ICT-2011-288048

EURECA

Enabling information re-Use by linking clinical Research and CAre

IP
Contract Nr: 288048

Deliverable: D9.1 Report on the development environment and on the available test data

Due date of deliverable: (10-31-2012)
Actual submission date: (01-28-2013)

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

Responsible WP: Philips

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
Cyril Krykwinski	IJB	cyril.krykwinski@bordet.be
Diane Van Vyve	IJB	diane.vanvyve@bordet.be
Norbert Graf	UdS	Norbert.Graf@uniklinikum-saarland.de
Andre Dekker	MAASTRO	andre.dekker@maastro.nl
Keyur Mehta	GBG	Keyur.Mehta@germanbreastgroup.de
Francesca Buffa	UOXF	francesca.buffa@imm.ox.ac.uk
Nikolaus Forgó	LUH	nikolaus.forgo@iri.uni-hannover.de
Magdalena Goralczyk	LUH	goralczyk@iri.uni-hannover.de
Bennet Lodzig	LUH	lodzig@iri.uni-hannover.de
Haris Kondylakis	FORTH	kondylak@ics.forth.gr
Lefteris Koumakis	FORTH	koumakis@ics.forth.gr
Kristof de Schepper	Custodix	kristof.deschepper@custodix.com

0.2 Documents history

Document version #	Date	Change
V0.1	01.10.2012	Starting version, template
V0.2	04.10.2012	Definition of ToC
V0.3	22.11.2012	First complete draft
V0.4	14.12.2012	Integrated version (send to WP members)
V0.5	02.01.2013	Updated version (send PCP)
V0.6	16.01.2013	Updated version (send to project internal reviewers)
Sign off		Signed off version (for approval to PMT members)
V1.0	28.01.2013	Approved Version to be submitted to EU

0.3 Document data

Keywords	
Editor Address data	Name: Cyril Krykwinski Partner: IJB Address: 121 Bd de Waterloo, 1000 Bruxelles Phone: +32 (0)2.541.36.37 Fax: E-mail: cyril.krykwinski@bordet.be
Delivery date	28.01.2013

0.4 Distribution list

Date	Issue	E-mailer
28.01.2012	1.0	Benoit.ABELOOS@ec.europa.eu
		INFSO-ICT-288048@ec.europa.eu

Table of Contents

List of Tables

1 Introduction

The objective of the first task of WP9 is to build the EURECA development and testing environment (e.g. “surrogate” databases) by coordinating all efforts that need to take place locally at each pilot site to provide access to suitable schema and instance-level datasets to be used by the project prototypes. These efforts need to be done early enough in the project implementation period in order to develop appropriate mappings between the canonical models and the EURECA core data sets, which are required input for WP4. This development is essential to coordinate efforts with the IT departments of all pilot sites, and to elaborate the information models characterising both electronic health record (EHR) systems and Clinical Trial (CT) systems of the pilot sites, but also the data necessary for testing and validating the interoperability framework and the EURECA tools and services.
These test data will be used in compliance with legal and ethical requirements described in WP7.

The principal pieces of information on datasets so far are the following ones:

- High level description of the dataset
- Data types, size of each data item, volume (size of dataset, i.e. number of items, number of patients), ranges
- Structure (free text, semi-structured, structured)
- Formats/ how the data will be exported/ standards
- Information on who is responsible for the data (e.g. lab), who manages the data, who enters data, where, how/who exports it
- Location of the data (now and in the future)/the systems when they are available
- Meta data (description data)

This report aims on describing the datasets available for the development environment and on the steps taken to generate the datasets and to comply with the legal EURECA framework.

1.1 Structure of the deliverable

Section 2 briefly introduces the pilot study by exemplifying several scenarios that have been chosen out of the scenarios that are described within WP1[1]. This pilot study supports the implementation of the consultation recommender service (2.1), the trial recruitment service (2.2) and the patient diary service (2.3). Additional scenarios will be selected in the second year of the project and implemented part of the pilot study. Section 3 describes the available test data of all pilot sites involved in the EURECA project, describing data from EHR systems as well as the TOP trial, a completed trial whose data are already available. Section 4 concisely describes the general legal framework, and the main steps that lead the clinical partners towards the approval of local legal authorities.

[1] EURECA project, “D1.2 Definition of relevant user scenarios based on input from users,” due in November 2012.

2 The pilot study

2.1 Data mining of consultation

- High-level description of the tool:

The data mining of consultation tool supports trial chairman in answering consultation requests both from clinicians as well as patients. The tool compares new, i.e. unanswered, consultation requests to old, i.e. already answered, consultation requests. For each unanswered consultation request the most similar already answered consultation requests will be determined and shown to the trial chairman. The answers to these top similar consultation requests might be in part or as a whole building blocks for the answer of the yet unanswered consultation request. The system is self adaptive and will learn form the feedback of the trial chairman to improve the quality of the suggestions.

- Types/amount of data that are needed for the development, the test and upcoming demonstrator of the tool:

At least 500 pairs of consultation requests and answers will be needed for the development and test of the tool. Prof. Dr. Norbert Graf from the hospital of the university of Saarland will provide this data. The data will be available as soon as it has been anonymized.

- How the tool will interact with such data:

The data will be stored in a data base. A description of the original data base is provided in chapter 3.2. The tool will query the data base to obtain the data, store the feedback from the trial chairman and the derived similarity models.

2.2 Trial recruitment

The recruitment of a set of eligible patients is crucial for the success of any clinical trial. Current implemented recruitment workflows are time-consuming because they require a lot of manual intervention steps executed by local site investigators. A semi-automated recruitment tool enables a complete identification of eligible patients and ensures both timely execution of the trial avoiding selection bias.

When starting the recruitment process for a specific trial, authorised investigators on local sites where the trial will run are notified by the trial chairman. These investigators will use the patient recruitment tool that is running on the site, to find eligible patients for the trial. The set of local site patients that the investigator is allowed to recruit is pseudonymised, and by using access control only the patients the investigator is allowed to see are included in this set. Next to this, the patients in the set need to have given consent for the recruitment. The patient data itself is retrieved through the semantic interoperability layer.

Each patient in the set will be checked against predefined matching criteria included in the selected trial. This will give the investigator a recruitment suitability indication for each given patient. The tool offers functionality to the investigator to easily filter the patient set. Several widgets are available for altering the patient set. For example a free-text query widget will be available to search for certain keywords in the patient's data. This widget will trigger a free-text service that offers a search interface connected to the EHR.

Additional patient information can be consulted by the investigator, where he can make recruitment decisions for each individual patient. This patient information will also be retrieved from the EHR through the semantic interoperability layer.
Once the investigator has composed a final set of eligible patients, the recruitment process can be finalised. The treating physician of each patient is notified, informing that the patient is eligible for the selected trial.

Type of data	**Need for these data**
EHR DW	• List of patients on the local site will be requested • Specific patient information will be fetched, in order to give the investigator a detailed view
Trial metadata	• The available trials will be listed • The trial criteria are run against the patient's medical information

2.3 Patient diary

- High-level description of the tools in the scenario

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 aldo 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 scenario assumes the existence of the following components:

- A PHR system (in our cas IndivoX)
- A PHR system (in our case IndivoX)
- A EURECA-compatible DWH
- A Clinical Trial (CT) management system (ObTiMA in our case)
- A National Registry for patients

The scenario will allow patient information from EHR/PHR/CT/National Registries to be pushed to the DWH. Then appropriate mechanism will allow the reuse of patient information (such as prefilling of eCRFs) withing PHR and CT systems

- Types/amount of data that are needed for the development, the test and upcoming demonstrator of the tool
  - Patient data should be available in the PHR system
  - Clinical Trials should be available in the CT system and the necessary eCRFs should be available.
  - Patient data should be available in a CT system
  - National Registries with patient data should be available as well (we do not expect this to be demonstrated in the first review)

- How the tool will interact with such data
  - Patient data will be exported from the PHR system and loaded to the DWH through
    Extract-Transform-Load (ETL) tools.
  - CT system will issue SPARQL queries to the DWH to retrieve information for patients

3 Available test data

3.1 Institut Jules Bordet

This section describes both Electronic Health Record and clinical trial systems that are used at Institut Jules Bordet and whose patient data could be used within the EURECA project.

- Number of patients that are available for sharing:

Until now (January 2013), 236 patients already gave their informed consent so that we can use their data for research purposes according to legal and ethical requirements (*see Section 4*). This amount of patient is meant to increase all along the period of the project.

- Anonymisation process:

Demographic data are used for the de-identification process in other types of unstructured data that IJB will provide to the project (e.g. consult and discharge report).

3.1.1 Electronic Health Record

Institut Jules Bordet have designed and developed his own Electronic Health Record system, called Oribase.

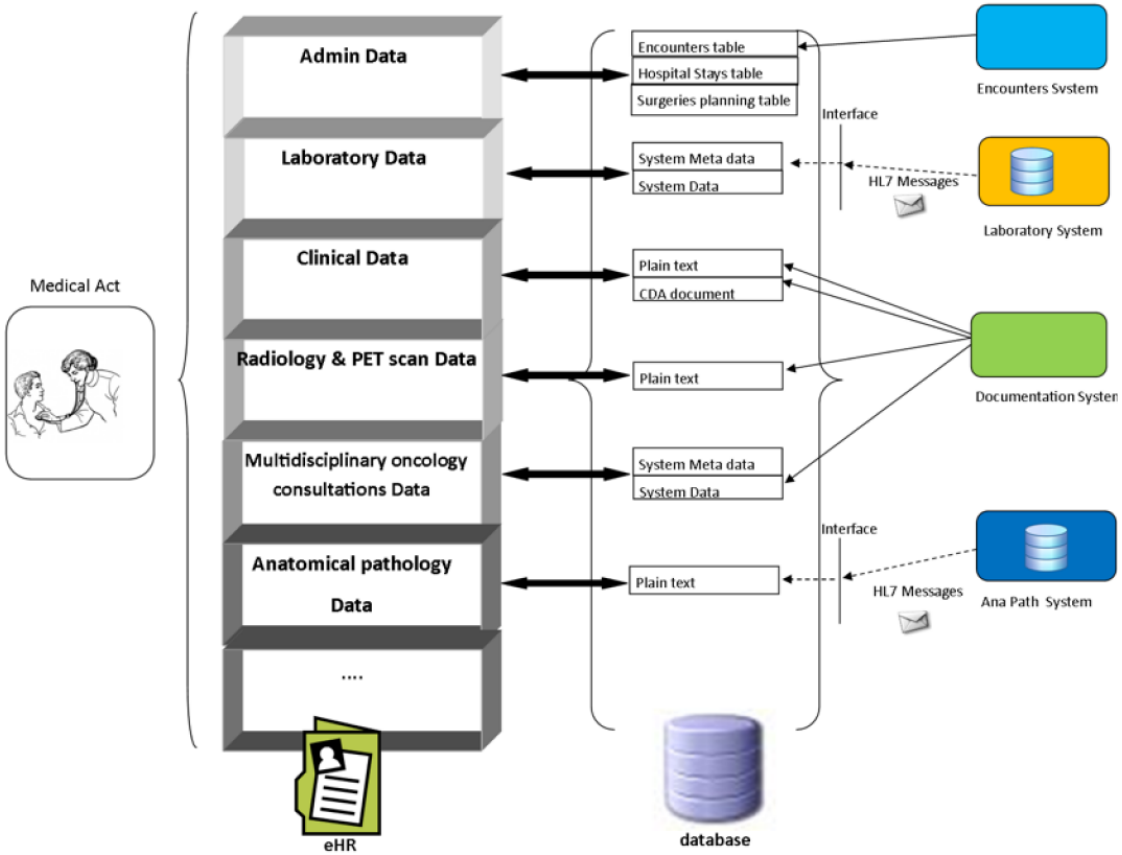


Figure 1 - A non-exhaustive EHR system view

3.1.1.1 Administrative data

- High level description of the dataset:

The Admissions, Discharge, and Transfer (ADT) subsystemcontains demographic data like patient ID, patient name, date of birth (see *Table 1*). ADT subsystems also stores medical appointments record (e.g. patient ID, name of the attending physician, consultation type, comments, date/time) (see *Table 2*) and hospital stays information (e.g. patient ID, admission, discharge date, location, name of the attending physician) (see *Table 3*).

Demographic data
eHR patient ID
first name
last name
gender
marital status
street Adress
postal code
city
country
phone number
national registry
…

Table 1 – Demographic data

Appointment record
apointment ID
eHR Patient ID
physician ID
consultation type
comments
appointment date
appointment time
appointment type
appointment location
…

Table 2 – Appointment record data

Hospital stay info
hospital stay ID
eHR Patient ID
physician ID
admission date
discharge end date
nursing ID
…

Table 3 – Hospital stay information

- Structure of the data:

All these data are structured and don't contain any free text unstructured data.

3.1.1.2 Laboratory data

- High level description of the dataset:

  Laboratory data are structured data that are sent as HL7 messages through the laboratory system and are stored in internal tables (see typical laboratory entries in *Table 4*).

Laboratory data
unique request ID
eHR ID
requesting physician
sampling date
sampling hour
result reception date
result reception hour
sample source
sample type
sample ID
test result value
test result Unit

test status (F = Completed; X = Canceled; I = Intermediate) threshold values test description comments test category description category test short name test long name …

Table 4 - Laboratory databases entries

3.1.1.3 Clinical data (Consult and discharge reports)

- High level description of the dataset:
  The documentation system is designed to use a common structure for persistent document like Clinical Document Architecture[2] (CDA), a semi-structured XML-based markup standard which intends to specify the encoding, structure and semantics of clinical documents for exchange. This standard allows the representation and machine processing of clinical documents in a way that the document textual informaiton remains both human readable and machine processable as it is structured based on LOINC and SNOMED-CT vocabularies (see *Table 6*) that provides well-defined meaning for clinical specific terms.

Clinical data
Nature of the document: • Laboratory • Clinical studies • Anotomopathology • Consultations • Radiography • Day hospital • Chemotherapy, etc. Reception date of the document Patient's Name Gender Birth date User ID Hospital unit Treating physician …

Table 5 – Patient documents

[2] EURECA project, "D2.1 State-of-the-Art report on standards," due in September 2012.

```
144 <observation classCode="OBS" moodCode="EVN">
145 <code code="50595-8" codeSystem="2.16.840.1.113883.6.1" codeSystemName="LOINC" displayName="Pathologie" />
146 <id root="11" extension="39" />
147 <value value="Cancer du sein" />
148 </observation>
149 </entry>
150 <entry>
151 <observation classCode="OBS" moodCode="EVN">
152 <code code="125725006" codeSystem="2.16.840.1.113883.6.96" codeSystemName="SNOMED CT" displayName="Etat civil" />
153 <id root="11" extension="20" />
154 <value value="Veuf" />
155 </observation>
156 </entry>
157 <entry>
158 <observation classCode="OBS" moodCode="EVN">
159 <code code="29553-5" codeSystem="2.16.840.1.113883.6.1" codeSystemName="LOINC" displayName="Age" />
160 <id root="11" extension="34" />
161 <value value="75" unit="ans" />
162 </observation>
163 </entry>
164 <entry>
165 <observation classCode="OBS" moodCode="EVN">
166 <code code="263495000" codeSystem="2.16.840.1.113883.6.96" codeSystemName="SNOMED CT" displayName="Sexe" />
167 <id root="11" extension="33" />
168 <value value="F" />
169 </observation>
```

Table 6 - Example of coding clinical terms with LOINC and SNOMED-CT

Examples of CDA documentation that are used in the EHR of the *Institut Jules Bordet*:

- Consultation note
- Surgery procedure report
- Echography report
- Endoscopy report
- Hospital Discharge Summary report
- Radiology report

Concepts that are too ambiguous and/or not enough accurate to use terms that already exist within a clinical vocabulary, are defined using an internal codification system, which is based on a unique ISO identifier which has been attributed to IJB. These internal ObjectID (see *Table 7*) codes are hierarchically defined and are used in the coding sections of CDA of our Institute.

oid	Explication	Mnémonique
2.25.299518904337880959076241620201932965147.9	Types de rencontres	encounter
2.25.299518904337880959076241620201932965147.9.2	Séjour en hospitalisation	hospitalstay
2.25.299518904337880959076241620201932965147.9.7	Consultation	consultation
2.25.299518904337880959076241620201932965147.9.7.1	Rencontre de type "urgence"	emergency
2.25.299518904337880959076241620201932965147.9.7.19	Consultation de cardiologie	consultationCardio
2.25.299518904337880959076241620201932965147.9.200	Concertation (cmo)	cmo
2.25.299518904337880959076241620201932965147.9.16	Intervention chirurgicale	surgery

Table 7 - Examples of IJB internal ObjectID codes

3.1.1.4 Multidisciplinary Team (MDT) data

- High level description of the dataset:
  Multidisciplinary cancer consultations' meeting objectives are to assemble a team of complementary specialists in all oncology disciplines (e.g. medical oncologists, radiation oncologists, urologists, pulmonary specialists, radiologists, pathologists, surgeons, general physicians, nurses and social workers) to discuss the best course for a patient's cancer treatment, or give their opinion and interpretation of an on-going treatment plan.

  Entered data for the multidisciplinary consultations are structured to simplify their manipulation. Structured data are also stored using HL7 v3 CDA standard.

3.1.1.5 Anatomical Pathology data

- High level description of the dataset:
  The anatomical pathology system is used by the anatomical pathology unit in order to send information from a local database to the global EHR system of the institute. The anatomical pathology databases are related to medical studies and their results. All reports follow the HL7 structure but contain unstructured text except for the tumor codification in SNOMED.

3.1.1.6 Chemotherapy prescriptions

- High level description of the dataset:
  Chemotherapy prescriptions contain structured data with drug prescriptions.

3.1.1.7 The cancer registry

- High level description of the dataset:

  "A cancer registry is a file concerning all the cases of cancer appearing in a defined population. In this file, the characteristics of cancerous patients are listed. It also contains the clinical and anatomopathological data which are collected from different sources of information".[3]

  The cancer registry has been set-up since 2000 at *Institut Jules Bordet*. It records structured information about all cancers diagnosed or treated at the Institute, and represents more than 2,500 tumours a year, that is currently more than 33,000 cases in total.
  Its objectives are to describe patient profiles and tumour characteristics, to assess the Institute's medical activity and patient prognosis, and to contribute to research and to the National Cancer Registry.

3.1.1.8 Picture Archiving and Communication System (PACS) and nuclear medicine

- High level description of the dataset:
  A Picture Archiving and Communication System (PACS) is a medical imaging technoglogy which combines software and hardware, and which is dedicated to the storage, transmission, display and printing of digital diagnostic images.
  The images are received and transmitted in the DICOM format (Digital Imaging and Communications in Medicine).

[3] http://ww.bordet.be/registre/register.htm

3.1.2 Clinical trial data – The TOP Trial

The Trial Of Principle (TOP) study[4] was started for the prospective evaluation of Topoisomerase II Alpha (TOP2A) gene amplification and protein overexpression as markers predicting the efficacy of Epirubicin in the primary treatment of breast cancer patients [1]. The TOP trial included 149 patients, 139 of whom were evaluable for response prediction analyses. The primary end point was pathologic complete response (pCR). TOP2A and gene expression profiles were evaluated using pre epirubicin biopsies. Gene expression data from ER-negative samples of the EORTC (European Organisation for Research and Treatment of Cancer) 10994/BIG (Breast International Group) 00-01 and MDACC (MD Anderson Cancer Center) 2003-0321 neoadjuvant trials were used for validation purposes.

Here are the inclusion and exclusion criteria that were defined:

Inclusion Criteria:

- Histologically-confirmed breast cancer (operable, locally advanced or inflammatory)
- Age less than 70 years
- Female patient
- Tumor size 2 cm at ultrasound examination.
- ER-negative tumors defined according to immunohistochemistry (i.e. < 10% of positive cells after immunostaining).
- Multifocal and multicentric breast tumors are allowed if all foci are ER-negative.
- Fixed and frozen samples from the primary tumor, obtained before treatment with epirubicin, must be available for evaluation of biological markers (TOP2A gene and protein, HER-2 gene, p53 gene, oligonucleotides microarrays).
- Written informed consent before study registration.
- Performance status 0 or 1 (ECOG scale)
- Normal CBC, hepatic and renal functions
- Normal left ventricular ejection fraction by echocardiography or muga scan
- Negative pregnancy test for all women of childbearing potential. Patients of childbearing potential must implement adequate non-hormonal measures to avoid pregnancy during treatment.

Exclusion Criteria:

- Metastatic breast cancer
- Serious medical conditions like:
  - Congestive heart failure or unstable angina pectoris, previous history of myocardial infarction within 1 year from study entry, uncontrolled arrhythmias.
  - History of significant neurologic or psychiatric disorders
  - Active uncontrolled infection
- Concomitant contralateral invasive breast cancer
- Concurrent treatment with hormonal replacement therapy
- Concurrent treatment with any other anti-cancer therapy
- Previous treatment with anthracyclines for breast cancer

The schedule of assessments is documented in *Table 8*.

[4] http://clinicaltrials.gov/ct2/show/NCT00162812?terl=top&rank=6

Mandatory Exams	Baseline < 28 days before 1st inclusion	Epirubicin treatment period	Post-Epirubicin treatment
Medical history	X		
Physical examination + clinical tumor assessment	X	X	X
Breast biopsy (TRU-CUT) + measurement of hormone receptors	X	(X)	
Serum sampel	X	X	X
Whole blood sample	X		
Haematology and biochemistry Red blood cells Hemoglobin Platelets WBC ANC Total bilirubin Serum creatinine GOT/GPT Alkaline phosphatase	X	X	X
ECG	X		
LVEF (US or MUGA)	X		
Chest X-Ray	X		
Bone scan	X		
Bone ultrasound	X		
Bilateral mammography	X		X
Breast ultrasound	X		X
Informed consent	X		

Table 8 - Schedule of assessments

There are different type of data that are available from the TOP trial:

- Case Report Forms,
- Aggregated clinical data,
- Genomic data.

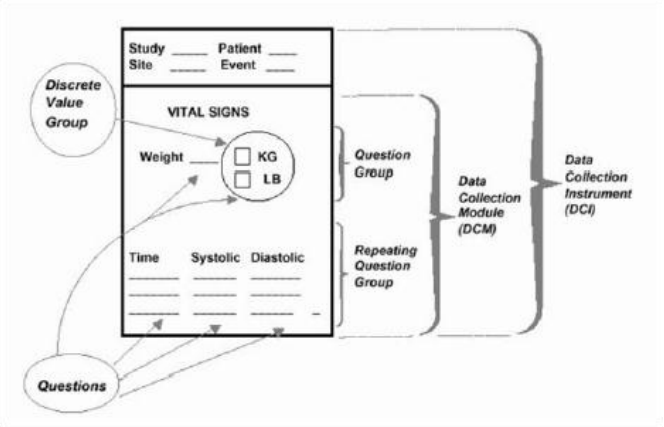
Case Report Forms

The Case Report Forms are used by sponsors of the clinical trials to collect data from the participating sites. It can be a paper or an electronic questionnaire. The CRFs contain all data of a patient collected during a trial, including eligibility criteria and adverse events. Oracle Clinical was used to capture the data from the CRF's for the TOP trial. The CRF data is provided to EURECA in excel format and follows the structure of the Oracle Clinical database.

Figure 2 - Oracle Clinical structure

Figure 2 shows an overview of the structure of the Oracle Clinical database. Question represents a (particular) question on a CRF and its anwer. When the answer can be

populated from a set of possible answers (e.g. a code list), the answers are store in DVG, a Discrete Value Group.
Questions are grouped in QG – Question group – which can be used to group (medically) related questions (which can be handy in order to reuse groups of questions of different CRFs). The question groups are used in DCM – Data Collection Module – which represent (the sections of) the CRF screens that are used to collect the data and these sections should be answered during a single clinical visit.
DCI – Data Collection Instrument – corresponds to a CRF. Typically, 1 DCI corresponds to 1 DCM, but can include more that one DCM. The DCI construct allows for CRFs which collect data during multiple visits.



Finally, a DCI Book specifies the order of the DCI's.

The above description describes the Oracle Clinical database structure. In addition to the database structure, Oracle Clinical allows for the definition of validation and derivation procedures. Validation procedures validate data entry. Derivation procedures create additional variables in the database to assist in analysis. The prototypical example would be a derivation procedure which would calculate the age based on the birth date (as entered on a CRF) and a visit date (also entered on a CRF).

Patient's Characteristics
DCM=QG=VS New

- Height |__|__|__| (cm) ***HGT*** ***HGT_UN(dft 'CM')***
- Weight |__|__|___,__| (Kg) ***WGT*** ***WGT_UN (dft 'KG')***
- BSA |___,___| (m^2) ***BSA*** ***(NUMBER 4,3) BSA_UN (dft 'M2')***
- Menopausal status: ***REP_STAT(DVG REP_STAT#New)***
  - ☐1 premenopausal (< 6 months since last menstrual period (LMP) and no prior ovariectomy and no estrogen replacement therapy)
  - ☐2 postmenopausal (prior bilateral ovariectomy, or > 12 months since LMP with no prior hysterectomy and not receiving LH-RH analog)
  - ☐3 above category not applicable and < 50
  - ☐4 above category not applicable and ≥ 50

Figure 3 - Annotated CRF excerpt

The excel export contains a sheet per DCM. *Figure 3* shows an excerpt from an annotated page of the TOP trial CRF. Data filled in this CRF would be exported onto an

excel sheet named “VS New”. *Table 9* shows the (transposed) excel export with the questions from the CRF excerpt in bold.

Variable name	value	value	value
PT	XXX	XXX	XXX
STUDY	TOP	TOP	TOP
DCM SUBNM	VS	VS	VS
PATIENT_POSITION_ID	XXX	XXX	XXX
CPEVENT	SCREENING	SCREENING	SCREENING
DCMNAME	VS	VS	VS
SUBSETSN	1	1	1
DCMDATE			
DOCNUM	XXX	XXX	XXX
ACCESSTS	16-09-2004 10:35	16-09-2004 11:04	16-09-2004 13:40
LOGINTS	16-09-2004 10:35	16-09-2004 11:04	16-09-2004 13:40
LSTCHGTS	16-09-2004 10:35	16-09-2004 13:19	22-05-2006 11:17
LOCKFLAG	N	N	N
DCMTIME			
ACTEVENT	1	1	1
SUBEVENT_NUMBER	0	0	0
VISIT_NUMBER	1	1	1
QUALIFYING_VALUE	6	6	6
QUALIFYING_QUESTION	20007	20007	20007
REPEATSN	1	1	1
FIRST_BOOK_PAGE	6	6	6
RECEIVED_DCM_STATUS_CODE	PASS 1 COMPLETE	PASS 1 COMPLETE	PASS 1 COMPLETE
HGT	**153**	**163**	**165**
HGT_UN	**CM**	**CM**	**CM**
REP_STAT	**PREMENOPAUSAL**	**PREMENOPAUSAL**	**PREMENOPAUSAL**
WGT	**50**	**52**	**95**
WGT_UN	**KG**	**KG**	**KG**
BSA	**1.45**	**1.55**	**2.2**
BSA_UN	**M2**	**M2**	**M2**

Table 9 - Excel export

Aggregated clinical data

The aggregated clinical data comprise information on tumor size, auxiliary lymph node status, tumor grade, biomarker expression status (estrogen receptor, progesterone receptor, HER2, TOP2A), and several clinical endpoints such as pathological complete response, distant metastasis-free survival and overall survival.

Genomic data

Three kinds of genomic information are available: whole human genome expression array data, SNP data, and methylation data.

- The whole human genome expression array data are Affymetrix GeneChip® Human Genome U133 Plus 2.0 Arrays[5]. This microarray contains probes for more than

[5] http://www.affymetrix.com/browse/products.jsp?productId=131455&categoryId=35760#1_1

38,500 transcripts corresponding to well-characterized genes and Unigene genes, giving a full-genome view of gene expression. The analysis will start from the “raw” .cel files that contain probe-level intensity data. This allows various schemes of data normalization and probeset data aggregation. The .cel files also contain the necessary information for array and hybridization quality assurance. The size of a .cel file for this microarray is around 32MB. Data is available for 120 patients from the TOP trial.

- The SNP data are Affymetrix SNP 6.0 Arrays[6]. This microaaray contains probes for more than 906,600 single nucleotide polymorphisms (SNPs) and more than 946,000 probes for the detection of copy number variation (CNV). This corresponds to a median inter-marker distance in the genome of less than 700 nucleotides. Again, the analysis will start from the .cel files, which allows maximum flexibility in the choice of the algorithms for CNV genotyping. The size of a .cel file for this array is around 61MB. Data is available for 70 patients from the TOP clinical trial.

- The methylation data are Illumina Infinium HumanMethylation27 BeadChips[7]. This array allows interrogating the methylation status of 27,578 highly informative CpG sites located in the proximal promoters of 14,475 protein coding genes. This corresponds to an average of two interrogated CpGs per genes although a subset of >200 cancer-related genes have 3-20 interrogated CpGs. The Infinium array uses a pair of probes for every CpG, with one probe measuring the level of the methylated CpG and the other probe measuting the level of the unmethylated CpG. The methylation of the CpG is then often expressed as a beta value, which is the ratio of the methylated signal on the sum of the methylated CpG to 1.0 for a fully methylated CpG. The data is available for (at least) 34 patients from the TOP trial.

3.2 Universität des Saarlandes

It is proposed that the following data will be provided:

1. Clinical trial data of the SIOP[8] Nephroblastoma trial and study of GPOH[9]
2. Data from the Hospital Information System of the University Hospital of the Saarland
3. Data from the Cancer Registry of the Saarland.

3.2.1 SIOP Nephroblastoma trial and study

The SIOP Nephroblastoma trial and study database contains clinical and consultation data of 2880 patients with kidney tumours diagnosed in Germany, Austria and Switzerland between 1989 and 2012. This database is part of the database for kidney tumours of the SIOP-RTSG (SIOP Renal Tumour Study Group). The database does not include DICOM data or molecular or biology data.

[6] http://www.affymetrix.com/browse/products.jsp?productId=131533&navMode=34000&navAction=jump&aId=productsNav#1_1
[7] http://www.illumina.com/products/infinium_humanmethylation27_beadchip_kits.ilmn
[8] SIOP: International Society of Paediatric Oncology
[9] GPOH: German Society for Paediatric Oncology and Haematology

The database is composed of different tables as shown in *Figure 4*. Most of the data fields are structured with only few free text data fields. The whole database is anonymized. This was done by Custodix using their CAT tool.

3.2.1.1 Schema of the SIOP Clinical database

For each of the tables in the database the name and data types of every column is given in the tables that follow.

Red tables and columns	Cleared from anonymous output
Blue columns	Anonymised, replaced by random identifier
Green columns	Make date relative to random date of birth
Purple columns	Free text anonymisation
Dark blue columns	Free text not to be anonymised as it does not contain any names or personal data.

Table 10 - Anonymisation legend

The database will be available via Custodix for those partners that have signed the contracts. LUH will be informed about this (*see 4.1*).

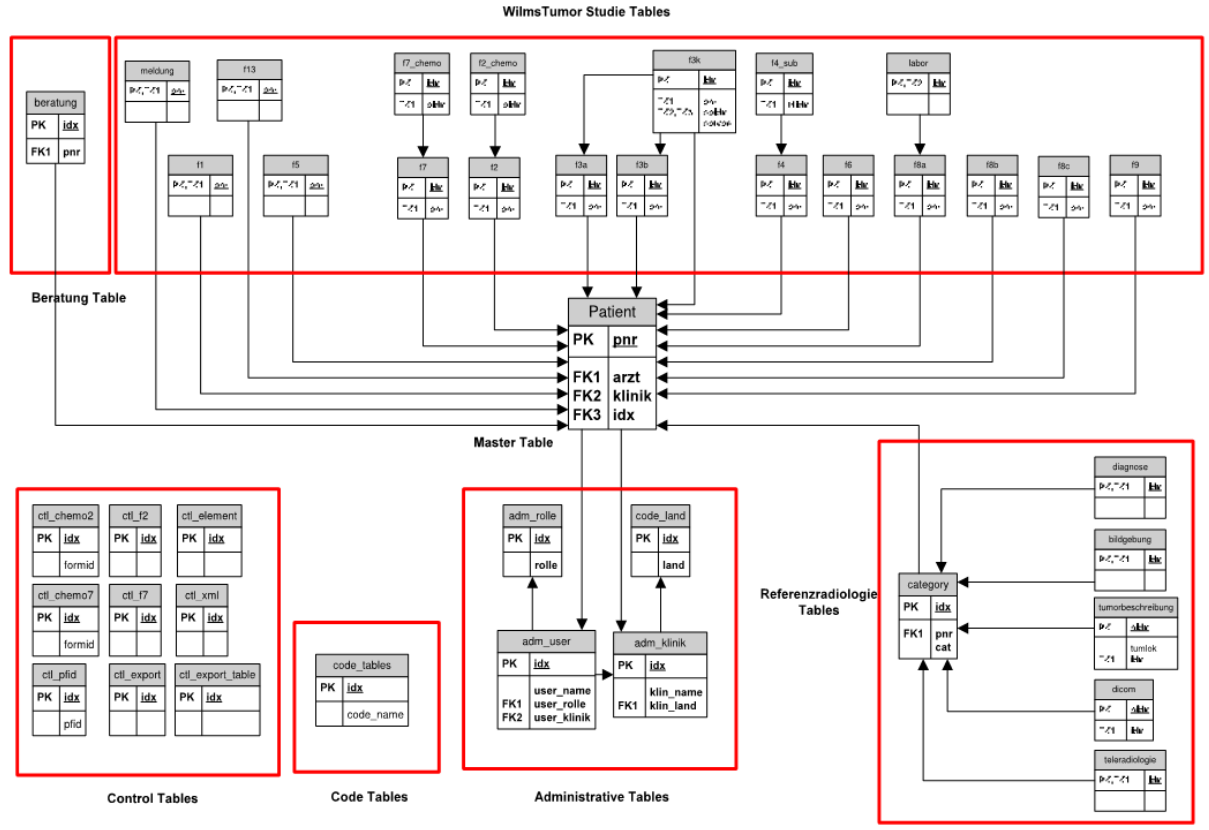


Figure 4 - Main relations of the SIOP Nephroblastoma database

3.2.1.2 Master Table

Name	Data type	
pnr	Long Integer	Patient pseudonym
pfreez	Long Integer	
pfid	Memo	
siopnr	Long Integer	
gpohpid	Memo	
izcode	Memo	
molid	Memo	
molmainz	Long Integer	
pat_ref	Long Integer	
pat_berat	Long Integer	
pat_stud	Long Integer	
pat_best	Long Integer	
pat_molek	Long Integer	
pat_path	Long Integer	
pat_op	Long Integer	
pat_chemo	Long Integer	
pat_ber	Long Integer	
pat_wilms	Long Integer	
pat_sts	Long Integer	
name	Memo	
vorname	Memo	
sex	Long Integer	
gebdat	DateTime	
pat_str	Memo	
pat_plz	Long Integer	
pat_ort	Memo	
geb_ort	Memo	
land	Long Integer	
diag_klinik	Long Integer	
klinik	Long Integer	
arzt	Long Integer	Physician pseudonym

Table 11 - Patient columns

Name	Data type	
pnr	Long Integer	Patient pseudonym
sts	Long Integer	
name5b	Memo	
alter	Long Integer	
unilat	Long Integer	
keine_met	Long Integer	
praeop_chem	Long Integer	
stad2_3	Long Integer	
echokard	Long Integer	
postop_ther	Long Integer	
nachbeob	Long Integer	
einv	Long Integer	
opdat	DateTime	
rand_erg	Long Integer	
rand_dat	DateTime	

Table 12 - Rand columns

3.2.1.3 Consultation Tables

Name	Data type	
idx (0)	Long Integer	Beratung pseudonym
sts	Long Integer	
anf_erf	Long Integer	
dat_anf (3)	DateTime	
anf_tag	Long Integer	
anf_uhr (5)	DateTime	
anf_fei	Long Integer	
beant_zeit	Long Integer	
anz_ruech_klin	Long Integer	
anz_rueck_ref	Long Integer	
dat_beant (10)	DateTime	
beant_uhr (11)	DateTime	
beant_dauer	Long Integer	
beant_erf	Long Integer	
name_berat	Long Integer	
name_berat_sonst	Memo	
pos_berat	Long Integer	
pos_berat_sonst	Memo	identifying?
klinik	Long Integer	
beruf	Long Integer	
beruf_sonst	Memo	identifying?
berat_pat	Long Integer	
berat_pat_dk	Long Integer	
berat_pat_de	Long Integer	
berat_pat_op	Long Integer	
berat_pat_rand	Long Integer	
berat_pat_chemo	Long Integer	
berat_pat_kom	Long Integer	
berat_pat_rez	Long Integer	
berat_pat_man	Long Integer	
berat_pat_stra	Long Integer	
berat_pat_gen	Long Integer	
berat_pat_sonst	Long Integer	
berat_pat_sonst_com	Memo	Is this relevant?
pnr (34)	Long Integer	Patient pseudonym
pat_gem	Long Integer	
berat_stu_dok	Long Integer	
berat_stu_web	Long Integer	
berat_stu_rde	Long Integer	
berat_stu_feh	Long Integer	
berat_stu_stud	Long Integer	
berat_stu_rand	Long Integer	
berat_zus_pro	Long Integer	
berat_zus_lit	Long Integer	
berat_anf_teil	Long Integer	
berat_anf_wilm	Long Integer	
berat_anf_pro	Long Integer	
berat_anf_log	Long Integer	
berat_anf_sonst	Long Integer	
berat_anf_sonst_com	Memo	
fragestellung (50)	Memo	
bew_in_klin	Long Integer	
bew_dring	Long Integer	
anf_dient	Long Integer	
anf_dient_sonst_com	Memo	Non indentifying
anf_weiter	Long Integer	
stud_disk	Long Integer	
Klinikintern	Long Integer	
Studienpro	Long Integer	
bean_lit	Long Integer	
bean_dat	Long Integer	
bean_anf	Long Integer	
weit_refstrah	Long Integer	
weit_refrad	Long Integer	
weit_refchir	Long Integer	
weit_refpath	Long Integer	
weit_gess	Long Integer	
int_studie	Long Integer	
weit_sonst	Long Integer	
weit_sonst_com	Memo	Non identifying
beraten (70)	Memo	
kontakt	Long Integer	
kon_ja	Long Integer	
kon_sonst_com	Memo	Contains physician names
wer_berat	Long Integer	
wer_berat_sonst_com	Memo	Non identifying
berat_erf	Long Integer	

Table 13 - Beratung columns

3.2.1.4 Referenzradiologie Tables

Name	Data type	
idx	Long Integer	Category pseudonym
pnr	Long Integer	Patient pseudonym
sts	Long Integer	
bildvon	Long Integer	
us_bilddat	DateTime	
mrt_bilddat	DateTime	
ct_bilddat	DateTime	
refdat	DateTime	
alterj	Long Integer	
alterm	Long Integer	
zeitbisref	Long Integer	
cat	Long Integer	
fr_ther	Long Integer	

fr_tumvol	Long Integer	
fr_rez	Long Integer	
fr_prog	Long Integer	
fr_sonst	Long Integer	
fr_sonst_com	Memo	safe
rezauf	Long Integer	
rezsonst_com	Memo	safe
fr_bes_com	Memo	

Table 14 - Category columns

Name	Data type	
idx (0)	Long Integer	Category pseudonym
sts	Long Integer	
us_hard	Long Integer	
us_cdrom	Long Integer	
us_online	Long Integer	
ct_hard	Long Integer	
ct_cdrom	Long Integer	
ct_online	Long Integer	
mrt_hard	Long Integer	
mrt_cdrom	Long Integer	
mrt_online	Long Integer	
us_dicom	Long Integer	
ct_dicom	Long Integer	
mrt_dicom	Long Integer	
us_format_sonst	Long Integer	
us_format_sonst_com	Memo	safe
ct_format_sonst	Long Integer	
ct_format_sonst_com	Memo	safe
mrt_format_sonst	Long Integer	
mrt_format_sonst_com	Memo	safe
us_scan	Long Integer	
ct_scan	Long Integer	
mrt_scan	Long Integer	
mrtab	Long Integer	
t1_ax	Long Integer	
t1_cor	Long Integer	
t1_sag	Long Integer	
t1km_ax	Long Integer	
t1km_cor	Long Integer	
t1km_sag	Long Integer	
t2_ax	Long Integer	
t2_cor	Long Integer	
t2_sag	Long Integer	
fatsat_com	Memo	safe
mrt_sonst_com	Memo	safe
mrt_ausw	Long Integer	
mrt_ir	Memo	safe (empty or number)
mrt_quali	Long Integer	
ctab	Long Integer	
ctab_nativ	Long Integer	
ctab_km	Long Integer	
ctab_spiral	Long Integer	
ctab_dicke	Double	
ctab_kv	Double	
ctab_mas	Double	
ctab_quali	Long Integer	
us	Long Integer	
us_tum	Long Integer	
us_nier	Long Integer	
us_kont	Long Integer	
us_vc	Long Integer	
us_lk	Long Integer	
us_le	Long Integer	
us_sonst	Long Integer	
us_sonst_com	Memo	safe
us_format	Long Integer	
us_quali	Long Integer	
Aug	Long Integer	
Mibg	Long Integer	
Cavo	Long Integer	
Angio	Long Integer	
Sonst	Long Integer	
sonst_com	Memo	safe
Roetx	Long Integer	
roe_paap	Long Integer	
roe_seit	Long Integer	
roe_quali	Long Integer	
Cttx	Long Integer	
cttx_nativ	Long Integer	
cttx_km	Long Integer	
cttx_dicke	Double	
cttx_kv	Double	
cttx_quali	Long Integer	
Ausreich	Long Integer	
Weitere	Long Integer	
bef_thx	Long Integer	
bef_us	Long Integer	
bef_ctab	Long Integer	
bef_mrt	Long Integer	
bef_cttx	Long Integer	
klin_an	Long Integer	
klin_nachfr	Long Integer	
erg_mrt_t1	Long Integer	
erg_mrt_t1km	Long Integer	
erg_mrt_t2	Long Integer	
erg_mrt_sonst	Long Integer	
erg_mrt_sonst_com	Memo	safe
erg_ctab_native	Long Integer	
erg_ctab_km	Long Integer	
erg_us_ab	Long Integer	
erg_us_nier	Long Integer	
erg_us_kont	Long Integer	
erg_us_vc	Long Integer	
erg_us_lk	Long Integer	
erg_us_le	Long Integer	
erg_us_sonst	Long Integer	
erg_us_sonst_com	Memo	safe
erg_roe_paap	Long Integer	
erg_roe_seit	Long Integer	
erg_cttx_nativ	Long Integer	
erg_cttx_km	Long Integer	
erg_mibg	Long Integer	
erg_sonst	Long Integer	

erg_sonst_com	Memo	safe

Table 15 - Bildgebung columns

Name	Data type	
Idx (0)	Long Integer	Category pseudonym
sts	Long Integer	
diag_klin	Long Integer	
diag_klin_com (3)	Memo	Just a short diagnose
Refdiagwie	Long Integer	
diag_refrad	Long Integer	
diag_refrad_com (6)	Memo	Just a short diagnose
va_neph	Long Integer	
Diaggleich	Long Integer	
lok_nier	Long Integer	
lok_lk	Long Integer	
lok_vc	Long Integer	
lok_lu	Long Integer	
lok_le	Long Integer	
lok_kno	Long Integer	
lok_hirn	Long Integer	
lok_sonst	Long Integer	
lok_sonst_com	Memo	empty
comment_ext	Memo	investigate
comment_int	Memo	investigate
ther_durchf	Long Integer	
primop	Long Integer	
patart	Long Integer	

Table 16 - Diagnose columns

Name	Data type	
idx	Long Integer	Category pseudonym
sts	Long Integer	
catidx	Long Integer	
tumlok	Long Integer	
lok_soli	Long Integer	
lok_intra	Long Integer	
lok_cra	Long Integer	
lok_cau	Long Integer	
lok_zen	Long Integer	
lok_ven	Long Integer	
lok_dor	Long Integer	
lok_dif	Long Integer	
lok_sonst	Long Integer	
lok_sonst_com	Memo	safe
lok_anz	Long Integer	
tumvol	Long Integer	
mrt_a	Double	
mrt_b	Double	
mrt_c	Double	
mrt_v	Double	
ct_a	Double	
ct_b	Double	
ct_c	Double	
ct_v	Double	
us_a	Double	
us_b	Double	
us_c	Double	
us_v	Double	
morph_hom	Long Integer	
morph_eizys	Long Integer	
morph_typ	Long Integer	
morph_extra	Long Integer	
morph_ober	Long Integer	
morph_gef	Long Integer	
morph_verk	Long Integer	
morph_cys	Long Integer	
morph_sonst	Long Integer	
morph_const_com	Memo	safe
ausdeh	Long Integer	
nekrosen	Long Integer	
einblut	Long Integer	
subkaps_fl	Long Integer	
va_rup	Long Integer	
fl_ab	Long Integer	
inf_pso	Long Integer	
inf_le	Long Integer	
inf_zwerch	Long Integer	
inf_sonst	Long Integer	
inf_sonst_com	Memo	safe
intrathor	Long Integer	
sonst	Long Integer	
sonst_com	Memo	safe
us_echo_nie	Long Integer	
us_echo_leb	Long Integer	
us_echo_sonst	Long Integer	
us_echo_sonst_com	Memo	safe
us_hom	Long Integer	
ctnat_echo_nie	Long Integer	
ctnat_echo_leb	Long Integer	
ctnat_echo_sonst	Long Integer	
ctnat_echo_sonst_com	Memo	safe
ctnat_hom	Long Integer	
ctkm_aufn	Long Integer	
ctkm_aufnst	Long Integer	
ctkm_hom	Long Integer	
mrtt1nat_echo_nie	Long Integer	
mrtt1nat_echo_leb	Long Integer	
mrtt1nat_echo_sonst	Long Integer	
mrtt1nat_echo_sonst_com	Memo	Safe
mrtt1nat_hom	Long Integer	
mrtt1km_aufn	Long Integer	
mrtt1km_aufnst	Long Integer	
mrtt1km_hom	Long Integer	
mrtt1km_echo_nie	Long Integer	
mrtt1km_echo_leb	Long Integer	
mrtt1km_echo_sonst	Long Integer	
mrtt1km_echo_sonst_com	Memo	safe
mrtt2_echo_nie	Long Integer	
mrtt2_echo_leb	Long Integer	
mrtt2_echo_sonst	Long Integer	
mrtt2_echo_sonst_	Memo	safe

com		
mrtt2_hom	Long Integer	
mrtsonst_echo_nie	Long Integer	
mrtsonst_echo_leb	Long Integer	
mrtsonst_echo_sonst	Long Integer	
mrtsonst_echo_sonst_com	Memo	safe
mrtsonst_hom	Long Integer	
tumstruk_com	Memo	safe
nier_erh	Long Integer	
nematose	Long Integer	
harnst	Long Integer	
path_sonst	Long Integer	
path_sonst_com	Memo	safe
kont_unauf	Long Integer	
kont_reste	Long Integer	
kont_nematose	Long Integer	
kont_hyp	Long Integer	
kont_harnst	Long Integer	
kont_dysp	Long Integer	
kont_zyst	Long Integer	
kont_sonst	Long Integer	
kont_sonst_com	Memo	safe

Table 17 - Tumourbeschreibung comuns

3.2.1.5 Wilms Tumor Study Tables

Name	Data type	
idx	Long Integer	Chemo2 pseudonym
pidx	Long Integer	f2 pseudonym
cday	Long Integer	
cweek	Long Integer	
dat	DateTime	
act	Double	
vcr	Double	
dox	Double	
gewicht	Double	
tvr	Double	
tvl	Double	
dosred	Long Integer	
grund_red	Long Integer	
dosred_com	Memo	

Table 18 - Chemo2 columns

Name	Data type	
idx	Long Integer	Chemo7 pseudonym
Pidx	Long Integer	f7 pseudonym
Cday	Long Integer	
Cweek	Long Integer	
Dat	DateTime	
Act	Double	
Vcr	Double	
Dox	Double	
Carbo	Double	
vp16	Double	
Cpm	Double	
Cyclo	Double	
Gewicht	Double	
Dosred	Long Integer	
grund_red	Long Integer	
dosred_com	Memo	

Table 19 - Chemo7 columns

Name	Data type	
pnr	Long Integer	Patient pseudonym
sts	Long Integer	
studie_teil	Long Integer	
studie	Long Integer	
studie_com	Memo	safe
w_ther	Long Integer	
w_ther_com	Memo	Locations
v_andklin	Long Integer	
v_andklin_com	Memo	Locations
beh_klin	Long Integer	
gr_malign	Long Integer	
gr_malign_com	Memo	safe
n_sym	Long Integer	
n_sym_com	Memo	safe
vortum	Long Integer	
vortum_com	Memo	safe
syndr	Long Integer	
anir	Long Integer	
wagr	Long Integer	
urofehl	Long Integer	
drash	Long Integer	
Emg	Long Integer	
Hemihyp	Long Integer	
Perl	Long Integer	
famwilms	Long Integer	
gerinn	Long Integer	
hypert	Long Integer	
rr_sys	Long Integer	
rr_dias	Long Integer	
and_syn	Long Integer	
and_syn_com	Memo	safe
fambel	Long Integer	
fam_elt	Long Integer	
fam_elt_com	Memo	safe
fam_ges	Long Integer	
fam_ges_com	Memo	safe
fam_sonst	Long Integer	
fam_sonst_com	Memo	safe
anzges	Long Integer	
mehrling	Long Integer	
meh_art	Long Integer	
meh_ent	Long Integer	
gebjmut	Long Integer	
gebjvat	Long Integer	
az	Long Integer	
diagdat	DateTime	

therdat	DateTime	
thermit	Long Integer	
thermit_com	Memo	safe
lokal	Long Integer	
metast	Long Integer	
lunge	Long Integer	
lu_nachw	Long Integer	
mediast	Long Integer	
leber	Long Integer	
exablk	Long Integer	
abdomen	Long Integer	
knochen	Long Integer	
weicht	Long Integer	
gehirn	Long Integer	
met_and	Long Integer	
met_and_com	Memo	safe
roe	Long Integer	
thor_ct	Long Integer	
ab_ct	Long Integer	
us	Long Integer	
mrt	Long Integer	
sonst	Long Integer	
sonst_com	Memo	safe
roe_anz	Long Integer	
ct_anz	Long Integer	
dur_lu_met	Long Integer	
kat_urin	Long Integer	
ref_rad	Long Integer	
us_r	Long Integer	
ct_r	Long Integer	
mrt_r	Long Integer	
us_l	Long Integer	
ct_l	Long Integer	
mrt_l	Long Integer	
us_a_r	Double	
us_b_r	Double	
us_c_r	Double	
us_v_r	Double	
ct_a_r	Double	
ct_b_r	Double	
ct_c_r	Double	
ct_v_r	Double	
mrt_a_r	Double	
mrt_b_r	Double	
mrt_c_r	Double	
mrt_v_r	Double	
bild_sp_r	Long Integer	
anztum_r	Long Integer	
tumstruk_r	Long Integer	
biop_r	Long Integer	
biop_r_spez	Long Integer	
gauche_r	Long Integer	
biopdat_r	DateTime	
us_a_l	Double	
us_b_l	Double	
us_c_l	Double	
us_v_l	Double	
ct_a_l	Double	
ct_b_l	Double	
ct_c_l	Double	

ct_v_l	Double	
mrt_a_l	Double	
mrt_b_l	Double	
mrt_c_l	Double	
mrt_v_l	Double	
bild_sp_l	Long Integer	
anztum_l	Long Integer	
tumstruk_l	Long Integer	
biop_l	Long Integer	
biop_l_spez	Long Integer	
gauche_l	Long Integer	
biopdat_l	DateTime	
protpat	Long Integer	
alter	Long Integer	
prim_op	Long Integer	
op_grund	Long Integer	
and_grund	Long Integer	
and_grund_com	Memo	safe
vorbeh	Long Integer	
beh	Long Integer	
beh_com	Memo	safe
bilat	Long Integer	
andtum	Long Integer	
andtum_com	Memo	safe
fup	Long Integer	
fup_com	Memo	safe
patlebt	Long Integer	
comment	Memo	

Table 20 - f1 columns

Name	Data type	
pnr	Long Integer	Patient pseudonym
sts	Long Integer	
diag	Long Integer	
rec_vcr	Long Integer	
rec_act	Long Integer	
rec_sonst	Long Integer	
rec_sonst_com	Memo	safe
dur_preop	Long Integer	
dat_preop	DateTime	
prog_left	Long Integer	
prog_right	Long Integer	
bet_surg	Long Integer	
bet_vcr	Long Integer	
bet_act	Long Integer	
bet_doxo	Long Integer	
post_surg	Long Integer	
post_av2	Long Integer	
post_avd	Long Integer	
post_cdcv	Long Integer	
post_sonst	Long Integer	
post_sonst_com	Memo	safe
radio_pat	Long Integer	
dat_treat	Memo	safe
comment	Memo	

Table 21 - f13 columns

Name	Data type	
idx	Long Integer	f2

		pseudonym
pnr	Long Integer	Patient pseudonym
sts	Long Integer	
formid	Long Integer	
gewicht	Double	
groesse	Long Integer	
oberfl	Double	
tox	Long Integer	
vod	Long Integer	
respdat_r	DateTime	
us_a_r	Double	
us_b_r	Double	
us_c_r	Double	
us_v_r	Double	
ct_a_r	Double	
ct_b_r	Double	
ct_c_r	Double	
ct_v_r	Double	
mrt_a_r	Double	
mrt_b_r	Double	
mrt_c_r	Double	
mrt_v_r	Double	
tumstruk_r	Long Integer	
respdat_l	DateTime	
us_a_l	Double	
us_b_l	Double	
us_c_l	Double	
us_v_l	Double	
ct_a_l	Double	
ct_b_l	Double	
ct_c_l	Double	
ct_v_l	Double	
mrt_a_l	Double	
mrt_b_l	Double	
mrt_c_l	Double	
mrt_v_l	Double	
tumstruk_l	Long Integer	
lunge_roe	Long Integer	
lunge_ct	Long Integer	
mediast	Long Integer	
leber	Long Integer	
abdomen	Long Integer	
knochen	Long Integer	
gehirn	Long Integer	
andmet	Long Integer	
andmet_com	Memo	safe
diag_met	Long Integer	
diag_met_anz	Long Integer	
prae_met	Long Integer	
prae_met_anz	Long Integer	
status2	Long Integer	
thora_ct	Long Integer	
thora_ct_com	Memo	safe
comment	Memo	

Table 22 - f2 columns

Name	Data type	
idx	Long Integer	f3 pseudonym

pnr	Long Integer	Patient pseudonym
sts	Long Integer	
opdat	DateTime	
operateur	Long Integer	
chirurg	Long Integer	
ber_op	Long Integer	
klin_op	Long Integer	
anzop_kh	Long Integer	
anzop_chir	Long Integer	
praeop_beh	Long Integer	
lokal	Long Integer	
lokal_prob	Long Integer	
cav_aus_diag	Long Integer	
cav_aus_ct	Long Integer	
gefaess_chir	Long Integer	
op_zugang	Long Integer	
op_zugang_com	Memo	safe
opart	Long Integer	
nephrec	Long Integer	
radikal	Long Integer	
part_neph	Long Integer	
stadop	Long Integer	
lk_makr	Long Integer	
per_aus	Long Integer	
per_biop	Long Integer	
per_biop_com	Memo	safe
per_radikal	Long Integer	
nv_aus	Long Integer	
nv_biop	Long Integer	
nv_biop_com	Memo	safe
nv_radikal	Long Integer	
nv_throm	Long Integer	
vc_aus	Long Integer	
vc_biop	Long Integer	
vc_biop_com	Memo	safe
vc_radikal	Long Integer	
vc_throm	Long Integer	
vc_byp	Long Integer	
vc_pro	Long Integer	
kap_aus	Long Integer	
kap_rup	Long Integer	
kap_rup_wann	Long Integer	
kap_rup_art	Long Integer	
lk_aus_reg	Long Integer	
lk_aus_ex	Long Integer	
lk_ent_reg	Long Integer	
lk_ent_ex	Long Integer	
lk_ent_com	Memo	safe
lk_rup	Long Integer	
lk_ent_anz	Long Integer	
un_int	Long Integer	
un_int_com	Memo	safe
nn_aus	Long Integer	
nn_ent	Long Integer	
fett_aus	Long Integer	
fett_ent	Long Integer	
ur_aus	Long Integer	
ur_ent	Long Integer	
le_aus	Long Integer	

le_ent	Long Integer	
pso_aus	Long Integer	
pso_ent	Long Integer	
zwe_aus	Long Integer	
zwe_ent	Long Integer	
mi_aus	Long Integer	
mi_ent	Long Integer	
pa_aus	Long Integer	
pa_ent	Long Integer	
co_aus	Long Integer	
co_ent	Long Integer	
kont_aus	Long Integer	
kont_ent	Long Integer	
and_aus	Long Integer	
and_ent	Long Integer	
and_com	Memo	safe
comment	Memo	

Table 23 - f3a columns

Name	Data type	
idx	Long Integer	f3 pseudonym
pnr	Long Integer	patient pseudonym
sts	Long Integer	
metopdat	DateTime	
operateur	Long Integer	
chirurg	Long Integer	
ber_op	Long Integer	
klin_op	Long Integer	
indikation	Long Integer	
praeopchem	Long Integer	
praeoprad	Long Integer	
praeopmetop	Long Integer	
met_lunge	Long Integer	
met_knochen	Long Integer	
met_zns	Long Integer	
met_leber	Long Integer	
met_weich	Long Integer	
met_sonst	Long Integer	
met_com	Memo	safe
met_chron	Long Integer	
seite	Long Integer	
mekt_erf	Long Integer	
verwachs	Long Integer	
met_ent	Long Integer	
met_ent_kpl	Long Integer	
met_ent_ink	Long Integer	
mekt_gen	Long Integer	
w_keine	Long Integer	
w_chemo	Long Integer	
w_radio	Long Integer	
w_reop	Long Integer	
w_stamm	Long Integer	
comment	Memo	

Table 24 - f3b columns

Name	Data type	
idx	Long Integer	f3 pseudonym
pnr	Long Integer	Patient pseudonym
sts	Long Integer	
opidx	Long Integer	
optype	Long Integer	
op_kompl_dat	DateTime	
kompl	Long Integer	
tumrupmi	Long Integer	
tumrupma	Long Integer	
blut	Long Integer	
druckabf	Long Integer	
herz	Long Integer	
gefaess	Long Integer	
darm	Long Integer	
darmver	Long Integer	
milz	Long Integer	
leber	Long Integer	
sonst	Long Integer	
sonst_com	Memo	safe
postop_kompl	Long Integer	
post_blut	Long Integer	
vc_obs	Long Integer	
adhaes	Long Integer	
invag	Long Integer	
wundinf	Long Integer	
wunddeh	Long Integer	
narbenh	Long Integer	
zwerchfh	Long Integer	
post_sonst	Long Integer	
post_sonst_com	Memo	safe
resand	Long Integer	
resorg	Long Integer	
resorg_com	Memo	safe
resgrund	Long Integer	
kompmed	Long Integer	
kompchir	Long Integer	
reop	Long Integer	
reopdat	DateTime	
tod	Long Integer	
tod_com	Memo	safe
spaetf	Long Integer	
spaetf_com	Memo	safe
verzoeg	Long Integer	
verzanz	Long Integer	
comment	Memo	

Table 25 - f3k columns

Name	Data type	
idx	Long Integer	f4 pseudonym
pnr	Long Integer	patient pseudonym
sts	Long Integer	
pdate	DateTime	
lokal_prob	Long Integer	

Table 26 - f4 columns

Name	Data type	
idx	Long Integer	f4_sub pseudonym

f4idx	Long Integer	f4 pseudonym
rel_his	Long Integer	
pathtyp	Long Integer	
berdat	DateTime	
eingnr	Memo	
lok_path	Long Integer	
klin_path	Long Integer	
anz_sch	Long Integer	
anz_blo	Long Integer	
chir_inf	Long Integer	
prim_beh	Long Integer	
lokal	Long Integer	
tum_mat	Long Integer	
gew_praep	Long Integer	
durchm	Double	
praep_intakt	Long Integer	
kaps_intakt	Long Integer	
ob_tusch	Long Integer	
multifok	Long Integer	
multifok_com	Memo	safe
rand_mak	Long Integer	
rand_mik	Long Integer	
rand_com	Memo	safe
thromb_nv_mak	Long Integer	
thromb_nv_mik	Long Integer	
proz_mak	Long Integer	
proz_mak_gen	Long Integer	
proz_his	Long Integer	
proz_his_gen	Long Integer	
proz_blas	Long Integer	
proz_epi	Long Integer	
proz_stroma	Long Integer	
reg_niere	Long Integer	
reg_perih	Long Integer	
reg_perir	Long Integer	
reg_lymph	Long Integer	
reg_res	Long Integer	
vit_niere	Long Integer	
vit_perih	Long Integer	
vit_perir	Long Integer	
vit_lymph	Long Integer	
vit_res	Long Integer	
inf_niere	Long Integer	
inf_hilus	Long Integer	
inf_venen	Long Integer	
inf_cava	Long Integer	
reste	Long Integer	
restyp	Long Integer	
ana_subtyp	Long Integer	
histotyp	Long Integer	
histo_sonst	Memo	safe
lk_tum	Long Integer	
lk_com	Memo	safe
lk_ent_anz	Long Integer	
vit_tum_anz	Long Integer	
stadlok	Long Integer	
stadlok_com	Memo	safe
stud_ass	Long Integer	
stud_vers	Long Integer	
mirrbl	Long Integer	
comment	Memo	

Table 27 - f4_sub columns

Name	Data type	
idx	Long Integer	f6 pseudonym
pnr	Long Integer	Patient pseudonym
sts	Long Integer	
nameradio	Long Integer	
klin_radio	Long Integer	
best_organ	Long Integer	
best_organ_sonst	Memo	safe
best_beg	DateTime	
best_end	DateTime	
geraet	Long Integer	
geraet_com	Memo	safe
prim_feld	Long Integer	
sat_aus	Long Integer	
sat_aus_com	Memo	safe
boost	Long Integer	
boost_com	Memo	safe
boost_sat_aus	Long Integer	
ges_dos	Double	
einz_dos	Double	
anz_best	Long Integer	
dauer	Long Integer	
unterbr	Long Integer	
unterbr_com	Memo	safe
leb_dos	Double	
kont_dos	Double	
boost_dos	Double	
b_anz_sitz	Long Integer	
b_dauer	Long Integer	
b_unterbr	Long Integer	
b_unterbr_com	Memo	safe
uebel	Long Integer	
erbr	Long Integer	
hepatox	Long Integer	
hepatox_com	Memo	safe
andtox_com	Memo	safe
hb	Double	
leuko	Long Integer	
neutro	Long Integer	
thrombo	Long Integer	
prim_ant_l	Double	
prim_ant_b	Double	
prim_post_l	Double	
prim_post_b	Double	
prim_and_l	Double	
prim_and_b	Double	
boost_ant_l	Double	
boost_ant_b	Double	
boost_post_l	Double	
boost_post_b	Double	
boost_and_l	Double	
boost_and_b	Double	
comment	Memo	

Table 28 - f6 columns

Name	Data type	
idx	Long Integer	f7 pseudonym
pnr	Long Integer	Patient pseudonym
sts	Long Integer	
formid	Long Integer	
stadlok	Long Integer	
histo1	Long Integer	
histo2	Long Integer	
volumen	Long Integer	
patrand	Long Integer	
random	Long Integer	
random_com	Memo	safe
sollther	Long Integer	
ther_dox	Long Integer	
ther_dox_com	Memo	safe
gcsf	Long Integer	
tox	Long Integer	
vod	Long Integer	
tox_verst	Long Integer	
status_ta	Long Integer	
postop_kompl	Memo	safe
grund_indiv	Memo	safe
gewicht	Double	
groesse	Long Integer	
oberfl	Double	
stad1_spez	Long Integer	
stad2_spez	Long Integer	
stad3_spez	Long Integer	
stad4_spez	Long Integer	
stad5_spez	Long Integer	
comment	Memo	

Table 29 - f7 columns

Name	Data type	
idx	Long Integer	f8a pseudonym
pnr	Long Integer	Patient pseudonym
f8acat	Long Integer	
catdose	Double	
sts	Long Integer	
dat	DateTime	
arrhyth	Long Integer	
arrhyth_ther	Long Integer	
herz_klin	Long Integer	
infdauer	Double	
echo	Long Integer	
sf_wert	Double	
esws	Double	
dias_path	Long Integer	
digital	Long Integer	
diuret	Long Integer	
ckmb	Long Integer	
blut_vers	Long Integer	
puls	Long Integer	
anaemie	Long Integer	
fieber	Long Integer	
syst	Long Integer	
diast	Long Integer	
comment	Memo	

Table 30 - f8a columns

Name	Data type	
idx	Long Integer	f8b pseudonym
pnr	Long Integer	Patient pseudonym
sts	Long Integer	
n_comment	Memo	safe
event_com	Memo	safe
meddra_llt_code	Long Integer	
toxgrad	Long Integer	
beginn	DateTime	
ende	DateTime	
weiterbest	Long Integer	
and_urs	Long Integer	
ther_urs	Long Integer	
verlauf	Long Integer	
verlauf_2	Long Integer	
comment	Memo	

Table 31 - f8b columns

Name	Data type	
idx	Long Integer	f8c pseudonym
pnr	Long Integer	Patient pseudonym
sts	Long Integer	
begdat	DateTime	
enddat	DateTime	
gpt	Double	
bili	Double	
ascit	Long Integer	
gewicht_zu	Double	
hepmeg	Long Integer	
heppain	Long Integer	
lebgroe	Double	
splmeg	Long Integer	
splpain	Long Integer	
milzgroe	Double	
actvorvod	Long Integer	
datact	DateTime	
dosact	Double	
gewicht_act	Double	
pat_bestr	Long Integer	
leb_bestr	Long Integer	
ther_prae	Long Integer	
ther_post	Long Integer	
ther_zeit	Long Integer	
ther_zeit_com	Memo	safe
and_fakt	Long Integer	
and_fakt_com	Memo	safe
comment	Memo	

Table 32 - f8c columns

Name	Data type	
idx	Long Integer	f9 pseudonym
pnr	Long Integer	Patient

		pseudonym
sts	Long Integer	
l_unt	DateTime	
ther_end	Long Integer	
status	Long Integer	
rezmet	Long Integer	
rem	Long Integer	
rez	Long Integer	
rezdat	DateTime	
met	Long Integer	
metdat	DateTime	
met_lu	Long Integer	
met_le	Long Integer	
met_ab	Long Integer	
met_zns	Long Integer	
met_kno	Long Integer	
met_kno_com	Memo	safe
met_lk	Long Integer	
met_lk_com	Memo	safe
met_weich	Long Integer	
met_weich_com	Memo	safe
met_sonst	Long Integer	
met_sonst_com	Memo	
zweitrem	Long Integer	
zweitremdat	DateTime	
kont	Long Integer	
symp	Long Integer	
rout_unt	Long Integer	
klin_unt	Long Integer	
bild	Long Integer	
us	Long Integer	
ct	Long Integer	
mrt	Long Integer	
vor_rueck	DateTime	
roe_thor	DateTime	
sono_ab	DateTime	
zweittum	Long Integer	
zweit_com	Memo	safe
zweit_diagdat	DateTime	
zweit_malig	Long Integer	
zweit_lok	Memo	safe
zweit_bestgeb	Long Integer	
spaet_kompl	Long Integer	
spaet_herz	Long Integer	
spaet_nier	Long Integer	
spaet_skel	Long Integer	
spaet_com	Memo	safe
totdat	DateTime	
autopsie	Long Integer	
tod_com	Memo	safe
todurs	Long Integer	
comment	Memo	

Table 33 - f9 columns

Name	Data type	
idx	Long Integer	f8a pseudonym
sts	Long Integer	
gewicht	Double	
groesse	Long Integer	
oberfl	Double	
datdox	DateTime	
doxein	Long Integer	
dox	Double	
infdauer	Double	
doxkum	Double	
blut_vor	Long Integer	
blut_end	Long Integer	
blut24	Long Integer	
blut48	Long Integer	
blut5	Long Integer	
blut21	Long Integer	
dox_vor	Double	
dox_end	Double	
tro_vor	Double	
tro24	Double	
tro48	Double	
tro5	Double	
bnp_vor	Double	
bnp21	Double	
ein_klinik	Long Integer	

Table 34 - labor columns

Name	Data type	
pnr	Long Integer	Patient pseudonym
sts	Long Integer	
auf_nr	Memo	
diagnose	Memo	safe
stad	Memo	safe
malign	Memo	safe
dgdat	DateTime	
lokal	Memo	safe
dgsich	Long Integer	
seite_mb	Long Integer	
stud	Long Integer	
studname	Long Integer	
einv	Long Integer	
efb	Long Integer	
dat	DateTime	

Table 35 - meldung columns

3.2.1.6 Further Tables

Name	Data type
pt_code	Long Integer
hlt_code	Long Integer
hlgt_code	Long Integer
soc_code	Long Integer

pt_name	Text
hlt_name	Text
hlgt_name	Text
soc_name	Text
soc_abbrev	Text

bull_field	Text
pt_soc_code	Long Integer
primary_soc_fg	Text

Table 36 - md_hierarchy columns

Name	Data type
idx	Long Integer
sts	Long Integer
nematose	Long Integer
anzherde_re	Long Integer
herde_re_diff	Long Integer
anzherde_li	Long Integer
herde_li_diff	Long Integer
herd_re_a	Long Integer
herd_re_b	Long Integer
herd_re_c	Long Integer
herd_re_v	Long Integer
herd_li_a	Long Integer
herd_li_b	Long Integer
herd_li_c	Long Integer
herd_li_v	Long Integer
artmessung	Long Integer
thrombus	Long Integer
thr_lokal	Long Integer
thr_morph	Long Integer
cavacomp	Long Integer
lkmet	Long Integer
lkmet_hilre	Long Integer
lkmet_hilli	Long Integer
lkmet_intab	Long Integer
lkmet_extab	Long Integer
lkmet_sonst	Long Integer
lkmet_sonst_com	Memo
lumetbilder	Long Integer
va_lumet	Long Integer
lumet	Long Integer
lumet_roetxp	Long Integer
lumet_roetxp_sol	Long Integer
lumet_roetxp_anzli	Long Integer
lumet_roetxp_anzre	Long Integer
lumet_roetxs	Long Integer
lumet_roetxs_sol	Long Integer
lumet_cttx	Long Integer
lumet_cttx_sol	Long Integer
lumet_cttx_anzli	Long Integer
lumet_cttx_anzre	Long Integer
metdiaggleich	Long Integer
lemet	Long Integer
lemet_anz	Long Integer
andmetbilder	Long Integer
va_andmet	Long Integer
andmet	Long Integer
knomet	Long Integer

3.2.1.7 Code Tables

Code tables do not have to be anonymised.

hirnmet	Long Integer
sonstmet	Long Integer
sonstmet_com	Memo
metas_response	Long Integer
lok_response	Long Integer
neu_metas	Long Integer
neu_lokrez	Long Integer

Table 37 - metastasen columns

Name	Data type
soc_code	Long Integer
hlgt_code	Long Integer

Table 38 - vsoc_hlgt_dupes columns

Name	Data type
hlgt_code	Long Integer
hlgt_name	Text
soc_code	Long Integer

Table 39 - v_meddra_hlgt columns

Name	Data type
hlt_code	Long Integer
hlt_name	Text
hlgt_code	Long Integer

Table 40 - v_meddra_hlt columns

Name	Data type
llt_code	Long Integer
llt_name	Text
pt_code	Long Integer

Table 41 - v_meddra_llt columns

Name	Data type
pt_code	Long Integer
pt_name	Text
hlt_code	Long Integer

Table 33 - v_meddra_pt columns

Name	Data type
soc_code	Long Integer
soc_name	Text

Table 42 - v_meddra_soc columns

Name	Data type
SwitchboardID	Long Integer
ItemNumber	Integer
ItemText	Text
Command	Integer
Argument	Text

Table 43 - Switchboard Items columns

Name	Data type
idx	Long Integer
name	Memo

Table 44 - code_abtklin columns

Name	Data type

idx	Long Integer
alter	Memo

Table 45 - code_alter columns

Name	Data type
idx	Long Integer
ana_subtyp	Memo

Table 46 - code_ana_subtyp columns

Name	Data type
idx	Long Integer
anf_dient	Memo

Table 47 - code_anf_dient columns

Name	Data type
idx	Long Integer
ankuend	Memo

Table 48 - code_ankuend columns

Name	Data type
idx	Long Integer
anlass	Memo

Table 49 - coe_anlass columns

Name	Data type
idx	Long Integer
anzop	Memo

Table 50 - code_anzop columns

Name	Data type
idx	Long Integer
anztum	Memo

Table 51 - code_anztum columns

Name	Data type
idx	Long Integer
artklin	Memo

Table 52 - code_artklin columns

Name	Data type
idx	Long Integer
artmess	Memo

Table 53 - code_artmess columns

Name	Data type
idx	Long Integer
name	Memo

Table 54 - code_asse columns

Name	Data type
idx	Long Integer
aus	Memo

Table 55 - code_aus columns

Name	Data type
idx	Long Integer
name	Memo

Table 56 - code_ausdeh columns

Name	Data type

idx	Long Integer
austl	Memo

Table 57 - code_austl columns

Name	Data type
idx	Long Integer
ausw	Memo

Table 58 - code_ausw columns

Name	Data type
idx	Long Integer
az	Memo

Table 59 - code_az columns

Name	Data type
idx	Long Integer
bean_anf	Memo

Table 60 - code_bean_anf columns

Name	Data type
idx	Long Integer
beant_zeit	Memo

Table 61 - code_beant_zeit columns

Name	Data type
idx	Long Integer
bef	Memo

Table 62 - code_bef columns

Name	Data type
idx	Long Integer
bef_subj	Memo

Table 63 - code_bef_subj columns

Name	Data type
idx	Long Integer
befund	Memo

Table 64 - code_befund columns

Name	Data type
idx	Long Integer
befzeit	Memo

Table 65 - code_befzeit columns

Name	Data type
idx	Long Integer
begruendung	Memo

Table 66 - code_begruendung columns

Name	Data type
idx	Long Integer
beh_klin	Memo

Table 67 - code_beh_klin columns

Name	Data type
idx	Long Integer
berat_erf	Memo

Table 68 - code_berat_erf columns

Name	Data type

idx	Long Integer
berat_pat	Memo

Table 69 - code_berat_pat columns

Name	Data type
idx	Long Integer
beruf	Memo

Table 70 - code_beruf columns

Name	Data type
idx	Long Integer
beurt	Memo

Table 71 - code_beurt columns

Name	Data type
idx	Long Integer
bild_sp	Memo

Table 72 - code_bild_sp columns

Name	Data type
idx	Long Integer
bildvon	Memo

Table 73 - code_bildvon columns

Name	Data type
idx	Long Integer
biop_spez	Memo

Table 74 - code_biop_spez columns

Name	Data type
idx	Long Integer
cat	Memo

Table 75 - code_cat columns

Name	Data type
idx	Long Integer
category	Memo

Table 76 - code_category columns

Name	Data type
idx	Long Integer
cav_aus	Memo

Table 77 - code_cav_aus columns

Name	Data type
idx	Long Integer
color	Memo

Table 78 - code_color columns

Name	Data type
idx	Long Integer
ctc	Memo

Table 79 - code_ctc columns

Name	Data type
idx	Long Integer
dgsich	Memo

Table 80 - code_dgsich columns

Name	Data type

idx	Long Integer
name	Memo

Table 81 - code_diag columns

Name	Data type
idx	Long Integer
diag_klin	Memo

Table 82 - code_diag_klin columns

Name	Data type
idx	Long Integer
diag_refrad	Memo

Table 83 - code_diag_refrad columns

Name	Data type
idx	Long Integer
diaggleich	Memo

Table 84 - code_diaggleich columns

Name	Data type
idx	Long Integer
dok	Memo

Table 85 - code_dok columns

Name	Data type
idx	Long Integer
due_nm	Memo

Table 86 - code_due_nm columns

Name	Data type
idx	Long Integer
due_per	Memo

Table 87 - code_due_per columns

Name	Data type
idx	Long Integer
due_schluess	Memo

Table 88 - code_due_schluess columns

Name	Data type
idx	Long Integer
due_von	Memo

Table 89 - code_due_von columns

Name	Data type
idx	Long Integer
due_weg	Memo

Table 90 - code_due_weg columns

Name	Data type
idx	Long Integer
echo	Memo

Table 91 - code_echo columns

Name	Data type
idx	Long Integer
einv	Memo

Table 92 - code_einv columns

Name	Data type

idx	Long Integer
erf	Memo

Table 93 - code_erf columns

Name	Data type
idx	Long Integer
name	Memo

Table 94 - code_f8acat columns

Name	Data type
idx	Long Integer
geeig	Memo

Table 95 - code_geeig columns

Name	Data type
idx	Long Integer
geraet	Memo

Table 96 - code_geraet columns

Name	Data type
idx	Long Integer
gr_malign	Memo

Table 97 - code_gr_malign columns

Name	Data type
idx	Long Integer
grund	Memo

Table 98 - code_grund columns

Name	Data type
idx	Long Integer
grundl_bef	Memo

Table 99 - code_grundl_bef columns

Name	Data type
idx	Long Integer
hepatox	Memo

Table 100 - code_hepatox columns

Name	Data type
idx	Long Integer
histo1	Memo

Table 101 - code_histo1 columns

Name	Data type
idx	Long Integer
histo2	Memo

Table 102 - code_histo2 columns

Name	Data type
idx	Long Integer
name	Memo

Table 103 - code_histosubtyp columns

Name	Data type
idx	Long Integer
histotyp	Memo

Table 104 - code_histotyp columns

Name	Data type

idx	Long Integer
hom	Memo

Table 105 - code_hom columns

Name	Data type
idx	Long Integer
indikation	Memo

Table 106 - code_indikation columns

Name	Data type
idx	Long Integer
kap_rup_art	Memo

Table 107 - code_kap_rup_art columns

Name	Data type
idx	Long Integer
kap_rup_wann	Memo

Table 108 - code_kap_rup_wann columns

Name	Data type
idx	Long Integer
klin_an	Memo

Table 109 - code_klin_an columns

Name	Data type
idx	Long Integer
km_aufn	Memo

Table 110 - code_km_aufn columns

Name	Data type
idx	Long Integer
km_aufnst	Memo

Table 111 - code_km_aufnst columns

Name	Data type
idx	Long Integer
kontakt	Memo

Table 112 - code_kontakt columns

Name	Data type
idx	Long Integer
land_name	Memo

Table 113 - code_land columns

Name	Data type
idx	Long Integer
lk_ent	Memo

Table 114 - code_lk_ent columns

Name	Data type
idx	Long Integer
lk_tum	Memo

Table 115 - code_lk_tum columns

Name	Data type
idx	Long Integer
lok_intra	Memo

Table 116 - code_lok_intra columns

Name	Data type

idx	Long Integer
lok_lu	Memo

Table 117 - code_lok_lu columns

Name	Data type
idx	Long Integer
lok_nier	Memo

Table 118 - code_lok_nier columns

Name	Data type
idx	Long Integer
lok_soli	Memo

Table 119 - code_lok_soli columns

Name	Data type
idx	Long Integer
lokal	Memo

Table 120 - code_lokal columns

Name	Data type
idx	Long Integer
name	Memo

Table 121 - code_lokal_prob columns

Name	Data type
idx	Long Integer
maengel	Memo

Table 122 - code_maengel columns

Name	Data type
idx	Long Integer
massn_subj	Memo

Table 123 - code_massn_subj columns

Name	Data type
idx	Long Integer
meh_art	Memo

Table 124 - code_meh_art columns

Name	Data type
idx	Long Integer
meh_ent	Memo

Table 125 - code_meh_ent columns

Name	Data type
idx	Long Integer
mekt_erf	Memo

Table 126 - code_mekt_erf columns

Name	Data type
idx	Long Integer
mekt_gen	Memo

Table 127 - code_mekt_gen columns

Name	Data type
idx	Long Integer
met_chron	Memo

Table 128 - code_met_chron columns

Name	Data type

idx	Long Integer
metbild	Memo

Table 129 - code_metbild columns

Name	Data type
idx	Long Integer
name	Memo

Table 130 - code_nephrec columns

Name	Data type
idx	Long Integer
nier_erh	Memo

Table 131 - code_nier_erh columns

Name	Data type
idx	Long Integer
op_grund	Memo

Table 132 - code_op_grund columns

Name	Data type
idx	Long Integer
op_zugang	Memo

Table 133 - code_op_zugang columns

Name	Data type
idx	Long Integer
op3a	Memo

Table 134 - code_op3a columns

Name	Data type
idx	Long Integer
op3b	Memo

Table 135 - code_op3b columns

Name	Data type
idx	Long Integer
opart	Memo

Table 136 - code_opart columns

Name	Data type
idx	Long Integer
name	Memo

Table 137 - code_organ columns

Name	Data type
idx	Long Integer
orgaus	Memo

Table 138 - code_orgaus columns

Name	Data type
idx	Long Integer
part_neph	Memo

Table 139 - code_part_neph columns

Name	Data type
idx	Long Integer
pat_gem	Memo

Table 140 - code_pat_gem columns

Name	Data type

idx	Long Integer
patadmin	Memo

Table 141 - code_patadmin columns

Name	Data type
idx	Long Integer
Patart	Memo

Table 142 - code_patart columns

Name	Data type
idx	Long Integer
name	Memo

Table 143 - code_pathtyp columns

Name	Data type
idx	Long Integer
per_aus	Memo

Table 144 - code_per_aus columns

Name	Data type
idx	Long Integer
pos_berat	Memo

Table 145 - code_pos_berat columns

Name	Data type
idx	Long Integer
praeop_beh	Memo

Table 146 - code_praeop_beh columns

Name	Data type
idx	Long Integer
prim_beh	Memo

Table 147 - code_prim_beh columns

Name	Data type
idx	Long Integer
prim_feld	Memo

Table 148 - code_prim_feld columns

Name	Data type
idx	Long Integer
name	Memo

Table 149 - code_prog columns

Name	Data type
idx	Long Integer
proz	Memo

Table 150 - code_proz columns

Name	Data type
idx	Long Integer
quali	Memo

Table 151 - code_quali columns

Name	Data type
idx	Long Integer
radikal	Memo

Table 152 - code_radikal columns

Name	Data type

idx	Long Integer
radikal2	Memo

Table 153 - code_radikal2 columns

Name	Data type
idx	Long Integer
radikal3	Memo

Table 154 - code_radikal3 columns

Name	Data type
idx	Long Integer
rand	Memo

Table 155 - code_rand columns

Name	Data type
idx	Long Integer
rand_erg	Memo

Table 156 - code_rand_erg columns

Name	Data type
idx	Long Integer
name	Memo

Table 157 - code_rand_erg_2 columns

Name	Data type
idx	Long Integer
name	Memo

Table 158 - code_rand_erg_3 columns

Name	Data type
idx	Long Integer
random	Memo

Table 159 - code_random columns

Name	Data type
idx	Long Integer
refdiagwie	Memo

Table 160 - code_refdiagwie columns

Name	Data type
idx	Long Integer
resgrund	Memo

Table 161 - code_resgrund columns

Name	Data type
idx	Long Integer
resorg	Memo

Table 162 - code_resorg columns

Name	Data type
idx	Long Integer
response	Memo

Table 163 - code_response columns

Name	Data type
idx	Long Integer
name	Memo

Table 164 - code_restyp columns

Name	Data type

idx	Long Integer
rezauf	Memo

Table 165 - code_rezauf columns

Name	Data type
idx	Long Integer
name	Memo

Table 166 - code_rueck columns

Name	Data type
idx	Long Integer
ruecks	Memo

Table 167 - code_ruecks columns

Name	Data type
idx	Long Integer
sat_aus	Memo

Table 168 - code_sat_aus columns

Name	Data type
idx	Long Integer
seite	Memo

Table 169 - code_seite columns

Name	Data type
idx	Long Integer
seite_mb	Memo

Table 170 - code_seite_mb columns

Name	Data type
idx	Long Integer
sex	Memo

Table 171 - code_sex columns

Name	Data type
idx	Long Integer
softw	Memo

Table 172 - code_softw columns

Name	Data type
idx	Long Integer
soli	Memo

Table 173 - code_soli columns

Name	Data type
idx	Long Integer
sortname	Memo

Table 174 - code_sort columns

Name	Data type
idx	Long Integer

Name	Data type
idx	Long Integer
thora_ct	Memo

Table 185 - code_thora_ct columns

Name	Data type
idx	Long Integer
thr_lokal	Memo

Table 186 - code_thr_lokal columns

stadium	Memo

Table 175 - code_stadium columns

Name	Data type
idx	Long Integer
status	Memo

Table 176 - code_status columns

Name	Data type
idx	Long Integer
status_ta	Memo

Table 177 - code_status_ta columns

Name	Data type
idx	Long Integer
status2	Memo

Table 178 - code_status2 columns

Name	Data type
idx	Long Integer
sts	Memo
stslock	Long Integer
stscolor	Long Integer

Table 179 - code_sts columns

Name	Data type
idx	Long Integer
name	Memo

Table 180 - code_stud_disk columns

Name	Data type
idx	Long Integer
studie	Memo

Table 181 - code_studie columns

Name	Data type
idx	Long Integer
tag	Memo

Table 182 - code_tag columns

Name	Data type
idx	Long Integer
ther_dox	Memo

Table 183 - code_ther_dox columns

Name	Data type
idx	Long Integer
thermit	Memo

Table 184 - code_thermit columns

Name	Data type
idx	Long Integer
thr_morph	Memo

Table 187 - code_thr_morph columns

Name	Data type
idx	Long Integer
throm	Memo

Table 188 - code_throm columns

Name	Data type
idx	Long Integer
todurs	Memo

Table 189 - code_todurs columns

Name	Data type
idx	Long Integer
toxgrad	Memo

Table 190 - code_toxgrad columns

Name	Data type
idx	Long Integer
tum_mat	Memo

Table 191 - code_tum_mat columns

Name	Data type
idx	Long Integer
tumlok	Memo

Table 192 - code_tumlok columns

Name	Data type
idx	Long Integer
tumstruk	Memo

Table 193 - code_tumstruk columns

Name	Data type
idx	Long Integer
tumvol	Memo

Table 194 - code_tumvol columns

Name	Data type
idx	Long Integer
uebel	Memo

Table 195 - code_uebel columns

Name	Data type
idx	Long Integer
urs	Memo

Table 196 - code_urs columnse

Name	Data type
idx	Long Integer
us_format	Memo

Table 197 - code_us_format columns

Name	Data type
idx	Long Integer
v_aus	Memo

Table 198 - code_v_aus columns

Name	Data type
idx	Long Integer
va_neph	Memo

Table 199 - code_va_neph columns

Name	Data type
idx	Long Integer
verlauf	Memo

Table 200 - code_verlauf columns

Name	Data type
idx	Long Integer
verlauf_2	Memo

Table 201 - code_verlauf_2 columns

Name	Data type
idx	Long Integer
volumen	Memo

Table 202 - code_volumen columns

Name	Data type
idx	Long Integer
vorl	Memo

Table 203 - code_vorl columns

Name	Data type
idx	Long Integer
yn	Memo

Table 204 - code_yn columns

Name	Data type
idx	Long Integer
yne	Memo

Table 205 - code_yne columns

Name	Data type
idx	Long Integer
yni	Memo

Table 206 - code_yni columns

Name	Data type
idx	Long Integer
ynn	Memo

Table 207 - code_ynn columns

Name	Data type
idx	Long Integer
zweit_malig	Memo

Table 208 - code_zweit_malig columns

3.2.1.8 Administration Tables

Administration tables are not exported

Name	Data type
idx	Long Integer
r_berat	Long Integer

Table 209 - v_adm_rolle columns

Name	Data type
idx	Long Integer
name	Memo
abfdat	DateTime

rolle	Long Integer
sqlstr	Memo

Table 210 - adm_abfrage columns

Name	Data type
idx	Long Integer
klin_name	Memo
klin_str	Memo
klin_plz1	Long Integer
klin_plz2	Long Integer
klin_ort	Memo
klin_art	Long Integer
klin_land	Long Integer
klin_abt	Long Integer
klin_abt_sonst	Memo
klin_tel	Memo
klin_fax	Memo
klin_pbox	Memo

Table 211 - adm_klinik columns

Name	Data type
idx	Long Integer
rolle	Memo
patadmin	Long Integer
patsiop9	Boolean
patsiop93	Boolean
patsiop2001	Boolean
patberat	Boolean
patrefrad	Boolean
patmolek	Boolean
patbest	Boolean
patpath	Boolean
patop	Boolean
patchemo	Boolean
recht_stud	Boolean
recht_berat	Boolean
recht_export	Boolean
recht_admin	Boolean
recht_wilms	Boolean
recht_ref	Boolean
recht_cpat	Boolean
recht_apat	Boolean
recht_delpat	Boolean
recht_freez	Boolean
recht_siopnr	Boolean
recht_sts	Boolean
recht_code	Boolean
recht_bugfix	Boolean
recht_meld	Boolean
meldung	Boolean
f1	Boolean
f2	Boolean
f3a	Boolean
f3b	Boolean
f3k	Boolean
f4	Boolean
f5	Boolean
f6	Boolean
f7	Boolean
f8a	Boolean
f8b	Boolean
f8c	Boolean
f9	Boolean
f13	Boolean
patup	Boolean
frag	Boolean
bild	Boolean
diag	Boolean
tube	Boolean
meta	Boolean
tele	Boolean
dicom	Boolean

Table 212 - adm_rolle columns

Name	Data type
idx	Long Integer
user_dat	DateTime
user_einv	Long Integer
user_einvdat	DateTime
user_lock	Long Integer
user_cxlogin	Memo
user_name	Memo
user_vorname	Memo
user_rolle	Long Integer
user_sex	Long Integer
user_pas	Memo
user_klinik	Long Integer
user_abteilung	Long Integer
user_abteilung_sonst	Memo
user_beruf	Long Integer
user_beruf_sonst	Memo
user_title	Memo
user_tel1	Memo
user_tel2	Memo
user_fax	Memo
user_email	Memo
user_komm_mit	Long Integer
user_protokol	Long Integer

Table 213 - adm_user columns

Name	Data type
idx	Long Integer
pnr	Long Integer
beg_user	Long Integer
pfid	Memo
beg_dat	DateTime
beg	Long Integer
beg_rueck	Long Integer
pro_beg	Memo

Table 214 - begruendung columns

3.2.1.9 Control Tables

Control tables are not exported.

Name	Data type
idx	Long Integer
formid	Text
did	Long Integer
wid	Long Integer
act	Boolean
vcr	Boolean
dox	Boolean
gewicht	Boolean
tv	Boolean

Table 215 - ctl_chemo2 columns

Name	Data type
idx	Long Integer
formid	Text
did	Long Integer
wid	Long Integer
act	Boolean
vcr	Boolean
dox	Boolean
carbo	Boolean
vp16	Boolean
cpm	Boolean
cyclo	Boolean
gewicht	Boolean

Table 216 - ctl_chemo7 columns

Name	Data type
idx	Long Integer
valuename	Text
valueid	Long Integer
org_value	Text
exp_value	Text

Table 217 - ctl_convert_value columns

Name	Data type
idx	Long Integer
table_nr	Long Integer
att_name	Text

Table 218 - ctl_export columns

Name	Data type
idx	Long Integer
table_nr	Long Integer
att_name	Text

Table 219 - ctl_export_n columns

Name	Data type
idx	Long Integer
table_name	Text
att_count	Long Integer
table_sql	Memo

Table 220 - ctl_export_table columns

Name	Data type
idx	Long Integer
formid	Text
week	Long Integer
vf1	Boolean
tf1	Memo

Table 221 - ctl_f2 columns

Name	Data type
idx	Long Integer
formid	Text
week	Long Integer
vf1	Boolean
vf2	Boolean
vf3	Boolean
vf4	Boolean
vf5	Boolean
vf6	Boolean
vf7	Boolean
vf8	Boolean
vf9	Boolean

Table 222 - ctl_f7 columns

Name	Data type
idx	Long Integer
fidx	Long Integer
fieldname	Memo
fieldobj	Memo

Table 223 - ctl_fields columns

Name	Data type
idx	Long Integer
name	Memo

Table 224 - ctl_forms columns

Name	Data type
idx	Long Integer
pfid	Memo
meldung	Boolean
f1	Boolean
f2	Long Integer
f3a	Boolean
f3b	Boolean
f3k	Boolean
f4	Boolean
f5	Boolean
f6	Boolean
f7	Long Integer
f8a	Boolean
f8b	Boolean
f8c	Boolean
f9	Boolean
f13	Boolean
protokol	Long Integer

Table 225 - ctl_pfid columns

Name	Data type
idx	Long Integer
table_name	Memo
form_name	Memo

tablepk	Memo
reg_nr	Long Integer

Table 226 - ctl_tables columns

obj_name	Memo
node_name	Memo

Table 227 - ctl_treeview_main columns

Name	Data type
idx	Long Integer
nodeidx	Long Integer
att_name	Memo

Name	Data type
idx	Long Integer
node_name	Memo

Table 228 - ctl_treeview_main_master columns

3.2.2 Data from the Hospital Information System (HIS) of the University Hospital of the Saarland

The usage of data from the HIS database of the University of the Saarland is still under negotiation with the Hospital administration and the data protection officer of the Hospital and the Ministry of Health of the State of Saarland.

After approval of data usage the data will be defined that are needed for specific scenarios of EURECA.

For the Microbiology scenario there are also data used coming from databases of the Department of Microbiology and from Pharmacy of the Saarland University. These data are combined with data of the HIS in a separate database that will not leave the Saarland University.

3.2.3 Data from the Cancer Registry of the Saarland

The usage of data from the Cancer Registry of the Saarland is still under negotiation with the data protection officer of the Ministry of Health of the State of Saarland.

After approval of data usage the data will be defined that are needed for specific scenarios of EURECA.

3.3 Maastro clinic

The data consists of patients treated with radiation therapy.

Data types, size of each data item, volume (size of dataset, i.e. number of items, number of patients), ranges:

Data item	Type
EMD	MySQL
PACS	MySQL+DICOM
euroCAT	XML+DICOM
ZyLAB	XML+*.txt+*.doc(x)+*.tiff

MAASTRO Clinic's Information and Services department handles all data requests.
Location of the data (now and in the future) / the systems are available inside MAASTRO IT network.

3.3.1 EMD (clinical)

EMD is the Electronic Hospital Record.

- Structure of the data: Structured and free text.
- Formats/ how the data will be exported/ standards

3.3.2 PACS (imaging)

The PACS is an imaging archive with DICOM objects.

- Structure of the data: Structured metadata, and images.
- Formats/ how the data will be exported/ standards

3.3.3 ARIA

ARIA is a record and verify database with treatments delivered to patients.

3.3.4 EuroCAT (clinical & imaging)

Euro-CAT is a de-identified data warehouse for research purposes containing EMD and PACS (and soon Aria) data with an ontology mapping module to NCI concepts. Using this module, local concepts can be mapped to NCI concepts.

- Structure of the data: Structured, free text and images.
- Formats/ how the data will be exported/ standards

3.3.5 ZyLAB (clinical – OCR scans)

Zylab is a database of documents linked to patients. It contains letters\forms\reports being sent to MAASTRO and scanned as TIFF and OCR text and it has letters in MS Word format that were sent by MAASTRO physicians. It contains XML files with metadata.

- Structure of the data: XML(metadata), *.tiff, *.txt and *.doc(x)
- Formats/ how the data will be exported/ standards

3.4 University of Oxford (UOXF)

Clinical information in Oxford University Hospitals (NHS OUH trust) Trust is gathered by accessing various clinical information systems and databases. These are mainly comprised of patient-generated clinical data, spanning across pathology case notes, imaging, patient history, surgery information, clinical notes; and more recently to these data genomic datasets have been added for combined analysis, such as DNA sequencing data.

A crucial step of the scenarios based on UOXF datasets is to gather clinical information from different sources andinvestigate the feasibility of an integration solution to connect the clinical systems. At the moment all data are collected in a semi-automated fashion from different systems for entry into the research databases (e.g. a Sarcoma database for the Sarcoma classifier scenario). The data are curetted by a data manager who performs logging-on on various systems and finds/validates the relevant patient information.

The aim of this project is to increase the automation in this clinical data gathering steps. It is under investigation whether to establish regular dumps of the clinical databases or live queries that an end user can request on the fly. We have carried out a survey in the NHS OUH to understand what are the systems in place and thus what are the requirements for such activity.The information gathered during this survey is presented below.

The NHS OUH datasets consists of both structured and unstructured data. For instance images from radiology systems and pathology reports will be mostly structured, although some free text is still required for pathology reports. Clinical notes are likely to be free text. However the structure of the received data will largely depend on the exporting functionality.

3.4.1 Data types and formats

The clinical systems under investigation have the potential to provide interfaces to export data in a standardised format. The HL7 messaging standard family has been widely adopted by vendors to provide an interface for extracting clinical data. DiCOM as an imaging standard enables lab information systems to export images. Direct database dump may be employed where interfaces are not available in the system. Table 1 gives some example of the data format from the clinical systems.

Data format	Description
HL7 version 2 message	ASCII, delimited
HL7 version 3+	Reference information model (RIM)
CDISC	XML encoded, Operational data model (ODM)
Proprietary format	XML or binary
Genomic data	text file

Table 229 - Selected data formats for collecting clinical data

3.4.2 Data Sources

Table 2 and Table 3 below summarize some of the data sources which will be considered to the Diagnostic classifier scenario in Eureca. In order to connect and query the data in these systems potential interfaces need to be established. Table 2 lists the databases that are being

populated by the clinical domain. Table 3 shows the systems clinicians interact with, and may provide interfaces for the semi-automated process
for data collection.
In the next sections we will describe some of these systems.

Databases	description
Pathology database, OUH trust	FileMaker database used by physicians
PACS radiology OUH trust	image database
Clinical notes	Oncology department database
Clinical notes OUH trust	NHS database for patient notes

Table 230 - Databases description

Clinical information systems	description
Pathology "CaseNote", OUH trust	NHS access to pathology case notes
Cerner EPR PowerChart	NHS access to patient history
Aria	Oncology department radiotherapy system

Table 231 - Commercial products and NHS systems to access clinical information

3.4.3 Electronic patient record

The Cerner Millennium EPR system is the main hospital information system that stores patient administrative information and general records that are associated with patient management. The data are associated with the care-flow of the patients; however a lot of this information has a limited use in oncology research. The Orbit system is a sub-system that connects with the main EPR system to manage patients in OUH. The Cerner Millennium EPR product has not been fully adopted to perform all its functionalities[10] and its adoption has been carried out in multiple step phases. The IT team within the NHS Trust is continuously working on extending this functionality andthis system will either link or in replace existing custom software and databases in different departments of the OUH trust.

3.4.4 PACS (imaging)

The NHS Picture Archiving and Communications Systems store images that are used in various clinical areas. Images can be viewed simultaneously at multiple locations within the trust. The PACS currently in use in the trust support the DiCOM standard for querying and retrieving image resources.

3.4.5 ARIA

The ARIA system is used for scheduling chemotherapy for patients in the OUH trust. This information system is a comprehensive information and image management solution for management of various aspects of oncology care. It is currently used in the context of radiotherapy and chemotherapy.

[10] Vender website: http://www.cerner.com/solutions/Hospitals_and_Health_Systems/

The information system has the potential to provide connectivity to other systems via HL7 interfaces. The manufacturer Varian[11] and the NHS IT team could potentially support the development of data access points.

3.4.6 Pathology databases

There are several Pathology databases, these are relational databases and are primarily made and interconnected using FileMaker Pro. These databases can be accessed by NHS staff via the "CaseNote" system, a web interface. There is on-going intensive development to enable the system to communicate with other systems for secondary research and clinical use.

3.4.7 Retrospective clinical studies databases

During the past decade, dedicated relational databases have been created in UOXF to support specific retrospective clinical, molecular pathology and genomic studies in cancer and other disciplines. The clinical and pathology data are typically generated from selected patients with specific consent, and they are imported into the databases in a mostly manually curate fashion. However automation of some tasks exist. These databases tend to be relational databases; in oncology they have been built mostly using FileMakerPro instances. Genomic raw data for these studies are often stored as flat files.

3.4.8 Biobank

The Oxford Musculoskeletal Biobank (OMB) stores the tissue and samples that can be retrieved and used by oncology researchers. Specific protocols for consenting the research and analysis that are associated with the Sarcoma scenario are being discussed and finalised.

[11] Vender website: http://www.varian.com/us/oncology/radiation_oncology/aria

3.5 German Breast Group Forschungs GmBH

3.5.1 Closed trial databases

High Level Description of Data:
This data comprises Baseline and Toxicity Data of selected GBG Studies that are now closed

Data Types:
Number of Patients in the study: 156
Number of Variables exported: 172
Size of Data (SPSS file): 52 Kb

Format of Variables: String /Numeric
Structure: structured derived data
Format: Data will be exported in spss or sas format
Person Responsible for data: Keyur Mehta (Statistician GBG Forshungs GmbH) Keyur.mehta@germanbreastgroup.de
Location of the Data: GBG Forschungs GmbH

Detailed Description of Data:

Name	Type	Width	Decimals	Label	Measurement
index	Numeric	8	2		Scale
meno_status	String	14	0	Menopausal status	Nominal
rad_surgical_type	String	17	0	Most radical surgery type	Nominal
height	Numeric	3	0	Height	Scale
weight	Numeric	6	2	Weight	Scale
KI	Numeric	3	0	Karnofsky index, %	Nominal
pT	Numeric	1	0	pT at 1st diagnosis	Nominal
pN	Numeric	1	0	pN at 1st diagnosis	Nominal
M	Numeric	1	0	M at 1st diagnosis	Nominal
Grade	Numeric	1	0	Grade	Nominal
ER_PgR	Numeric	1	0	Hormone receptor status	Nominal
Her2	Numeric	1	0	Her2	Nominal
locoregional_BL	Numeric	1	0	locoregional_BL	Nominal
liver_BL	Numeric	1	0	liver_BL	Nominal
lung_BL	Numeric	1	0	lung_BL	Nominal
bone_BL	Numeric	1	0	bone_BL	Nominal
CNS_BL	Numeric	1	0	CNS_BL	Nominal
other_BL	Numeric	1	0	other_BL	Nominal
haemoglobin_BL	Numeric	1	0	Haemoglobin, baseline	Nominal
leuko_BL	Numeric	1	0	Leucocytes, baseline	Nominal
neutro_BL	Numeric	1	0	Neutrophils, baseline	Nominal
thrombo_BL	Numeric	1	0	Thrombocytes, baseline	Nominal
AP_BL	Numeric	1	0	AP, baseline	Nominal

SGOT_BL	Numeric	1	0	SGOT, baseline	Nominal
SGPT_BL	Numeric	1	0	SGPT, baseline	Nominal
Bili_BL	Numeric	1	0	Bilirubin, baseline	Nominal
crea_BL	Numeric	1	0	Serum Creatinine, baseline	Nominal
ECG	Numeric	1	0	ECG	Nominal
echo	Numeric	1	0	Echocardiography	Nominal
LVEF	Numeric	2	0	LVEF	Scale
age	Numeric	2	0	age	Scale
chemo	Numeric	1	0	Any chemotherapy	Nominal
chemo_adj	Numeric	1	0	Chemotherapy adjuvant or neo-adjuvant	Nominal
chemo_met	Numeric	1	0	Chemotherapy palliative	Nominal
anthra	Numeric	1	0	Anthracycline-containing chemotherapy	Nominal
anthra_adj	Numeric	1	0	Anthracycline-containing chemotherapy adjuvant or neo-adjuvant	Nominal
anthra_met	Numeric	1	0	Anthracycline-containing chemotherapy palliative	Nominal
tax	Numeric	1	0	Taxane-containing chemotherapy	Nominal
tax_adj	Numeric	1	0	Taxane-containing chemotherapy adjuvant or neo-adjuvant	Nominal
tax_met	Numeric	1	0	Taxane-containing chemotherapy palliative	Nominal
endocrine	Numeric	1	0	Endocrine therapy	Nominal
endocrine_adj	Numeric	1	0	Endocrine therapy adjuvant	Nominal
endocrine_met	Numeric	1	0	Endocrine therapy palliative	Nominal
radio	Numeric	1	0	Radiotherapy	Nominal
radio_adj	Numeric	1	0	Radiotherapy adjuvant	Nominal
radio_met	Numeric	1	0	Radiotherapy palliative	Nominal
herceptin	Numeric	1	0	Trastuzumab	Nominal
herceptin_adj	Numeric	1	0	Trastuzumab adjuvant	Nominal
herceptin_met	Numeric	1	0	Trastuzumab palliative	Nominal
bisphosph	Numeric	1	0	Bisphosphonate treatment	Nominal
bisphosph_adj	Numeric	1	0	Bisphosphonate treatment adjuvant	Nominal
bisphosph_met	Numeric	1	0	Bisphosphonate treatment palliative	Nominal
other_ther	Numeric	1	0	Other treatments	Nominal
other_adj	Numeric	1	0	Other treatments adjuvant	Nominal
other_met	Numeric	1	0	Other treatments palliative	Nominal
duration	Numeric	3	0	Duration of trastuzumab	Scale
nausea	Numeric	1	0	Nausea	Nominal
nausea_34	Numeric	1	0	Nausea, grade 3-4	Nominal
vomiting	Numeric	1	0	Vomiting	Nominal
vomiting_34	Numeric	1	0	Vomiting, grade 3-4	Nominal
diarrhoea	Numeric	1	0	Diarrhoea	Nominal

diarrhoea_34	Numeric	1	0	Diarrhoea, grade 3-4	Nominal
mucositis	Numeric	1	0	Mucositis	Nominal
mucositis_34	Numeric	1	0	Mucositis, grade 3-4	Nominal
constipation	Numeric	1	0	Constipation	Nominal
constipation_34	Numeric	1	0	Constipation, grade 3-4	Nominal
gastro	Numeric	1	0	Other gastrointestinal disorders	Nominal
gastro_34	Numeric	1	0	Other gastrointestinal disorders, grade 3-4	Nominal
anorexia	Numeric	1	0	Anorexia, loss of appetite	Nominal
anorexia_34	Numeric	1	0	Anorexia, loss of appetite, grade 3-4	Nominal
allergic	Numeric	1	0	Allergic reactions	Nominal
allergic_34	Numeric	1	0	Allergic reactions, grade 3-4	Nominal
oedema	Numeric	1	0	Oedema	Nominal
oedema_34	Numeric	1	0	Oedema, grade 3-4	Nominal
fatigue	Numeric	1	0	Asthenia (fatigue)	Nominal
fatigue_34	Numeric	1	0	Asthenia (fatigue), grade 3-4	Nominal
alopecia	Numeric	1	0	Alopecia	Nominal
skin	Numeric	1	0	Skin changes (including HFS)	Nominal
skin_34	Numeric	1	0	Skin changes (including HFS), grade 3-4	Nominal
handfoot	Numeric	1	0	Hand-foot-syndrome	Nominal
handfoot_34	Numeric	1	0	handfoot_34	Nominal
nail	Numeric	1	0	Nail changes	Nominal
nail_34	Numeric	1	0	Nail changes, grade 3	Nominal
sensory	Numeric	1	0	Sensory neuropathy	Nominal
sensory_34	Numeric	1	0	Sensory neuropathy, grade 3-4	Nominal
neurol_other	Numeric	1	0	Other neurological disorders	Nominal
neurol_other_34	Numeric	1	0	Other neurological disorders, grade 3-4	Nominal
pain	Numeric	1	0	Pain	Nominal
pain_34	Numeric	1	0	Pain, grade 3-4	Nominal
infection	Numeric	1	0	Infection (including pneumonia)	Nominal
infection_34	Numeric	1	0	Infection, grade 3-4	Nominal
fever	Numeric	1	0	Fever	Nominal
fever_34	Numeric	1	0	Fever, grade 3-4	Nominal
thromboembolic	Numeric	1	0	Thromboembolic events	Nominal
thromboembolic_34	Numeric	1	0	Thromboembolic events, grade 3-4	Nominal
dyspnea	Numeric	1	0	Dyspnoea	Nominal
dyspnea_34	Numeric	1	0	Dyspnoea, grade 3-4	Nominal
respiratory	Numeric	1	0	Other respiratory or pulmonary disorders	Nominal
respiratory_34	Numeric	1	0	Other respiratory or pulmonary disorders, grade 3-4	Nominal
cardiac	Numeric	1	0	Cardiac events	Nominal
cardiac_34	Numeric	1	0	Cardiac events, grade 3-4	Nominal
renal	Numeric	1	0	Renal and urinary disorders	Nominal

renal_34	Numeric	1	0	Renal and urinary disorders, grade 3-4	Nominal
eye	Numeric	1	0	Eye disorders	Nominal
eye_34	Numeric	1	0	Eye disorders, grade 3-4	Nominal
hot_flushes	Numeric	1	0	Hot flushes and sweating	Nominal
hot_flushes_34	Numeric	1	0	Hot flushes and sweating, grade 3	Nominal
muscle_skel	Numeric	1	0	Musculosceletal disorders	Nominal
muscle_skel_34	Numeric	1	0	Musculosceletal disorders, grade 3-4	Nominal
liver	Numeric	1	0	Hepatobiliary disorders	Nominal
liver_34	Numeric	1	0	Hepatobiliary disorders, grade 3-4	Nominal
other	Numeric	1	0	other	Nominal
other_34	Numeric	1	0	Other, grade 3-4	Nominal
FN	Numeric	1	0	Febrile neutropenia	Nominal
anaemia	Numeric	1	0	Anaemia, any grade	Nominal
anaemia_34	Numeric	1	0	Anaemia, grtade 3-4	Nominal
leuko	Numeric	1	0	Leukopenia, any grade	Nominal
leuko_34	Numeric	1	0	Leukopenia, grade 3-4	Nominal
neutro	Numeric	1	0	Neutropenia, any grade	Nominal
neutro_34	Numeric	1	0	Neutropenia, grade 3-4	Nominal
thrombo	Numeric	1	0	Thrombopenia, any grade	Nominal
thrombo_34	Numeric	1	0	Thrombopenia, grade 3-4	Nominal
AP	Numeric	1	0	Alkiline phosphatase, any grade	Nominal
AP_34	Numeric	1	0	Alkiline phosphatase, grade 3-4	Nominal
ASAT	Numeric	1	0	ASAT, any grade	Nominal
ASAT_34	Numeric	1	0	ASAT, grade 3-4	Nominal
ALAT	Numeric	1	0	ALAT, any grade	Nominal
ALAT_34	Numeric	1	0	ALAT, grade 3-4	Nominal
bili	Numeric	1	0	Bilirubin, any grade	Nominal
bili_34	Numeric	1	0	bilirubin, grade 3-4	Nominal
crea	Numeric	1	0	Serum creatinine, any grade	Nominal
crea_34	Numeric	1	0	Serum creatinine, grade 3-4	Nominal
nausea_12	Numeric	1	0	nausea_12	Nominal
vomiting_12	Numeric	1	0	vomiting_12	Nominal
diarrhoea_12	Numeric	1	0	diarrhoea_12	Nominal
mucositis_12	Numeric	1	0	mucositis_12	Nominal
constipation_12	Numeric	1	0	constipation_12	Nominal
gastro_12	Numeric	1	0	gastro_12	Nominal
anorexia_12	Numeric	1	0	anorexia_12	Nominal
allergic_12	Numeric	1	0	allergic_12	Nominal
oedema_12	Numeric	1	0	oedema_12	Nominal
fatigue_12	Numeric	1	0	fatigue_12	Nominal
skin_12	Numeric	1	0	skin_12	Nominal
handfoot_12	Numeric	1	0	handfoot_12	Nominal
nail_12	Numeric	1	0	nail_12	Nominal
sensory_12	Numeric	1	0	sensory_12	Nominal

neurol_other_12	Numeric	1	0	neurol_other_12	Nominal
pain_12	Numeric	1	0	pain_12	Nominal
infection_12	Numeric	1	0	infection_12	Nominal
fever_12	Numeric	1	0	fever_12	Nominal
thromboembolic_12	Numeric	1	0	thromboembolic_12	Nominal
dyspnea_12	Numeric	1	0	dyspnea_12	Nominal
respiratory_12	Numeric	1	0	respiratory_12	Nominal
cardiac_12	Numeric	1	0	cardiac_12	Nominal
renal_12	Numeric	1	0	renal_12	Nominal
eye_12	Numeric	1	0	eye_12	Nominal
hot_flushes_12	Numeric	1	0	hot_flushes_12	Nominal
muscle_skel_12	Numeric	1	0	muscle_skel_12	Nominal
liver_12	Numeric	1	0	liver_12	Nominal
other_12	Numeric	1	0	other_12	Nominal
anaemia_12	Numeric	1	0	anaemia_12	Nominal
leuko_12	Numeric	1	0	leuko_12	Nominal
neutro_12	Numeric	1	0	neutro_12	Nominal
thrombo_12	Numeric	1	0	thrombo_12	Nominal
any_tox	Numeric	1	0	Any toxicity	Nominal
any_tox_34	Numeric	1	0	Any toxicity, grade 3-4	Nominal
any_haemtox	Numeric	1	0	Any haematological toxicity	Nominal
any_haemtox_34	Numeric	1	0	Any haematological toxicity, grade 3-4	Nominal
any_biochem	Numeric	1	0	Any biochemistry toxicity	Nominal
any_biochem_34	Numeric	1	0	Any biochemistry toxicity, grade 3-4	Nominal
any_other	Numeric	1	0	Any other toxicity	Nominal
any_other_34	Numeric	1	0	Any other toxicity, grade 3-4	Nominal

4 Legal requirements on the available test data

4.1 Data Sharing

Sharing of highly sensitive patient data is a great challenge. Made possible by the technical revelation in the information technology era, risks to patient privacy rights due to intensive processing of vast amounts of personal data constitute a major concern to sharing of patient data. In order to preserve patient's rights by, at the same time, putting as little constraint on medical research as possible, LUH has developed a legal framework that balances out patients and researchers interests and rights and makes data sharing in a safe environment possible.[12] Within the EURECA project, this legal framework is adopted in the project building phase to allow efficient data sharing between project partners to advance the technical EURECA infrastructure build up.

4.1.1 Legal Data Sharing Framework in the Project Building Phase[13]

The legal framework rests upon contracts that all partners of the project dealing with processing of patient data have to sign. These contracts split up into the 'Data Transfer Agreement' which relates to the export of data to the EURECA project in the project building phase and which is signed by the data exporting hospitals and the Center for Data Protection (CDP), the central data controller in EURECA, and the 'Contract on Data Protection and Data Security within EURECA', which relates to the processing of project data by project end users and which is signed by all project partners who process data within the project building phase and the CDP. The Data Transfer Agreement obliges the data exporting hospitals to transfer to the project only data which has been collected, stored and transferred in a manner which is compliant to the Data Protection Directive[14] and the relevant data protection requirements of the member states. Furthermore, the data exporters oblige themselves to only transfer data to the project that has been stripped off all direct identifiers using the CUSTODIX 'anonymisation' tool CATS or any other 'state of the art' computing tool guaranteeing an equivalent high standard of data coding. After all direct identifiers have been removed from sensitive patient data the remaining medical information is stored in a project database which can be located at hospital site.[15] Data coding is personal data processing according to Art 2 (b) of the Data Protection Directive[16]. This may only be carried out on the grounds of informed consent given by the patient, Art 8 (2a) of the Directive[17], or on the grounds of national member

[12] For detailed information on the Legal Framework see: Forgó, Kollek et al., Ethical and Legal Requirements for Transnational Genetic Research, 2010.
[13] For further information on the Legal Framework see Deliverable 2.2 as well as Deliverable 7.1 (due in 31.12.2012).
[14] Data Protection Directive 95/46/EC.
[15] Forgó et al, ibid, p.103.
[16] Art 2 (b) of the Directive reads: 'processing of personal data' shall mean any operation or set of operations which is performed upon personal data, whether or not by automatic means, such as collection, recording, organization, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, blocking, erasure or distruction.
[17] Art 8 (2a) of the Directive reads: Paragraph 1 (Member states shall prohibit the processing of (...) data concerning health (...)) shall not apply where the data subject has given his explicit consent to the processing of those data (...).

state law allowing data processing for scientific research purposes, Art 8 (4)[18] and Recital 34[19] of the Directive. The data exporting hospitals will therefore be supplied with project specific consent forms which each patient participating in the project shall sign. The encoded data, called EURECA project data, will be pushed to the EURECA project infrastructure, meaning a central data warehouse[20] or any other clinical research institution, called project end user.
The encoded medical information can only be regarded as legally anonymous data within a closed project user group who signed to the EURECA 'Contract on Data Protection and Data Security', as reidentification of patients from transferred sensitive (genomic) data is possible [2].[21] Within the 'Contract on Data Protection and Data Security' the project end users oblige themselves to handle all received data in compliance with applicable data protection law. Furthermore they oblige themselves not to attempt to reidentify any patient from project data. This obligation is backed by a penalty clause. The penalty clause kicks in if an end user attempts to identify a patient from EURECA project data. This economic sting will keep project end users from attempting to breach data privacy rights of participating patients. The Art 29 Working Party figured that the integration of an economic sting is an effective mechanism to ensure patient privacy rights.[22] The contract penalty has to be paid to the EURECA consortium. As a central data warehouse is installed within the EURECA project for building purposes, the CDP takes over full responsibility within the contractual framework to ensure privacy and security of data stored in the central data warehouse as data controller. Furthermore the CDP is also under the obligation not to attempt to reidentify any patient from project data. This obligation is similarly backed by a penalty clause. Within the EURECA project infrastructure, data is therefore legally anonymous to all participating end users and the CDP. As data can be regarded as being anonymous, the legal restrictions of the Data Protection Directive 95/46/EC do not apply. Data can easily and efficiently be shared for building purposes.

Within the legal framework, each entity is responsible for complying with the relevant data protection regulations and is data controller in accordance with Art 2 (d) of the Directive[23]. This means that the data exporting hospitals have to supply compliantly collected, stored and transferred data to the data warehouse or any other end user. The CDP is in charge of data stored in the warehouse. The end users have to compliantly

[18] Art 8 (4) of the Directive reads: Subject to the provision of suitable safeguards, Member States may, for reasons of substantial public interest, lay down exemptions in addition to those laid down in paragraph (2) either by national law or by decision of the supervisory authority.
[19] Recital 34 of the Directive reads: Recital 34 of the Data Protection Directive reads: Whereas member states must also be authorized, when justified by grounds of important public interest, to derogate from the prohibition on processing sensitive categories of data where important reasons of public inertest so justify in areas such as public health and social protection – especially in order to ensure the quality and cost-effectiveness of the procedures used to settling claims for benefits and services in the health insurance system – scientific research and government statistics; whereas it is incumbent on them, however, to provide specific and suitable safeguards so as to protect the fundamental rights and the privacy of individuals.
[20] Located at the premises of UPM.
[21] See: Forgó, Kollek et al., Ethical and Legal Requirements for Transnational Genetic Research, 102ff.
[22] WP Opinion 05/2012, p.11.
[23] Art 2 (d) of the Directive reads: 'Controller' shall mean the natural or legal person, public authority, agency or any other body which alone or jointly with others determines the purposes and means of processing of personal data (…).

process data which is taken out of the warehouse or transferred from any data exporter to their institution. Towards this project data they are data controller.

4.1.2 CDP as central Data Controller

Within the outlined data sharing framework, the CDP[24] acts as a central data controller. By implementing a central data controller the responsibility for sensitive medical data processing is clearly allocated.[25] The CDP is a central contact point for all patients participating in the EURECA project for privacy concerns. It has to ensure that medical information is processed fairly and lawfully, Art 6 (1a) of the Directive, in particular that it is only collected for specified, explicit and legitimate purposes and not further processed in a way incompatible with those purposes, Art 6 (1b) of the Directive. The CDP has to implement appropriate technical and organisational measures to protect sensitive medical data against accidental or unlawful destruction or loss, alteration, unauthorised disclosure and access.[26] It is therefore of utmost importance that the CDP functions as an independent body designated to protect patients' privacy rights.[27] Several precautions have been taken to ensure independence of the CDP. The CDP is chaired by two natural persons, Professor Nikolaus Forgó as president, and Brecht Claerhout as vice-president. It functions as a non-profit organisation established under Belgian Law. Its actions are non-commercially driven. The CDP and its founding members do not obtain any monetary compensation, neither from the project consortium or any third institution, for the activities undertaken as CDP in the project. No party constituting the CDP has any personal interest in linking anonymised project information to individual patients participating in the project. The CDP is furthermore integrated into the contractual data protection framework. The CDP can be held liable by any patient participating in the EURECA project or any project partner for each breach of patient privacy (data disclosure). To avoid any privacy breach is therefore a financial necessity to the members of the CDP. The concerns of colliding interests within the CDP are therefore eliminated by the contractual framework.[28] Lastly, the parties behind the CDP will not act as Trusted Third Party within the EURECA data sharing framework. This is owed to the fact that all direct identifiers will be erased from sensitive patient information before it is shared for infrastructure building purposes.

4.1.3 Summary

The legal framework allows sharing of anonymous data in the project building phase between partners who signed the relevant data protection contracts supplied by LUH for project building purposes (eg. building of tools etc). Data can be shared easily and efficiently between partners, whilst patient privacy rights are preserved. It is unnecessary to conclude specific bilateral data transfer agreements for each transfer of data. This clearly decreases administrative efforts. The integration of a CDP in the legal project framework, with whom data protection contracts are concluded by all project partners, increases the reciprocal trust of partners that data protection requirements are met at different processing stages in the project. The partners can rely upon not being held liable for any breach of patient privacy that was not responsibly caused by them. This clearly decreases the risk that data sharing constitutes for individual partners participating in the EURECA project.

[24] Centre for Data Protection Protection (CDP), http://www.privacypeople.org/
[25] For this requirement see: WP Opinion 01/2010, p.16.
[26] Forgó, Kollek et al., Ethical and Legal Requirements for Transnational Genetic Research, p.106f.
[27] EURECA DOW, p.150.
[28] For further information please see D7.1.

4.2 Internal process by clinical partners

4.2.1 Lawful processing of personal data at the Institut Jules Bordet

Before medical data are transferred to the EURECA data warehouse and therefore be embedded in the EURECA Data Protection Framework, in order to process such medical data from patients of the *Institut Jules Bordet*, some legal and internal requirements are to be fulfilled.

4.2.1.1 Requirements of the Belgian Law on Data Protection (Law of 8 December 1992):

1. The processing of personal data must be notified to the Belgian Data Protection Authority (article 17§1). Such publication is then publicly accessible on the Belgian Data Protection Authority website (for the 'pilot' phase of the EURECA project (transfer of data from IJB to Xerox), the notification has been published on July 5, 2012). This notification is not an authorization. The Belgian Data Protection Authority does not verify whether the processing of personal data is in conformity with the Belgian law.

2. In principle, the processing of medical data is prohibited (article 7§1), except if such processing satisfies one of several conditions (article 7§2) as foreseen by the Belgian Law.

    - One exception is the consent of the data subject/patient (article 7§2, a). For living patients, the Institut Jules Bordet has developed a broad informed consent (the « Patient Information and consent to use biological and medical data for scientific research ») which has been given positive opinion of the Institut Jules Bordet Ethics Committee on July 7, 2011. This consent is explicit but not specific. However, even if not specific, the processing is compatible with the finality of an integrated mono-specialized cancer public institution where a patient can reasonably foresee that such processing could occur. The Belgian law remaining silent about the processing of personal data from deceased patients, this issue will be addressed in section 2.

    - Another exception to this prohibition to process medical data is scientific research (article 7§2, k). If the processing is necessary to scientific research, data may be processed (upon conditions). Scientific research not being defined by the Belgian Law, the Belgian Data Protection Authority has interpreted it in a broad way as any research based on an objective method (objective observations and measurements and statistical analysis) that has a scientific purpose. The EURECA project may satisfies this condition as well.

4.2.1.2 Requirements of the Institut Jules Bordet:

In addition to these legal requirements, the Institut Jules Bordet has defined internal policies that researchers must comply with. These are not legally binding.

1. Each employee of the Institut Jules Bordet processing personal data from patients shall be bound by an obligation of confidentiality.

2. Each processing of personal data must be assessed and given a positive opinion of the Institut Jules Bordet Ethics Committee. In its review process, the Ethics Committee shall take into consideration the interests at stake balancing the scientific and ethical aspects of the project. For the 'pilot' phase of the EURECA project (transfer of data from IJB to Xerox), the Ethics Committee has been given its positive opinion on March 23, 2012 on the processing itself, the information to the living patients and the processing of personal data from deceased patients.

3. The researchers shall process coded/pseudomymised personal data and shall take all the necessary and appropriate IT and security measures to protect such personal data. For the pilot phase, the personal data will be pseudonymised trough an encrypting algorithm and the data transfer secured trough a VPN encrypting.

4. Each transfer of personal data shall be covered by a Data Transfer Agreement between the provider and the recipient.

5 Conclusion

This report gives an overview of all potentially available test data sources, together with general legal requirements, which allow the implementation of all the tools and services of the pilot study for the firt phase of the project. These clinical data are aimed to be provided by clinical partners for the development of these tools. But, as sharing patient data is a task that requires a certain amount of legal and ethical requirements, this document presented the different legal steps that were necessary to complete these requirements.

The development and then the implementation of the scenarios and use cases within WP1 are made posible by the amount of information and data that were collected from all clinical sites. We will then continue to work on adding continuously additional dataset to the project to make progress within all work packages and achieve the project's purposes in term of development of tools that will mainly enable interoperability between research and care.

6 References

[1] C. Desmedt, A. Di Leo, E. de Azambuja, D. Larsimont, B. Haibe-Kains, J. Selleslags, S. Delaloge, C. Duhem, J.-P. Kains, B. Carly, M. Maerevoet, A. Vindevoghel, G. Rouas, F. Lallemand, V. Durbecq, F. Cardoso, R. Salgado, R. Rovere, G. Bontempi, S. Michiels, M. Buyse, J.-M. Nogaret, Y. Qi, F. Symmans, L. Pusztai, V. D'Hondt, M. Piccart-Gebhart, and C. Sotiriou, "Multifactorial approach to predicting resistance to anthracyclines," *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, vol. 29, no. 12, pp. 1578–86, Apr. 2011.

[2] P. Ohm, "Broken promises of privacy: responding to the surprising failure of anonymization," pp. 1–64, 2009.