient's case the relevant data out of the overwhelmingly large amounts of heterogeneous patient data and treatment information.

The project aims to achieve semantic interoperability between EHR and clinical trial systems that is consistent with existing standards despite the heterogeneity of the various sources. EURECA will develop solutions that fulfil the data protection and security needs and the legal, ethical and regulatory requirements related to linking research and EHR data.

We are interested in your opinion about the project in general and specifically we would like to know from you how to improve the EURECA environment.

We would like to thank you in advance for participation in the interview.

Questions

What is your affiliation?

O clinics O basic research O pharmacy O IT

O patient

In which country do you live? _____

Are you aware of any problems with electronic health records (EHR)?

Which problems are you facing in your research in relation to EHRs?

What are your specific needs in this area?

- Which of them are the most relevant needs?

What are general needs for you in this area?

What needs to be changed?

Can you describe tools that you would like to work with in relation to EHR?

- Which requirements are necessary for such tools?

Can you suggest other people we should contact?