Deliverable No. 2.2

Definition on scenarios and use cases and report on scenario based user needs and requirements

Grant Agreement No.:	270089
Deliverable No.:	D2.2
Deliverable Name:	Definition on scenarios and use cases and report on scenario based user needs and requirements
Contractual Submission Date:	30/09/2011
Actual Submission Date:	30/09/2011

Dissemination Level		
PU	Public	X
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

COVER AND CONTROL PAGE OF DOCUMENT	
Project Acronym:	***p-medicine***
Project Full Name:	From data sharing and integration via VPH models to personalized medicine
Deliverable No.:	D 2.2
Document name:	Definition on scenarios and use cases and report on scenario based user needs and requirements
Nature (R, P, D, O)[1]	R
Dissemination Level (PU, PP, RE, CO)[2]	PU
Version:	1.0
Actual Submission Date:	30/09/2011
Editor: Institution: E-Mail:	Norbert Graf USAAR graf@uks.eu

ABSTRACT:

This deliverable successfully identified, elaborated and specified the end-user needs and requirements for the p-medicine project's technological, methodological and clinical research infrastructures/frames. The initial focus on system requirements in the form of 'use cases/scenario' has been effectively extended by presenting detailed overviews of p-medicine platform architecture, IT components specifications and other research activities useful for further Work Packages.

KEYWORD LIST: user needs, requirements, use cases, scenarios

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n^{o} 270089.

[1] **R**=Report, **P**=Prototype, **D**=Demonstrator, **O**=Other

[2] **PU**=Public, **PP**=Restricted to other programme participants (including the Commission Services), **RE**=Restricted to a group specified by the consortium (including the Commission Services), **CO**=Confidential, only for members of the consortium (including the Commission Services)

MODIFICATION CONTROL			
Version	Date	Status	Author
0.1	03/04/2011	Draft	Norbert Graf
0.2	26/04/2011	Draft	Norbert Graf
0.3	15/05/2011	Draft	Norbert Graf
0.4	06/06/2011	Draft	Norbert Graf
0.5	04/07/2011	Draft	Ruslan David
0.6	13/08/2011	Draft	Norbert Graf
0.7	17/08/2011	Draft	Norbert Graf
0.8	08/09/2011	Draft	Norbert Graf
0.9	13/09/2011	Draft	Norbert Graf
1.0	25/09/2011	Final	Norbert Graf

List of contributors

- Norbert Graf, USAAR
- Holger Stenzhorn, USAAR
- Ruslan David, USAAR
- Stefan Kiefer, FhG IBMT
- Gabriele Weiler, FhG IBMT
- Marie-Luise Christ-Neumann, FhG IAIS
- Stefan Rüping, FhG IAIS
- Simona Rossi, SIB
- Alberto Anguita, UPM
- Marian Taylor, UOXF
- Danny Burke, Ecancer
- Manolis Tsiknakis. FORTH
- Lefteris Koumakis, FORTH
- Kristof de Schepper, Custodix
- Elias Neri, Custodix
- Christian Ohmann, UDUS
- Wolfgang Kuchinke, UDUS
- Georgios Stamatakos, ICCS
- Martin Stanulla, CAU
- Martin Zimmermann, CAU

Contents

1 EXECUTIVE SUMMARY 8
PERSPECTIVES ON USER NEEDS AND REQUIREMENTS 9
2 INTRODUCTION AND PROJECT BACKGROUND 10
2.1 Requirement analysis, scenarios, architecture and workflows *11*
2.1.1 Technological perspective on needs and requirements *12*
2.1.2 Scenarios and requirements analysis *12*
2.1.3 Architecture *13*
2.1.4 Applets *20*
2.1.5 Workflows *20*
2.2 Introduction to p-medicine Interoperability Specifications *21*
2.3 Data *21*
2.3.1 Availability of retrospective data *28*
2.3.2 Prospective data *40*
3 THE IDENTIFICATION OF USER NEEDS 42
INTRODUCTION 42
3.1 Process of collecting EUNs *43*
3.2 End User Needs *44*
3.2.1 'Healthcare providers' EUNs *44*
3.2.2 'Patients' EUNs *46*
3.2.3 'Researchers' EUNs *49*
3.2.4 'Clinical Research Organisations' EUNs *51*
3.2.5 EUNs and use cases / scenarios *52*
4 CONTEXT SCENARIOS FOR USABILITY TESTING 54
4.1 Overview of usability testing process *54*
4.2 Key questions for describing and structuring user performance in context *56*
4.3 Cancer patients' questionnaire *58*
4.3.1 Questions of the survey *59*
4.3.2 The resulting patients' needs *60*
4.3.3 Recommendations for the presentation of all relevant information *61*
4.4 Conclusion regarding ISO 9241 ..... *62*
5 VPH SCENARIOS 63
5.1 VPH Toolbox Scenario *63*
6 SECURITY SCENARIOS 65
INTRODUCTION 65
6.1 Single Sign-on Scenario *65*
6.2 Single Sign-out Scenario *65*
6.3 Access Rights Scenario *65*
6.4 User Enrolment Scenario *65*
7 CLINICAL SCENARIOS 66
INTRODUCTION 66
7.1 Nephroblastoma *66*
7.1.1 Pathway Scenario *68*
7.1.2 Imaging Scenario *68*
7.1.3 (Severe) Adverse Event ((S)AE) Prediction Scenario *68*
7.1.4 Tumour Marker Scenario *68*
7.1.5 Oncosimulator Scenario *68*
7.2 Breast Cancer *68*
7.2.1 Breast Cancer Scenarios *69*
7.2.2 Oncosimulator Scenario *70*
7.3 Acute Lymphoblastic Leukaemia *70*
7.3.1 Oncosimulator Scenario *72*
7.3.2 Biobank scenario *73*

8 PATIENT EMPOWERMENT SCENARIOS 74
INTRODUCTION 74
8.1 Search for running clinical trials in Europe *75*
8.2 Consent and Re-consent Scenario *75*
8.2.1 Informed Consent (Patient's Perspective) *76*
8.3 Own Data Scenario *76*
8.4 Access to Biobanks Scenario *76*
8.5 Summarize the history of the disease in an understandable way and increase patient-doctor understanding *76*
9 BIOBANKING SCENARIOS 77
INTRODUCTION 77
10 CLINICAL TRIALS 78
INTRODUCTION 78
10.1 IT support for clinical trials *81*
10.2 Analysis of clinical trial data and across clinical trials *82*
10.3 EUNs of different stakeholders in clinical trials *83*
11 OBTIMA SCENARIOS 85
INTRODUCTION 85
11.1.1 Pseudonymization Scenario *86*
11.1.2 Data Entry of Prospective Clinical Trial Dato *86*
11.1.3 Data Manager of Prospective Clinical Trials *86*
11.1.4 eCRF Developer for Prospective Clinical Trials *86*
11.1.5 Data Synchronization with HIS during running trial in ObTiMA *86*
11.1.6 SAE/SUSAR Scenario *86*
11.1.7 Drug interaction Scenario *87*
11.1.8 DICOM Scenario *88*
11.1.9 Consultation Scenario *88*
11.1.10 Trial Development Scenario *88*
11.1.11 Trial Outline Builder Scenarios *88*
11.1.12 Participating Centres Scenario *88*
11.1.13 Patient Access to his/her trial data and Diary Scenario *89*
11.1.14 Repository Scenario *89*
11.1.15 Semantic interoperability Scenario *89*
11.1.16 Reporting Scenario *89*
11.1.17 Sync and Push services (see 13.2) *89*
12 DOCTOREYE SCENARIOS 91
INTRODUCTION 91
12.1 Nephroblastoma Scenario *91*
12.1.1 Segmenting Nephroblastoma from MRI images *91*
12.2 Signal Intensity Scenario *93*
12.2.1 Introduction *93*
12.2.2 Motivation *93*
12.2.3 Data Description *94*
12.2.4 The Method *94*
12.2.5 Conclusion *98*
13 P-MEDICINE IT-COMPONENTS SCENARIOS 99
INTRODUCTION 99
13.1 Scenarios dealing Interoperability Specifications *99*
13.1.1 Summary Documents HL7 CCD *99*
13.1.2 Laboratory Reports and Messages *100*
13.1.3 Sharing Imaging Results *102*
13.1.4 Medical Knowledge Retrieval *103*
13.1.5 Transfer of Documents on Media *103*
13.1.6 Patient Demographics Query *104*
13.1.7 Manage Sharing of Documents *105*
13.1.8 Consult and History and Physical Note *105*
13.1.9 Patient ID Cross-Referencing *106*

13.1.10 Notification of Document Availability ... *107*
13.1.11 Clinical Research Interoperability Specification ... *108*
13.2 Push Scenario – Usage of clinical data from hospital information systems (see 11.1.5) ... *109*
13.3 Ontology-Based Semantic Search Framework ... *109*
13.4 P-medicine portal scenario ... *109*
14 CERTIFICATION OF TOOLS, SOFTWARE, SERVICES AND MODULES ... 111
INTRODUCTION ... 111
14.1 Good Clinical Practice (GCP) Compliance ... *111*
15 SCENARIOS FOR EDUCATION AND TRAINING ... 112
INTRODUCTION ... 112
15.1 Help patients understand the IEmS ... *112*
15.2 Teach health care professionals when to use the p-medicine tools ... *112*
15.3 Impart understanding of the p-medicine environment ... *112*
15.4 Scenario for Education and Training ... *112*
16 EVALUATION AND VALIDATION OF SCENARIOS ... 114
INTRODUCTION ... 114
16.1 Workflow Scenario: Anonymize local clinical databases and upload them in p-medicine environment ... *114*
16.1.1 Evaluation process ... *114*
16.1.2 Validation Process ... *115*
16.2 VPH Toolbox Scenario: statistical analysis of cancer samples with associated gene expression and clinical features ... *115*
16.2.1 Evaluation Process ... *115*
16.2.2 Validation process ... *117*
16.3 Obtima Scenario: eCRF developer for prospective clinical trials ... *117*
16.3.1 Evaluation Process ... *117*
16.3.2 Validation Process ... *118*
17 CONCLUSION ... 119
INTRODUCTION ... 119
17.1.1 Use Case Scenario Success ... *119*
17.1.2 End-User Needs Challenge ... *121*
17.1.3 Linkage to other Deliverables and Work Packages ... *121*
APPENDICES ... 122
Appendix 1 – Abbreviations and acronyms ... *122*
Appendix 2 – Context scenarios ... *126*
Context Scenario (Bioinformatician) ... *126*
Context Scenario (Data Manager) ... *136*
Context Scenario (Biostatistician) ... *149*
Context Scenario (Chairman and Clinician) taken from project ACGT ... *155*
Context Scenario (Senior Fellow Scientist in Clinic) ... *165*
Appendix 3 – Patients' answers of the questionnaire presented in tables ... *181*
Appendix 4 – Template for use cases/scenarios ... *198*
Appendix 5 – Use Cases ... *201*
VPH Scenarios ... *201*
VPH Toolbox Scenario ... *201*
Security Scenarios ... *204*
Single Sign-on Scenario ... *204*
Single Sign-out Scenario ... *210*
Access Rights Scenario ... *213*
User Enrolment Scenario ... *214*
Clinical Scenarios ... *218*
Pathway Scenario for Nephroblastoma ... *218*
Imaging Scenario ... *222*
(Severe) Adverse Event ((S)AE) Prediction Scenario ... *228*
Tumour Marker Scenario for nephroblastoma ... *233*
Oncosimulator Scenario for nephroblastoma ... *236*
Oncosimulator Scenario for Breast Cancer ... *242*

Oncosimulator Scenario for ALL *247*
Patient Empowerment Scenarios *252*
Search for running clinical trials in Europe *252*
Consent and Re-consent Scenario *255*
Own Data Scenario *258*
Access to Biobanks Scenario *261*
Summarize the history of the disease in an understandable way and increase patient-doctor understanding *264*
Biobanking Scenarios *268*
Linking the own biomaterial data repository to the p-medicine biobank access framework for the collaboration with specific user groups *268*
Managing patient's biomaterial and related data in clinical trials with ObTiMA *272*
Offering human biomaterial to a closed and/or open clinical research community for research *276*
Requesting specific human biomaterial within a closed and/or an open clinical research community for research purposes *279*
Clinical Trials Scenarios *282*
Statistical Analysis of cancer samples with associated gene expression data and clinical data *282*
Data management in international clinical trials by ECRIN *287*
Use of data mining to improve study feasibility *290*
Improved Patient Recruitment in oncological clinical trials *293*
Use of p-medicine platform with decision support to conduct oncological clinical trials by ECRIN *296*
Increased Subject Retention rates in oncological clinical trials *299*
ObTiMA Scenarios *302*
Pseudonymization Scenario *302*
Data Entry of Prospective Clinical Trial Data *306*
Data Manager of Prospective Clinical Trials *309*
eCRF Developer for Prospective Clinical Trials *313*
Data Synchronization with HIS during running trial in ObTiMA *317*
SAE/SUSAR Scenario *320*
Drug interaction Scenario *324*
DICOM Scenario *327*
Consultation Scenario *331*
Trial Development Scenario *334*
Trial Outline Builder Scenarios *340*
Participating Centres Scenario *346*
Patient Access to his/her trial data and Diary Scenario *349*
Repository Scenario *352*
Semantic interoperability Scenario *356*
Reporting Scenario *360*
p-medicine IT-Components Scenarios *363*
Push Scenario – Usage of clinical data from hospital information systems (see 11.1.5) *363*
Ontology-Based Semantic Search Framework *369*
p-medicine portal scenario *373*
Scenario for Education and Training *377*
Appendix 6 – Ethical Approval *380*

1 Executive Summary

This deliverable is rather large with 380 pages. For readability context scenarios, specific use cases and scenarios, and results from a questionnaire are found in Appendices. This allows gathering the main information in the first part of the deliverable (120 pages). Every reader, who wishes to go into detail, can do so, by reading the appendices. To make the context scenarios even better usable FhG IAIS will provide a web service allowing searching for specific information in the context scenarios.

The main objective of this deliverable is to elaborate the user needs and requirements for the proposed p-medicine's technological and clinical research infrastructure. By concluding that understanding the end users is the key to success, 'D2.2 Definition on scenarios and use cases and report on scenario based user needs and requirements' appears to be one of the most important research topics of the p-medicine project with a deep impact on the proposed activities and research strategies.

The p-medicine infrastructure aims to deliver a state of the art technological platform in order to facilitate the personalized health care and decision support. The project's technical solution is to embrace the current and future web and its technologies and build its architecture under the tenets of the Service Oriented design. The advantages/requirements, mentioned in the project description, are:

- **Ubiquitous availability** (“anywhere, anytime, any device”) enables mobility, easiness of use, and low cost access and use of the platform;
- **Collaboration** - social networking and other “Web2.0” features are inherent qualities of the proposed solution that make possible the building of virtual communities of users to promote interactivity, research, and education;
- **Software as a Service (SaaS)** - central registration and on demand availability of software tools to healthcare professionals and researchers enable the provision of software as a commodity while strengthening interoperability and standardisation of the shared code base.

The proposed/envisaged p-medicine technological platform features are:

- Access to the correlated repositories of experimental and research data from public sources, research projects
- Access to the correlated repositories of tools, services and models (VPH ToolKit)
- Advanced search and discovery capabilities
- Automated and secure upload of patient's data and correlation with publicly available data (PubMed repository)
- Extensive tissue, disease, and compound ontologies, standards and interoperability features to ensure advanced and accurate correlations
- An intuitive, web-based interface resulting in quick adoption by the End Users (Healthcare providers, patients, researchers, etc.)

- Collaboration capabilities streamlining communication, networking, information sharing and education
- State-of-the-art security and data protection
- SaaS (Software as a Service) delivery model requiring minimal IT support
- Demonstrate the usefulness and openness to the whole VPH community
- Demonstrate that the infrastructure mechanisms and services are compatible with VPH-Share

Perspectives on user needs and requirements

The p-medicine platform has different and complex user needs and requirements and in order to overcome the complexity of the proposed for implementation project's goals all user needs and requirements have been aligned according to three main 'user needs and requirements' pillars:

- Technological perspective;
- End users' perspective;
- Clinical/Medical perspective.

Technological perspective on user needs and requirements (described bellow) will have an important impact on p-medicine platform's requirements, nevertheless it should be able to accept with 'flexibility' the 'End users needs and requirements' perspective as well as the 'Clinical Medical' perspective.

This deliverable will focus exclusively on Perspectives of End Users but the linkage between Technological and Clinical/Medical Perspectives as well will be underlined and described.

2 Introduction and Project Background

It is the purpose of p-medicine to deliver an architecture that will allow to drive medicine to more individualized treatments based on exploiting the vast amount of heterogeneous data of single patients by software, services, tools and models that will support physicians in decision making in their daily care of patients. Today we are facing a paradigm shift in medicine going from hospital and clinical based care to a new standards approach, which is not yet completely defined. Comparing changes in other areas of daily life they can be described as consumer controlled compared to producer controlled in former times. Nowadays the producer needs to respond to the consumer (fig. 2.1). Translating this to healthcare patient empowerment cannot be neglected anymore and will influence healthcare in all dimensions.

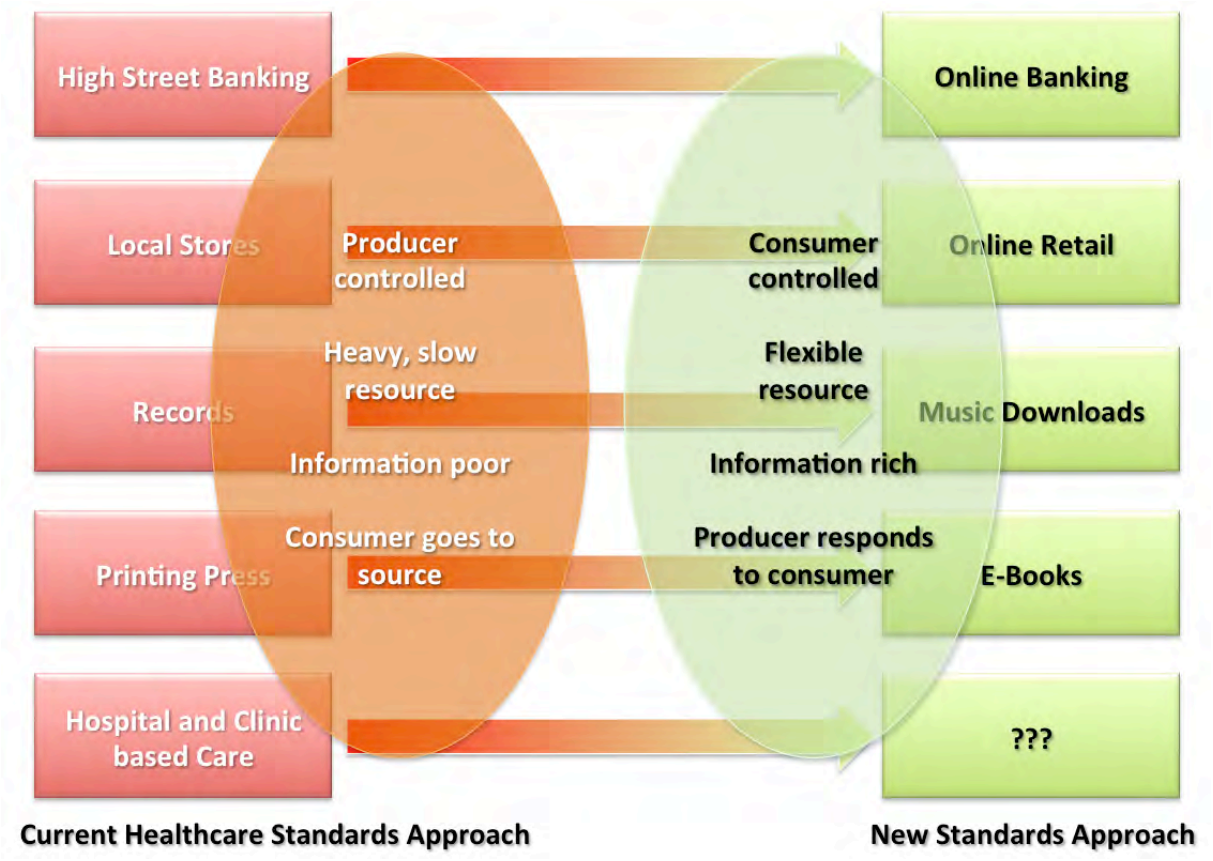


Figure 2.1: Paradigm shift in daily life including healthcare. (Adapted from Ken Lunn, CMLS Network Annual Symposium, London, 23rd June 2011)

In connection with the scientific/technical dimensions of the work *p-medicine* will develop a data warehouse and a workbench with a tools repository. Heterogeneous pseudonymized/anonymized data from different origins will be stored in a data warehouse for further use by the scientific community. Clinical data will be exploited coming from hospital information systems and clinical trials. The legal framework of the project, which is based on the results of ACGT (Advancing Clinico-genomic trials[3]), will be further developed and will guarantee data privacy and security. Most important for *p-medicine* are validated tools and services that provide interfaces to allow interoperability with biobanks, genetic databases,

[3] http://eu-acgt.org/

and medical imaging systems and data warehouses. These tools have to meet requirements to be used in large, international multicentre clinical GCP conform trials and need to be able to be integrated into existing systems used by ECRIN and other communities. This includes aspects like data security by adopting the legal and ethical framework based on international requirements and approved concepts for anonymization and pseudonymization including validation. Previous R&D work done in European funded projects like ACGT, ContraCancrum and ECRIN (European Clinical Research Infrastructures Network) fit perfectly into this approach and will be heavily drawn on. The following figure (fig. 2.2) shows the main components and their interdependency of the *p-medicine* system architecture from a clinical perspective.

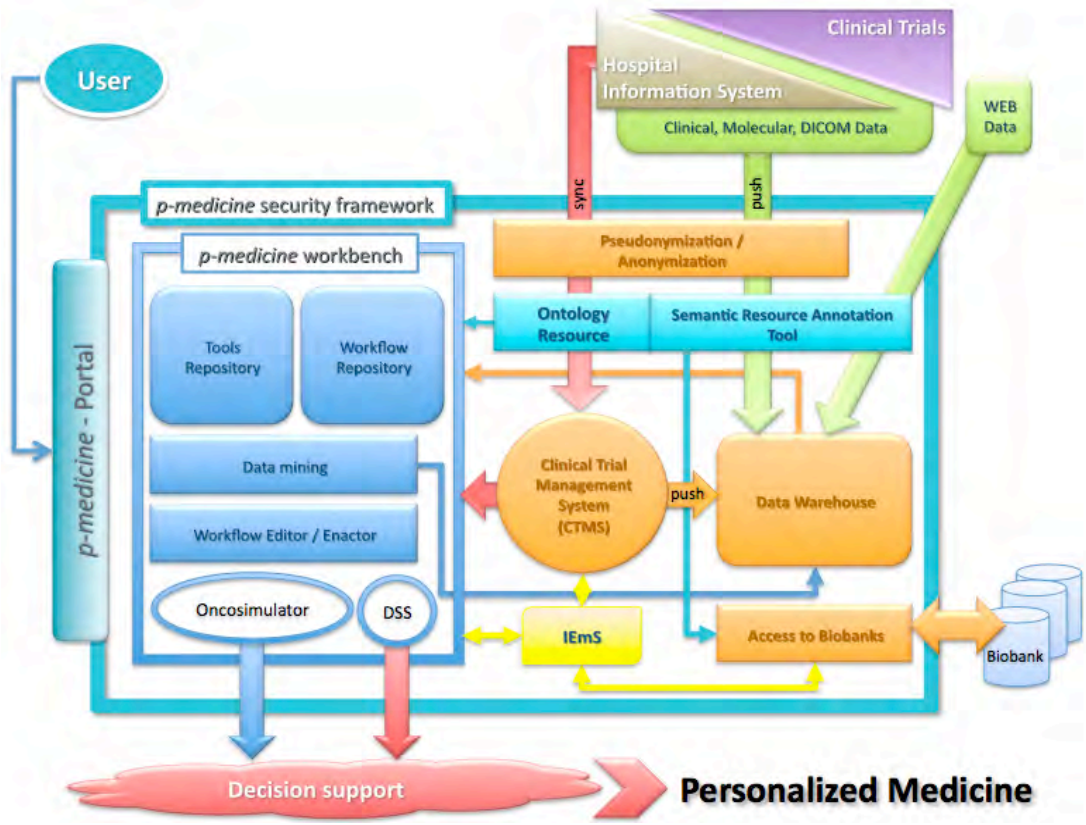


Figure 2.2: The architecture of p-medicine from a clinical perspective

2.1 Requirement analysis, scenarios, architecture and workflows

A user will be able to get access to *p-medicine* via a secure portal to use tools and workflows from the *p-medicine* workbench to execute his models by mining data from the data warehouse. Data from Hospital Information Systems (HIS) or the integrated Clinical Trial Management Systems (CTMS) via a push service feed the data warehouse. The CTMS can synchronize with the HIS using a sync service. Data entering the *p-medicine* environment will be pseudonymized/ anonymized and semantically annotated. Access to external biobanks will be established and freely available data from the web can be stored in the data warehouse. Depending on the scenario users are able to execute models with the *p-medicine* Oncosimulator or they can use the Decision Support System (DSS). In both cases results will lead to personalized medicine via decision support. Patients as users of *p-medicine* can interact with the *p-medicine* environment via the Interactive Empowerment Service (IEmS) that will be developed in the project's lifetime. As the Oncosimulator is a main component it is described in more detail in upcoming deliverables of WP12.

2.1.1 Technological perspective on needs and requirements

The technological perspective on needs and requirements of the p-medicine platform ideates the core and state-of-the-art elements of the software development process. Additionally, the 'waterfall model', which represents a sequential design process, used in software development processes fits perfectly to the proposed project's objectives. In classical 'waterfall model' the software development progress is seen as flowing steadily downwards (like a waterfall) through the phases of:

- Requirements
- Design
- Implementation
- Verification
- Maintenance.

A challenging task of p-medicine platform is defining the requirements and requirements analysis. In order to assure the development of a functional and state-of-the art system the bellow main requirements are proposed for implementation:

- Software as a Service (SaaS)
- Interoperability
- Flexibility
- Modularity
- Security and granular access for end users
- Social networking frames

2.1.2 Scenarios and requirements analysis

Conceptually, requirements analysis includes three types of activity:

- **Requirements gathering**: the task of communicating with users to determine what their requirements are.
- **Analysing requirements**: determining whether the stated requirements are complete, implementable, ambiguous, or contradictory, and then resolving these issues.
- **Recording requirements**: Requirements might be documented in various forms, such as natural-language documents, use cases/scenario, user stories, or process specifications.

Requirements analysis in the frames of p-medicine project will be a continuous process due the technological platform complexity and its modular infrastructure. Nevertheless, the main focus will be on recording requirements in the form of 'use cases/scenario' and if it would be applicable in the form of 'process specifications'.

Project contributors will employ several techniques to elicit the requirements and user needs. In general, this will include such activities as holding interviews, or holding focus groups (requirements workshops, meetings) and creating requirements lists in the form of '**use cases/scenario**' as well as continuous scientific literature reviews. A template for **use cases/scenarios** is provided for all kind of users (Appendix 4). At later stages and according to the elaborated '**use cases/scenario**' project activities will focus on prototyping.

2.1.3 Architecture

The architecture of p-medicine, developed in WP3, is primarily a modular based one following the well established paradigm of Service Oriented Architecture (SOA) providing functionalities in self-contained service modules with clearly defined and delineated functionality, interoperability and interface descriptions. For end-users, e.g. clinicians and others, the architecture is hidden behind a GUI. This GUI needs to be intuitive and user-friendly. The end-user will enter via a portal after initial registration, where his roles and rights are fixed. According to his/her roles and rights the GUI will only display those functionalities the end-user is able to work with. In principle there are several layers of the architecture of which a clinician as an end-user will not get aware. He will only be able to see the front end via the GUI. The access to deeper layers is regulated by the roles and rights management system (fig. 2.3).

Fig. 2.3: Different layers of the architecture.

Most important for the architecture of p-medicine is the mentioned modularity of the system. All developed software, tools and services should be as granular and modular as possible and provide standardized, open interfaces and functionality descriptions (e.g. via something similar to WSDL (Web Services Description Language)), so that a user can easily build new models as a composition of existing granular tools (fig. 2.4). As an example one needs only once to develop a tool that will link gene expression data of a tumour with the KEGG database. If this tool is as generic as possible and if the interface between the gene expression data and the KEGG database is standardized one will be able to use this tool in different settings and models, independent of the underlying tumour or disease. Such an approach needs to standardize interfaces between different tools and tools and data. A description of such specifications needs to be done, that allows different research groups how to standardize their data and what are preconditions to run such composed models. It is important that for each granular tool a standardized interface to data needs to be defined. The developed models will then be able to be used in scenarios (fig. 2.5).

Model a)

Model b)

Fig. 2.4: The composition of two different modules [a) and b)] out of granular tools A, B, C, D and E]. The different tools are connected with different data.

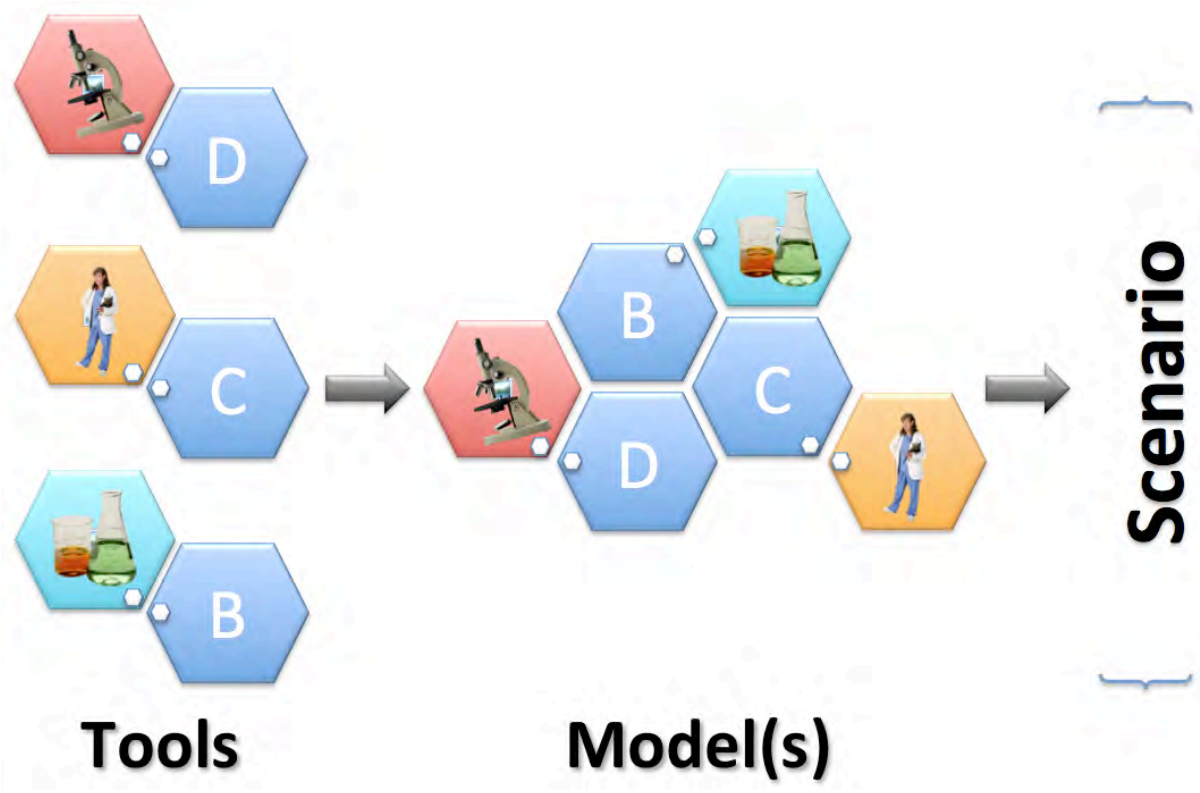


Fig. 2.5: From Tools to models to scenarios

The development of the architecture for p-medicine needs to take into account, how tools and models will be developed within in the project. A tool will process input data to produce a result. Such result or output data might be input data for another tool (fig. 2.6a and b).

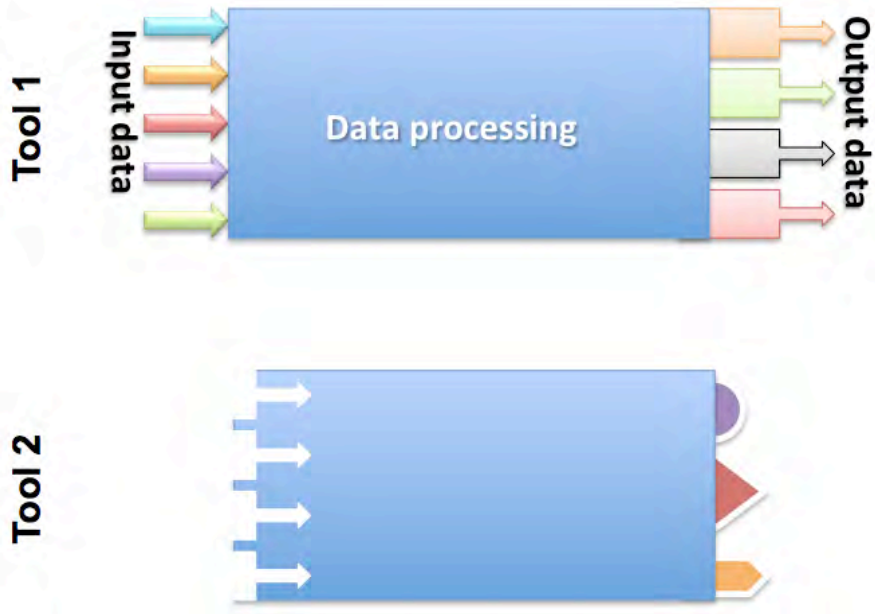


Fig. 2.6a: Tools and input and output data

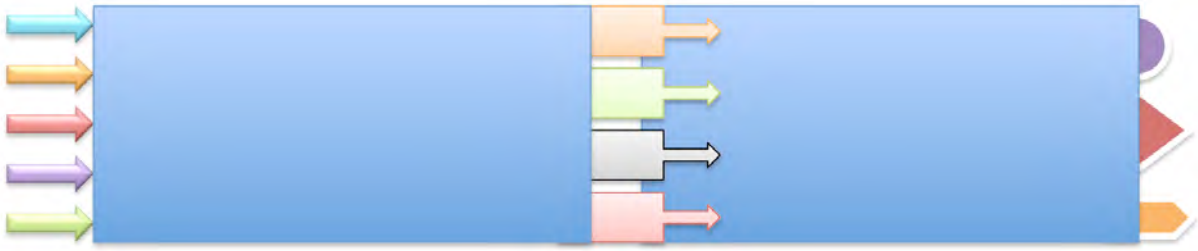


Fig. 2.6b: Tool 1 and Tool 2 can be combined for a new tool

To combine tools for making more complex tools it is useful to categorize tools into four different levels:

1. Level for fundamental tools
2. Level for basic tools
3. Level for modular tools
4. Level for domain specific tools, models, services

Fundamental tools are such tools that are fundamental for the architecture. This level includes mainly IT-tools that can be used in all models. The basic level will contain only such tools that are domain and scenario unspecific, e.g. pseudonymization tool, curation tool for data, etc. Modular tools are scenario specific but not domain specific, e.g. a tool for patient empowerment, etc.. At a higher level tools, models services are domain specific. Fundamental and basic tools can be re-used in different scenarios and domains. For this purpose interoperability and standardization is of utmost importance to avoid building each tool again and again from scratch. Even an interface tool as a basic tool can be developed that will handle issues regarding interoperability of interfaces of tools and data. Such a basic interface tool (fig. 2.7) can also be used as a tool for data import, if it is developed in a very generic way to also handle databases.

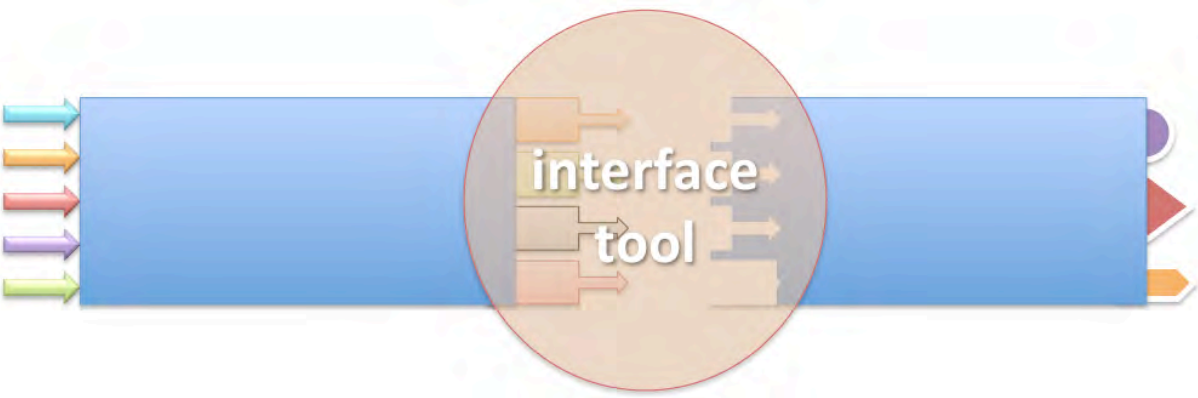


Fig. 2.7: Interface tool to handle interoperability between tools or databases.

According to the classification of tools scenarios will be classified in the following levels (fig. 2.8):

1. Domain specific scenarios
2. Domain unspecific scenarios
3. Basic scenarios
4. Technical scenarios

Technical scenarios are part of other scenarios including basic scenarios. Domain unspecific scenarios can be composed of basic scenarios and will be able to be used in domain specific scenarios. By doing so, scenarios do not need to be developed from scratch and it will foster the development of standards and interoperability. If interoperability and standards are developed the integration of external scenarios into the p-medicine framework is possible.

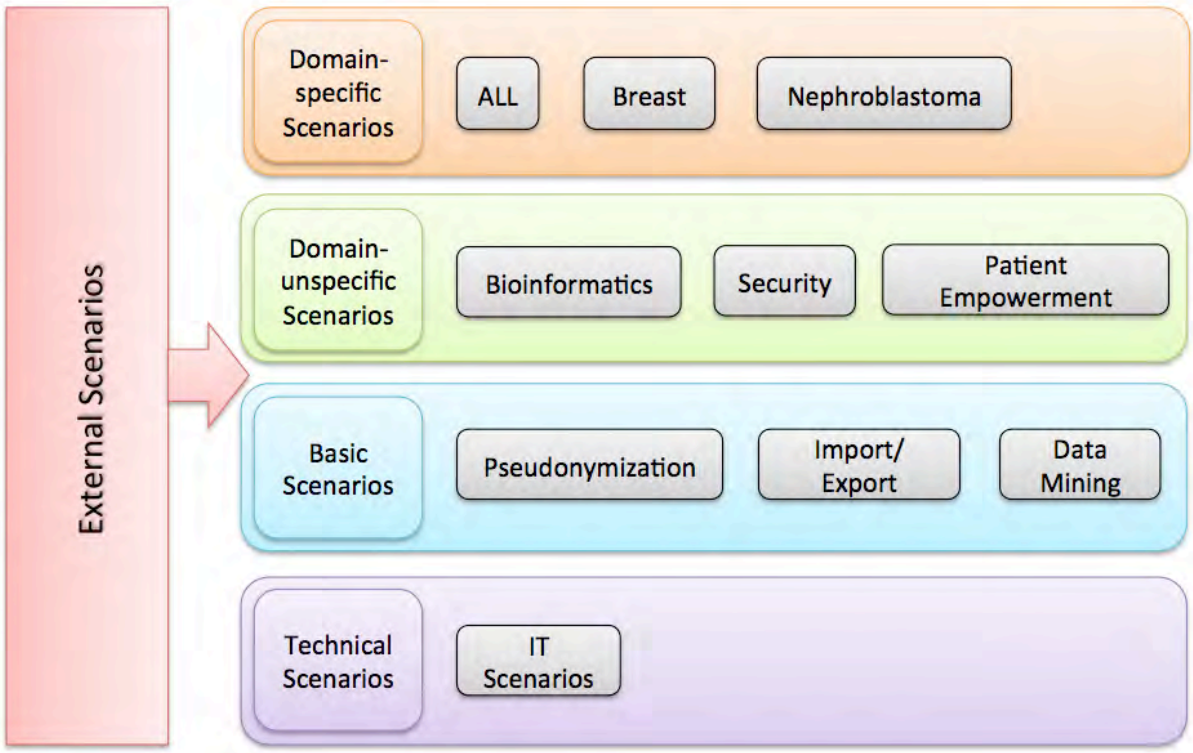


Fig. 2.8: Levels of scenarios giving examples on all levels. Interoperability and standards will allow the integration of external scenarios

The building of such a modular architecture with tools and models categorized into different levels is a major factor contributing to the sustainability of the architecture beyond the funding period of p-medicine.

The development of tools and models and also 'Decision Supporting Services' (DSS) has to be done in several steps. The backbone of all tools, models and services is a system biology approach. Therefore in a first step end-user driven use cases have to be defined describing clinically driven scenarios. These use cases form the basis for the building of tools. Before the programming of a tool a mock-up will be made used for evaluation of usability by end-users. This will be done in the workbench of p-medicine using fake data (fig. 2.9). For the evaluation of the developed tool a testbed needs to be set up where retrospective data will be used. This testbed (including tools and data) will be fixed to allow repeated runs of the tools. If the tools are used in clinical settings with prospective data a testbed with the legal framework for prospective data will be set up allowing the curation and update of the data. This testbed will be named "clinbed". The tools used in this clinbed are validated and will be certified.

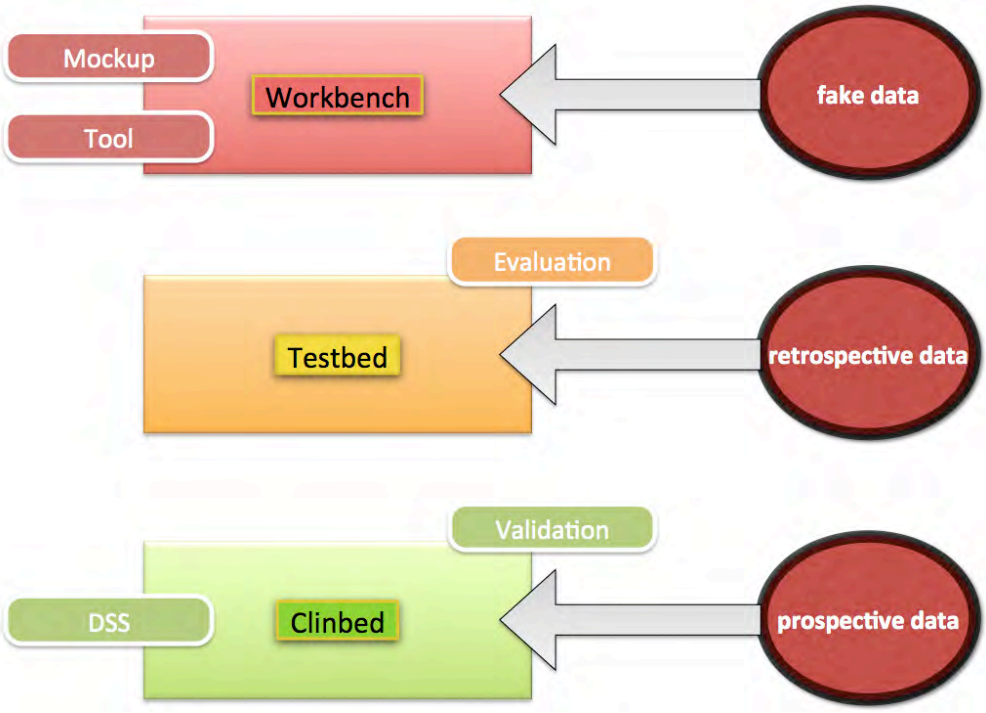


Fig. 2.9: The hierarchy of the architecture for developing tools from workbench, to testbed to clinbed.

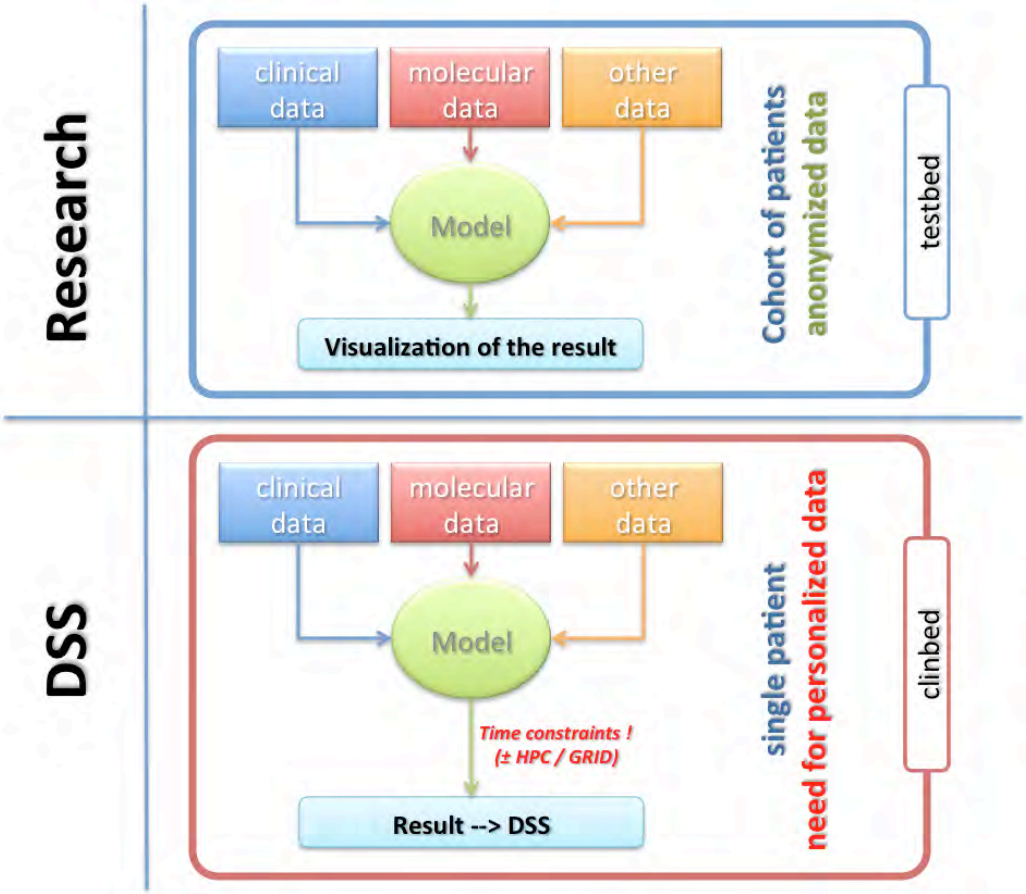


Fig. 2.10: Tools, services, models used for research or DSS. (DSS: Decision support service; HPC: High performance computing)

All tools, services and models that will be developed can be used twofold (fig. 2.10):

- in research
- for decision support services (DSS)

In research the execution of tools, services and scenarios can be done in the testbed without the use of personal data. Results will be visualized according to the specification of the use case. In contrast models for DSS always need personal data. The results are needed within a short timeframe to allow physicians to treat specific patients in time according to the results of the DSS. Therefore logistics need to be set up along a timeline including the analysis of biomaterial, DICOM data, etc. and the execution of the model. On the IT site high performance computing or GRID computing might be necessary (fig. 2.11).

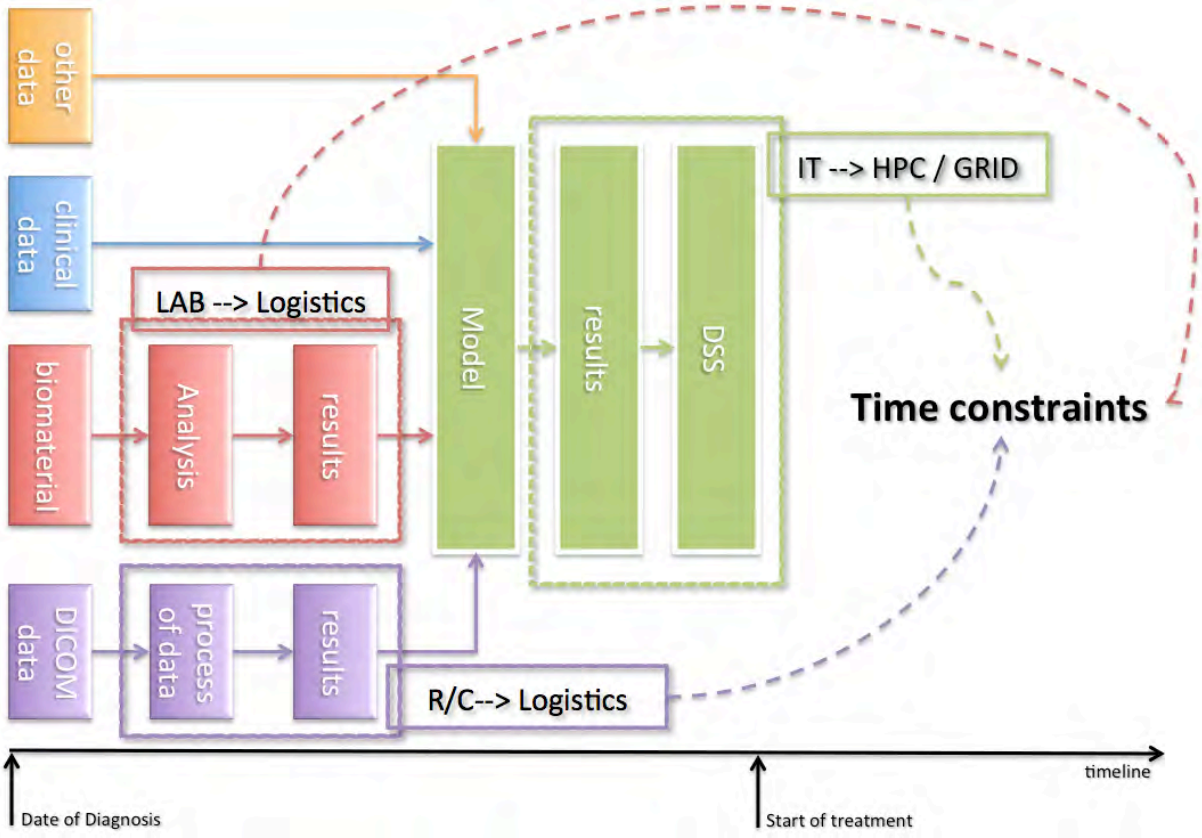


Fig. 2.11: Time constraints in Decision Supporting Services (DSS). (LAB: Laboratory; R/C: Radiologists/Clinicians)

The roles and rights management system regulates which data and tools, services and models an end-user can access and work with.

Of high importance for the architecture of the p-medicine platform is to facilitate data exchange with other health care systems in accordance with the legal framework of p-medicine. Otherwise it will become an “information island” that contain different patient’s data sets, isolated from other information about the patients, with limited access and value. As a result, the p-medicine platform should interoperate with other systems throughout the entire health and clinical studies information environment.

At a minimum, p-medicine should export anonymized data to and import-export data from other systems in a standardized (and interoperable) way. To provide interoperability, the p-medicine platform should support from the very beginning communications, messaging, and content encoding standards as other health information systems (HIS) or EHR.

2.1.4 Applets

Most important for all developed Tools, Models, and services is their user friendliness. Usability issues will be addressed as early as possible starting by the developmental process as described above in 2.1.3. Starting with mock-ups it is guaranteed that only such tools will be built were usability is evaluated by end-users. To increase the usability all developed tools, models and services should be represented by applets (fig. 2.12). An end-user can install these applets on different devices and will be able to execute them by simply clicking on the applet. This will be part of the clinbed.

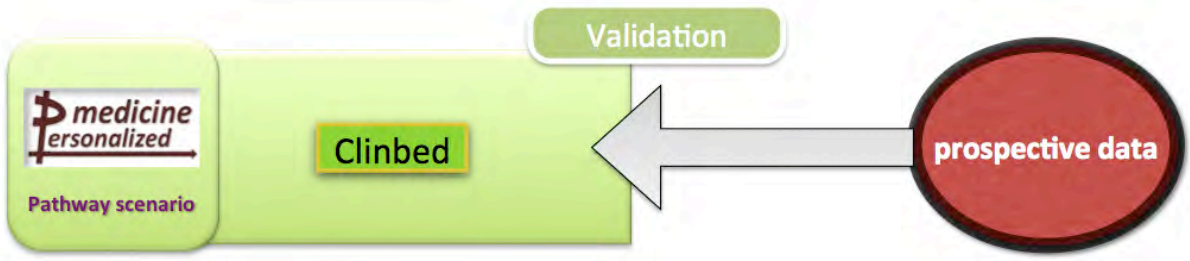


Fig. 2.12: A tool in the clinbed is represented by a specific applet.

2.1.5 Workflows

Workflow is usually defined as the progression of steps (tasks, events, interactions, etc.) that comprise a work process, involve two or more persons, and create or add value to the organization's (project, process, etc.) activities. In a sequential workflow, each step is dependent on occurrence of the previous step; in a parallel workflow, two or more steps can occur concurrently.

The workflow editor is a necessary end user component that should have a user friendly Graphical User Interface (GUI) for graph based modelling. With respect to the integrated environment the user should be allowed to:

- Build a new or to modify an existing workflow
  - The modified workflow should be stored as a new one
  - There should be a link between the old and the new workflow
- Have a visual representation of the status of the workflow enactment process
  - The workflow execution should be possible to suspend and to continue again
- Set breakpoints in the execution of a workflow to examine intermediate results
  - The notification about the breakpoint should be realized by a visual indication
- Select a workflow for research or for decision support depending on the data selected for execution

Each workflow needs additional metadata associated with it. A workflow is seen as a service that accepts workflow parameters and returns workflow results.

The workflow editor developed in ACGT will be refined for the use in p-medicine.

2.2 Introduction to p-medicine Interoperability Specifications

On 17 December 2010, Vice-President of the European Commission Neelie Kroes and United States Secretary for Health and Human Services (HHS) Kathleen Sebelius signed a Memorandum of Understanding (MoU)[4] in Washington to promote a common approach on the interoperability of electronic health records and on education programmes for information technology and health professionals. Common standards and interoperability stand to create huge growth opportunities for the eHealth industry as well as having a positive impact on the safety and quality of care. The Memorandum stresses the need for a joint vision on internationally recognised and utilised interoperability standards for electronic health record systems and increased competences and mobility of IT professionals. Such common standards are important to achieve widespread interoperable eHealth services so that eHealth can reach its full global market potential.

By following the need for a joint vision on internationally recognised and utilised interoperability standards for p-medicine project are recommended the Healthcare Information Technology Standards Panel (HITSP)[5] standards specifications. HITSP is a cooperative partnership between the public and private sectors from the United States. The Panel was formed for the purpose of harmonizing and integrating standards that will meet clinical and business needs for sharing information among organizations and systems.

Appropriate and recommended for p-medicine project are:

- **IS 08 Personalized Healthcare** - The Personalized Healthcare Interoperability Specification describes family history and genetic/genomic lab order and results, which are used to provide personalized treatment specific to genetic makeup.

- **IS 158 Clinical Research** - The Clinical Research Interoperability Specification covers clinical research in all its forms as it interoperates with healthcare systems, particularly EHRs. The specification spans two industries, healthcare and clinical research, and incorporates standards from healthcare (HL7 and IHE) and research (CDISC). The design leverages existing HITSP constructs and communication methodologies where applicable, and lays out new constructs as needed. The design also leverages the current players in the clinical research industry such as Electronic Data Capture (EDC) systems and research registries.

It is not the purpose of p-medicine to develop electronic or personal health records but to develop a system that will be compatible in the future with EHRs. The interoperability specifications are defined in D2.1 and WP4. It is of most importance that infrastructure mechanisms and services are compatible with other VPH infrastructures, e.g. with VPH-Share. Therefore a close collaboration between p-medicine and VPH-Share has already started and will be warranted throughout the lifetime of p-medicine.

2.3 Data

Multi-level data collection within clinico-genomic trials and interdisciplinary analysis by clinicians, molecular biologists and others involved in life science is mandatory to further

[4] Official press release - RAPID - Europa, http://europa.eu/rapid/pressReleasesAction.do?reference=IP/10/1744&format=HTML&aged=0&language=EN&guiLanguage=en

[5] http://www.hitsp.org

improve the outcome of cancer patients. It is essential to merge the research results of biomolecular findings, imaging studies, scientific literature and clinical data from patients and to enable users to easily join, analyse and share even great amounts of data.

The problem of sharing clinical data is a major hurdle for the facilitation of research using that data. There is the need to gain access to distributed data sources in a routine, transparent way, following appropriate anonymization and security procedures, if patient specific medical simulations will be incorporated into clinical practice. While solutions exist to enable access to federate, distributed data sources, in many cases these are either not appropriate or acceptable to a hospital, or not generic enough to be used in anything other than the narrow scenarios for which they were developed.

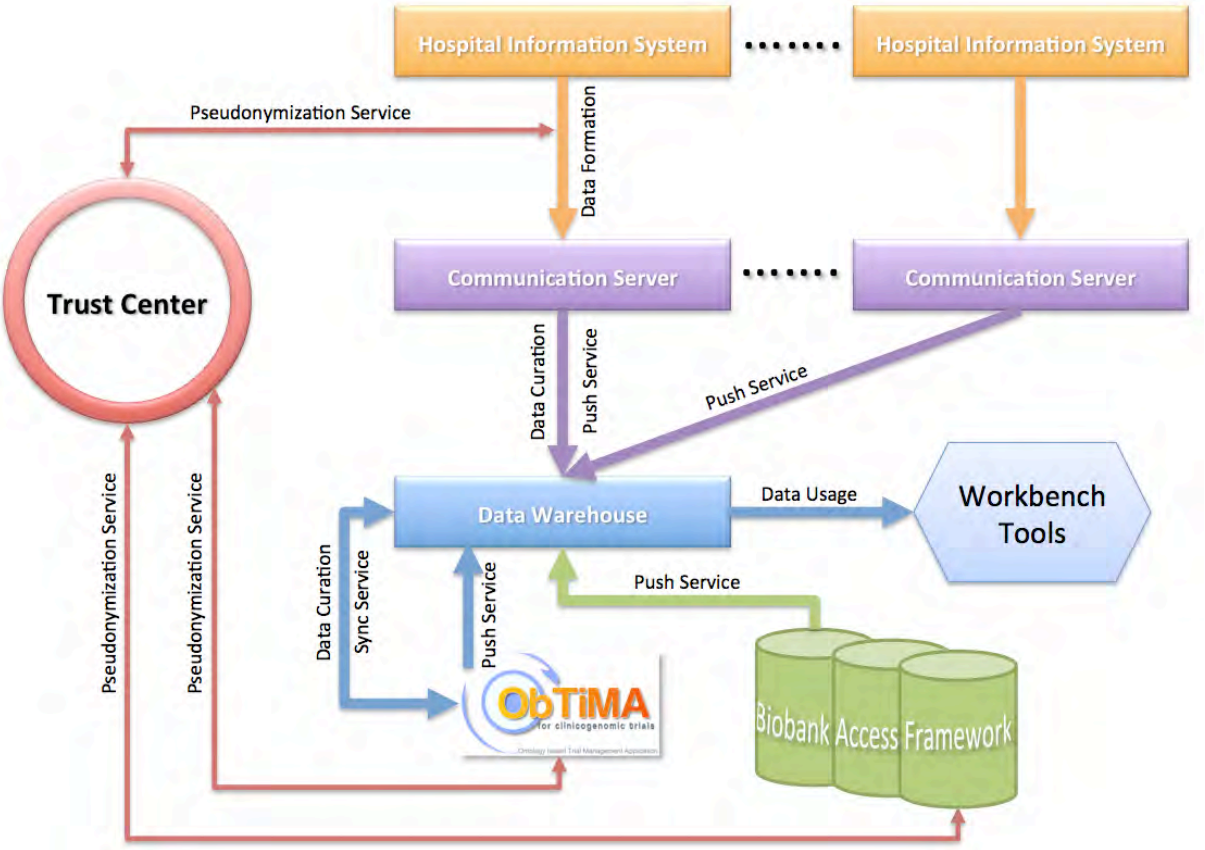


Fig. 2.13: Data flow from Hospital information systems via communication servers to the data warehouse including pseudonymization service, push service, sync service and data curation as well as data usage.

To get access to data from the Hospital Information System (HIS) (fig. 2.13) a tool needs to be used by the hospital, which will shift data to a communication server (CS) outside of the firewall of the HIS. This tool needs to integrate a pseudonymization service that will be used for pseudonymization of data in p-medicine. The data that will be copied to CS are previously defined. As an optimum all data of a patient stored in the HIS are requested. On the communication the format of the data is predefined so that the data from the HIS need to be mapped with these predefined items. Besides structured data other data like surgical reports, pathological reports and other text files as well as DICOM data should be shifted to the CS after pseudonymization of data. The pseudonymization of text files needs to be done as well. The tool to do this job will be primarily developed in a generalized way at the University Hospital in Homburg. The tool and the description of it will be provided to everybody who wants to copy data from their HIS to a CS. The data on the CS can be pushed by the push

service to the data warehouse where the data are annotated using HDOT (see WP4). With the help of a sync- service re-entry of data to specific items in ObTiMA will be avoided, as they can be stored automatically in ObTiMA. The push- as well as the sync- service is important for data curation. Access to data from biobanks is possible after pseudonymization and storage in the data warehouse. All data can be used from the data warehouse for usage in the workbench for developing tools.

There are also many problems related to the disparate nature of the data sources; data on the same clinical pathologies may be stored in different formats by different hospitals, with some data fields stored by some hospitals and not others. While some institutions may be in a position to impose some uniformity on the data imposed in its hospitals, where pan-European or international research is concerned, this is unlikely for many hospitals. The quality of data is another problem. In some clinical environments, certain data fields may not be stored routinely, leaving the available data incomplete.

A secure and scalable data warehouse will be built as a central research resource of *p-medicine* with respective services for collecting and sharing annotated anonymized clinical data and other research relevant data from diverse heterogeneous sources such as in particular clinical trials and electronic patient records from hospital information systems. The data warehouse will store and manage large data sets in an affordable manner and provides the main resource for new knowledge discovery, VPH modelling and simulation. A push concept will be implemented, which allows owners of data (i.e. clinicians, trial chairmen) to annotate and upload their data to the warehouse, in order to make so far unexploited data resources available to research.

A key challenge to *p-medicine* is to integrate heterogeneous and large amount of data from multiple sources (fig 2.14). In order to stock the infrastructure with data we need to be able to integrate clinical trial data from different clinical research centres, which most likely will use different semantics.

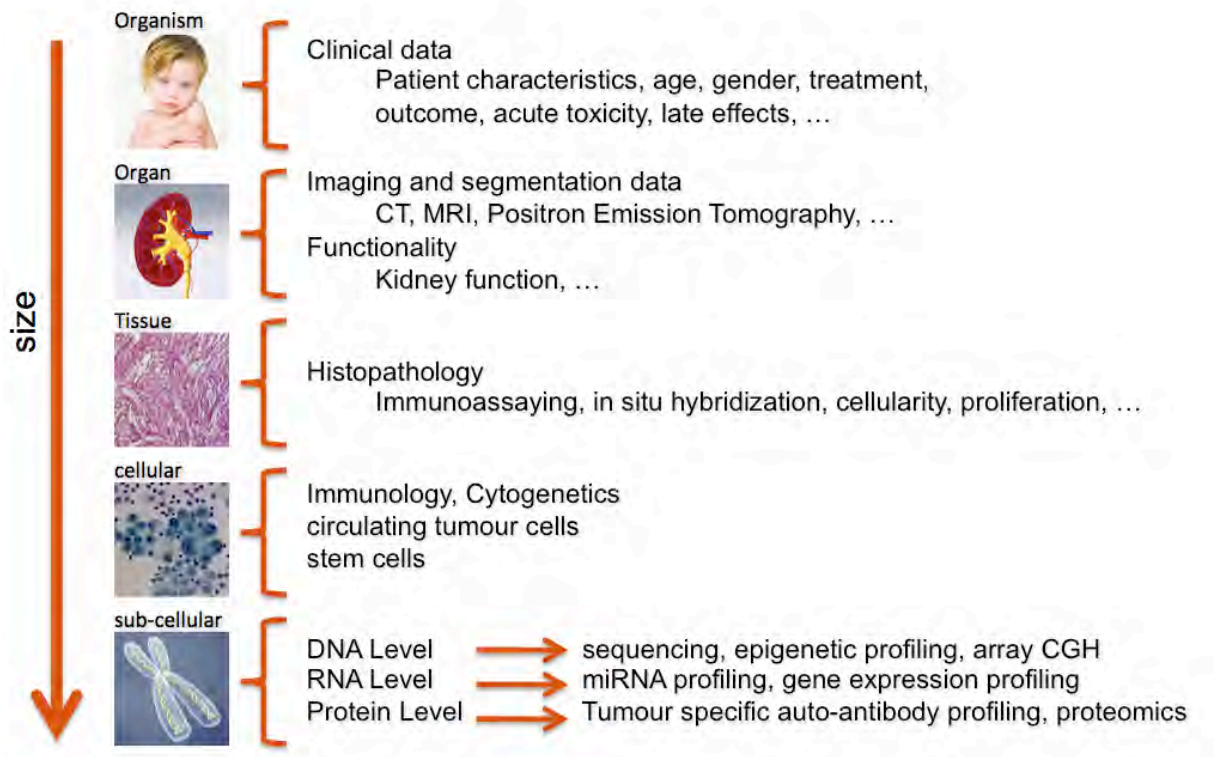


Fig. 2.14: The heterogeneity of data and their size increasing from organism to sub-cellular level

The use of data in p-medicine will be in compliance with the legal and ethical framework. Within the p-medicine platform only anonymized data are handled. No personal data will be used. Pseudonymized data are regarded as personal data as long as it is easy to get the link between a pseudonym and the patient. If the effort to do so is by far disproportional laborious in time, costs or workload pseudonymized data will be regarded as 'de facto' anonymized. To be in compliance with the legal framework the following rules have to be followed:

1. Use only personal data when needed
2. Anonymize personal data
3. Get informed consent from patients for the use and sharing of data
4. Do contracts between data providers and data users
5. Get ethical approval for the research to be done
6. Annotate tools in a way that they can only be executed if the needed data are anonymized.

The last rule (no. 6) is not mandatory in the p-medicine platform.

In p-medicine fake data, retrospective and prospective data will be used to build tools, services and models. To start as fast as possible with mock-ups for tools fake data will be used. Such fake data are available as soon as the structure of data from the different domains is known. The evaluation of the tools will be done with retrospective data, whereas prospective data are needed for Clinical decision support (fig. 2.15).

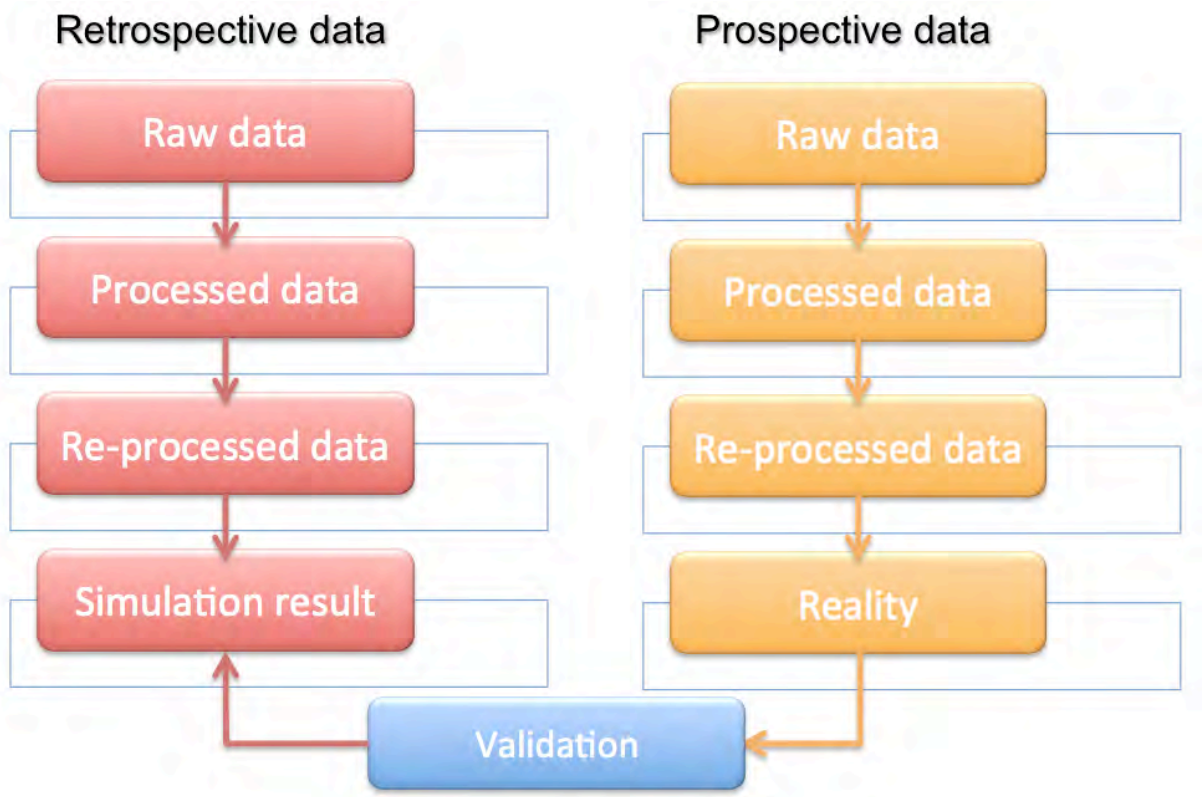


Fig. 2.15: The use of retrospective and prospective data

The use of the different data types reflects the hierarchy of the development of tools. Besides the legal requirements the following issues need to be kept in mind by using data:

1. Data need curation to rely on them
2. Data need to be annotated to use in a standardized way in different models
3. Data will be created not at one time point but during the course of a disease (fig. 2.16)
4. Data will be shared between different institutes and hospitals within a clinical trial (fig. 2.17)
5. Data might not always be as complete as expected
6. Raw data might need to be processed before they can be used in a model
7. The time point when data are acquired might have an influence on the result of a model (fig. 2.18)
8. The storage of data needs standardization and needs to be fixed to use them repetitively in models

Ad 1.
As data change over time, e.g. relapses can occur etc., it is of utmost importance that data curation is in place. If this is not the case decision supporting tools cannot be used in the clinical environment and research will give false results. IT tools alone cannot solve the curation of data. Logistics have to be set up to achieve this goal. This is time and money consuming. A further aspect that does not be neglected is the fact that clinical data coming from clinical trials are more reliable and precise than any other clinical data.

Ad 2.
If same data from heterogeneous sources will be used in a model this is only possible if the data are annotated or linked to Ontology. To achieve this goal WP4 deals with standardization and interoperability issues.

Ad 3.
Diagnosis, treatment and follow-up of patients will be done along a timeline. During that period at every time point new data will be created that might be needed in the model (fig. 2.16). It is important that all data of a patient will be available at the time a model will be executed. A process of automatic upload of all the heterogeneous data from different sources from a patient in the data warehouse is needed in a timely manner. Such an automated process can be started by the end-user before running the model.

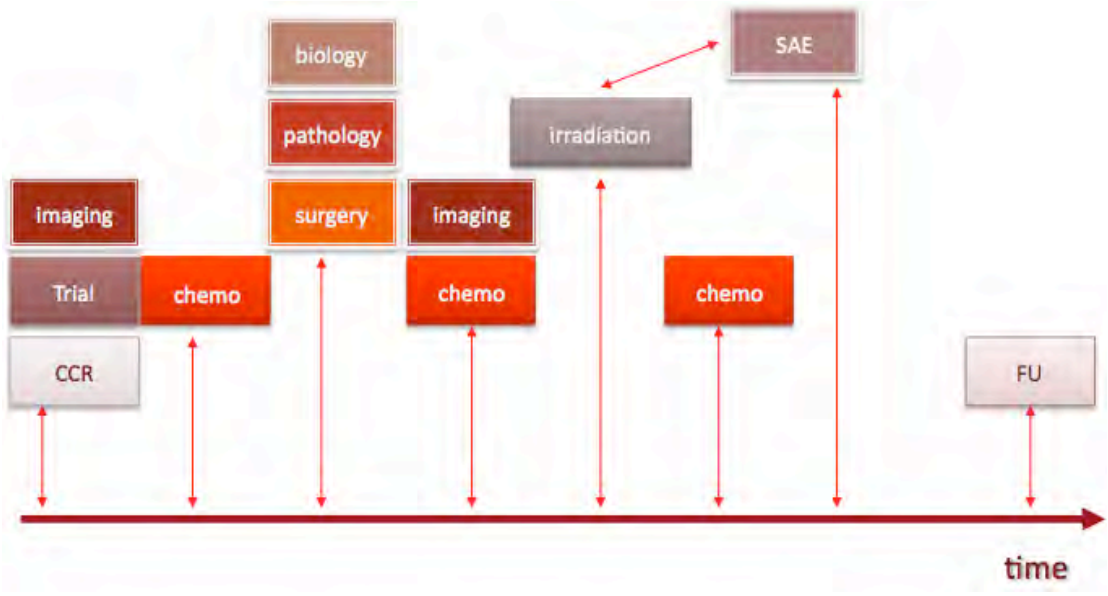


Fig. 2.16: Data Flow in clinical trials over time

Ad 4.
Analysing the data flow in clinical trials it is important to know that data will be shared between different institutes, hospitals, study centres, registries, regulatory bodies and others (fig. 2.17). This data flow is only allowed with anonymized data. If heterogeneous data from different sources need to be shared and combined the same pseudonym needs to be used for all data of single patients. This has to be supported by IT-tools. If feedback of data and results of models need to be given to patients, the pseudonym needs to be linked to the concrete patient via a trust centre. Such a feature of feedback is always needed in tools for patient empowerment.

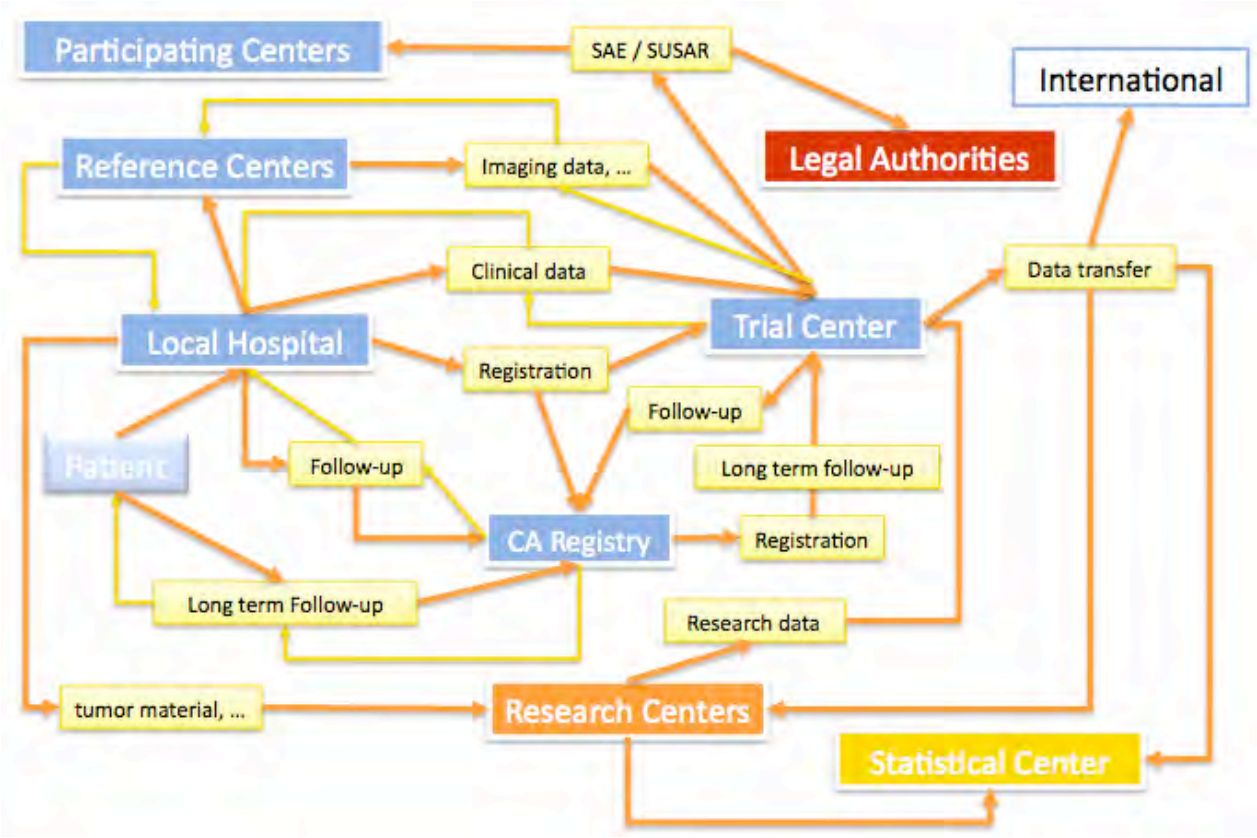


Fig. 2.17: Data Flow in clinical trials between institutes, hospitals, study centres, regulatory bodies etc.

Ad 5.
In a clinical setting one needs to know that data are not always complete. This is even true for data coming from prospective clinical trials. If, for example, a MRI is needed at the time of diagnosis, there might be patients in such a poor conditions, that an MRI is not possible to do. Such patients have to be included in research questions to avoid a bias in data analysing, if the sickest patients are excluded from analysis. In summary this means that tools, services and models have to run with incomplete data sets as well.

Ad 6.
Sometimes raw data have to undergo specific processes before they can be used in tools, models and services. To get the most out of data interaction between data producer/providers and data users is of utmost importance. Only the data provider knows the

limitations of the data. If he understands for which purpose the data are needed the better the data quality will be. Sometimes a processing of data, normalisation of data, etc. is needed before they can be entered in a model. The annotation of data falls in this category as well.

Ad 5.
The time point, when data are generated, is of importance for the interpretation of results of a model. As an example the tumour volume of a specific cancer in a specific patient is needed for data input into the model at different time points (fig. 2.18). The correct tumour volume over time is given by the red curve in figure 2.17. If in a patient an MRI is done at all 4 time points (1, 2, 3 and 4) the correct volume of the tumour can be described. If the MRIs at time point 2 and 3 are missing because they are not required the result of the analysis of the tumour volume will be, that there is a tumour response up to time point 4. But this is not reflecting the reality, as there is again a progression of the tumour with increasing tumour volume at time point 4. Such uncertainties of the data cannot be avoided. What can be done is to define data and time points for their collection as precise as possible and to base this decision on system biology models for selecting the optimal dataset over time. This will also help to validate the models.

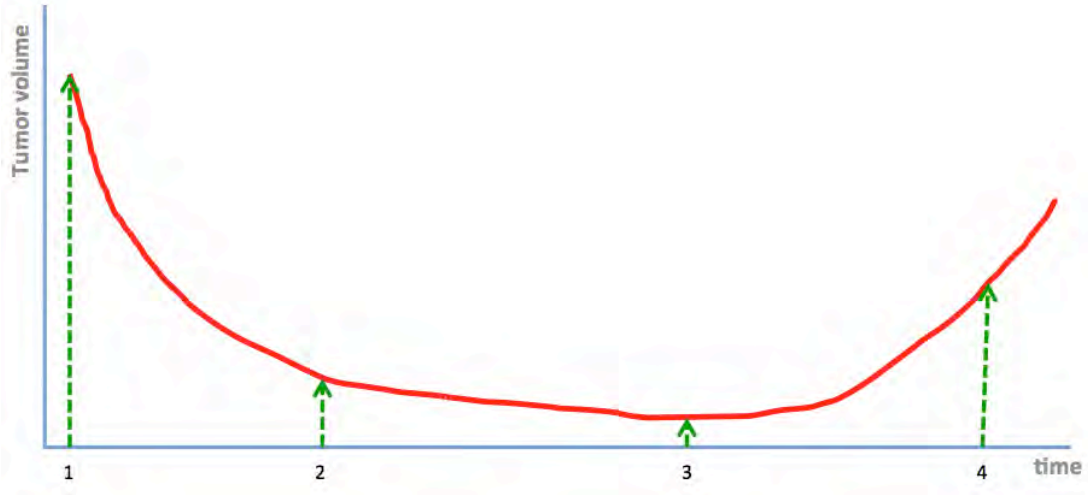


Fig. 2.18: The collection of data over time will influence the correctness of model predictions as described.

Ad 6.
The storage of data needs standardization (WP4). As soon as data are needed for decision supporting tools the source of the data needs to be fixed and needs to be unchanged in structure but curated over time to use them repetitively in models. Therefore a clinbed, as described above, is a precondition for fixing the source of data for models and the curation of data over time.

2.3.1 Availability of retrospective data

In the following section the availability of retrospective data is described. It is of utmost importance that before any data can be used the legal framework of p-medicine has to be in place. This means that only pseudonymized data can be used and that there are contracts between the data producer/providers and the data users are signed. These contracts are available and send to all partners of p-medicine.

2.3.1.1 Nephroblastoma

The International Society of Paediatric Oncology (SIOP) enrolled children with Wilms tumour into 6 studies up to now (SIOP 1, SIOP 2, SIOP 5, SIOP 6, SIOP 9, SIOP 93-01). Graf et al. give a review of these studies. Since 1994 more than 3000 patients with a kidney tumour are enrolled in the SIOP / GPOH studies and trials[6]. The 7th trial and study (SIOP 2001) started in 2002. The randomized question of this trial is stopped in December 2009 after reaching the proposed number of patients. The study continues as a registration study up to the end of 2013.

2.3.1.1.1 Data security

Data security will be handled according to the legal framework of p-medicine.

2.3.1.1.2 Ethical issues

Ethical approval for the use of data in p-medicine is given by the Ethical Committee of the 'Ärztekammer des Saarlandes' in Saarbrücken Germany (Appendix 3) at the 16th of March 2011.

2.3.1.1.3 Contracts

Contracts between data providers and data users will be signed before data can be shared.

2.3.1.1.4 Available data

As treatment in Wilms Tumours starts in the SIOP trials without histological proven diagnosis the prediction of a correct diagnosis and the response to preoperative chemotherapy is of highest clinical relevance. The following data are available:

- Molecular biology data from serum
  - Molecular biology data from serum
  - Autoantibodies against nephroblastoma
  - miRNA data
- Gene expression data
- Imaging studies with data from tumour rendering
- Clinical data

Molecular biological data:

1. Autoantibodies against Wilms tumour.

[6] Graf N, Tournade MF, de Kraker J: The Role of Preoperative Chemotherapy in the Management of Wilms Tumor - The SIOP Studies. Urologic Clinics of North America, 27:443-454, 2000

It is well known that tumours develop autoantibodies against tumour specific antigens. Our group could show this for nephroblastoma. First results are already presented to the scientific community[7,8].

2. miRNA in serum of patients with Wilms tumour

For many diseases miRNA serves as tumour specific markers. We are currently analysing miRNA in serum of patients with Wilms tumour. First results lend strong support to the idea of using specific miRNA profiling of human blood as a diagnostic tool[9].

2. Gene expression data

From part of the registered patients in SIOP 2001 gene expression data will be made available.

All molecular biological data will be available via the Data warehouse of p-medicine. A description and structure of the data is given in the following tables (tab. 2.1; 2.2).

Material			Identifier	Comments
touch preps	number	number: ~ 1-10, no order	10 Ids	
blood	number of vials	ca. 1-3 (with date)	3 Ids	
		type (EDTA, heparin, unknown)	per ID	
		volume (1-10 ml for each)	per ID	
		DNA extracted (yes/no)	per ID	
		amount DNA (microgram/ml)	per ID	a further table might be needed due to multiple entries
		free text	per ID	serum send, miRNA etc.

[7] Nourkami N, Fischer U, Leidinger P, Heisel S, Habel N, Hoppe A, Graf N, Meese E: Immune response pattern in Wilms Tumour patients: New biomarkers for early diagnosis of malignant childhood tumours. 7th International Meeting on the Biology of Childhood Renal Tumors. Banff; 1st – 3rd of March 2010

[8] Heisel S, Habel NC, Hoppe A, Keller A, Nourkami N, Berthold F, Lenhof HP, Gessler M, Graf N, Meese E: Identification of serological markers and generation of autoantibody signatures to improve differential diagnosis of Wilms and Non-Wilms tumours. 7th International Meeting on the Biology of Childhood Renal Tumors. Banff; 1st – 3rd of March 2010

[9] Keller A, Leidinger P, Bauer A, ElSharawi A, Haas J, Borries A, Wendschlag A, Giese N, Tjaden Ch, Nikolaus S, Ruprecht K, Huwer H, Huebers J, Jacobsen G, Rosenstiel P, Sina Ch, Wullich B, Graf N, Reichrath J, JagerSU, Staehler P, Staehler C, Beier M, Scheffler M, Buechler MW, Wischhusen J, Häusler S, Dietl J, Mueller-Quernheim J, Backes CH, Lenhof HP, Schreiber S, Katus HA, Rottbauer W, Meder B, Franke A, Hoheisel J, Meese E: miRNA signatures of human blood – promising biomarkers for human diseases. Submitted, 2010

normal kidney	number of samples	1-3 (with date)	3 Ids	
		amount per sample (ml)	per ID	
		frozen -80, -20, thawed, in culture medium	per ID	
		DNA extracted (yes/no)	per ID	
		amount DNA (microgram/ml)	per ID	a further table might be needed due to multiple entries
		DNA quality (text)	per ID	a further table might be needed due to multiple entries
		free text	per ID	serum send, miRNA etc.
tumor	number of samples	1-5 (with date)	5 Ids	
		amount per sample (ml)	per ID	
		frozen -80, -20, thawed, in culture medium)	per ID	
		additional identifiers (free text)	per ID	
		DNA extracted (yes/no)	per ID	
		amount DNA (sample ID)	per ID	a further table might be needed due to multiple entries
		DNA quality (text)	per ID	a further table might be needed due to multiple entries
		RNA extracted (yes/no)	per ID	
		amount RNA (sample ID)	per ID	a further table might be needed due to multiple entries
		RNA quality (text or RIN)	per ID	a further table might be needed due to multiple entries
material for culture		0-2 (+ free text)	2 ID	
		cultured, successful? (Y/N + text)	per ID	
		frozen in DMSO? (number of vials, text)	per ID	
additional material				

Tab. 2.1: The structure of the data for biomaterial

Analysis			Comments
Allele loss	chromosome arm 1p	marker1	NI/het/LOH; part.LOH; other; nd
		marker2	
		marker3	
		extendable, 0-10 marker	
	chromosome arm 11p	marker1	NI/het/LOH; part.LOH; other; nd
		marker2	
		marker3	
		extendable, 0-10 marker	
	chromosome arm 16q	marker1	NI/het/LOH; part.LOH; other; nd
		marker2	
		marker3	
		extendable, 0-10 marker	
	more chromosomes possible (realistic 3-10 over all)		
CTNNB1	exon3 size	ok/altered/nd	
	exon3 sequence	wt/het/hom (+text)	mutations key
WT1	deletion analysis	ok/altered/nd (+text)	
	mutation analysis	ok/altered/nd (+text)	
mRNA expression		yes/no (+text)	
miRNA expression		yes/no (+text)	
Gene expression		String file (+text)	
addl. tests			

Tab. 2.2: The structure of the data of analytical tests. Additional test are possible.

Imaging data:

All Imaging data are stored as DICOM Files and will be available via the Data warehouse. MRI (T1 with and without contrast enhancement, T2, T2 flair and diffusion weighted imaging) at the time of diagnosis and after 4 weeks of preoperative chemotherapy will serve as the input images. These data needs pre-processing before entering the Oncosimulator to get information of tumour volume and morphology. DoctorEye will be used to segment the tumour and to calculate histograms as described for the glioma scenario.

Clinical data:

All clinical data will be stored in ObTiMA and will be available via ObTiMA.

2.3.1.2 Breast Cancer

There are several retrospective studies enrolled in p-medicine. These studies are described in WP9 of Annex I. For the phase II trials bevacizumab retrospective data are available.

2.3.1.2.1 Data security

Data security will be handled according to the legal framework of p-medicine.

2.3.1.2.2 Ethical issues

The NHS National Research Ethics Service gave ethical approval for the Breast Cancer Avastin Trial as already mentioned in Annex I of p-medicine. Ethical approval for usage of these data in p-medicine is waiting.

2.3.1.2.3 Contracts

Contracts between data providers and data users will be signed before data will be shared.

2.3.1.2.4 Available data

Subject	Data Item	Alternative Naming	Units	Format	Typical Values	Number of Time-points	Timepoint details	Additional Notes
Demographics	Subject ID		str		OX01	1		Just to allow data queries etc
Demographics	Weight		kg	decimal, 1 place	40-100	1	At baseline, pre avastin administration	
Demographics	Height		cm	integer	150-200	1	At baseline, pre avastin administration	Not strictly needed as drug is administered perkg not per BSA
Demographics	Age		years	integer	18-70		At baseline, pre avastin administration	
Demographics	PS			integer	0-3	1	At baseline, pre avastin administration	Broad indication of well-being, according to WHO
Demographics	Menopausal Status			str	pre menopausal, post menopausal, peri menopausal		At baseline, pre avastin administration	
Disease Status at Baseline	Laterality			str1	L, R, B (left, right, bilateral)			potentially bilateral
Disease Status at Baseline	Diagnosis:Histology			str	ductal, lobular, mixed			
Disease Status at Baseline	Diagnosis Date		date	dd/mm/yyyy		1		
Disease Status at Baseline	Stage			str	IIIb	1		this is derived from the T,N,M fields
Disease Status at Baseline	T			integer	T1, T2, T3	1		
Disease Status at Baseline	N			integer	N0, N1, N2	1		
Disease Status at Baseline	M			integer	MO, M1	1		
Disease Status at Baseline	ER status			str	positive, negative	1		status positive is score >=3, derived from both percentage and intensity of staining

Disease Status at Baseline	ER score			integer	0-8	1		
Disease Status at Baseline	her2 status			str	positive, negative	1		status positive is score 3
Disease Status at Baseline	her2 score			integer	0-3	1		
Disease Status at Baseline	PR status			str	positive, negative	1		status positive is score >=3, derived from both percentage and intensity of staining
Disease Status at Baseline	PR score			integer	0-8	1		
Disease Status at Baseline	Site of Mass			str	right upper outer quadrant	1		
BaselineSignSymptoms	BaselineSignSymptom					1	Significant conditions present at baseline	described as per CTC toxicity criteria
BaselineSignSymptoms	BaselineSignSymptom StartDate		date	dd/mm/yyyy		1		
BaselineSignSymptoms	BaselineSignSymptom StopDate		date	dd/mm/yyyy		1		
BaselineSignSymptoms	BaselineSignSymptom continuous			str	continuous, intermittent	1		
BaselineSignSymptoms	BaselineSignSymptom CTCGrade			integer	1,2,3,4	1		
Trial Drug Administration	Bevacizumab Doseage		mg/kg	integer	15	1		
Trial Drug Administration	Bevacizumab Dose Administered		mg	integer	1025	1		
Trial Drug Administration	Bevacizumab Administration Date		date	dd/mm/yyyy		1		
DCE MRI	Date of DCE MRI		date	dd/mm/yyyy		2	Pre-Avastin and PostAvastin(15-22 days)	
Vital Signs	BP systolic			Integer		2		
Vital Signs	BP diastolic			Integer		2		

Vital Signs	Temperatue			decimal, 1 place		2		
Vital Signs	Pulse			integer		2		
Breast Physical Examination	sizes of axillary nodes		cm	decimal, 1 place, possibly repeating	1,2	2	Pre-Avastin and PostAvastin(15-22 days)	
Breast Physical Examination	status of axillary nodes			str	mobile, fixed	2	Pre-Avastin and PostAvastin(15-22 days)	
Breast Physical Examination	number supraclavicular nodes			integer	2	2	Pre-Avastin and PostAvastin(15-22 days)	
Breast Physical Examination	sizes of supraclavicular nodes		cm	decimal, 1 place, possibly repeating				
Breast Physical Examination	status of supraclavicular nodes			str	mobile, fixed	2	Pre-Avastin and PostAvastin(15-22 days)	
Haematology	Haematology Date		date	dd/mm/yyyy		2	Pre-Avastin and PostAvastin(15-22 days)	
Haematology	Haemoglobin		g/dl	decimal, 1 place	13.0-17.0	2	Pre-Avastin and PostAvastin(15-22 days)	
Haematology	White Cells		10^9/L	decimal, 2 places	4.00-11.00	2	Pre-Avastin and PostAvastin(15-22 days)	
Haematology	Platelets		10^9/L	integer	150-400	2	Pre-Avastin and PostAvastin(15-22 days)	
Haematology	Neutrophils		10^9/L	decimal, 2 places	2.00-7.00	2	Pre-Avastin and PostAvastin(15-22 days)	
Haematology	Lymphocytes		10^9/L	decimal, 2 places	1.00-4.00	2	Pre-Avastin and PostAvastin(15-22 days)	
Haematology	Monocytes		10^9/L	decimal, 2 places	0.20-1.00	2	Pre-Avastin and PostAvastin(15-22 days)	
Haematology	Eosinophils		10^9/L	decimal, 2 places	0.00-0.50	2	Pre-Avastin and PostAvastin(15-22 days)	
Haematology	Basophils		10^9/L	decimal, 2 places	0.00-0.10	2	Pre-Avastin and PostAvastin(15-22 days)	

Haematology	Haematocrit		l/l	decimal, 3 places	0.40-0.50	2	Pre-Avastin and PostAvastin(15-22 days)	
Haematology	Mean Cell Volume		fl	decimal, 1 place	83-105	2	Pre-Avastin and PostAvastin(15-22 days)	
Haematology	Red Cell Count		10^{12}/L	decimal, 2 places	4.50-5.50	2	Pre-Avastin and PostAvastin(15-22 days)	
Haematology	Mean Cell HGB		pg	decimal, 1 place	27.0-32.0	2	Pre-Avastin and PostAvastin(15-22 days)	
Haematology	Mean Cell HGB%		g/dl???	decimal, 1 place	31.5-34.5	2	Pre-Avastin and PostAvastin(15-22 days)	
Coagulation	Prothom Time	PT	seconds	decimal, 1 place	12.0-15.5	2	Pre-Avastin and PostAvastin(15-22 days)	
Coagulation	APTT		seconds	decimal, 1 place	24.0-34.0	2	Pre-Avastin and PostAvastin(15-22 days)	
Coagulation	INR			decimal, 1 place	0.8-1.2	2	Pre-Avastin and PostAvastin(15-22 days)	Ratio, compared with normal PT time
Biochemistry	Alanine Transaminase	ALT, ALAT, SGPT	IU/L	integer	10-45	2	Pre-Avastin and PostAvastin(15-22 days)	
Biochemistry	Sodium	Na	MMOL/L	integer	135-145	2	Pre-Avastin and PostAvastin(15-22 days)	
Biochemistry	Potassium	K	MMOL/L	decimal, 1 place	3.5-5.0	2	Pre-Avastin and PostAvastin(15-22 days)	
Biochemistry	Glucose		MMOL/L	decimal, 1 place	3.0-5.5	2	Pre-Avastin and PostAvastin(15-22 days)	
Biochemistry	Urea		MMOL/L	decimal, 1 place	2.5-6.7	2	Pre-Avastin and PostAvastin(15-22 days)	
Biochemistry	Creatinine		umol/L	integer	54-145	2	Pre-Avastin and PostAvastin(15-22 days)	
Biochemistry	Calcium	Ca	MMOL/L	decimal, 2 places	e.g. 2.39	2	Pre-Avastin and PostAvastin(15-22 days)	
Biochemistry	Phosphate		MMOL/L	decimal, 2 places	0.80-1.45	2	Pre-Avastin and PostAvastin(15-22 days)	

Biochemistry	Total protein		g/L	integer	60-80	2	Pre-Avastin and PostAvastin(15-22 days)	
Biochemistry	Albumin		g/L	integer	35-50	2	Pre-Avastin and PostAvastin(15-22 days)	
Biochemistry	Total Bilirubin		umol/L	integer	3-17	2	Pre-Avastin and PostAvastin(15-22 days)	
Biochemistry	AST		IU/L	integer	15-42	2	Pre-Avastin and PostAvastin(15-22 days)	
Biochemistry	LDH		IU/L	integer	100-190	2	Pre-Avastin and PostAvastin(15-22 days)	
Biochemistry	Alk. Phosphatase		IU/L	integer	95-290	2	Pre-Avastin and PostAvastin(15-22 days)	
Biochemistry	GGT		IU/L	integer	0-42	2	Pre-Avastin and PostAvastin(15-22 days)	
Biochemistry	Creatinine Clearance			decimal, 1 place	50-100	2	Pre-Avastin and PostAvastin(15-22 days)	calculated using Cockcroft Formula
Urinalysis	Date of Assessment		date	dd/mm/yyyy		2	Pre-Avastin and PostAvastin(15-22 days)	
Urinalysis	Leucocytes 1-2 minutes			str	neg,trace, small +, moderate ++, large +++	2	Pre-Avastin and PostAvastin(15-22 days)	
Urinalysis	Nitrate 60 seconds			str	neg, mild +, moderate ++, strong +++	2	Pre-Avastin and PostAvastin(15-22 days)	
Urinalysis	Urobilinogen 60 secons			str	normal, above normal	2	Pre-Avastin and PostAvastin(15-22 days)	
Urinalysis	Protein 60 seconds			decimal, 1 place	neg, trace, +, ++, +++	2	Pre-Avastin and PostAvastin(15-22 days)	
Urinalysis	pH 60 seconds			decimal, 1 place	5.0-7.5	2	Pre-Avastin and PostAvastin(15-22 days)	
Urinalysis	Bloods 60 seconds			str	neg,non-haemolysed trace, haemolysed trace, small +, moderate ++, large +++	2	Pre-Avastin and PostAvastin(15-22 days)	

Urinalysis	Specific gravity 45 seconds			decimal, 3 places	1,005	2	Pre-Avastin and PostAvastin(15-22 days)	
Urinalysis	Ketone 40 seconds			str	neg, trace, small +, moderate ++, large +++	2	Pre-Avastin and PostAvastin(15-22 days)	
Urinalysis	Bilirubin 30 seconds			str	neg, small +, moderate ++,large +++	2	Pre-Avastin and PostAvastin(15-22 days)	
Urinalysis	Glucose 30 seconds			str	neg, trace, +, ++, +++	2	Pre-Avastin and PostAvastin(15-22 days)	
AdverseEvents	AdverseEvent			str		1		Worsening or new significant conditions compared with baseline, described as per CTC toxicity criteria
AdverseEvents	AdverseEventStartDate		date	dd/mm/yyyy		1		COULD BE MANY ADVERSE EVENTS,reported in chronological order (same event may be present for more than one subject, occurring in different orders)
AdverseEvents	AdverseEventStopDate		date	dd/mm/yyyy		1		
AdverseEvents	AdverseEventcontinuous			str	continuous / intermittent	1		
AdverseEvents	AdverseEventCTCGrade			integer	1(mild),2(moderate), 3(severe), 4(severe)	1		
DCE MRI	**ktrans**		**mL/100mL/min**	**decimal, 1 place**		**2**	**Pre-Avastin and PostAvastin(15-22 days)**	*******This is derived from the flows calculated from hundreds os voxels. Individual readings over several timepoints per voxel exist.**
gene proliferation				**decimal, 2 places**		**2**	**Pre-Avastin and PostAvastin(15-22 days)**	******* for each gene, about 40000. A .cel file is generated.**

2.3.1.3 Acute Lymphoblastic Leukaemia (ALL)

Around 80% of minor ALL (Acute Lymphoblastic Leukaemia) patients in Germany are treated according to ALL-BFM studies, which are coordinated by the University Hospital Schleswig-Holstein in Kiel. Annually, about 500-550 new cases of minor ALL patients are reported in Germany. Biobanking data and samples for German patients are collected and processed in the study centre in Kiel. The data is stored in a home-grown data base management system (Postgres, Access). Data collection is paper-based. Paper-based forms are filled in from the treating hospitals and sent to Kiel, where the data is manually entered into the data base management system. Currently, approximately 600 parameters per patient are documented. Clinical data is collected in the trial management system Marvin.

When minor ALL patients, who were treated in an ALL-BFM study, have a relapse, they are treated according to ALL-REZ BFM studies, which are coordinated from the Charité in Berlin. Annually, ca. 60 relapses occur in Germany. Biobanking data and samples for German relapse patients are collected in the study centre in Berlin. The data is collected in a web-based biobanking management system that was tailored for the Charité based on the flexible data management framework Scopeland. Clinical data is collected in the trial management system Marvin.

The ALL-BFM and ALL-REZ BFM study groups participate in European multicentre clinical trials. In such trials each European partner has own solutions to store clinical and biobanking data. New ALL European wide studies will start at the beginning of the next year. It is foreseen that in these studies, European partners, who are still not satisfied with their solutions for biobanking data management, will use the Scopeland system.

Currently, ALL partners can only access clinical and biobanking data for their own patients that are stored in their own databases. It is e.g. not possible for clinicians or researchers in Berlin to access the biobanking data for their patients that is stored in Kiel, or pseudonymized sample data from other European partners. It needs to be pointed out that sharing of clinical and biobanking data between partners could help to find better therapies and improve patient treatment.

2.3.1.3.1 Data security

Data security will be handled according to the legal framework of p-medicine.

2.3.1.3.2 Ethical issues

Ethical approval for the ALL Trial is given as already mentioned in Annex I of p-medicine. Ethical approval for usage of these data in p-medicine is waiting.

2.3.1.3.3 Contracts

Contracts between data providers and data users will be signed before data will be shared.

2.3.2 Prospective data

Prospective data will be used for the validation of the tools. If models and tools are used for Decision support the architectural infrastructure as described above is needed. Besides data security data curation is of utmost importance.

2.3.2.1 Nephroblastoma

In addition to the mentioned retrospective data proteomic data will be used in p-medicine. A platform for identifying the disease-proteome signature will be established to define protein expression patterns that can identify specific phenotypes (diagnosis), establish a patient's specific outcome independent of treatment (prognosis) and predict a potential outcome from the effects of a specific therapy (prediction). For the benefit of a personalized medicine this platform requires the proteomic tools, the 'hardware', and the 'software', to extract meaningful statistical and biological information from samples, which are defined by hundreds or, thousands of measurements.

Advances in DNA/RNA-technologies including gene microarray analysis and genomic fingerprinting will be further pursued to rapidly screen for global and specific changes in gene/mRNA expression. However, compelling reasons argue for the approach focusing on the protein perspective. Proteins, rather than genes or mRNAs, are the functional output of the cell and therefore might be expected to provide the most relevant information, particularly when interpretation of their expression takes into account their dynamics in specific biological contexts. In addition, a number of reports have compared the steady-state levels of proteins with those of their corresponding mRNAs. Results from these studies have suggested that mRNA abundance is a poor indicator of the levels of the corresponding protein and, unsurprisingly, most licensed tests that are available for disease detection are protein- based assays. The enzyme-linked, immunosorbent assay (ELISA) system, for example, represents the most reliable, sensitive and widely available protein-based testing platform for the detection and monitoring of certain cancers. These assays require, to a certain extent, the identities of proteins of interest - the 'biomarkers' - for choosing suitable capture molecules. Capture is traditionally antibody based, requiring a specific antibody for each protein of interest. The protein of interest - the 'biomarker' - has to be identified before one can start to generate such antibodies and both procedures are painstaking efforts requiring meticulous validation processes.

Whereas ELISA or protein microarrays can reveal only changes in targeted proteins/ known biomarkers, proteomics approaches examine the collection of proteins to determine how, when and where they are expressed; they are particularly promising in the analysis of biological fluids and new biomarker identification. Plasma is among the most accessible biological materials available; at the same time the plasma proteome is challenging because of its complexity and vast dynamic range: Plasma contains several thousand proteins with concentrations ranging from as high as 30 to 50 mg/ml (serum albumin) to femtomolar concentrations for serum biomarkers. To overcome this challenge presented by the wide range of concentrations of plasma proteins, the proteomic 'hardware' requires

- separation techniques to remove high abundance proteins such as albumin and immunoglobulins, that interfere with the detection of less abundance proteins,
- enrichment by chromatographic/electrophoretic means to reduce complexity and to improve identification and
- high-end analysis mass spectrometry systems.

The latter includes the surface-enhanced laser desorption/ionization time-of-flight mass

spectrometry (SELDI-TOF-MS) technique, which is a high- throughput techniques for the analysis of complex biological specimens such as serum as well as the LTQ Orbitrap XL, a high performance hybrid LC-MS and MSn system to identify low level components in complex mixtures.

Using proteomic monitoring should allow defining individual protein profiles and - independent of the identity of the proteins or peptides - proteomic patterns to be used as diagnostic, prognostic and predictive paradigms on the way to a personalized molecular medicine.

The structure of the data as well how to store them is not yet solved.

3 The identification of User Needs

Introduction

This section will focus in special on end users needs (EUNs). The proposed EUNs of the p-medicine platform have been identified on p-medicine project description and as a result of scientific literature reviews. In general terms all are belonging to the below groups of end users:

1. Healthcare providers
2. Patients
3. Researchers including lawyers, ethicists, IT-people
4. Clinical Research Organisations (CRO)

It is important to mention that due to the p-medicine project interdisciplinary, flexibility and modular infrastructure these four major groups can be extended in future.

Healthcare providers – a health care provider is an individual or an institution that provides preventive, curative, promotional or rehabilitative health care services in a systematic way to individuals, families or communities. An individual health care provider (health worker) may be a health care professional, an allied health professional, a community health worker, or another person trained and knowledgeable in medicine, nursing or other allied health professions, or public/community health. Institutions include hospitals, clinics, primary health care centres and other service delivery points.

In this deliverable we will describe only the general needs of medical doctors with access to the p-medicine platform and in order to identify specific needs detailed use case scenarios will be provided from all clinical project partners. These use cases will be analysed in order to identify other (in special not mentioned or not described) EUNs.

As a result two conventional versions of EUNs are proposed for implementation:

- An Initial Version and
- The Final Version of EUNs.

This approach will assure to design a robust and state-of-the-art platform focused on EUNs and other requirements (technological and clinical/medical perspectives).

A similar workflow scenario will be applied to other groups of p-medicine platform end users.

Patients - a patient is any recipient of medical attention, care, or treatment. In the p-medicine project and in general on the p-medicine platform patients will play a central role due to the proposed “personalised medicine” objectives. EUNs of patients will be checked from existing Personal Health Record (PHR) systems, which demonstrated advanced acceptance and usage rates. Special attention is given on specific EUNs of patients with ALL, Wilms tumour and Breast Cancer.

EUNs related to patients will not follow the above proposed versioning scenario, due to the requirement to keep the flexibility in place. PHR environment is very flexible and do not accept (in special due to high competition) predefined (and usually not representative) and not flexible EUNs. One of the major p-medicine technological requirements should be ‘Flexibility’ it will play a crucial role in case of Patients as end users. In this context a flexible implementation strategy, a prototype focused on gathering end patients feedback will be proposed for implementation.

Researchers - a researcher is somebody who performs research, the search for knowledge or in general any systematic investigation to establish facts. Researchers can work in academic, industrial, government, or private institutions. The p-medicine project itself represents a research project with a network of high-skilled and professional researchers familiar with interdisciplinary research topics. Nevertheless the area of interest defined in the frames of EUNs for researcher will be mainly focused on research topics applied to ALL, Wilms tumour and Breast Cancer. Lawyers and ethicists guarantee data security in accordance with the legal framework. IT researchers provide the technological background of the platform to run the system smoothly.

Detailed use cases with related workflows from (ALL, Wilms tumour and Breast Cancer) research perspective will be defined below. An important activity has to be the benchmarking of defined Researchers' EUNs to the technological frames of the p-medicine platform. In order to assure a defined and realistic implementation of identified operational Researchers' EUNs a versioning approach is proposed for implementation

Clinical Research Organisations (CRO) - a contract research organization, also called a clinical research organization, (CRO) is a service organization that provides support to the pharmaceutical and biotechnology industries in the form of outsourced pharmaceutical research services. CROs range from large, international full service organizations to small, niche specialty groups and can offer their clients the experience of moving a new drug or device from its conception to FDA/EMA marketing approval without the drug sponsor having to maintain a staff for these services. EUNs of CROs are broad but the focus will be on aspects related to the goals of p-medicine, and in special to ALL, Wilms tumour and Breast Cancer. All activities related to the identification of CROs' EUNs will follow the proposed versioning scenario but the challenging task will be to track the feedback from an active CRO with the main goal to identify new 'needs' or strengthen the identified EUNs.

3.1 Process of collecting EUNs

The process of collecting and identifying EUNs is defined by the below listed activities:

- Initial definition of EUNs (Initial Version) – the main focus of this deliverable;
- Use Cases Scenario Requests – addressed to all project partners;
- Analysis of the Received Use Case Scenarios – addressed in details on further deliverables and other WPs related to p-medicine system requirements, architecture, design and prototyping;
- Final definition of EUNs (Final Version) – will be described and implemented in the frames of further deliverables (end users' manuals, guides and platform specifications).

EUNs	Versioning	'Flexible' approach
Healthcare providers	Initial Version / Final Version	No
Patients	No	Yes (feedback & requests tracking frames)
Researchers	Initial Version / Final Version	No
Clinical Research Organisations	Initial Version / Final Version	No

Only the EUNs applied to patients will follow a so-called 'Flexible' approach. An initial set of Patients' EUNs will be defined but it will require changes and adjustments throughout all implementation processes of the p-medicine platform. In order to assure a successful implementation EUNs feedback is needed. Only by following this approach the p-medicine platform will be widely accepted, explored and successfully used by patients.

3.2 End User Needs

This section describes in details the previously mentioned End Users Needs (EUNs) in the perspective of the below groups of end users: Healthcare providers; Patients; Researchers; Clinical Research Organisations (CRO).

The sub-sections will end with a table with a short description of the identified EUNs. If applicable the versioning approach is recommended for implementation. It will assure the tracking and documentation of all identified EUNs in special after analysing the collected use case scenarios.

3.2.1 'Healthcare providers' EUNs

A committee of the Institute of Medicine of the National Academies[10] has identified a set of 8 core care delivery functions that EHR systems should be capable of performing in order to promote greater safety, quality and efficiency in health care delivery. The eight core functions are:

- Health information and data,
- Result management,
- Order management,
- Decision support,
- Electronic communication and connectivity,
- Patient support,
- Administrative processes and reporting,
- Reporting and population health.

The report[11] was sponsored by the U.S. Department of Health and Human Services and is one part of a public and private collaborative effort to advance the adoption of EHR systems. The above core functions are in strong relationship with the 'Healthcare providers' EUNs.

In the perspective of 'Healthcare providers' EUNs, p-medicine platform, with all its functionalities and modular structure, cannot provide all these EUNs. More important is to guarantee interoperability with EHRs for future developments. The EUNs of p-medicine are mainly based on scenarios coming from ALL, Breast Cancer and nephroblastoma.

By underlining the functionalities of the proposed p-medicine platform and taking into account the interoperability to future EHR systems core functions of the EUNs of healthcare providers are summarized as follow:

[10] http://www.iom.edu

[11] Institute of Medicine. Key Capabilities of an Electronic Health Record System, 2003.

- Access to patient's heterogeneous health data from different sources
- Ability to join data and use them for research and decision support
- Access to tools, services and models of the p-medicine platform (Oncosimulator, DoctorEye, ObTiMA)
- Access to advanced and secure communication and connectivity (with patients and/or other healthcare professionals)

The identification of 'Healthcare providers' EUNs will follow the 'Versioning' approach. The initial version of the identified and proposed for implementation EUNs are presented in the table below. The final version will be defined and described in the frames of further deliverables as a result of analysis of the received and analysed use case scenarios.

Initial Version of the 'Healthcare Providers' EUNs

Healthcare Providers' EUNs	**Description**	**Comments**
Access to patient's heterogeneous health care data from different sources	• Clinical data • Imaging data • Pathological data • Laboratory data • Clinical trial data • Biobanking data • Research data • etc.	Data security is of utmost importance (legal framework) Notification of data should be available Secure access according to the legal framework
Ability to join data and use them for research and decision support	• Ontologies • Standardization and annotation • Curation of data • Access to HPC if needed • etc.	
Access to tools, services and models of the p-medicine platform (Oncosimulator, DoctorEye, ObTiMA)	• Interoperability • Annotation • Testbed • Clinbed • etc.	Decision support services SAE / SUSAR module DICOM module etc.
Access to advanced and secure communication and connectivity (with patients and/or other healthcare professionals)	• Communication tool in ObTiMA	

3.2.2 'Patients' EUNs

To date, little work has been conducted to identify the EUNs of patients with ALL, Wilms tumour and Breast Cancer. This section is an overview of synthesis of the findings from literature review and a general description of functionalities and features identified on PHR systems.

In general terms PHR includes different classes of information and tasks that users (Patients) can access and perform. Keeping in mind that most of the literature is generated by health care professionals the most widely described tasks of PHRs are listed below:

- Review/update medical records
- Make/change appointments
- Request referrals and prescriptions/refills
- Review laboratory results
- Email physicians and other health professionals
- Solicit and obtain generic and/or personalized health advice
- Participate in chats, online discussion and support groups
- Receive decision support for medical choices[12]

The Markle Foundation's Connecting for Health collaborative, a public-private endeavour, works toward an interoperable health information infrastructure defined PHR. In their report on the subject stated as: "An electronic application through which individuals can access, manage and share their health information, and that of others for whom they are authorized, in a private, secure, and confidential environment."[13]

Despites multiple advanced PHRs solutions with advanced functionalities one important conclusion is: "The adoption and effectiveness of PHRs will therefore depend as much on systems and user interfaces as on data in records"[14].

A recent (July, 2011) study[15], conducted by the IBM Institute for Business Value, indicates that "information seekers" - people who will increasingly turn to technology to help manage health-related challenges to reach their wellness goals, drive the growing demand for healthcare devices. The study surveyed more than 1,300 consumers currently using health and wellness devices and found that these consumers are demanding a new generation of health devices, greater simplicity and better information sharing. Users want the ability to connect with their caregiver and reduce office visits to their healthcare professionals and the added ability to collaborate online with a community of peers with similar issues and interests. According to the survey, users will expect devices to easily share information with their family or healthcare professionals. Additionally, they require:

[12] Marchionini G., Rimer B.K., and Wildemuth. Evidence Base for Personal Health Record Usability Final Report to the National Cancer Institute. University of North Carolina at Chapel Hill, February 10, 2007.

[13] Connecting for Health. The personal health working group final report. Markle Foundation; 2003 Jul 1.

[14] Marchionini G., Rimer B.K., and Wildemuth B.: Evidence Base for Personal Health Record Usability, Final Report to the National Cancer Institute, 2007.

[15] http://www.ehealthnews.eu/ibm/2673-ibm-study-identifies-new-generation-of-connected-health-devices. July, 2011.

- **Ease of use** - 96 percent said ease of use is the top factor in selecting one device over another.
- **Reasonable pricing** - Costs at or below $100 is a critical decision factor according to three quarters of users who consider price well ahead of features, customer support, warranty or stylish design.
- **Real-time information sharing** - 86 percent of consumers want real-time, easy-to-understand feedback from their devices.

Very interesting are the research results in the terms of interoperability from other surveys. Asked to weigh the relative importance of interoperability against other preferences, including technology type, PHR provider, and medical identification scheme, the quantitative survey respondents rated interoperability and portability factors as least important.[16] The research results suggest that interoperability and portability would serve as an additional acceptance and success factor of the p-medicine platform.

Patients' EUNs	Description	Technological background	Comments
Simplification of login process without any (or minimal) devise requirement to p-medicine platform	Access to p-medicine platform (patient's interface) without any complex technological requirement (web based and/or from mobile devises)	SaaS platform with granular access rights	
Giving eConsent and re-consent	Participating in a clinical trial or a research project patients can give consent and get information about the research carried out	Access via a secure website	Collaboration with CONTRACT[17]
Instant access to his own clinical data, including laboratory results, imaging data	Patient can access own data, within ObTiMA he can enter data and he can use a patient diary	Access should be via Data warehouse or ObTiMA Integrating Healthcare Enterprise (IHE) Radiology Technical Framework has released the Cross-Enterprise Document Sharing for Imaging (XDS-I) Integration Profile	Changing or deleting the data will not be allowed In the patient diary data can be edited by the patient, but all changes are recorded in the audit trail

[16] Lafky D.B., Horan T.A., Prospective Personal Health Record Use Among Different User Groups: Results of a Multi-wave Study. Proceedings of the 41st Hawaii International Conference on System Sciences, 2008.

[17] CONTRACT: Consent in a Trial and Care environment, FP7-HEALTH.2010.4.2-6, Grant agreement no: 261412

Search for data or information using search facilities including the use of natural language	Patient can search for his data or for other information about his disease in the Internet or for ongoing trials for his disease including enrolment guidelines or for basic information of available samples in biobanks of his biomaterial and what was done with it	Access via a secure website as for eConsent	It is important to integrate/implement advanced semantic search functionality
Medication Frames	Ability for patients to access medication prescriptions with links to further information including drugs description and drug interaction	Automatically telling drug interactions and contraindications or dosage modifications	
Advises, alerts, reminders, etc. for related information	Patients get alerts if a new research project starts and he needs to give consent, or he gets a reminder to enter data in his patient diary, or for other internal messaging from a trial he is enrolled, etc.	Data mining and semantic analysis existing patient data, plus benchmarking with existing knowledge base e-mail server for sending information	
Communication facilities	Patient can communicate with his physician, or with PI of a research project for getting a second opinion, or he can search for contact details of other patients with the same history, if these patients do allow it	Social networking frames would be required but with a special attention on data privacy.	
Security	Patients have to be sure that their data are secure and any not-authorised access is not possible.	Access to the system is only possible via the portal	
Feedback and Suggestions tracking frames	Patients should be able to suggest continuously improvements for p-medicine system or feedback/ticketing related functionalities should be available.	Feedback or Ticketing frames	

3.2.3 'Researchers' EUNs

Researchers as end users of the p-medicine platform are one of the most complex and powerful End Users in the terms of needed functionalities. The first (initial) version of 'Researchers' EUNs will take into account a couple of important topics shortly presented below:

1. **Modularity Related Frames** –the modular concept and the proposed interoperable integration of models, tools and services into the p-medicine platform is provided by the functionalities, features and metadata of the models, tools and services

2. **Time Related Frames/Perspectives** – maintenance and sustainability of the p-medicine platform will be guaranteed by an open access architecture and clearly defined interoperability features that allow the integration of data, tools, models and services from other projects

3. **Disease Related Frames** – the main focus of p-medicine platform are ALL, Wilms tumour and Breast Cancer. The infrastructure of p-medicine will be open to other diseases in the domain of cancer and beyond

4. **Semantic/Ontology Based Search Engine Frames** - of high importance for researchers is the availability of an integrated advanced semantic search engine including data mining features

5. **Clinical Research (Trial) Related Frames** – the conduct of clinical trials, sharing and joining of data and cross trial analysis in a secure framework to generate new knowledge or to develop decision support services is of utmost importance

6. **Legal Framework** – anonymization/pseudonymization of personal data, the secure storage and access of data needs to be guaranteed

7. **Other Research Related Frames** – other 'Researchers' EUNs, which could not be integrated into the above points.

Initial Version of the 'Researchers' EUNs

Researchers' EUNs	**Topic**							**Comments and/or linkage to other section, deliverable, or WP**
	1	**2**	**3**	**4**	**5**	**6**	**7**	
Secure access to p-medicine platform	+	+	+	+	+	+	+	Access via the p-medicine portal
Access to patient's health information and data without the possibility to change or alter data	+	+	+	+	+	+	+	Consent issues and contracts between data provider and data user according to the legal framework are conditions sine qua non
Create new and modify existing workflows	+	+	+	+	+	+	+	Use of data and tools keeping in mind modularity issues
Create new and modify existing tools, models, services	+	+	+	+	+	+	+	This functionality is solely for IT researchers

Access to data via a semantic search engine	+	+	+	+	+	+	+	p-medicine platform should act as an innovative discovery platform for anonymized life sciences data. Initial patient's data should be easily imported and integrated with public data, and explored within relevant biological and clinical context. As potential publically available data would be recommended the integration of: • PubMed Repository with > 20 Mil. Biomedical scientific abstracts • Clinical Trials Repositories (ClinicalTrials.gov18, EU Clinical Trials Register19) • Gene Ontology (data correlation with PubMed) • Drugs Description • Drug Interaction database (e.g. Medscape) • News, Announcements • Other publically available information
ObTiMA interface and functionalities	+	-	-	-	+	-	-	To conduct and to analyse clinically trials even across different trials
DoctorEye interface and functionalities	+	-	-	-	+	-	-	Use DoctorEye whenever needed within and outside of clinical trials
Oncosimulator interface and functionalities.	+	-	+	-	+	-	-	Usage for research and decision support
Access to advanced and secure communication and connectivity	-	+	-	-	-	+	+	A networking platform at least for registered researchers should be in place.
Access to p-medicine clinical trials frames (without the possibility to change or alter data!)	-	-	+	-	+	+	+	The possible Clinical Trials related research tasks and requirements are very complex due to specific needs of specific clinical trial, see ObTiMA
Other modules/tools [Example] A p-medicine application programming interface (API)								The need to integrate other modules, tools. In order to implement this feature an accessible and very well described particular set of rules and specifications of p-medicine platform (p-medicine API) should be available.

1. Modularity Related Frames; 2. Time Related Frames/Perspectives; 3. Disease Related Frames; 4. Semantic/Ontology Based Search Engine Frames; 5. Clinical Research (Trial) Related Frames; 6. Legal Framework; 7. Other Research Related Frames

[18] http://www.clinicaltrials.gov July 2011
[19] https://www.clinicaltrialsregister.eu July 2011

3.2.4 'Clinical Research Organisations' EUNs

Close to the trend of outsourcing clinical trials, another way to cut the costs of new drug development is to adopt new technologies to manage the huge amounts of patient information involved[20]. As clinical trials are more complicated, expensive, regulated, and monitored, clinical research organizations (CROs) are dealing specifically with data collection and monitoring[21]. This trend is evolving in parallel with the adoption of EHR by hospitals and physician practices, and it holds significant promise for clinical research.

The p-medicine platform should address directly all the needs of CROs as soon as all identified requirements will serve (at later stage) as a background for further exploitation and a wide acceptance of the p-medicine clinical research frames as well as the interoperable, integrated modules (ObTiMA).

One of the major needs for any CROs is and will remain patient enrolment in clinical trials. The following features are available after registration of the CRO:

- Access to a search tools for specification of the number patients with specific criteria (Age, disease, stage of disease, treatments, allergies, etc.) who might be able to enter a research project
- Access to freely available tools, services etc.
- Share suggestions, feedback and any other requests and/or comments regarding the usability and flexibility of p-medicine platform

After signing contracts between p-medicine and a CRO for usage of the p-medicine platform more features will be available depending on the contract.

Initial Version of the 'Clinical Research Organisations' EUNs

CROs' EUNs	Description	Technological background	Comments
Simplification of login process without any (or minimal) devise requirement to p-medicine platform	Access to p-medicine platform (CRO's interface) via the portal.	p-medicine portal	Access is depending on credentials. Free access with only limited possibilities. After a signed contract more features can be used, depending on the contract
eContract	Access to the p-medicine platform is restricted to those stakeholders having signed a contract with p-medicine	p-medicine should provide templates to build specific contracts	Tool like the eConsent tool

[20] Carlson P E, Clinical Research Industry Trends, National Center on Education and the Economy, January 2007

[21] Brooks K. CRO industry update: growth, expansion and new opportunities. Contract Pharma. Available at: http://www.contractpharma.com July 2011

Access to patient's anonymized data (clinical, imaging, laboratory, research, biobanking, etc.) including search facilities (see EUNs of researchers)	Access is limited according to the signed contract with a CRO		Only after signed contracts for usage of data
Communication facilities	Mainly communication with clinicians and researchers, if they agree in advance Communication with patients will not be allowed during the initial phase of p-medicine		Depending on consent from stakeholders and contract between CRO and p-medicine
Security	Patients have to be sure that their data are secure and any not-authorised access is not possible.	Access to the system is only possible via the portal	
Internal messaging and alerts	Flexible subscription frames to alerts and/or messages from p-medicine system.	Internal messaging frames (e-mail server)	
Flexibility and usability	CROs should be able to select topics, sections "of interest" by simple "drag-and-drop" approach.	p-medicine system should be able to keep CRO's preferences and advanced usability frames should be implemented.	
Feedback and Suggestions tracking frames	CROs should be able to suggest continuously improvements for p-medicine system or feedback/ticketing related functionalities should be available.	Feedback or Ticketing frames	

3.2.5 EUNs and use cases / scenarios

It is important to mention that use cases as well as scenarios are not different in some cases. The only difference will be related to the credentials a person has by entering the p-medicine platform. These credentials will allow him to do or not to do specific tasks that are possible within the use case.

Security issues are the same for all end users. A communication tool can be developed having several possibilities for usage. Contact to patients will not be allowed by CROs per se, as patients need to give consent.

A tool for feedback or the semantic search engine will be the same for all stakeholders but their functionality is depending on the role and rights an end user has.

Taking these into account the development of tools, services or models need to be generalized so that specific features can be used according to credentials of a user. The second point regarding generalization of tools is mentioned in paragraph 2.1.3 and fig. 2.10 and deals with fact that some tools might be able to be used as research tools or as tools for clinical decision support. This difference is based on the selection of data from multiple or a single one.

4 Context scenarios for usability testing

4.1 Overview of usability testing process

Usability plays an essential role in the whole development process of the project p-medicine. The main objective of the usability methodology[22] in the beginning of a project is to describe the task with the whole context of use of the end users. To assure that the software used in p-medicine will meet the high demands of the end users and that the platform fulfils the requirements for usability of the main target groups, the software has to be evaluated by the users throughout the development period. Taking user needs into account early in the project development can reduce implementation costs and avoid loss of time.

There are the following objectives to achieve in p-medicine:

First of all to identify the various user groups. Then to interview prospective end users to understand their task with the whole context of use to get the users' needs. With the whole context of use is meant the users' prior knowledge and qualification, his working environment, and his specific way of working. This procedure is an essential process for the usability engineer. The interviews are documented in form of context scenarios (Appendix 2). The resulted user requirements are no product properties but represent the bridge between problem and concrete solution[23] that the user is able to conduct his task with the support of the developed tools in p-medicine. The whole context of use with the task of the user will be described in detail in form of these context scenarios. It is necessary to get a common understanding of the user's task. This common understanding of the task must also have the software developer to assure a usable user interface that supports the end user to achieve his/her aim in an efficient, effective and satisfied way.

To support the patient in clinical care systems we prepared a list of relevant key questions (chapter 4.2) to get the patients needs. Ecancer produced the corresponding tables to present the patients' answers in a statistical way (Appendix 3). The evaluation of the answered questions is shown below. The information was not sufficient to write a complete context scenario of the results. Only the dialogue principles are taken into account to give a first specification of the patients' needs.

To define user needs and requirements for tools, methods and services for VPH research focused on clinical usage we have to revise the requirements in an iterative process many times, as they evolve further requirements during the development phase. It is necessary to enable the end user to work with the developed tools, so that he can conduct his task and achieve his aim in an efficient, effective and satisfied way[24].

The usability process we will use is described in D2.1. In contrast to the precursor project ACGT[25] cancer patients are also involved. For this user group the user interface must be very easy and comprehensible to use, regarding the various background knowledge of information technology and its handling.

The interviewed target groups who will use the software in their daily work are clinicians, trial managers, bioinformaticians, biostatisticians, data managers and patients. With one representative of each group interviews were taken exclusive patients. The standard key

[22] p-medicine Deliverable D2.1: State of the art review of the p-medicine environment

[23] Leitfaden Usability; available on DAkkS website (German's National Accreditation Body (former DATech) only in German language)
http://www.dakks.de/sites/default/files/71-SD-2-007_Leitfaden%20Usability%201.3.pdf

[24] Ergonomics of human-system interaction - Part 11: Guidance on Usability (ISO 9241 - 11:1996)

[25] http://eu-acgt.eu

questions, described in D2.1 were extended of some special key questions for the bioinformaticians and data managers, s. below. The other interviews were taken with the standard key questions[22]. They allow the definition of context scenarios describing and structuring user activities with the p-medicine platform.

The first interviews started in February 2011 of two target groups the bioinformatician and the data manager. The interviews of the other prospective user groups, biostatistician and clinician took place on the first progress meeting on Crete in June 2011. Another interview was taken with a biologist working in a medical hospital. His task is i.a. the management of clinical trials. For the patients there was generated a special questionnaire which was put online from the 14th of July until the 1st of August for contribution and participation (chapter 4.3). Twenty-six cancer patients participated and answered the questionnaire. This limited number of participants resulted in the limited amount of time. We got no knowledge about the patients' business and their age. From the patients' answers the usability engineer collected the user needs and tried to derive the system requirements.

The aim of the definition of the scenarios was to identify potential issues and to explicit the usage requirements and derive system requirements according to the dialogue principles as described in ISO 9241 – Part 110[26].

The interviews were documented in five context scenarios (Appendix 2) that have been sent first to the interviewees themselves for validation before the usability engineer derives the system requirements. These context scenarios will serve the basis for requirements specifications, the architecture design, and the system evaluation. Achieving a common understanding of the requirements is indeed a necessary step to enable the developer of a platform supporting efficient user activities, and user satisfaction.

In Appendix 2 the five context scenarios are listed. Each is structured in six chapters:

- introduction,
- assumptions,
- routine activities,
- special features during the working process
- organisational conditions and
- other comments to critical incidents which already occurred.

With "comments to critical incidents" we wanted to enable the user to give some feedback about the tools he/she uses and which problems occur during conducting the task. Additionally if the user has some visions how to conduct a task in an easier way, this can be mentioned here.

In the first column the user's task with the whole context of use is described. It can be read as a story of his/her daily work. With the dialogue principle in the second column the system requirements are derived, finally. The third column describes the resulted system requirements.

All these context scenarios of the various user groups of p-medicine should give the developer a common understanding of the user's task and show him which needs are essentially necessary to enable the user to achieve his/her aim in an efficient and satisfied way.

The evaluation of the cancer patients' answers (Appendix 3) is not written in form of context scenarios because of insufficient information. They are described according to the dialogue principles and the answers of the 26 patients.

[26] Ergonomics of human-system interaction - Part 110: Dialogue principles (ISO 9241 -110:2006)

The next step is to consolidate the implementation of the software tools in accordance to the requirement specification defined by the context scenarios. After the first prototypes were implemented real prospective users will have the opportunity to test the software. The first prototypes need not to have the complete functionality of the tool. It should give the user a first view of the interface and what is possible. The resulting use scenarios[22] are documented and will be described in detail in WP 15.

The following subchapters show the key questions for the bioinformaticians / biostatisticians, data managers and the questionnaire for cancer patients. The standard key questions are described in D2.1. The various context scenarios of one of a representative of the five user groups and the answers of the patients are listed in Appendix 2 and 3. For the developer it is necessary to read the user stories in the context scenarios with the derived system requirements very carefully to understand the users' task and daily work. Only with this knowledge in mind he/she is able to develop a usable and easy to use interface for the various user groups in p-medicine. During the whole developmental phase the developer should have a good cllaboration to the usability engineer and the end-user who has to be enabled to achieve his/her aim in an efficient, effective and satisfied way.

4.2 Key questions for describing and structuring user performance in context

These questions are adapted to the work of **bioinformaticians**, **biostatisticians** and **data managers**.

Introduction	1. Describe your work in one or two sentences. 2. From which tasks is your work composed (list typical key tasks, which are time-consuming or frequently occurring or very important)? Which of these key tasks the software should support? 3. How work is organised (e.g. as various tasks, as a sequence of tasks, as repetitive single task)?
Assumptions	4. What kind of qualification is needed for performing the tasks (for task completion / for using software)? What kinds of skills are missing? 5. Who or which event decides what to do? (Who selects your jobs? Jobs are performed autonomously, work is divided, data is needed from colleagues or external sources.) 6. Which software do you use for your work? What kinds of components are selected for your workflow? 7. From where are these components, repository, colleagues, self-made? 8. Are there standardized components, standardized workflows or sub-processes? Which of them are missing, which are desired additionally? 9. Which data sources are you using in your work? 10. Which information do you get from these sources and how do you store and annotate them?

	11. At which level are tools, components, workflows and data reused and transferred?
Routine activities	12. Which working steps are executed? 13. Which working steps are performed repeatedly? (Automated execution desired / necessary)? 14. Which working steps are executed by the software? Can you control the autonomous process / is control allowed / desired / required? 15. How do you concretely combine all information you got from different data sources (e.g. gene annotation, SNPs(Single Nucleotide Polymorphism), medical literature, public data repositories (GEO, ArrayExpress, SRA), clinical databases, …) to produce results? Which structure can be evolving? 16. What are the final products of your work? 17. Are several users working in parallel on the same object (e.g. transaction, file, document, data record)? 18. Is there a defined sequence of working steps? If so, how is it composed? (more flexibility needed / desired?) 19. Which overview do you have with respect to the overall workflow? 20. Which are the results / partial results and how are they used / continued? 21. Which kind of feedback do you get concerning your working results and effects?
Special features during the working process	22. How do you work and share results with your co-workers? 23. How could this be done more easily? 24. Which kind of interruptions appear? Why, when? (organisational / social / technical)? 25. How are mistakes reported back and solved (organisational / social / technical)? 26. Which important special cases have to be considered (respectively cross the user's mind spontaneously; e.g. division of work / collaboration)? 27. How can an eScience solution be setup from scratch? 28. Which phases exist when creating eScience workflows or analysis processes?

Organisational conditions	29. Which organisational aims are defined for the working tasks? 30. Are there mechanisms to control the efficiency of work? (If so, which ones? Are they necessary?) 31. Which kinds of changes are made to existing workflows? 32. Which is the most time consuming task (processing)? 33. Which is the most time consuming phase (construction)? 34. How long does it take to select tasks, tools or components? 35. How long does it take to configure or setup tasks, tools or components? 36. How long does it take to connect tasks, tools or components to others? 37. Which changes are expected or desired by the user considering the performance of work? Are there any suggestions from you? Visions!!! 38. Which results / working steps affect third parties (e.g. customers) directly? And which are the consequences? 39. Which are the stress factors and how are they handled? 40. What are the most annoying features in each of the software you use? 41. What would you change to make the software (workflow) more convenient to use and facilitate your work?
Other comments to critical incidents which already occurred	Put examples in here, when the interviewee tells something about critical incidents concerning the software during the interview. Usually such problems should be analysed within use scenarios. What can be done to make your work easier? Ideas & Visions for improving features, which are time-consuming & difficult to execute!

4.3 Cancer patients' questionnaire

The following questionnaire was prepared by ecancer. It should give a better understanding of what is needed from a new decision support tool that helps patients become more involved in decisions about their treatment. This survey was made available online in Bristol. Cancer patients getting this survey were informed in a talk with their doctors about the following points:

1. All the information they provide will be used to ensure that the designers of the tool will meet their needs and those of other patients.
2. All the data will be anonymized and the only people who will be allowed to view this information will be doctors and scientists who are involved in the project to create the tool.
3. The new tool will give cancer patients information about different aspects of their cancer. This will prepare them in a way to discuss their treatment options with their doctor much better.

After this consultation patients can become more involved in making decisions about their care, if they want. This process is called “Patient Empowerment”.

4.3.1 Questions of the survey

1. When discussing your treatment options with your doctor, please rate how important it is to have information on the following:
    a. The latest new treatment
    b. The treatment options available to me (in my particular circumstances)
    c. How effective the different suggested treatments are
    d. The survival rates of the suggested treatments
    e. Quality of life after the suggested treatments
    f. The side effects of the suggested treatments
    g. Opportunities to be involved in a clinical trial
    h. My disease
    i. The best questions to ask the doctor for the most relevant information
    j. How often the hospital treats the disease
    k. How successful my hospital has been in treating patients like me
    l. How often my doctor treats this disease
2. How important is it to be able to access all of the above information on the internet?
3. How important is it to have a record of what you have discussed with your doctor to refer back to at a later date?
4. After speaking with your doctor, how much of the information you are given do you usually understand?
5. How important do you think it is to be given printed information explaining more detail about what the doctor has said?
6. How important is it to be able to communicate with other patients who are affected by the same illness as you?
7. How important do you think it is that your doctor obtains information about your psychological well-being when discussing your diagnosis?
8. How likely would you be to join an internet based social media network of patients (like facebook)?
9. How important do you think it is to be given extra information at the following times:
    a. Before speaking to your doctor
    b. When speaking to your doctor
    c. When you return home after speaking to your doctor
10. Do you have access to a computer at home?
11. How often do you use the internet on any computer?
12. How often do you use email on any computer?
13. How often do you use Microsoft Word on any computer?
14. How often do you access the internet through your mobile phone?
15. If you own a smart phone, how often do you download applications?

The answers of the twenty-six cancer patients are prepared in tables by ecancer and are listed in Appendix 3

4.3.2 The resulting patients' needs

The research questionnaire organized by ecancer was set up online in Bristol. The answers of 26 cancer patients are not restricted of breast cancer patients.

All provided answers and comments were handled anonymously.

We don't know how old the cancer patients are and what kind of cancer disease they have. They all have access to a computer at home. 20 of them use the computer very often. For the usability engineer it is important to know what the task of the cancer patients is, which are the interested information to know about her disease?

The main tasks can be described to get information about:

- my disease
- the opportunities to be involved in a clinical trial
- the quality of life after the suggested treatments
- the side effects of the suggested treatment
- the latest new treatments
- the treatment options available to me (in my particular circumstances)
- how effective the different suggested treatments are
- the survival rates of the suggested treatments
- the best questions to ask the doctor for most relevant information
- how often the hospital treats this disease
- how successful my hospital has been treating patients like me
- how often my doctor treats this disease
- who are affected by the same illness as me

From the different background knowledge, profession and technical experience of the cancer patients it is absolutely necessary to develop the user interface in a very easy, comprehensible and self-descriptive way.

The dialogue principles of the ISO 9241 – 110 illustrate an approach to identify the most important usability aspects for the interaction of the user with the dialogue system. The adaptability of each principle depends on the user group and their context of use.

Suitability for the task is to support the user to conduct her task, i.e. if the functionality and the dialogue are based on the characteristic user's task rather than on the used technology for task completion. The dialogue should present only the information that is necessary for the user to conduct the task successfully. For the system means this to provide information concerning the disease or treatment in an easy and comprehensible way.

The registration process for getting more information about the own treatment and its progress should be easy and intuitive performed. It should be self-descriptive and controllable. All medical expressions should be explained in the user's language. The shortcuts should be also explained in a comprehensible way.

The patient is interested in more information about her own disease and the possible treatments and the side effects. This has to be presented in a clear way. The dialogue steps should be adapted to the work processes. The user has to know in each step where she is and how to do the next step.

If the patient got side effects from medical products or from chemotherapy she would be interested to have the possibility to exchange experiences with other cancer patients. It could also be helpful for her emotional / psychological support. A forum would be a possible

solution for this request where cancer patients could discuss their experiences, problems and treatment progress with other cancer patients in the same situation.

Cancer patients would like to inform about the competent rehabilitation centres. A list of competent centres could be presented and described so that the patient can chose the best one for her recovering. The system could present a list of specialists for the different kind of cancer as well as the best questions to ask the doctor for most relevant information.

All information the cancer patient does not get from the treated doctor should be presented in a clear and easy way to get answers on outstanding issues. There should also be a list of treatment options described in a comprehensible way. It should be self-descriptive.

In some situations it could be helpful for the cancer patient to have the relevant information about her treatment on paper. This would require a print button. Recording the talk with the treated doctor could also be helpful for the patient to hear everything once more at home.

There should also be the possibility to get information from cancer patients in other clinics and their treatment progress.

A list of specialists for the different kind of cancer should be available via the system.

4.3.3 Recommendations for the presentation of all relevant information

The quality of information[27] depends on the following:

- Clarity of information, i.e. the information content is conveyed quickly and accurately
- Discriminability, i.e. the presented information can be distinguished accurately
- Conciseness, i.e. only the necessary information is given to the users
- Consistency, i.e. the same information is presented in the same way through the whole application according to the user's expectation
- Detectability, i.e. user's attention is directed towards information required
- Legibility, i.e. all information is easy to read, clear structure of the content
- Comprehensibility, i.e. meaning is clearly understandable, unambiguous, interpretable and recognizable

The information should be presented in that way to enable the user to perform her task of getting better informed about the own disease, its treatment and the adverse side effects efficiently, effectively and with satisfaction. The structure of the information should be clearly arranged. To use self-descriptive pictures could help the user for better understanding.

The user would like to read and understand all important information that she expected. She would like to overlook the details quickly, to collect all important issues and not get distracted from basically through unnecessary information.

The patient should only get relevant information for her disease and treatment options. The dialogue should not present information that is irrelevant for conducting her task. The information should be readable and comprehensible. It should be consistent, i.e. all expressions should be used in the same way. The system should only use the vocabulary the user is familiar with or that the patient uses in relation to her knowledge and experience.

In all steps the user should be supported via action guiding information. The user knows in any situation where he/she is, from where she came and which steps is the next one. Very important information should be presented at a high level so that the user has not to search

[27] Ergonomics of human-system interaction - Part 12: Representation of Information (ISO 9241 – 2:1998)

in deeper levels and lose the orientation. Direct feedback from the system on the patient's actions is essential.

The user should be supported to detect input errors and to avoid them. If mistakes occur anyway, they should be described in the user's language to facilitate the elimination of such errors.

The user should have access only on his/her treatment data. He/she has no right to see other patient data.

4.4 Conclusion regarding ISO 9241

Considering all end-user groups a common portal will be developed to enable clinicians, bio-researchers, data managers and at least patients to use the software for conducting their task and achieve their aim in an efficient, effective and satisfied way. This can only be realized when the requirements of all user groups are taken into consideration. As part of a context analysis the actual usage requirements have been elevated with the prospective user groups. Interviews have been conducted with one of the various user groups, bioinformatician, biostatistician, data manager, biologist in the role of clinical trial manager and clinician/chairman. The described key questions help the usability engineer to get all relevant information about the user's task with the whole context of use.

On the basis of the existing context scenarios (Appendix 2) the hidden usage requirements are identified and described. These requirements are no product features or functionalities, but represent the bridge between the way of looking at a problem and a concrete solution[28]. The users' needs serve as a common understanding of the task and its context of use for the developer. With this knowledge the developer has the ability to develop a first prototype. For this prototype it is not essential to have the full functionality, it should only give the user a first view.

In the development process the task analysis will be the first step, the second step is the interaction design and the interface design will be the third one. This sequence reflects the steps in the software development process when usability will be applied.

In p-medicine the various user groups have different tasks and it is not useful to collect everything in a short summary. The developers of the different tools and portal have to read the written context scenarios (Appendix 2) of the various user groups very carefully to get a common understanding of their tasks when considering the ISO 9241, in particular the seven dialogue principles of part 110 and part 12 for the user interface design.

The user will not ask the question "Does it look great?" but the answer to the question will be "Which information has to be displayed to the user in which form and at which time?" The corresponding principles with recommendations are described in ISO 9241-12. With all this information a rough user interface in form of a first prototype can be designed. The developed product is now no random product but it builds consistently on validated usage requirements. The user will always prefer a more feasible than a user friendly product.

The developer of the portal has to consider that the registration process should be conducted in a very simple, self-descriptive and clear structured way. The user should be guided through the registration process as well as through the whole execution of his/her task until successful completion without loss of time.

[28] Leitfaden Usability; available on DAkkS website (German's National Accreditation Body (former DATech) only in German language) http://www.dakks.de/sites/default/files/71-SD-2-007_Leitfaden%20Usability%201.3.pdf

5 VPH Scenarios

5.1 VPH Toolbox Scenario

The Virtual Physiological Human (VPH) is synonymous with a programme in computational biomedicine, which aims to develop a framework of methods and technologies to investigate the human body as a whole.[29] The goal of the VPH is to achieve a more efficient and effective twenty-first century healthcare system and to create new economic opportunities for European healthcare industries. "The vision of a 'digital me' that contains all my healthcare information, safely managed for access by the various biomedical professionals with my approval, communicated with all my wearable and implanted technology to constantly monitor my health status and informing me, my family and friends, or my healthcare providers of alarming events, supporting the collaboration of various specialists around my complex systemic diseases, and used with all my data to predict the future development of my health in order to facilitate disease prevention and a fully self-aware lifestyle, is a powerful vision."[30]

Additionally, VPH is a major European e-Science initiative intended to support the development of patient-specific computer models and their application in personalized and predictive healthcare. The VPH Network of Excellence (VPH-NoE)[31] project is tasked with facilitating interaction between the various VPH projects and addressing issues of common concern. A key deliverable is the 'VPH ToolKit' - a collection of tools, methodologies and services to support and enable VPH research, integrating and extending existing work across Europe towards greater interoperability and sustainability.[32]

Researchers from the VPH-NoE project concluded that a single monolithic 'toolkit' is incapable of addressing the needs of the VPH. Rather, the VPH ToolKit should be considered more as a 'toolbox' of relevant technologies, interacting around a common set of standards. The latter apply as well to the information used by tools, including any data and the VPH models themselves, and also to the naming and categorising of entities and concepts involved.

Currently the VPH ToolKit encompasses many elements, reflecting the multi-faceted arena of VPH research, and some of the main developments and reported activities are related to:

- **Standards**: models, data, ontologies, and infrastructure interoperability

  VPH-NoE's standards working group (VPH-SWG) has been established, which is primarily coordinated by VPH-NoE stakeholders, and works in consultation with the broader VPH research community (academic, industrial, and clinical).

  - Ontology standards
  - Data standards
  - Modelling standards
  - Infrastructure interoperability standards

[29] Coveney PV, Diaz V, Hunter P, Kohl P, and Viceconti M, The Virtual Physiological Human, Interface Focus, June 6, 2011 1:281-285

[30] Hunter P, Coveney PV et al.: A vision and strategy for the virtual physiological human in 2010 and beyond. Phil. Trans. R. Soc. A (2010) 368, 2595–2614

[31] http://www.vph-noe.eu

[32] Cooper J, Cervenansky F, Fabritiis GD, Fenner J, Friboulet D, Giorgino T, Manos S, Martelli Y, Villà-Freixa J, Zasada S, Lloyd S, McCormack K, and Coveney PV: The Virtual Physiological Human ToolKit, Phil. Trans. R. Soc. A (2010) 368 , 3925-3936.

- **Imaging tools**

  One of the main objectives of VPH-NoE's imaging subgroup is to develop an online help tool called GUIDE (Guidelines for Image Development Environment) which will be part of the VPH ToolKit portal. The purpose of this tool is to guide users - developers, researchers, and clinicians - in choosing the proper biomedical image analysis tools for their work (software, libraries, etc.), and to provide support enabling their sharing and open use.

- **High performance computing**

  Computational infrastructure within the EU includes EGEE[33] providing low-end clusters, and DEISA[34] providing supercomputer class resources. The VPH-NoE has obtained access to both of these infrastructures for VPH-I researchers; to EGEE through the EGEE Biomedical Virtual Organisation, and to DEISA through a 'Virtual Community' allocation. The Partnership for Advanced Computing in Europe, PRACE, is a unique persistent pan-European Research Infrastructure for High Performance Computing (HPC). PRACE forms the top level of the European HPC ecosystem. PRACE-project is funded in part by the EU's 7th Framework Programme[35]. P-medicine will get access to PRACE if needed.

- **VHP ToolKit portal website[36]**

  ToolKit portal website is anticipated to be a key resource for the community.

The p-medicine project will benefit by focusing on synergies and frames for (re)using, implementing, exploiting, and integrating VPH-NoE's achievements and realisations. Besides this p-medicine will also store all developed tools, services and models in the VPH-Toolbox. Both usage of existing and storage of developed tools are part of the VPH Toolbox Scenario. It is the intention to build up an interactive collaboration within the VPH NoE to harmonize tools, methods and services in interlinking with the VPH Toolkit and/or VPH Toolbox. This links directly to task 2.3 of WP2 (User Needs and Requirements) named "User requirements and specifications for the collaboration of the p-medicine environment with other research infrastructure initiatives (VPH NoE, ECRIN, BBMRI, ENCCA, ESFRI, DEISA, etc.) data management systems".

[33] http://www.eu-egee.org
[34] http://www.deisa.eu
[35] http://www.prace-project.eu/
[36] http://toolkit.vph-noe.eu

6 Security Scenarios

Introduction

Security needs to be available in most of the components of the p-medicine Platform. There are some important security components that need to be implemented to offer a reliable and secure system. First a mechanism is needed that allows the users to authenticate themselves by providing personal credentials. In this way the users can confirm their identity on the different sites/services of the platform. Another important part of security is access control. A user may only see and manipulate resources of the p-medicine on which he has access rights. Other security components include: encrypted storage of data, pseudonymisation of patients and safe transmission of data (confidentiality and integrity).

6.1 Single Sign-on Scenario

The end-user needs to authenticate himself on different sites/services of the p-medicine Platform. An architecture where a user needs to provide his credentials for each site/service separately is not sustainable and not user-friendly. A better architecture uses a central Identity Provider (IdP), explained the following use case. This use case has two possible flows: Normal flow and alternative flow.

6.2 Single Sign-out Scenario

A user that is authenticated on one or more sites/services using SSO, may want to logout from all this sites/services. This logout should be user-friendly, making it possible to logout from all the sites/services in one simple action (Single Sign-Out). The steps that are needed for Single Sign-Out are explained in the following use case.

6.3 Access Rights Scenario

The sites/services of the p-medicine Platform are protected access control, meaning that every user needs to have access rights to view/manipulate resources of these sites/services. How these access rights are granted is explained in the next use case. *Note: this is currently a placeholder.*

6.4 User Enrolment Scenario

The registration of a user on a particular site/service of the p-medicine platform is not straightforward. A local site/service user account is not sufficient if Single Sign-On is used, an extra central IdP account is needed. These accounts need also to be linked. The following use case gives a vision on how the user enrolment can work.

7 Clinical Scenarios

Introduction

As p-medicine is clinically driven the clinical scenarios are centrally for the project. ALL, Breast Cancer and Nephroblastoma will serve as test cases for the p-medicine platform. The developed tools will be disease specific but they will be built in a way that they can easily be transferred to other cancer types and even to other domains. This will be made possible by the modular way tools are built and by keeping aspects of generalization in mind.

7.1 Nephroblastoma

Wilms tumour or Nephroblastoma is the second most common intraabdominal cancer of childhood and the fifth most common paediatric malignancy overall. It represents approximately six percent of all paediatric cancers and accounts for more than 95% of all tumours of the kidney in the paediatric age group.[37],[38],[39].

From the perspective of the 'Patients' as end-users, patients with Nephroblastoma are children with no access to the p-medicine platform (PHR p-medicine). This particularity needs to be taken into account for developmental strategies. The same is the case for acute lymphoplastic leukaemia (ALL, see 4.3). In both diseases the p-medicine platform has to accept new user registrations and data submission frames from the parents of patients with Nephroblastoma diagnosis. This needs to be considered on Patient Consent and Patient Empowerment scenarios as well.

Of particular interest in nephroblastoma is the Oncosimulator scenario starting during the lifetime of ACGT (Advancing Clinico-Genomic Trials) - an Integrated Project, partly funded by the EC (FP6-2005-IST-026996)[40]. The research has been focused on elaborating a state-of-art concept as an integrated software system simulating in vivo tumour response to therapeutic modalities within the clinical trial environment. The aim is to support clinical decision making in individual patients by predicting response to preoperative chemotherapy. In p-medicine the Oncosimulator will be refined and optimized by using more data from molecular biology. The main research findings refer to the technology of the system, the clinical requirements and the types of medical data needed[41]. Other use cases will be presented as well.

The *Oncosimulator* is at the same time a concept of multilevel integrative cancer biology, a complex algorithmic construct, a biomedical engineering system and eventually in the future a clinical tool which primarily aims at supporting the clinician in the process of optimizing cancer treatment in the patient individualized context through conducting experiments *in silico* i.e. on the computer. Additionally it is a platform for simulating, investigating better understanding and exploring the *natural phenomenon* of cancer, supporting the design and

[37] Pastore G, Znaor A, Spreafico F, et al. Malignant renal tumours incidence and survival in European children (1978–1997): report from the Automated Childhood Cancer Information System project. Eur J Cancer. 2006;42:2103–2114.

[38] Breslow N, Olshan A, Beckwith JB, et al. Epidemiology of Wilms tumor. Med Pediatr Oncol. 1993;21:172–181.

[39] Davidoff A, WILMS TUMOR, Curr Opin Pediatr. 2009 June; 21(3): 357–364.

[40] http://eu-acgt.org July, 2011

[41] Graf N, Hoppe A, Georgiadi E, Bellemann R, Desmedt C, Dionysiou D, Erdt M, Jacques J, Kolokotroni E, Lunzer A, Tsiknakis M, Stamatakos G: 'In Silico' oncology for clinical decision-making in the context of nephroblastoma. Klin Pädiatr 221:141-149, 2009.

interpretation of clinicogenomic trials and finally training doctors, researchers and interested patients alike[42,43,44].

A synoptic outline of the clinical utilization of a specific version of the *Oncosimulator*, as envisaged to take place following an eventually successful completion of its clinical adaptation, optimization and validation process is provided in the form of steps (Figure 5.1), which are described in detail in the DOW of WP12.

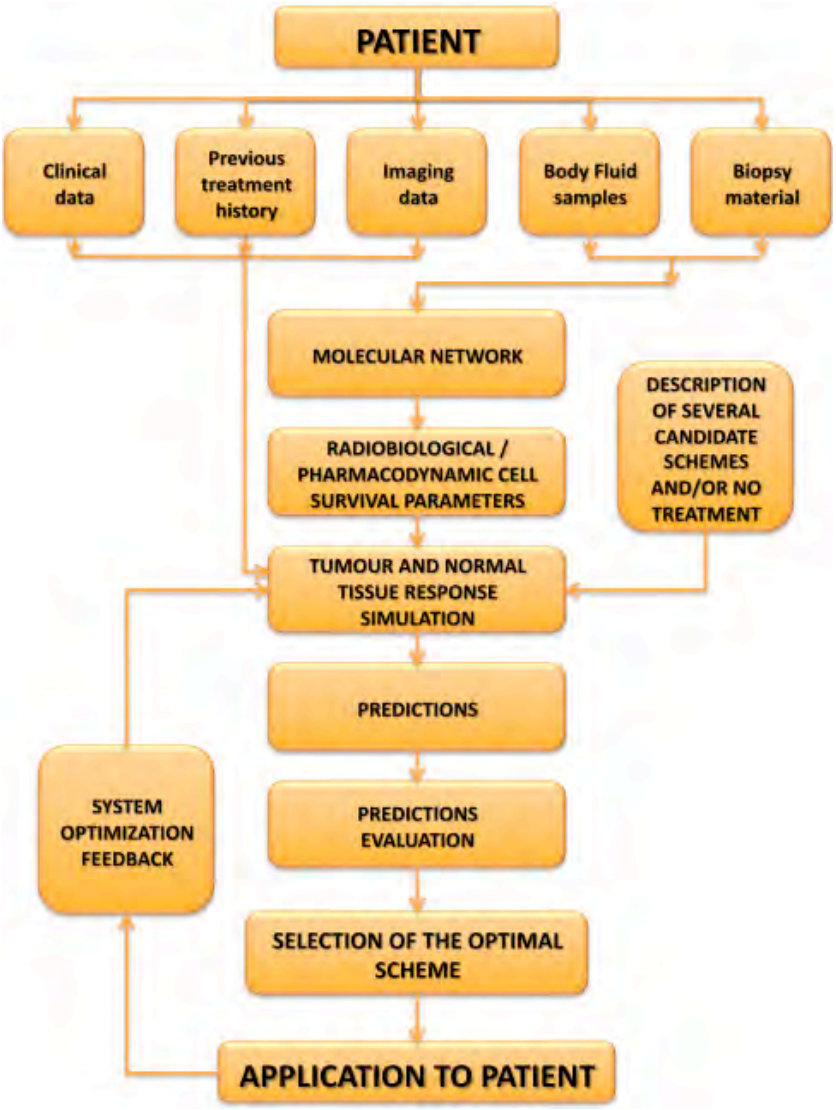


Fig 5.1: A synoptic outline of the Oncosimulator

[42] Stamatakos, G. S. and Uzunoglu, N. 2006b. Computer simulation of tumour response to therapy. In S. Nagl Ed. Cancer Bioinformatics: from therapy design to treatment. John Wiley & Sons Ltd, Chichester,UK. pp.109-125

[43] Stamatakos G.S., D.D. Dionysiou, N.M. Graf, N.A. Sofra, C. Desmedt, A. Hoppe, N. Uzunoglu and M. Tsiknakis. 2007a. The Oncosimulator: a multilevel, clinically oriented simulation system of tumor growth and organism response to therapeutic schemes. Towards the clinical evaluation of in silico oncology. Proc 29th Annual Intern Conf IEEE EMBS. Cite Internationale, Lyon, France Aug 23-26. SuB07.1: 6628-6631

[44] Graf, N., A. Hoppe , E. Georgiadi, R. Belleman, C. Desmedt, D. Dionysiou, M. Erdt , J. Jacques, E. Kolokotroni, A. Lunzer, M. Tsiknakis and G. Stamatakos. 2009. "In silico oncology" for clinical decision making in the context of nephroblastoma. Klin Paediatr 221: 141-149

7.1.1 Pathway Scenario

In the pathway scenario clinical, molecular and open source data are integrated to find those pathways that are mainly disrupted in Nephroblastoma in general, or in specific subtype of nephroblastoma, or in single patients. In single patients this finding can help to select specific drugs for the treatment of a specific patient and can serve as a basis for a decision support tool. A description of the scenario is given here:

7.1.2 Imaging Scenario

The imaging scenario will have two different features. DICOM data of patients with nephroblastoma need to stored in the data warehouse for further analysis and these imaging data need to be post-processed for usages in the Oncosimulator.

7.1.3 (Severe) Adverse Event ((S)AE) Prediction Scenario

The prediction of an SAE within a clinical trial would help to make treatment safer for patients. By extracting an individual patient profile from his data including pharmacogenomics data (if available) and performing data mining in literature, SAE/SUSAR databanks and clinical trials, in which the specific drug is used, the individual risk of possible (S)AEs will be predicted. Despite the fact that this use case deals with patients with nephroblastoma, it can be generalized to any other disease, if the disease domain is taken into consideration during data mining.

7.1.4 Tumour Marker Scenario

There are no serum tumour markers known in nephroblastoma predicting outcome or specific subtypes. This use case will define a pattern of miRNAs, tumour specific autoantibodies and other serum proteins as specific markers for nephroblastoma.

7.1.5 Oncosimulator Scenario

The development of the Oncosimulator for nephroblastoma did start in ACGT. IN p-medicine it will be further refined and used in a larger set of patients.

7.2 Breast Cancer

Breast cancer is the most common cancer in women worldwide, comprising 16% of all female cancers. It is estimated that 519 000 women died in 2004 due to breast cancer, and although breast cancer is thought to be a disease of the developed world, a majority (69%) of all breast cancer deaths occurs in developing countries[45].

p-medicine project will focus (in close collaboration with project partners) in special on targeted drugs, pathway and oncosimulator scenarios, nevertheless, one of the key message of WHO is: “Early detection in order to improve breast cancer outcome and survival remains the cornerstone of breast cancer control.”[46] Additionally, one of the WHO’s proposed actions

[45] WHO Global Burden of Disease, 2004

[46] http://www.who.int/cancer/detection/breastcancer/en/index.html July, 2011

for member states is the reorientation and strengthening of health systems by implementing and monitoring cost-effective approaches for the early detection of breast cancer.[47] It suggests that p-medicine platform due to its modular infrastructure and powerful tools could focus as well on yearly breast cancer detection. It is of high importance in special by taking into account that breast cancer treatment; prognosis and survival rate varies greatly depending on cancer type and staging.

This deliverable will not cover all clinical aspects related to breast cancer pathophysiology, treatment and/or genetic pathways but some of the major particularities (will be described in further deliverables) are presented below:

- Breast cancer staging using the TNM system;
- Breast cancer receptors status (Estrogen Receptor (ER), Progesterone Receptor (PR), and HER2/neu)
- Genome Mutations (p53, BRCA1, BRCA2) and Breast Cancer pathways
- Breast Cancer treatment and/or related clinical trials:
    - Task 9.2: Clinical trials;
    - Subtask 9.2.2: Breast Cancer phase II trial (Bevacizumab trial -1);
    - Subtask 9.2.3: Breast Cancer phase II pharmacodynamic trial (Bevacizumab trial-2)
    - Subtask 9.2.4: Breast Cancer (Circulating tumour cells (CTCs) trial)
    - Subtask 9.2.5: Breast Cancer Stem cell models
- Breast Cancer VPH Modelling and the Integrated Oncosimulator:
    - Task 12.1: Development of the Breast Cancer p-medicine Oncosimulator models
    - Task 12.2: Clinical adaptation, optimization and partial validation of the Oncosimulator models

7.2.1 Breast Cancer Scenarios

The Breast Cancer scenarios will be developed in close collaboration with p-medicine project partners enrolled in the breast cancer trials within WP12 (VPH modelling and integrated Oncosimulator). The specific scenario suggested for the breast cancer VPH will be to model the response to preoperative therapy using the available trials. This will be done within WP12 in two phases:

- Response to anti-angiogenic treatment
- Response to combined modalities of biological drugs with standard cytotoxic and/or hormonal therapies

The first phases will be the primary aim and will be validated within the duration of the project using the existing Bevacizumab phase II trials (Bevacizumab 1 and 2 trials, please explore WP9 for further information). Both of these trials address the same drug and the data from the trials will be merged in a single meta-entity to be used tuning and validation of the Oncosimulator breast cancer model. Thus, the primary aim would be to have a solid and validated modelling of angiogenesis and response to anti-angiogenic drugs. Furthermore,

[47] 2008-2013 Action Plan for the Global Strategy for the Prevention and Control of Noncommunicable Diseases, WHO

due to the high number of trials in breast cancer, we will explore the possibility of validating further combined therapies models using large-scale data-mining of published CTs. This will be done in collaboration with partners responsible for WP 7 and WP 11.

7.2.2 Oncosimulator Scenario

It is the intention of the Oncosimulator to predict the likely response of a given patient's breast cancer to one or more candidate treatment schemes while toxicological limitations are taken into account.

7.3 Acute Lymphoblastic Leukaemia

Leukaemia is the most common childhood malignancy. It accounts for 30% of all cancers diagnosed in children under 15 years of age in industrialized countries. Around 2000, the average incidence for this age group in the European Region was 46.7 cases per million per year, with a slightly lower level in eastern than in western European countries. European population-based cancer registries show an average increase in the incidence of childhood leukaemia of 0.7% per year between 1970 and 1999.[48]

There are various types of leukaemia with different geographical distribution patterns. In Europe, acute lymphoblastic leukaemia (ALL) accounts for around 80% of leukaemia among children aged 0-14 years.[49] ALL has an annual incidence of up to 40 cases per million children among industrialized western European countries and up to 30-35 cases per million in eastern European countries, but fewer than 20 per million in sub-Saharan Africa.[50] In developed countries, more than 80% of ALL is of the precursor B-cell subtype that is responsible for the pronounced peak of incidence in early childhood and largely accounts for the observed variation in the total incidence of childhood leukaemia among countries. [51] [52]

This deliverable is not focused on providing detailed and informative description of ALL clinical aspects, pathophysiology, treatment and/or genetic pathways but the major particularities related to the mutations of genes regulating B-lymphoid development in ALL are presented in the bellow table.

[48] WHO, INCIDENCE OF CHILDHOOD LEUKAEMIA, FACT SHEET 4.1, December 2009, CODE: RPG4_Rad_E1

[49] Coebergh J-W et al. Leukaemia incidence and survival in children and adolescents in Europe during 1978-1997. Report from the Automated Childhood Cancer Information System project. European Journal of Cancer, 2006, 42:2019-2036

[50] Parkin DM et al., eds. International incidence of childhood cancer, Vol. II. Lyon, International Agency for Research on Cancer, 1998 (IARC Scientific Publications No. 144)

[51] Greaves MF et al. Geographical distribution of acute lymphoblastic leukaemia subtypes: second report of the collaborative group study. Leukemia, 1993, 7:27-34.

[52] Stiller C, ed. Childhood cancer in Britain: incidence, survival, mortality. Oxford, Oxford University Press, 2007.

Selected recurring regions of DNA copy number alteration in pediatric ALL[53]

Cytoband	Start (Mb)	End (Mb)	Size (Mb)	B-ALL N (%)	T-ALL N (%)	Gene(s) in region
Deletions						
1p33	47.440	47.479	0.039	0	3 (6.0)	*TAL1*
2p21	43.337	43.624	0.287	2 (1.04)	1 (2.0)	*THADA*
3p14.2	60.064	60.318	0.254	8 (4.17)	0	*FHIT*
3q13.2	113.538	113.686	0.148	13 (6.77)	0	*CD200, BTLA*
3q26.32	Various			7 (3.13)	0	*TBL1XR1*
4q25	109.393	109.442	0.049	3 (1.56)	4 (8.0)	*LEF1*
4q31.23	150.055	150.200	0.145	6 (3.13)	1 (2.0)	None; telomeric to *NR3C2*
5q31.3	142.760	142.847	0.087	9 (4.69)	3 (6.0)	*NR3C1, LOC389335*
5q33.3	Various			8 (4.17)	3 (6.0)	*EBF1*
6p22.22	26.345	26.368	0.023	13 (6.77)	0	*HIST1H4F, HIST1H4G, HIST1H3F*, *HIST1H2BH*
6q16.2-3	99.852	102.492	2.640	10 (5.21)	5 (10)	16 genes including *CCNC*
6q21	109.347	109.435	0.088	11 (5.73)	4 (8.0)	*ARMC2, SESN1*
7p12.2	50.193	50.241	0.048	17 (8.85)	1 (2.0)	*IKZF1* (Ikaros)
8q12.1	60.195	60.289	0.094	7 (3.65)	0	Immediately 5′ *TOX*
9p21.3	Various			65 (33.85)	36 (72.0)	*CDKN2A*
9p13.2	Various			57 (29.69)	5 (10)**	*PAX5*
10q23.31	89.666	89.728	0.062	0	3 (6.0)	*PTEN*
10q24.1	97.879	98.057	0.178	2 (1.04)	0	*BLNK*
10q25.1	111.772	111.850	0.078	9 (4.69)	0	*ADD3*
11p13	33.874	34.029	0.155	1 (0.52)	4 (8.0)	5′ of *LMO2*
11p12	36.575	36.583	0.008	4 (2.08)	2 (4.0)	*RAG2, LOC119710*
12p13.2	Various	11.808	0.020	51 (26.56)	4 (8.0)	*ETV6*
12q21.33	90.786	91.039	0.253	13 (6.77)	0	3′ of *BTG1*
13q14.11	43.758	43.895	0.137	10 (5.21)	3 (6.0)	*C13orf21, LOC400128*
13q14.2	47.885	47.968	0.083	9 (4.69)	6 (12.0)	*RB1*
13q14.2-3	49.471	50.360	0.889	12 (6.25)	3 (6.0)	Includes *MIRN16-1, MIRN15A*,
15q15.1	39.045	39.837	0.792	6 (3.13)	0	18 genes including *LTK* and *MIRN626*
17q11.2	26.090	26.259	0.169	4 (2.08)	2 (4.0)	7 genes including NF1
17q21.1	35.185	35.230	0.045	3 (1.56)	0	*IKZF3* (*ZNFN1A3*, Aiolos)
19p13.3	0.229	1.531	1.302	17 (8.85)	0	*TCF3* to 19ptel
20p12.1	10.370	10.405	0.035	9 (4.69)	1 (2.0)	*C20orf94*
21q22.12	35.350	35.354	0.004	3 (1.56)	0	Immediately distal to *RUNX1*

[53] Mullighan CG, Downing JR, Global Genomic Characterization of Acute Lymphoblastic Leukemia, Semin Hematol. 2009 January; 46(1): 3-15.

21q22.2	38.706	38.729	0.023	5 (2.60)	0	*ERG*
Amplifications						
1q23.3-q44	161.491	qtel	81.326	16 (8.33)	0	*PBX1*to 1qtel
6q23.3	135.556	135.714	0.158	0	5 (10)	*MYB, MIRN548A2, AHI1*
9q34.12-q34.3	130.687	qtel	7.676	3 (1.56)	0	155 genes telomeric of *ABL1*, including 3′ region of *ABL1*
21q22.11-q22.12	32.896	35.199	2.303	6 (3.125)	0	33 genes including *RUNX1*
22q11.1-q11.23	ptel	21.888	21.888	3 (1.56)	0	277 genes telomeric (5′) of *BCR*, including 5′ region of *BCR*

7.3.1 Oncosimulator Scenario

In p-medicine project Christian Albrecht University (CAU) is taking part mainly in WP9 and WP12 and deals with the user requirements for p-medicine from a clinical perspective. In addition it will provide clinical trial and care data that will be used VPH modelling and decision support. CAU will also be a pilot site for validating these tools.

In view of the growing depth of information at different levels in ALL, new approaches that go beyond the statistical approaches currently applied in data mining may be helpful in gaining new perspectives on treatment strategies for clinical application especially in those patients with a dismal response to treatment. In particular, the increasing dimensionality and complexity of available clinical and genetic/genomic data demands more comprehensive solutions in order to resolve the bottleneck of data interpretation.

Therefore, in this task two scenarios have been chosen for VPH modelling in childhood ALL:

1) an MRD and
2) a disease recurrence scenario. For this purpose, the following data will be made available for three different patient groups from trial ALL-BFM 2000 (data on basic characteristics at diagnosis, treatment, response and outcome, only, are available for more than 4000 patients):

Data of a representative cohort of 664 patients will be used:

1) Basic data:

   gender, age at diagnosis, white blood cell count at diagnosis, blood blast count, hemoglobin levels and platelet counts at diagnosis, FAB classification, complete immunophenotyping data, ploidy status, status for prognostic relevant chromosomal translocations (ETV6/RUNX1, BCR/ABL, MLL/AF4, E2A/PBX1), percentage of bone marrow blasts, extramedullary disease (CNS, testis, and others).
2) Treatment data:

   risk group stratification, cumulative drug doses, information on HSCT and cranial irradiation, information on time frame for the application of treatment phases.
3) Response data:

prednisone response, blast percentages in the bone marrow on treatment days 15 and 33, MRD analyses on treatment days 33 and 78.

4) Outcome data:

   relapse, treatment-related mortality, secondary malignancy.

5) Gene expression data:

   low-density array of 95 genes previously associated with treatment response and/or outcome.

7.3.2 Biobank scenario

It has been identified that the CRIP concept[54] seems to be a good solution to integrate and share the biobanking data and necessary clinical data within the ALL studies. CRIP is a meta biobank that is maintained and further developed at IBMT. For adapting CRIP to the ALL scenario, a core data set describing the data that is necessary to share needs to be developed. Furthermore, interfaces to biobanking data management systems need to be specified and implemented.
It was discussed that adapting CRIP for ALL partners should start from a minimal scenario, utilizing an initial data set and integrating firstly e.g. only 2 biobanking management systems (GENICA, Italy, and UK) from different partners including the Scopeland system. The approach can then be extended to integrate data from more biobanking or trial management systems and extend the data set according to the needs of the users. For further details see chapter 9 and appendix 5 (use cases).

[54] http://www.crip.fraunhofer.de/en/ethics_policy/privacy_regime July, 2011

8 Patient Empowerment Scenarios

Introduction

The patient empowerment tools must feel easy and comfortable for patients. To help ensure this, there are a few guiding principles for the creation of the tools

- There should be one tool composed of different sub-tools and not a series of different tools
- The tool should be cloud based
- Patients and professionals will both be users of the different elements of the tool, with access via the p-medicine portal according to their rights and roles
- The tool must communicate with patients using language they are comfortable with
- The tool must be totally secure giving patients the confidence to share their data
- Touch screen technology should be used where possible

Elements of the patient empowerment tool will be used in the Clinical Decision Support work package; therefore it is very important that the WPs work together closely to find an integrated solution.

Patients are typically seen as the recipients of care. An important ideal of personalized medicine is to better enable patients themselves to be participants and guides in their own health care. The role of patients will be strengthened in *p-medicine* by allowing them to decide at any time what kind of research is allowed to be done with their data and their own biomaterial. Patient empowerment is based on information coming from research. Only by using this information to educate patients shared decision support is possible. This will enhance transparency for patients in the healthcare system and will convince patients to use their data for research purposes as shown in figure 6.1.

Fig 6.1: The circuit of patient empowerment from research to decision support and back to research. The green arrow indicates the necessity of tools for patients to provide feedback to enhance clinical research. Adapted from: "The Patients and Consumers Perspective"; eHealth Conference, Barcelona, 15th March 2010.

This work deals with the development of the Interactive Empowerment Service (IEmS). The aim in providing IEmS is twofold:

- Help the patient to understand her/his medical documentation.
- Empower the patient to make informed choices.

In line with the aim to develop a personalized medicine, the empowerment tool will aim at enabling the patients understanding of the whole data set that the hospital has collected. This process implies that patients are able to understand medical statements, as well as legal and ethical considerations. Thus, the empowerment tool must not only represent data in a convenient format, but data must also be translated into a language that is understandable to the patient. Of course, this does not only entail the wording of the information, but there is the need to come up with ways to organize the data in a manner that makes it easier to decide for the patient what is of interest to him/her at the moment. This statement is consistent with a second goal of the empowerment tool: to give a patient a chance to make an informed choice. In order to build the IEmS the patient view is of utmost importance. Task 14.2 of the DoW will provide the necessary linguistic analysis to develop the Patient View.

Use cases for patient empowerment that will be supported and tested within *p-medicine* are the following:

1. Search for running clinical trials in Europe
2. Consent and re-consent
3. Usage of the own data and own biomaterial
4. Summarize the history of the disease in an understandable way and increase patient-doctor understanding

These use cases will increase the compliance of patients to their treatment and will improve the quantity and the quality of data for research purposes. Transparency in data handling, augmentation of the patient's knowledge about his/her disease and participation as an active partner in a shared decision process in the management of his/her disease increases trust in the Health Care System including data handling and demands for more research by patients allowing the use of his/her individual data to solve his/her personal medical problem.

8.1 Search for running clinical trials in Europe

The search for the best treatment for a given patient has to get access to running trials in Europe (Eudract database) by selecting those trials that fit the best to the patients disease characterized by the individual data of the disease of the single patient. Data mining tools should also be used to search other databases, literature and results of closed trials and patient cohorts treated outside of trials. Such a tool should suggest those treatments with the highest survival rates or the lowest toxicity, or other characteristics that can be chosen. The tool should be useable for patients but also for physicians. The result given to patients must be given in a patient understandable language whereas it can be in more detail displayed for clinicians, giving also the references for further information.

8.2 Consent and Re-consent Scenario

Data created from a clinical trial should be securely stored in the data warehouse as done for other clinical or molecular data. The analysis of the individual profile of the patient might serve as a discriminator in an econsent tool as part of the IEmS. Such an approach will lead to an individualized econsent adjusted to the patients needs. For the future such an approach would mean that patients primarily have to answer a questionnaire online

(part of the tool) before they are guided to the individual consent form to sign. The signature should be possible to do electronically as well as paper based (possibility to print the individual consent form). To create such a form automatically data about the disease, the treatment etc. are needed as well. Such informed consent can be done for patients within or outside trials. Functionality for re-consent needs to be implemented. Access to the informed consent by different stakeholders has to be considered. The patient needs to get the possibility to reject informed consent at any time, or to restrict consent to only specific items, etc. Further functionalities of such an IT tool is described in more detail in the corresponding scenario.

8.2.1 Informed Consent (Patient's Perspective)

Informed consent from Patient's Perspective should be clearly visible and accessible for all p-medicine end users (patients and/or patient's relative). Patients need to be aware about the term "informed consent" in an easy understandable way. A close collaboration with the EU project CONTRACT[55] is given.

8.3 Own Data Scenario

As needed for the consent tool, patients should have the possibility to se which of their data are stored electronically. He might also be able to validate his own data, as well as giving input to missing data. Even eCRFs for patients can be built. Such a tool might be built in ObTiMA.

8.4 Access to Biobanks Scenario

Patients will be able to access the biobank data stored on them with the data "translated" into a patient friendly format and language.

8.5 Summarize the history of the disease in an understandable way and increase patient-doctor understanding

This use case summarizes the usage of data mining and knowledge discovery tools that are able to summarize the history of a patient's specific disease with all relevant information and in a language understandable by patients. Patients will complete a questionnaire that will allow a psycho-cognitive profile to be developed. This profile will then be displayed to the doctor as part of the suit of clinical decision support tools aiding the appropriate decision making for each individual patient.

[55] http://www.contract-fp7.eu/

9 Biobanking Scenarios

Introduction

A biobank, also known as a bio-repository, is a place that collects, stores, processes and distributes biological materials and the data associated with those materials. These may include human bio-specimens such as tissue or blood and related clinical information pertaining to the donor of that bio-specimen.

A special focus of p-medicine project is the Biobanking and Biomolecular Resources Research Infrastructure (BBMRI)[56], one of the first European Research Infrastructure projects funded by the European Commission (EC). The EC-funded preparatory phase of BBMRI came to its end in January 2011. During the past 3 years BBMRI has grown into a 53-member consortium with over 280 associated organisations (largely biobanks) from over 30 countries, making it the largest research infrastructure project in Europe. During the preparatory phase the concept of a functional pan-European biobank was formulated and has now been presented to Member States of the European Union and for associated states for approval and funding.

BBMRI proposes to form an interface between specimens and data and top-level biological and medical research. BBMRI will be implemented under the ERIC (European Research Infrastructure Consortium) legal entity. BBMRI-ERIC foresees headquarters (central coordination) in Graz, Austria, responsible for coordination of the activities of National Nodes established in participating countries. BBMRI is in the process of submitting its application to the European Commission for a legal status under the ERIC regulation, with an expected start date at the end of 2011.

According to the available BBMRI project description WP 4[57] (WP4 - Biomolecular Tools and Resources) will develop a concept to integrate existing biomolecular resources, technologies, standards and know-how into the operational concept of BBMRI, and provide molecular tools for interrogation of bio-banked samples.

In p-medicine the following Biobank Access use cases will be provided:

- Linking the own biomaterial data repository to the p-medicine biobank access framework for the collaboration with specific user groups
- Managing patient's biomaterial and related data within p-medicine infrastructure for clinical trials
- Offering human biomaterial to a closed and/or open clinical research community for research
- Requesting specific human biomaterial within a closed and/or an open clinical research community for research purposes

[56] http://www.bbmri.eu
[57] http://www.bbmri-wp4.eu

10 Clinical Trials

Introduction

Clinical trials are essential to achieve better treatments for patients. As a result of the Clinical Trials Directive 2001/20/EC the conduct of clinical trials throughout Europe has changed[58],[59]. The directive, aimed largely at holding pharmaceutical companies to higher standards, has tied up academic clinical research, particularly large trials, with redundant paperwork, liability tangles and unending bureaucracy[52].

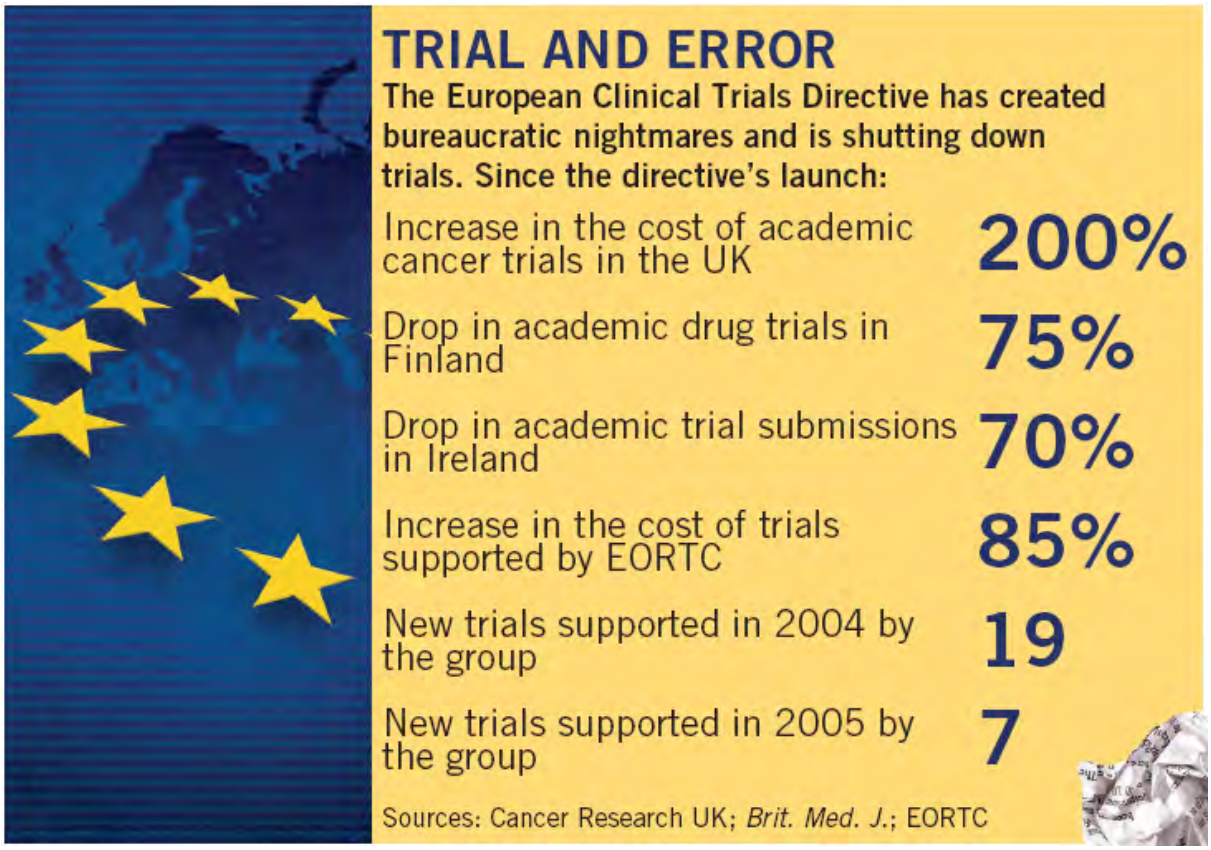


Fig 8.1: The Impact of the European Clinical Trials Directive 2001/20/EC (52; figure taken from the article)

Brandon Keim writes in Nature Medicine: "The cost of academic cancer trials has doubled since 2004**,** according to Cancer Research UK, the country's largest sponsor of academic cancer research. The European Organization for the Research and Treatment of Cancer estimates that expenses have risen by 85% and says the number of trials it supports has dropped by 63%. The Save European Research campaign, which represents more than 3,000 scientists, says academic drug trials have dropped by 70% in Ireland and 25% in Sweden. The number of Finnish academic drug trials shrunk by 75%" [52]. One of the biggest bottlenecks is the directive's requirement that each trial has to have a single sponsor who is

[58] Keim B: Tied up in red tape, European trials shut down. Nature Medicine 13:110, 2007

[59] Pritchard-Jones K: Clinical trials for children with cancer in Europe – Still a long way from harmonisation: A report from SIOP Europe. European Journal of Cancer 44:2106-2111, 2008

fully liable for all legal and financial issues. For trials running in different European Countries the problem of a single sponsor is not solved yet. Kathy Pritchard-Jones summarizes key issues for Cancer Trials in the European Journal of Cancer[53]. Though this article deals with clinical trials for children, most of these points are relevant for clinical trials in adults.

Scenarios and structures that help to run more clinical trials and to bridge the gab between treatment given to patients today and research to find better treatment for patients is of utmost importance.

Issue	Experience of European paediatric study groups running investigator-led ('non-commercial') trials in childhood cancers
Definition of an interventional clinical trial	'Standard of care' regimens often include medicines used 'off label' Variation in acceptance by national regulatory authorities of such use as 'background medicine' or whether it falls outside the definition of an 'interventional clinical trial'
Sponsorship	National variation in whether a single European sponsor is required or a national co-sponsorship arrangement is accepted Complex contractual negotiations required between partners
Insurance and Indemnity	Large variation in costs and in whether 'no fault' indemnity is required Insurance costs increased 100-fold with no perceptible change in risks between consecutive trials of the same study group Premiums may be paid by fundraising efforts of childhood cancer parents' associations
Definition of an IMP	Hugely variable for use of old drugs with no or limited paediatric information in their marketing authorisations IMP definition has major impact on bureaucracy of pharmaco-vigilance
Pharmaco-vigilance	Hugely bureaucratic with no noticeable improvement in patient safety (which was in any case very good in childhood cancer trials) National variation in onward reporting requirements for SUSARs when drug is used in more than one trial Inconsistency in inspection findings of regulatory processes for the same trial
Sponsor obligation to provide free drug	Large national variations in how this is absorbed into national health insurance schemes or whether this must be paid for by sponsor Required for IMPs, whose definition is also variable
Drug formulations adapted for children	Lack of appropriate formulations for young children for many oral anti-cancer drugs Strict definition of 'manufacturing' excludes young children from some clinical trials when no appropriate formulation exists
Ethical considerations	Ethical committees need appropriate expertise to evaluate appropriateness of new drug trials in children Timelines to receive the 'single' national ethical approval highly variable Institutions have created other hurdles to opening a trial, variably labelled 'R & D' approval

Table 8.1: Key issues for Paediatric Cancer Trials in relation to the EU Clinical Trial Directive 2001/20/EC[53]

In detail the following problems in clinical care of patients do exist today:

- There is a time lack for physicians being kept informed about all the new developments in medicine, even in their specialized field. Every week hundreds of new papers are published. To find the most relevant, to read them all and to judge them as important for the own work is impossible.
- Today teamwork is of utmost importance. No physician is able to treat a patient with cancer by his own. He always has to communicate and work together with other specialists in medicine. As a result a lot of so called Cancer Comprehensive Centres are established to facilitate the interdisciplinary work. But up to now no IT infrastructure is supporting this by storing all relevant data in a database, so that every treating physician will have immediate access to the history, diagnosis, treatment and other relevant data of patients in an anonymous and secure way.

- Physicians do not get feedback of how efficient they are working. They do not have any statistics regarding the survival of their patients compared to the survival of all patients with that kind of cancer. There is no benchmarking telling them they are doing good or bad.
- Physicians do not know about the possibilities of modern IT technologies that could help them to support them in daily care of patients, or in developing new clinical trials. The lack of this knowledge leads to a lack of requests and requirements to IT people for the creation of new and user friendly tools in this respect.
- Only a minority of patients are enrolled in prospective clinical trials. The reason for this is manifold:
    - Physicians do not (want to) enter patients in clinical trials because
        - they fear the burden of workload by entering patients (documentation, regulatory and administrative necessities, etc.)
        - they are not well informed about the meaning and impact of clinical trials (fear of experiments with their patients, simply not used to enrol patients in clinical trials, etc.)
        - in most curricula of Medical Schools Clinical trials are missing, so that students will not learn about the benefits of clinical trials
    - Patients do not want to enter a clinical trial
        - they are not informed at all about clinical trials
        - they are not well informed about the meaning and impact of clinical trials (fear of taking part in an experiment, etc.)
    - There is no financial and/or administrative support to cover the overhead of clinical trials
        - the burden of European regulations contrasts the available resources to increase the number of new clinical trials
        - infrastructures in hospitals or outpatient facilities are lacking (no data manager, etc.)
- Today patients do use the internet to get information about their disease. There is no way how a patient can trust such information. Often information is contrary and alienates patients.
- Even if patients do find relevant information, they may not understand the medical language used in these information.
- More patients are asking for second opinions regarding their disease. This is time consuming for physicians, expensive for the health care system and often unsatisfying for patients. They often get different and contrary answers.

10.1 IT support for clinical trials

IT support for clinical trials can be given during different phases in the conduct of a clinical trial. Such phases provides the framework of clinical trial uses cases. They are listed here:

- Planning of a new trial
- Trial Management
- Trial closure
- Analysis of the trial and reporting of results

In each of these phases many use cases can be described. An excellent overview gives the Clinical Trials Tool Kit from the Department of Health and the Medical Research Council in UK[60]. The following 2 figures show the roadmap for the planning of a new trial and the management and closure of a clinical trial as provided by the Clinical Trials Tool Kit.

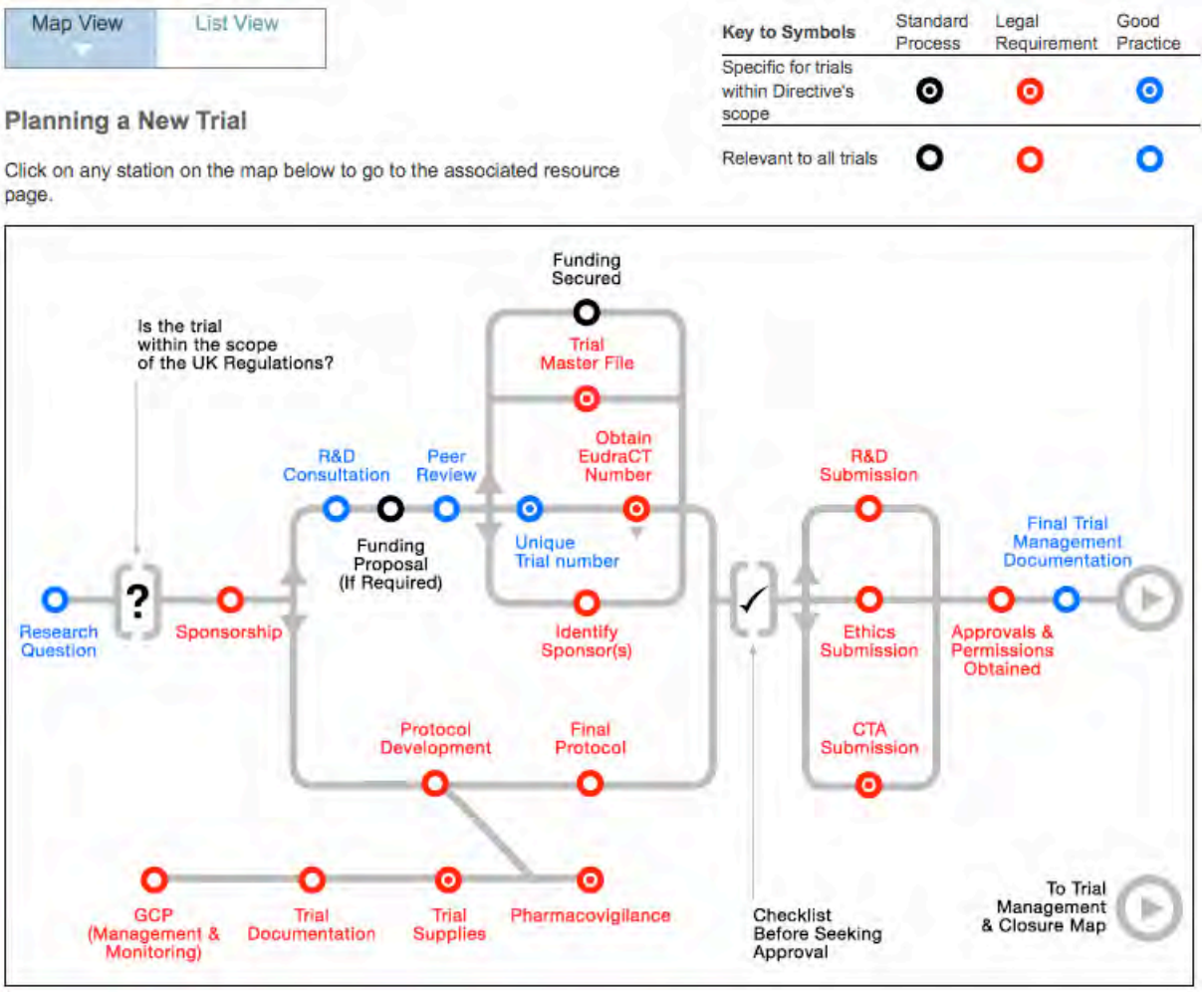


Fig 8.2: Planning of a new trial (from Clinical Trials Tool Kit: http://www.ct-toolkit.ac.uk/)

[60] http://www.ct-toolkit.ac.uk/

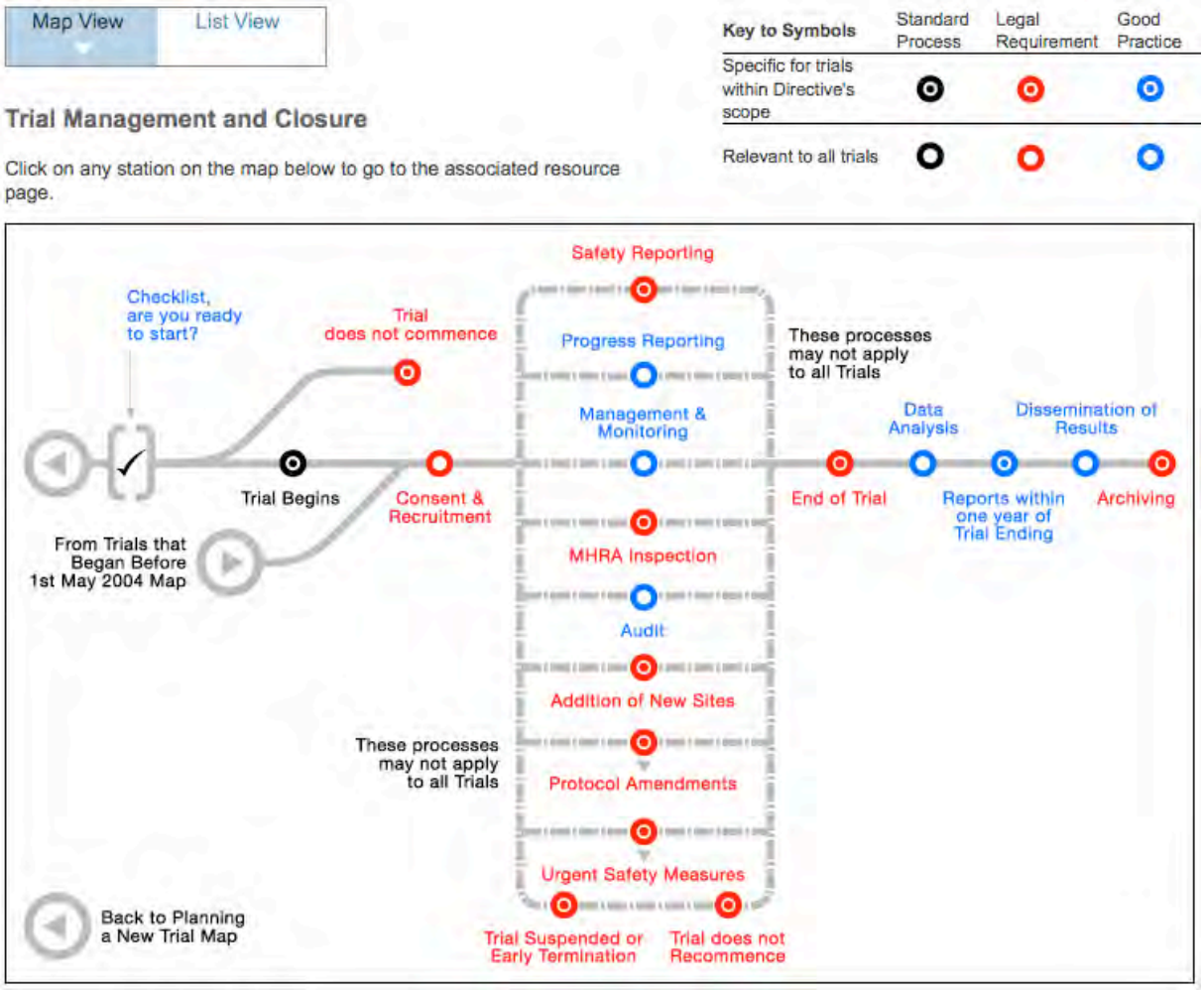


Fig 8.2: Management and closure of a trial (from Clinical Trials Tool Kit: http://www.ct-toolkit.ac.uk/)

10.2 Analysis of clinical trial data and across clinical trials

Tools for analysis of clinical trials will be developed in ObTiMA in level 4 of the Trial Outline Builder (TOB). To perform analyses across clinical trials or joining data from the clinical trial with research data in other databases the need for standardization of clinical trial data is of utmost importance. In June 2008, the American Health Information Community (AHIC) approved a recommendation to develop a Clinical Research Use Case. Taking into account feedback from interested private and public stakeholders this Use Case was developed[61]. It refers first of all on the use of EHR in clinical research. It has been driven by the ANSI-convened Clinical Research Value Case Workgroup to represent the AHIC prioritization process and provide context for the national (US) agenda activities, beginning with the selection of harmonized standards by the Healthcare Information Technology Standards Panel (HITSP).[62]

[61] Clinical Research Value Case Workgroup, Use of Electronic Health Records in Clinical Research: Core Research Data Element Exchange, Detailed Use Case, April 23rd, 2009.

[62] http://www.hitsp.org

As a result the p-medicine platform needs to be open to standards related to both clinical trials and EHRs. Identification, development, and harmonization of standards to support interoperability associated with clinical research are addressed in details in the above-mentioned document of the AHIC regarding the clinical use case. This document is further addressed in WP 4 (Standardisation, Semantic Interoperability and Data Integration). In p-medicine different use cases will need the use of standardized data for exchange, e.g. exchange of SAEs with regulatory bodies etc.

10.3 EUNs of different stakeholders in clinical trials

Due to the complexity of clinical trials different stakeholders do have different user needs. The following table gives an overview of most relevant stakeholders of investigator-initiated trials.

Stakeholder	**Description**	**Access to p-medicine platform** (tracking of all activities via audit trail)	**Needs & Requirements**
Sponsor (Clinical Research Sponsor)	Clinical trials are sponsored by government agencies, private organizations (pharmaceutical, biotechnology and medical devices companies), and individual researchers	Access to p-medicine platform and in special to Clinical Research Frames	Level of access is restricted to the clinical trial sponsored
Principal investigator	Person responsible for running the clinical trial	Access to p-medicine platform via the portal	ObTiMA use cases
Clinical trial physician	Local physician taking part in a clinical trial	Access to p-medicine platform via the portal	ObTiMA use cases
Data Manager	Person managing the data of a clinical trial	Access to p-medicine platform via the portal	ObTiMA use cases
Basic researcher including person running a biobank	Researchers analysing biomaterial	Access to p-medicine platform via the portal	e.g. Biobanking use cases
Laboratory Department(s) and/or Laboratory Information	The LIMS is a software-based laboratory and information	Automated access to the p-medicine platform	Data exchange with CTMS / ObTiMA

Management System (LIMS)	management system that offers a set of key features to support modern laboratories		
Study Subjects	Members of the public who have volunteered to participate in a clinical trial study.	Access to p-medicine via the portal Informed consent	ObTiMA use cases
Patient(s)	A patient is any recipient of medical attention, care, or treatment. (Nephroblastoma, Breast Cancer, AAL)	Access to p-medicine via the portal	ObTiMA use cases
Patient's Relatives	Parents or relatives of patients (children) with Nephroblastoma, ALL	Access to p-medicine via the portal Need for informed consent.	ObTiMA use cases
Regulatory Agencies	European Medicines Agency (EMA)[63] Food and Drug Administration (FDA)[64]	No access to p-medicine platform	Level of access is restricted. Data flow is only from p-medicine to EMA
CROs	Clinical Research Organisations	Limited access to the p-medicine platform Need of contracts	See paragraph 3.2.4

[63] http://www.ema.europa.eu
[64] http://www.fda.gov

11 ObTiMA Scenarios

Introduction

ObTiMA[65], an ontology-based clinical trial management system, has been developed in special as a proof-of-concept application to highlight the possibilities of ontology based creation and managing of clinical trials within the ACGT (Advancing Clinico-Genomic Trials on Cancer)[66] project.

ObTiMA is modular developed with a core basic module for data management of clinical trials. Different other modules are under development in p-medicine. ACGT started[67] to make ObTiMA GCP conformant and to build the basis for certification of ObTiMA to use in GCP conform Trials. Interoperability issues between the p-medicine platform and ObTiMA are of utmost importance.

In order to overcome the interoperability obstacles standards mentioned in D2.1 are implemented. Important standards used are:

Standard	Short Description
The HL7 Study Design Standard*	The HL7 Study Design Standard captures information on the design, analysis process and intent of an individual study. The study design standard transports trial design and eligibility criteria information in a standardized format. Specifically the study design standard covers arms, epochs, subject assignment, planned encounters (visits), planned interventions, planned observations (assessments), eligibility criteria and study characteristics.
The Clinical Data Acquisition Standards Harmonization (CDASH) Standard version 1.1	CDASH Version 1.1 was developed via CDISC's consensus-based standards development process that included comments from organizations in all three ICH regions (US, Europe and Japan). It describes the basic recommended (minimal) data collection fields for 18 domains, including common header fields, and demographic, adverse events, and other safety domains that are common to all therapeutic areas and phases of clinical research. CDASH V 1.1 also includes implementation recommendations and best practice guidelines, regulatory references and other information on the CDASH project.[68]
LOINC	The LOINC database provides a set of universal names and ID codes for identifying laboratory and clinical test results.[69]

* an on-going project within Health Level Seven (HL7), sponsored by both the Clinical Data Interchange Standards Consortium (CDISC) and the Food and Drug Administration (FDA), to develop HL7 version 3 messages for structured study information.

[65] http://www.obtima.org

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

[67] Report on ObTiMA as a GCP conformant software application, http://eu-acgt.org/uploads/media/ACGT_USAAR_D2_6_final_01.pdf

[68] xml.coverpages.org/CDISC-**CDASH**-v10-2008-10-01.pdf

[69] loinc.org/downloads/files/LOINCManual.pdf

11.1.1 Pseudonymization Scenario

Pseudonymization in ObTiMA needs to work on the fly, meaning that a local user treating a patient always works with the real personal data whereas every other user will never see personal data in ObTiMA. To make this possible a trust centre needs to be enrolled in this scenario.

11.1.2 Data Entry of Prospective Clinical Trial Dato

Clinical trials of investigational medicinal products must be carried out in accordance with ICH GCP and national legislation. Requirements for the data capture include quality control, quality assurance. The end user requires clear instructions and prompts, drop-down lists etc. to help with speed and accuracy of data input.

11.1.3 Data Manager of Prospective Clinical Trials

According to GCP criteria and legal regulations of clinical trials an end user needs the facility to raise data clarification queries within the ObTiMA software, and allocate status to queries (e.g. close them when satisfied); the role plays an important part in demonstrable quality assurance.

11.1.4 eCRF Developer for Prospective Clinical Trials

Users need to design electronic case report forms (eCRFs) that are carried out in accordance with ICH GCP and national legislation.

11.1.5 Data Synchronization with HIS during running trial in ObTiMA

During a running trial, a clinician or data clerk can import data from a hospital information system (HIS) to fill patient CRFs in ObTiMA.

11.1.6 SAE/SUSAR Scenario

Reporting and handling of SAEs and SUSARs in clinical trials has o be done according to GCP criteria. All needed information can be found at the website of EudraVigilance: http://eudravigilance.ema.europa.eu/human/index.asp: “EudraVigilance is a data processing network and management system for reporting and evaluating suspected adverse reactions during the development and following the marketing authorisation of medicinal products in the European Economic Area (EEA). The first operating version was launched in December 2001[70].

[70] http://eudravigilance.ema.europa.eu/human/index.asp

EudraVigilance supports in particular the:

- Electronic exchange of suspected adverse reaction reports (referred to as Individual Case Safety Reports) between the European Medicines Agency (EMA), national Competent Authorities, marketing authorisation holders, and sponsors of clinical trials in the EEA;
- Early detection of possible safety signals associated with medicinal products for Human Use;
- Continuous monitoring and evaluation of potential safety issues in relation to reported adverse reactions;
- Decision making process, based on a broader knowledge of the adverse reaction profile of medicinal products especially in the frame of Risk Management.

Taking into account the pharmacovigilance activities in the pre- and post- authorisation phase, EudraVigilance provides two reporting modules:

- The EudraVigilance Clinical Trial Module (EVCTM) to facilitate the electronic reporting of Suspected Unexpected Serious Adverse Reactions (SUSARs) as required by Directive 2001/20/EC[71].
- The EudraVigilance Post-Authorisation Module (EVPM) designed for post-authorisation ICSRs, Regulation (EC) No 726/2004, Directive 2001/83/EC as amended, and Volume 9A of the "Rules Governing Medicinal Products in the European Union: Pharmacovigilance for medicinal products for human use"[72].

EudraVigilance is also one of the main pillars of the European Risk Management Strategy[73], a joint effort between the EMA and national Competent Authorities to strengthen the conduct of pharmacovigilance in the EEA. EudraVigilance facilitates the process of risk management at several levels including aspects of risk detection, risk assessment, risk minimisation and risk communication. Consequently, EudraVigilance contributes to the protection and promotion of public health in the EEA and provides a powerful tool for the EMEA and national Competent Authorities in monitoring the safety of medicinal products and in minimising potential risks related to suspected adverse reactions.

The reporting obligations of the various stakeholders are defined in the Community legislation, in particular Regulation (EC) No 726/2004, Directive 2001/83/EC as amended and Directive 2001/20/EC."

11.1.7 Drug interaction Scenario

Often patients do receive more than one drug. It is very difficult for a physician to know all interactions between different drugs[74]. Therefore for safety reasons and interaction checker is very useful. Such interaction checkers are even freely available as a web-service or as an applet, e.g. the Interaction Checker from Medscape:

http://reference.medscape.com/drug-interactionchecker

[71] http://eudravigilance.ema.europa.eu/human/docs/Directives/Dir2001-20_en.pdf

[72] http://ec.europa.eu/health/documents/eudralex/vol-9/index_en.htm

[73] http://eudravigilance.ema.europa.eu/human/evRiskManagement.asp

[74] Lei Zhang, Yuanchao (Derek) Zhang, Ping Zhao, and Shiew-Mei Huang: Predicting Drug–Drug Interactions: An FDA Perspective. AAPS J. 2009 June; 11(2): 300–306

Such a service or tool would be beneficial for patients and physician can be integrated into the IEmS and into ObTiMA for physicians as a service.

The tool should help to find dangerous interaction between two drugs that are prescribed to a patient. A physician should do this check always before subscribing drugs. If all the drugs a patient gets are stored in CRFs in ObTiMA then such a service can automatically check for interaction and send a warning to the treating physician, announcing that there is incompatibility between drugs. In addition this service names the drugs and gives information about what are the risk for the patient. This use case can be combined with the use case for the prediction of an SAE (see chapter 7.1.3, use case: PSN_3).

11.1.8 DICOM Scenario

The usage of DICOM files within clinical trials for reference and for research is high. Therefore such a use case is of utmost importance.

This use-case describes how DICOM data can be send from a local hospital to the data warehouse after automatic pseudonymization of the data. In a second step it describes how DICOM data can be downloaded for reviewing or post-processing.

11.1.9 Consultation Scenario

In this scenario a local physicians can ask for consultation of a patient treated within a clinical trial.

11.1.10 Trial Development Scenario

As a result of the regulatory regulations the development of a trial is very complex, bureaucratic and time consuming. From trial to trial the same procedures need to be followed. IT can help to standardize the development of trials by guiding a chairman through the process of fulfilling all regulations and writing the trial protocol with the help of templates. See also chapter 10.1.

Templates will guide the trial chairman or people responsible for writing a new trial through all needed tasks according to legal, ethical and GCP regulations. There are also templates available for writing a standardized trial protocol.

11.1.11 Trial Outline Builder Scenarios

There will be two scenarios described as use cases: Statistical toolbox and Gene expression parallel coordinates.

11.1.12 Participating Centres Scenario

In clinical trials the selection of participating centres is of utmost importance. The trial chairman needs to know which centres are compliant with GCP criteria and which physicians can work as trial investigators from a centre. Such information can be stored in

a database, which needs regular updates. Such a process can be automatized. A graphical view or representation of participating centres on a map is beneficial. Researchers to include research institutes can also use this tool.

11.1.13 Patient Access to his/her trial data and Diary Scenario

If patients are enrolled in clinical trials, they are allowed to see there stored data and might be able to write data into a specific diary CRF. This will allow to check and validate data of patients as well as enhance data curation. The patient is not allowed to change data in the database, but he is allowed to comment to data. He can only write in the diary CRF. The expected benefits will be better validated and curated data within clinical trials. This transparency will increase patient empowerment.

11.1.14 Repository Scenario

An end user can store parts as well as an entire CRFs into a (centralized) repository This end user or others can subsequently retrieve, (re)assemble and reuse those full or partial CRFs in other new trials or studies.

11.1.15 Semantic interoperability Scenario

Data from both external as well as internal data sources should be integrated and used along with the data collected using the CRFs within ObTiMA.

11.1.16 Reporting Scenario

The end user receives a summary report of the data collected of a patient. The end user can be a physician but also the patient him/herself. Therefore the look and content of the report should be adaptable in relation to the end user.

11.1.17 Sync and Push services (see 13.2)

Data stored in hospital information systems (HIS), clinical trial management systems and trial repositories provide a precious source for clinical research, especially in the field of personalized medicine. However, it is difficult to exploit such data for VPH modelling, data mining or decision support applications, because the data sources are mostly heterogeneous, unstructured and the semantics is often not defined unambiguously. The aim of p-medicine is to integrate the data from these sources syntactically and semantically in a data warehouse, in order that tools and services can exploit the data seamlessly.

Therefore, in p-medicine tools are required that allow data managers of hospital information systems and clinical trial management systems to push data from their systems in a common format into the data warehouse. Furthermore, in this process they need to be enabled to annotate their data with a shared ontology to describe the data semantically. To enable such a scenario, push services will be developed in p-medicine that allow to push data into the data warehouse.

Furthermore, in p-medicine sync services will be developed that avoid redundant data entry into the clinical trial management system ObTiMA, when the data is already available in hospital information systems. The sync services will allow retrieving data for patient CRFs in ObTiMA from hospital information systems during a running clinical trial.

12 DoctorEye Scenarios

Introduction

DoctorEye is a flexible, clinically driven and easy-to-use annotation platform for quick and precise identification and delineation of tumors in medical images. By using the platform the clinician can efficiently and intuitively annotate large number of 3D tomographic datasets. Both manual and well-known semiautomatic segmentation techniques are available in the platform allowing clinician to annotate multiple regions of interest at the same session. Additionally, it includes contour drawing, refinement and labelling tools that can effectively assist in the delineation of tumors. Furthermore, segmented tumor regions can be annotated, labelled, deleted, added and redefined. The platform has been tested over several MRI datasets to assess usability, extensibility and robustness with promising results[75].

DoctorEye platform is proposed for flexible and modular integration (with focus on interoperability) into p-medicine platform. It will serve as a next development activity of DoctorEye platform and as one of the core p-medicine modules able to enrich the proposed for implementation Sharing Imaging Results interoperability specifications.

12.1 Nephroblastoma Scenario

12.1.1 Segmenting Nephroblastoma from MRI images

Segmentation of abdominal tumors, such as nephroblastoma, constitutes a challenging task, mainly due to the inherent complexity and variability of tumour structures. As in the vast majority of tumor cases, this complexity is directly mirrored into their radiological appearance in medical images: usually they do not have a constant grey level, their boundaries are often poorly defined and, also, they may contain small, sharp-edged heterogeneities. The lack of symmetry and clear distinction between the different tissue structures in the abdominal area, as opposed to e.g. the brain, leads to the inevitable use of semi-automatic segmentation techniques, such as region growing[76], graph-cuts[77] and active contours[78].

A novel snake-based semi-automatic segmentation technique has been integrated into the DoctorEye platform, which was tested and validated on a plethora of nephroblastoma tumor images, providing substantially improved results, compared to traditional snakes and region growing approaches[79,80].

A traditional snake is a deformable model that is driven to the boundary of an image shape (internally or externally) by finding an equilibrium between three counterbalancing energies:

---

[75] Skounakis E, Sakkalis V, Marias K, Banitsas K, Graf N. DoctorEye: A multifunctional open platform for fast annotation and visualization of tumors in medical images. Conf Proc IEEE Eng Med Biol Soc. 2009;2009:3759-62.

[76] Adams R, Bischof L. Seeded region growing. *IEEE Trans Pattern Anal Mach Intell*,1994, 16:641–647

[77] Boykov Y, Jolly M. Interactive graph cuts for optimal boundary and region segmentation of objects in ND images. In*: Int Conf Comp Vis*, 2001, pp 105–112

[78] Kass M, Witkin A, Terzopoulos D. Snakes: Active contour models. *Int J Comp Vis*, 1998, 1: 321-331

[79] Farmaki C, Marias K, Sakkalis V, and Graf N. A spatially adaptive active contour method for improving semi-automatic medical image annotation. *Proc Int Congr Med Phys Biomed Eng,* Munich, Germany, 7-12 Sept. 2009

[80] Farmaki C, Marias K, Sakkalis V, and Graf N. Spatially adaptive active contours: a semi-automatic tumor segmentation framework. Int J Comput Assist Radiol Surg, 2010, 5(4): 369-84

the internal energy, which controls the snake's elasticity and curvature, the image energy, which depends on the gradient of the image, and the balloon energy, which is a force energy that pushes the snake to either expand or shrink. These three types of energy are controlled by specific global parameters of the snake model. The key point of the spatially adaptive active contours method is the discrimination of image regions according to underlying characteristics, so that the snake doesn't exhibit the same behaviour over the entire image. Ideally, it is desirable that the snake should pass by all the 'insignificant' internal small blobs lying inside of the tumor boundary, during its deformation, and delineate accurately the true boundary of the pathology. To this end, a snake should be very rigid inside the region to be segmented, while a large expanding force should be applied, in order to push the snake to evolve fast and securely towards the boundary. On the contrary, a weaker force should be applied on a flexible snake around the object boundary, so that it smoothly adjusts to the true edges, instead of being pushed over them.

The proposed approach accomplishes that by dividing the image pixels into two different groups and assigning a different parameter set to each one, thus allowing the snake to topologically adapt its behaviour according to the characteristics of each pixel region. For the efficient determination of those image regions, a binary mask is produced, where the white pixels correspond to the regions where we wish the snake to be flexible, and the black pixels indicate the regions where we want the snake to be rigid. The extraction of this binary mask is based on gradient and corner features. Instead of using global parameter values, a different set of parameter values is assigned to each one of the extracted regions. Therefore, the improved algorithm is able to spatially adapt the snake's behaviour to the image and include, or not, small high-contrast regions (which, in the case of tumor segmentation, could be important necrotic areas), according to image features, while, at the same time, it can detect accurately boundary details. Figure 12.1 demonstrates three different cases of nephroblastoma, where the tumor boundary was correctly extracted using the spatially adaptive active contour algorithm.

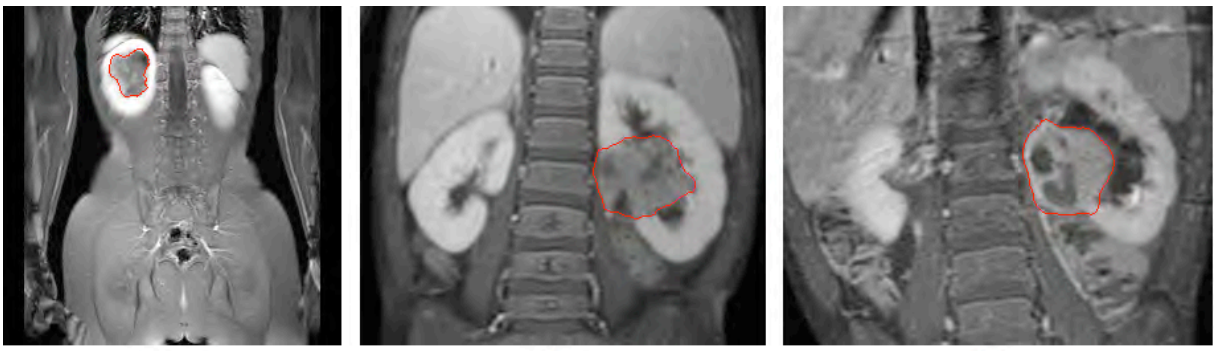


Fig. 12.1: Application of Spatially Adaptive Active Contours module of DoctorEye platform, on images containing nephroblastoma

This efficient segmentation approach has been integrated into the DoctorEye platform, as an independent module. The user has to define an initial draft contour inside the tumor boundary, by clicking on a few points around this contour (even three points, leading to a triangular shaped contour, are enough). The model parameters are set by the algorithm on-the-fly, according to gradient and corner features of the image, so that the user only needs to click on "Run", and the contour starts evolving toward the true tumor boundary.

12.2 Signal Intensity Scenario

12.2.1 Introduction

This scenario encapsulates a signal analysis framework for assessment of temporal tumor changes in nephroblastoma that will be also implemented in the integrated DrEye environment. The aim of this method is to identify, localize and quantify any malignant area changes present in a 3D MRI using histogram analysis on the entire volume. The histogram analysis detects the distribution of the tumour, and quantitatively models its growth or shrinkage offering the potential to assist clinicians in objectively assessing subtle changes during therapy. The proposed method has been applied to the glioma cases and due to the flexibility of the technique, can be generalized to any type of cancer where medical imaging is routinely used to characterize tumor response over time, including the nephroblastoma case. In the next sections the analysis follows the glioma case since initial results have been obtain using this data[81].

12.2.2 Motivation

In recent clinical work (fig. 12.2)**Error! Reference source not found.** with glioma data, it was shown that histograms of signal intensities between cerebrospinal fluid (CSF), vital tumor, necrotic and cystic areas within the tumor vary consistently with patient response to therapy in all modalities analyzed. Using this imaging biomarker information, it might become possible to describe quantitative histogram biomarker changes in the tumor during the follow-up of single patients that are correlated to treatment response or progression. The results of this study indicated that the higher the standardized median and mean values of signal intensities in T1 during the follow-up of a single patient, the more likely the patient suffers from progression of disease. If these values are reducing it is more likely that a tumor response can be established, as shown in the following figure.

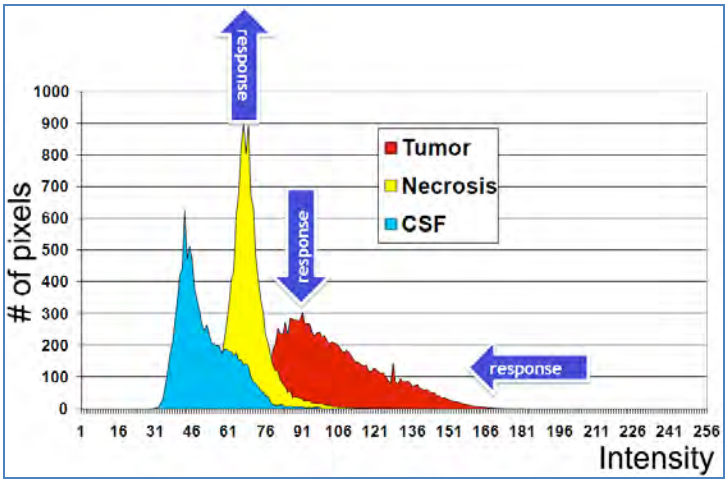


Fig. 12.2: Histogram changes with regard to treatment response, as observed by clinicians.

[81] J. Zepp, N. Graf, E. Skounakis, et al., "Tumor segmentation: The impact of standardized signal intensity histograms in glioblastoma," 4th International Advanced Research Workshop on In Silico Oncology and Cancer Investigation, 2010.

These observations led us to the hypothesis that temporal histogram analysis framework can potentially provide objective differential information concerning brain tissues by using the characteristics of its distributions. A brief description of the overall framework is given below. The proposed differential histogram analysis framework takes no account of the spatial registration of 3D volumes and may avoid/surpass such constraints focusing only in the processing of the signal intensities of the 3D MRI volume.

12.2.3 Data Description

Due to the highly invasive nature of glioma in the vast majority of cases patient are operated after diagnosis. For this reason, it is rarely the case that temporal cancer data before and after therapy is available. From a pool of brain glioma datasets three patient datasets P1, P2 and P3 were used. The data was acquired by different sequence modalities on distinct follow-up times as shown in the following table. For subjects P2 and P3 the time presented under study date field denote the acquisition time after surgery. Examinations were acquired on a 1.49 Tesla MR Siemens scanner with 5mm slice thickness Areas of CSF, Cyst, Tumor, Necrosis and Edema were identified and annotated by a radiologist to use for validation purposes.

Subject	P1	P2	P3
Modality	Gd-enhanced T1	T1 & T2-FLAIR	Gd-enhanced T1
StudyDate	1) On Diagnosis, Before Surgery	1) 3½ months 2) 5 months 3) 8 months	1) 4½ months 2) 7½ months 3) 8½ months

Tab. 12.1: Description of examined data

12.2.4 The Method

The objective of this analysis, when applied to glioma cases, was initially focused on the identification of the malignant areas. Specifically, in many cases tumor was present in only one of the two hemispheres (Type I) whereas in other cases malignant tissue clearly occupied regions from both hemispheres (Type II). Depending on the location of the tumor volume, different techniques embodied in the same histogram analysis framework were applied. The data was acquired by different sequence modalities (T1 Gd-enhanced, T1 and T2 Flair) on distinct follow-up times.

In case of Type I, the 3D volume of the brain hemisphere containing no malignant tissue was marked as BaselineArea (BA), whereas the 3D volume of the brain hemisphere containing malignant tissue was marked as CriticalArea (CA). Since the acquisition modalities used can ensure that malignant areas appear in high intensities, the two hemispheres can easily be distinguished to BA and CA only considering the intensity distributions. An example of Type I malignant tissue identification is depicted in the following figure 12.3.

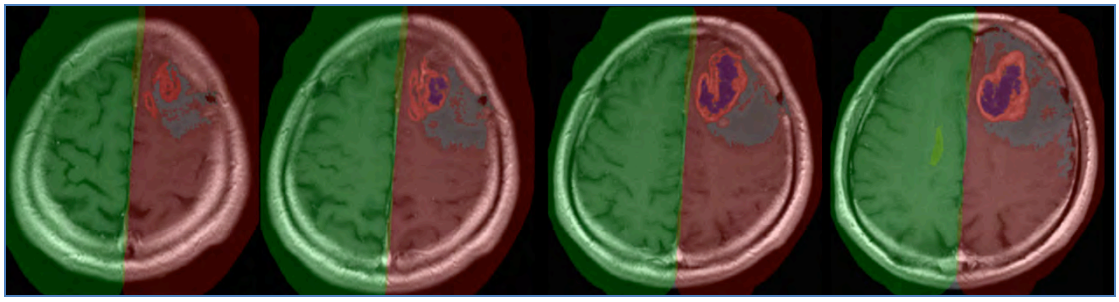


Fig. 12.3: Malignant tissue identification Type I. BaselineArea (BA) is marked with green and CriticalArea (CA) with red.

In case where malignant areas exist in both hemispheres (Type II), further patient's examinations were used. The brain volume of the first (in time) MR examination was marked as BA and constituted the reference examination. Then, each one of the follow-up examination volumes was marked as CA and histograms of BA and CA were subtracted to form the SA distribution (see Fig. 12.4).

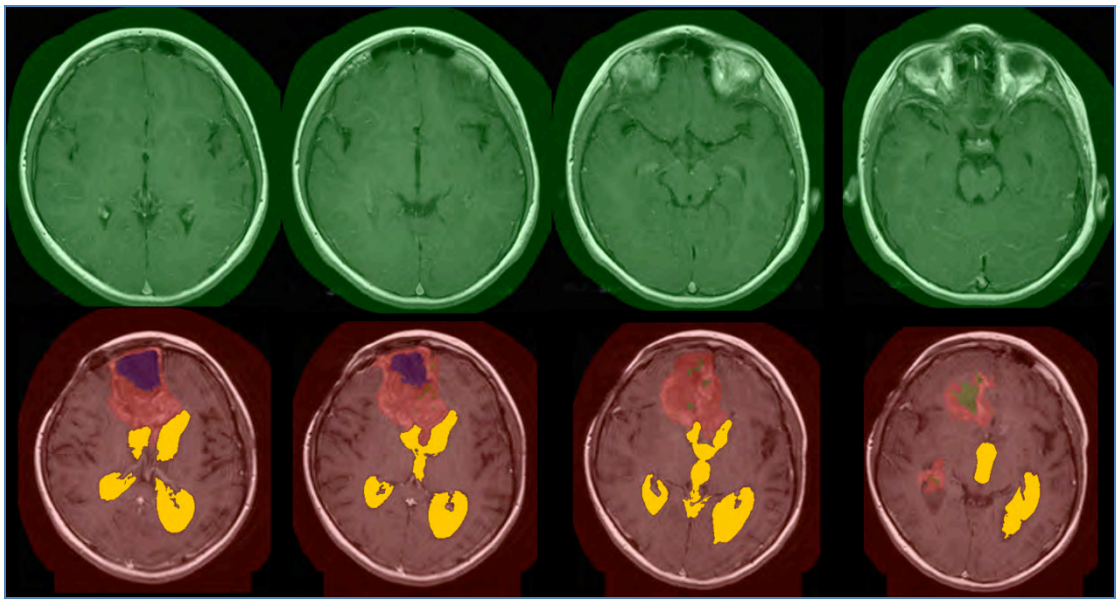


Fig. 12.4: Malignant tissue identification Type II. BaselineArea (BA) is marked with green and CriticalArea (CA) with red.

After histograms BA and CA were obtained, they were subtracted to form the StudyArea (SA) histogram as in:

$$SA(i) = \max[CA(i) - BA(i), 0]$$

The SA distribution reveals the intensity distribution of malignant areas. In case of complete absence of malignant tissue in the 3D MRI at Type I, the left and right hemispheres are similarly and equally depicted in the histogram; a subtraction of the two hemisphere histograms would result in a negligible spectrum. However, when one of two hemispheres actually contains malignant tissue, the histogram distributions of the hemispheres differ significantly and their simple difference can identify the intensity range and distribution of malignant areas.

Gaussian Mixture Modeling (GMM) with Expecation Maximization (EM) was then applied for curve fitting on the StudyArea (SA) histogram distribution. A set of mixed distributions were accurately distinguished by applying individual Gaussian distributions to the observed data. GMM curve fitting is then applied, using two Gaussian curves and rejecting the low and high intensity values. A maximization algorithm (EM) was used to estimate the component parameters of the mixture model distribution. The two components of the outputted model were used to determine intensity boundaries and separate the tumoral histogram distribution from other malignant area distribution. Using the two-component GMM model, two intensity regions were identified on every SA histogram. The region covering the highest intensity values was isolated to simulate tumor's progress and histogram based criteria of this region were extracted for comparison. In addition, histogram specific metrics were calculated for assessing histogram distribution changes during follow-up.

Representative results of malignant identification type I on T1 Gd-enhanced MRI and T1 MRI data are shown in Fig. ***12.***12.5 (a-c). In the top subplot of each study case, the identified SA histograms are shown in gray area and two distinct intensity areas are clearly depicted. The clinical expert's annotations are also shown in the plots for evaluation purpose. Results of the two-component Gaussian curve fitting as applied on SA histogram distribution are shown in the bottom subplot of each study case. PDF 1 curve marks the histogram intensity ranges containing tumoral data, whereas PDF 2 curve identifies the intensity ranges of other malignant areas present in the brain volume. Notice in subfigure (a) that PDF 2 covers the necrotic intensity area, in (b) it covers the cystic area and in (c) the edema area. The proposed framework was also tested on T2-FLAIR examination data and the GMM curve fitting model was able to successfully discriminate tumoral from other malignant intensity areas, in Fig. ***12.***12.5 (d).

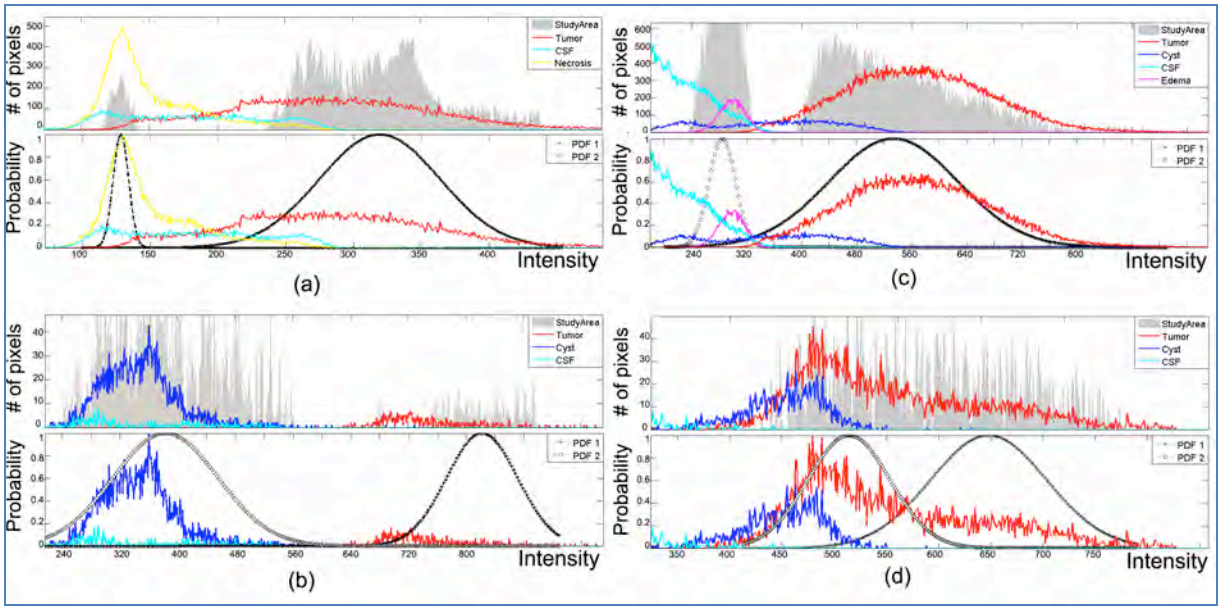


Fig. 12.5: GMM curve fitting illustration. In top subplots SA histogram is shown in gray area; in bottom subplots the two-component GMM fits SA where tumoral area is identified by histogram data under PDF 1 (cross-dashed line) and other malignant area by data under PDF 2 (circle-dashed line). Annotations are also shown including tumor (red), CSF (cyan), necrosis (yellow), cyst (blue) and edema (magenta).

The tumor volume change assessment methodology was applied to several glioma cases. Tumoral areas were outputted by identification type I and identification type II, for all available follow-up examinations. The tumoral volume change is illustrated in Fig. 12.6 where in the top subplot of each subject, SA histogram areas are shown; in the bottom subplot tumor volume change is depicted together with tumor annotation.

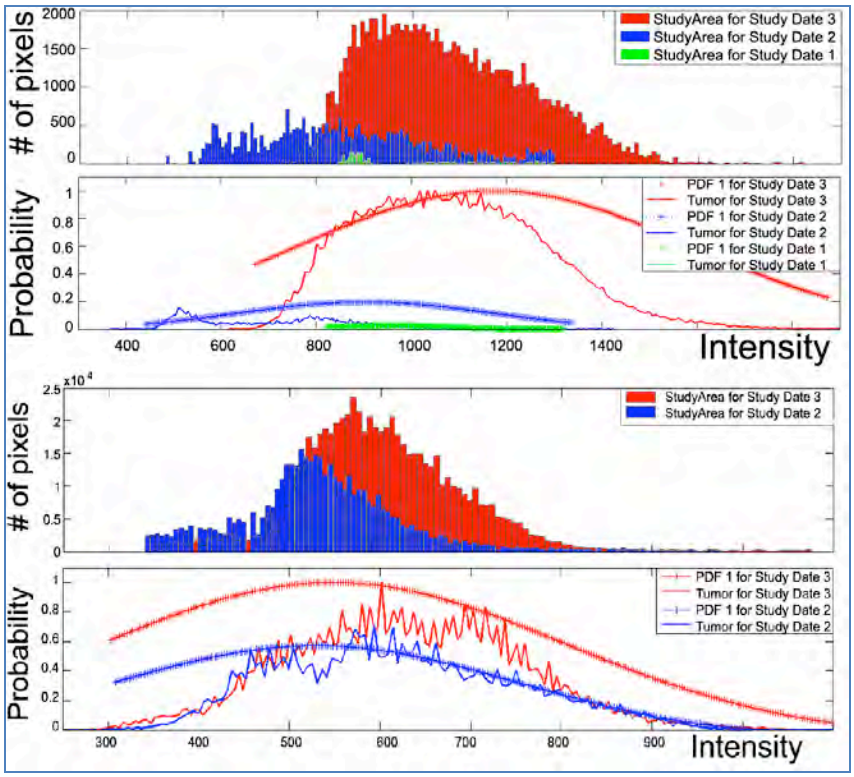


Fig. 12.6: Tumor volume change illustration. Different colors are used to distinct study dates. In each top subplot, SA histogram areas are shown; in each bottom subplot, the corresponding GMM fit is depicted (cross-dashed line) in comparison with doctor's tumor annotation (solid line).

For the evaluation of tumor's volume change, SA histogram distributions under the two-component Gaussian mixture were extracted. Histogram area of follow-up examinations under PDF 1 were compared and assessed for volume change using Kullback-Leibler Divergence (KLD), Earth Mover's Distance (EMD) and two-sample Kolmogorov-Smirnov test (KS-test). EMD is an important perceptually meaningful metric for comparing histogram-distribution changes, which measures the minimum cost required to perform a histogram matching between two histogram distributions. Therefore, it was applied as a metric to directly evaluate the distance between the entire SA and tumor annotations through time, respectively. KS-test is a non-parametric method, which uses the maximal distance between cumulative frequency distributions (CDF) in order to determine if two datasets differ significantly, and returns the maximum difference between the CDF curves. KLD measures the distance between two density distributions and equals to zero value if and only if the two distributions are equal. Quantitative results of temporal tumor volume change were provided through the statistical measures mentioned above.

12.2.5 Conclusion

Summarizing, malignant identification was applied to different acquisition schemes, and the proposed framework was able to discriminate tumoral from other malignant areas in case of glioma tumor. Automated identification, quantification and volume change of tumoral and other malignant areas was achieved through the follow-up, with expected impact on the personalization of cancer treatment strategies. GMM fit curves were able to characterize the intensity areas found in image slices, leading to a back-projection 3D segmentation of the malignant areas.

Therefore, we strongly believe that histogram analysis implemented in this work can be highly beneficial for the nephroblastoma case in order to better understand the actual response of the patient in successive studies. This in turn, can be a more objective way to validate any given model developed in p-medicine since it is often the case that subtle changes in follow-up tumour volume estimation can be difficult to assess objectively by the clinician.

13 p-medicine IT-Components Scenarios

Introduction

As the p-medicine infrastructure needs to be compatible with other VPH projects, mainly VPH-Share, IT scenarios will focus on interoperability, modularity and flexibility requirements. The ‘scenarios' presented below will serve as a core background for technical implementations/specifications and in special for prototyping related activities.

The technical infrastructure and security framework of the p-medicine platform will be built in accordance with legal and ethical regulations, best practice cases in other EU research projects to guarantee an infrastructure that will be able to serve other VPH projects. Scenarios dealing with user needs and requirements in the IT sector are manifold. Most important for p-medicine are the following:

- Summary Documents HL7 CCD
- Laboratory Reports and Messages
- Sharing Imaging Results
- Medical Knowledge Retrieval
- Transfer of Documents on Media
- Patient Demographics Query
- Manage Sharing of Documents
- Consult and History and Physical Note
- Patient ID Cross-Referencing
- Notification of Document Availability
- Clinical Research Interoperability Specification

13.1 Scenarios dealing Interoperability Specifications

This section describes interoperability specifications of IT scenarios in detail with references to standards. Interoperability within the p-medicine platform is an absolute need not only for an integrative IT architecture in VPH but also a corner stone for certification.

13.1.1 Summary Documents HL7 CCD

The ‘Summary Document’ describes the summary of the patient's current medical status. It can include a variety of information as administrative data (registration, demographics, insurance, etc.) and clinical data (history, diagnosis, medication list, allergies, test results, reports etc.). The selection of the data is possible to predefine.

Using HL7 Continuity of Care Document (CCD) standard it will be possible to exchange the information between the p-medicine platform and other IT infrastructures, including Electronic Health Record (EHR) in the future. Additionally, the ‘Summary Document’ would serve as an integration and interoperable solution between other modules of the p-medicine platform.

Selected/Suggested Standards:

Standard	Short Description
The HL7 Clinical Document Architecture (CDA®)	The CDA® Release 2.0 provides an exchange model for clinical documents (such as discharge summaries and progress notes) - and brings the healthcare industry closer to the realization of an electronic medical record. By leveraging the use of XML, the HL7 Reference Information Model (RIM) and coded vocabularies, the CDA makes documents both machine-readable - so they are easily parsed and processed electronically - and human-readable - so they can be easily retrieved and used by the people who need them. CDA documents can be displayed using XML-aware Web browsers or wireless applications such as cell phones. While Release 2.0 retains the simplicity of rendering and clear definition of clinical documents formulated in Release 1.0 (2000), it provides state-of-the-art interoperability for machine-readable coded semantics. The product of 5 years of improvements, CDA R2 body is based on the HL7 Clinical Statement model, is fully RIM-compliant and capable of driving decision support and other sophisticated applications, while retaining the simple rendering of legally authenticated narrative.
HL7 Implementation Guide: CDA Release 2 - Continuity of Care Document (CCD), April 01, 2007	The Continuity of Care Document implementation guide describes constraints on the HL7 Clinical Document Architecture, Release 2 (CDA) specification in accordance with requirements set forward in ASTM E2369-05 Standard Specification for Continuity of Care Record (CCR). The resulting specification, known as the Continuity of Care Document (CCD), is developed as a collaborative effort between ASTM and HL7. It is intended as an alternate implementation to the one specified in ASTM ADJE2369 for those institutions or organizations committed to implementation of the HL7 Clinical Document Architecture.
Integrating the Healthcare Enterprise (IHE) Exchange of Personal Health Record Content (XPHR)	The Exchange of Personal Health Record Content (XPHR) integration profile describes the content and format of summary information extracted from a PHR system used by a patient for import into healthcare provider information systems, and visa versa. The purpose of this profile is to support interoperability between PHR systems used by patients and the information systems used by healthcare providers. This profile does not address all the data exchange requirements of PHR systems.

(adapted from the Healthcare Information Technology Standards Panel (HITSP) Summary Documents Using HL7 Continuity of Care Document (CCD) Component - V2.5)

13.1.2 Laboratory Reports and Messages

Laboratory Reports and Messages are generated/operated by Laboratory Systems, which represent information systems supporting the testing, analysis, and information management for laboratory organizations. Message formats and value sets/code tables (e.g., diagnosis type, gender, patient class, result status, specimen collection method, abnormal flags, observation result status codes interpretation, timestamp format) are contained in the HL7 Version 2.5.1 Messaging Standard.

Flexible electronic laboratory data in Electronic Medical Records (EMR) have many advantages. Users can view, sort, and pool laboratory information to support trend analysis and clinical decision-making. Laboratory data can also be used to trigger clinical decision support systems such as alerts and reminders.[82]

Selected/Suggested Standards:

Standard	Short Description
Health Level Seven (HL7) U.S. Realm - Interoperability Specification: Lab Result Message to EHR (ORU^R01) (HL7 Version 2.5.1) September, 2007	This guide contains the necessary specifications for clinical laboratory results reporting to EHRs for use in the U.S. Realm.
International Health Terminology Standards Development Organisation (IHTSDO) Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT®)	SNOMED CT consists of a technical design, core content architecture, and Core content. SNOMED CT Core content includes the technical specification of SNOMED CT and fully integrated multi-specialty clinical content. The Core content also includes a concepts table, description table, relationships table, history table, ICD-9-CM mapping, and Technical Reference Guide. Additionally, SNOMED CT provides a framework to manage language dialects, clinically relevant subsets, qualifiers and extensions, as well as concepts and terms unique to particular organizations or localities.
Health Level Seven (HL7) Version 2.5.1 Messaging Standard	The HL7 Version 2.5.1 Messaging Standard is an application protocol for electronic data exchange in healthcare. It and prior versions have widespread use in the U.S. and internationally. Both message formats and value sets/code tables (e.g., diagnosis type, gender, patient class, result status, specimen collection method, abnormal flags, observation result status codes interpretation, timestamp format) are contained in the standard.
Health Level Seven (HL7) Clinical Document Architecture Release 2 (CDA R2)	The HL7 Clinical Document Architecture is an XML-based document mark-up standard that specifies the structure and semantics of clinical documents for the purpose of exchange. CDA is one instantiation of HL7's Version 3.0 Reference Information Model (RIM) into a specific message format. Of particular focus are message formats for Laboratory Results and Continuity of Care (CCD) documents. Release 2 of the HL7 Clinical Document Architecture (CDA) is an extension to the original CDA document mark-up standard that specifies the structure and semantics of clinical documents for the purpose of exchange. CDA R2 includes a prose document in HTML, XML schemas, data dictionary, and sample CDA documents. CDA R2 further builds upon other HL7 standards beyond just the Version 3.0 Reference Information Model (RIM) and incorporates Version 3.0 Data Structures,

[82] Neil R. Kudler1 and Liron Pantanowitz. Overview of laboratory data tools available in a single electronic medical record. J Pathol Inform. 2010; 1: 3.

	Vocabulary, and the XML Implementation Technology Specifications for Data Types and Structures.
Logical Observation Identifiers Names and Codes (LOINC®)	A database of universal identifiers for laboratory and other clinical observations. The laboratory portion of the LOINC database contains the usual categories of chemistry, haematology, serology, microbiology (including parasitology and virology), and toxicology; as well as categories for drugs and the cell counts typically reported on a complete blood count or a cerebrospinal fluid cell count. Antibiotic susceptibilities are a separate category. The clinical portion of the LOINC database includes entries for vital signs, hemodynamic, intake/output, EKG, obstetric ultrasound, cardiac echo, urologic imaging, gastro endoscopic procedures, pulmonary ventilator management, selected survey instruments, and other clinical observations.

(adapted from the Healthcare Information Technology Standards Panel (HITSP) Lab Result Message Component – V2.3, HITSP Send Laboratory Result Message Transaction – V2.4, HITSP Lab Report Document Component - V2.3)

13.1.3 Sharing Imaging Results

Integrating Healthcare Enterprise (IHE) Radiology Technical Framework has released the Cross-Enterprise Document Sharing for Imaging (XDS-I) Integration Profile. It specifies actors and transactions that allow users to share imaging information across enterprises. This profile depends on the IHE IT-Infrastructure Cross-Enterprise Document Sharing (XDS) profile. XDS for Imaging (XDS-I) defines the information to be shared such as sets of DICOM instances (including images, evidence documents, and presentation states).

Selected/Suggested Standards:

Standard	Short Description
Integrating the Healthcare Enterprise (IHE) Radiology Technical Framework Revision 10.0	Final Text Version: • Volume 1: Integration Profiles Volume 2: Transactions • Volume 3: Transactions (continued) • Volume 4: National Extensions These documents provide specification of the following profiles: Radiology Scheduled Workflow (SWF) Patient Information Reconciliation (PIR) Consistent Presentation of Images (CPI) Presentation of Grouped Procedures (PGP) Access to Radiology Information (ARI) Key Image Note (KIN) Simple Image and Numeric Report (SINR) Charge Posting (CHG) Post-processing Workflow (PWF) Reporting Workflow (RWF) Evidence Documents (ED) Portable Data for Imaging (PDI) Nuclear Medicine Image **Cross-enterprise Document Sharing for Imaging (XDS-I)** Mammography Image

	Import Reconciliation Workflow (IRWF) Teaching File and Clinical Trial Export (TCE)

13.1.4 Medical Knowledge Retrieval

The Retrieval of Medical Knowledge Transaction has as a background the HITSP Interoperability Specification[83] and represents a description of the request and receipt of additional knowledge about a medical term/concept based on specific context parameters. This Transaction does not prescribe the knowledge content of the message returned but provides the specifications for the query for and receipt of additional knowledge.

Selected/Suggested Standards:

Standard	Short Description
Health Level Seven (HL7) Version 3.0 Context-Aware Information Retrieval Specification: URL Implementation Guide	To support the integration of knowledge resources into CISs, the Clinical Decision Support Work Group (CDS WG) has been developing a set of standard specifications for context-aware knowledge retrieval. The first of these specifications, entitled Context-Aware Knowledge Retrieval (Info button), Knowledge Request Standard, was approved in September 2010 as a normative ANSI/ISO HL7 standard. This specification provides a standard mechanism for clinical information systems to submit knowledge requests to knowledge resources. In addition, a URL-based implementation guide has been developed to specify knowledge request implementations using the HTTP protocol.[84]

(adapted from the Healthcare Information Technology Standards Panel (HITSP) Retrieval of Medical Knowledge Transaction (HITSP/T81) – V1.1)

13.1.5 Transfer of Documents on Media

Cross-Enterprise Document Media Interchange (XDM) - provides document interchange using a common file and directory structure over several standard media. This permits the patient to use physical media to carry medical documents. This also permits the use of person-to-person email to convey medical documents.

The XDM solution is intended to be easy to implement with pre-existing email clients, CD burners and USB ports. XDM does not include any additional reliability enhancements. XDM requires that the recipient be able to support human intervention in order to manually control the importing of the data (patient ID reconciliation, selection of patient of interest from possibly multiple patients' documents on the media).

Selected/Suggested Standards:

Standard	Short Description
Digital Imaging and Communications in Medicine (DICOM) Part 3.12: Media Formats	This DICOM Standard describes the services and the data necessary for the interchange of information between digital imaging computer systems found in health care settings. PS 3.12 of the DICOM Standard articulates the structure between the

[83] HITSP Retrieval of Medical Knowledge Transaction, HITSP/T81, July 8, 2009, Version 1.1

[84] http://wiki.hl7.org/index.php?title=Product_Infobutton

and Physical Media for Media Interchange	Media Storage Model and specific media. Media physical characteristics are also covered.
Integrating the Healthcare Enterprise (IHE) IT Infrastructure Technical Framework (ITI-TF) Revision 5.0 or later, Cross-Enterprise Document Media Interchange (XDM) Integration Profile	Provides document interchange using a common file and directory structure over several standard media types. This permits the patient to use physical media to carry medical documents. This also permits the use of person-to-person email to convey medical documents. XDM supports the transfer of data about multiple patients within one data exchange.

(adapted from the Healthcare Information Technology Standards Panel (HITSP) HITSP Transfer of Documents on Media Transaction (HITSP/T33) – V1.3)

13.1.6 Patient Demographics Query

The Integrating the Healthcare Enterprise (IHE) Patient Demographics Query (PDQ) Integration Profile transaction is intended for use wherever Health Level Seven (HL7) messages are suitable to identify patients from a list of potentials. Due to its complexity and modularity p-medicine platform will benefits from the implementation of the IHE PDQ Integration Profile Transaction, which involves a request by a Patient Demographics Consumer for demographic information about patients, whose demographic data matches data contained in the query. The process flows in the IHE PDQ Integration Profile transaction are shown in the IHE IT Infrastructure Technical Framework, Volume 2 (IHE-ITI TF-2), Section 3.21.4.

Selected/Suggested Standards:

Standard	Short Description
Integrating the Healthcare Enterprise (IHE) IT Infrastructure Technical Framework (ITI-TF) Revision 5.0 or later, Patient Demographics Query (PDQ) Integration Profile	Provides ways for multiple distributed applications to query a central patient information server for a list of patients, based on user-defined search criteria, and retrieve a patient's demographic (and, optionally, visit or visit-related) information directly into the application.
Health Level Seven (HL7) Version 2.5, Chapter 2 - Control, Chapter 3 - Patient Administration, Chapter 5 - Query	The HL7 Version 2.5 Messaging Standard is an application protocol for electronic data exchange in healthcare. It and prior versions have widespread use in the U.S. and internationally. Both message formats and value sets / code tables (e.g., diagnosis type, gender, patient class, result status, specimen collection method, abnormal flags, observation result status codes interpretation, timestamp format) are contained in the standard.

(adapted from the Healthcare Information Technology Standards Panel (HITSP) Patient Demographics Query Transaction (HITSP/T23) – V2.4)

13.1.7 Manage Sharing of Documents

Manage sharing of documents is one of the major p-medicine requirements. To assure the functionality of an interoperable system Cross-Enterprise Document Sharing (XDS) is recommended for implementation (IHE IT Infrastructure Technical Framework, Volume 1 (ITI TF-1): Integration Profiles).

Cross-Enterprise Document Sharing enables a number of healthcare delivery organizations belonging to an XDS Affinity Domain (e.g., a community of care) to cooperate in the care of a patient by sharing clinical records in the form of documents. Federated document repositories and a document registry create a longitudinal record of information about a patient within a given XDS Affinity Domain. This profile is based upon ebXML Registry standards and SOAP. It describes the configuration of an ebXML Registry in sufficient detail to support Cross Enterprise Document Sharing.

Selected/Suggested Standards:

Standard	Short Description
Integrating the Healthcare Enterprise (IHE) IT Infrastructure Technical Framework (ITI-TF) Revision 4.0 or later, Section 10 Cross-Enterprise Document Sharing (XDS.a)	The IHE IT Infrastructure Technical Framework defines specific implementations of established standards to achieve integration goals that promote appropriate sharing of health information to support optimal patient care. Section 10, Cross-Enterprise Document Sharing facilitates the registration, distribution and access across health enterprises of patient electronic health records. IHE Integration Profiles offer a common language that healthcare professionals and vendors may use in communicating requirements for the integration of products. The current version of the ITI-TF, rev. 7.0, specifies the IHE transactions defined and implemented as of August 10, 2010. Vol. 1 (ITI TF-1): Integration Profiles Vol. 2: Transactions - Volume 2 is divided into three separate sub-volumes: Vol. 2a (ITI TF-2a): Transactions ITI-I through ITI-28. These transactions are used in the following profiles CT, PSA, EUA, PIX, RID, XDS, ATNA, PDQ, PWP, NAV Vol. 2b: (ITI TF-2b): Transactions (cont'd) ITI-29 through ITI-50. These transactions are used in the following profiles PAM, XDM, XUA, XDS ◦ Vol. 2x (ITI TF-2x): Appendices A through W and Glossary • Vol. 3 (ITI TF-3): Contains Section 4 Cross-Transaction Specifications and Section 5 IHE Content Specifications

(adapted from the Healthcare Information Technology Standards Panel (HITSP) Manage Sharing of Documents Transaction Package (HITSP/TP13) – V2.6)

13.1.8 Consult and History and Physical Note

The HL7 Health Level Seven (HL7) Implementation Guide for CDA Release 2 describes in details and with examples the Consultation Note document. History and Physical (H&P) Notes are described in the Health Level Seven (HL7) Implementation Guide for CDA Release 2.

Selected/Suggested Standards:

Standard	Short Description
Health Level Seven (HL7) HL7 Version 3 Standard: Clinical Document Architecture (CDA), Release 2	The HL7 Clinical Document Architecture is an XML-based document mark-up standard that specifies the structure and semantics of clinical documents for the purpose of exchange. CDA is one instantiation of HL7's Version 3.0 Reference Information Model (RIM) into a specific message format. Of particular focus for Interoperability Specifications are message formats for Laboratory Results and Continuity of Care (CCD) documents. Release 2.0 of the HL7 Clinical Document Architecture (CDA) is an extension to the original CDA document mark-up standard that specifies the structure and semantics of clinical documents for the purpose of exchange. CDA R2 includes a prose document in HTML, XML schemas, data dictionary, and sample CDA documents. CDA R2 further builds upon other HL7 standards beyond just the Version 3.0 Reference Information Model (RIM) and incorporates Version 3.0 Data Structures, Vocabulary, and the XML Implementation Technology Specifications for Data Types and Structures.
Health Level Seven (HL7) Implementation Guide for CDA Release 2.0: Consultation Note	The HL7 Implementation Guide for CDA Release 2.0: Consultation Note defines additional constraints on the CDA Header and Body used in a Consultation document in the U.S. realm, and provides examples of conforming fragments in the body of the document and an example of a conforming XML instance.
Health Level Seven (HL7) Implementation Guide for CDA Release 2.0: History and Physical (H&P) Notes	The HL7 Implementation Guide for CDA Release 2.0: History and Physical (H&P) Notes defines additional constraints on the CDA Header and Body used in a History and Physical document in the U.S. realm, and provides examples of conforming fragments in the body of the document and an example of a conforming XML instance.

(adapted from the Healthcare Information Technology Standards Panel (HITSP) Consult and History & Physical Note Component (HITSP/C84) – V1.1)

13.1.9 Patient ID Cross-Referencing

Patient ID Cross-Referencing Transaction, based on HITSP package, is used for identifying and cross-referencing different attributes for the same patient. It contains a query for cross-reference and patient identity feed transactions. These transactions are used to identify patients from a list of potentials, and/or to communicate patient demographic data (for further information, please explore Patient Demographics Query section).

Selected/Suggested Standards:

Standard	**Short Description**
Health Level Seven (HL7) Version 2.3.1 Chapter 2 – Control, Chapter 3 – Patient Administration	The HL7 Version 2.3.1 Messaging Standard is an application protocol for electronic data exchange in healthcare. It and prior versions have widespread use in the U.S. and internationally. Both message formats and value sets/code tables are contained in the standard.
Health Level Seven	The HL7 Version 2.5 Messaging Standard is an application

(HL7) Version 2.5, Chapter 2 – Control, Chapter 3 – Patient Administration, Chapter 5 - Query	protocol for electronic data exchange in healthcare. It and prior versions have widespread use in the U.S. and internationally. Both message formats and value sets/code tables (e.g., diagnosis type, gender, patient class, result status, specimen collection method, abnormal flags, observation result status codes interpretation, timestamp format) are contained in the standard.
IHE IT Infrastructure Technical Framework Supplement Patient Identifier Cross-Reference HL7 V3 (PIXV3) and Patient Demographic Query HL7 V3 (PDQV3) August 10, 2010	This supplement provides a new version of the Patient Identifier Cross-Referencing and Patient Demographics Query profiles leveraging HL7 version 3 and SOAP-based web services. The scope of the Patient Identity Feed, the PIX Query, the PIX Update Notification, and the Patient Demographics Query is identical as that for the HL7 v2.5 messages (i.e. same transaction semantics, same message constraints). In this version IHE is providing more details for 185 implementers of the individual transactions, and the new 2007 DSTU of the HL7 V3 Patient Topic as the basis of the messages in the transaction are used. The actual changes to the format compared to the previous year are minimal, as the message content only changes the focal class from identified entity to patient.

(adapted from the Healthcare Information Technology Standards Panel (HITSP) Patient ID Cross-Referencing Transaction Package (HITSP/TP22) – V2.4)

13.1.10 Notification of Document Availability

The Transaction is based on the Integrating the Healthcare Enterprise (IHE) IT Infrastructure Technical Framework (TF) Supplement - Notification of Document Availability (NAV).

Selected/Suggested Standards:

Standard	Short Description
Health Level Seven (HL7) Version 2.3.1 Chapter 2 – Control, Chapter 3 – Patient Administration	The HL7 Version 2.3.1 Messaging Standard is an application protocol for electronic data exchange in healthcare. It and prior versions have widespread use in the U.S. and internationally. Both message formats and value sets/code tables are contained in the standard.
Health Level Seven (HL7) Version 2.5, Chapter 2 – Control, Chapter 3 – Patient Administration, Chapter 5 - Query	The HL7 Version 2.5 Messaging Standard is an application protocol for electronic data exchange in healthcare. It and prior versions have widespread use in the U.S. and internationally. Both message formats and value sets/code tables (e.g., diagnosis type, gender, patient class, result status, specimen collection method, abnormal flags, observation result status codes interpretation, timestamp format) are contained in the standard.
IHE IT Infrastructure Technical Framework Supplement Notification of Document Availability (NAV), August 10, 2010	The Notification of Document Availability Profile (NAV) introduces a mechanism allowing notifications to be sent point-to-point to systems within a Cross-Enterprise Document Sharing affinity domain (See IHE IT Infrastructure XDS Integration Profile), eliminating the need for manual steps or polling mechanisms for a Document Consumer to be aware that documents that may be of interest have been registered with an XDS Document Registry Actor.

(adapted from the Healthcare Information Technology Standards Panel (HITSP) Notification of Document Availability Transaction (HITSP/T29) – V2.4)

13.1.11 Clinical Research Interoperability Specification

Clinical Research Interoperability Specification starts with a conclusion that in eHealth domain is a lack of harmonized standards including consistent terminology, nomenclature and semantics used to exchange clinical research data and to assure the interoperable exchanges of that information. Clinical research related information should be standardized and in special harmonized with clinical care. The harmonization of clinical care and research requires compatible information models and clinical researchers should remain abreast of (and participate in) development in standards for clinical care data and systems. Coordination is required to ensure that standardization movements in both the health care and the clinical research domains evolve in tandem.[85] These processes are not specifically addressed in this document. They may be addressed in future research activities.

In order to achieve the p-medicine project's goals the below Clinical Research specifications/standards are recommended for implementation. A special interest would represent "HITSP Clinical Research Interoperability Specification, HITSP/IS158, Version 1.0" publication where Clinical Research case studies are described in details: protocol-driven sponsored research scenario; registry reporting scenario; research network scenario. Close to clinical research case studies, HITSP describes in details the information exchange requirements and the design specification.

Selected/Suggested Standards:

Standard	Short Description
The HL7 Study Design Standard* * - an on-going project within Health Level Seven (HL7), sponsored by both the Clinical Data Interchange Standards Consortium (CDISC) and the Food and Drug Administration (FDA), to develop HL7 version 3 messages for structured study information.	The HL7 Study Design Standard captures information on the design, analysis process and intent of an individual study. The study design standard transports trial design and eligibility criteria information in a standardized format. Specifically the study design standard covers arms, epochs, subject assignment, planned encounters (visits), planned interventions, planned observations (assessments), eligibility criteria and study characteristics.
The Clinical Data Acquisition Standards Harmonization (CDASH) Standard version 1.1	CDASH Version 1.1 was developed via CDISC's consensus-based standards development process that included comments from organizations in all three ICH regions (US, Europe and Japan). It describes the basic recommended (minimal) data collection fields for 18 domains, including common header fields, and demographic, adverse events, and other safety domains that are common to all therapeutic areas and phases of clinical research. CDASH V 1.1 also includes implementation recommendations and best practice guidelines, regulatory references and other information on the CDASH project.

(adapted from the Healthcare Information Technology Standards Panel (HITSP) Notification of Document Availability Transaction (HITSP/T29) – V2.4)

[85] Rachel L. Richesson, Jeffrey Krischer. Data Standards in Clinical Research: Gaps, Overlaps, Challenges and Future Directions. J Am Med Inform Assoc. 2007 Nov–Dec; 14(6): 687–696.

13.2 Push Scenario – Usage of clinical data from hospital information systems (see 11.1.5)

13.2.1.1 Data translation for PUSH services

When a user pushes his data into the p-medicine data warehouse (DW), this needs to translate it into HDOT format. The DW invokes the translation services in the semantic layer, providing the data received and an ontology annotation that permits to translate that data. The semantic layer returns the data in HDOT format.

13.2.1.2 Ontology annotation of external databases

Annotation of external databases in terms of the HDOT ontology is necessary for data to be stored and integrated in the p-medicine Data Warehouse. The tool will offer data managers a graphical interface to perform this annotation. The interface should be intuitive enough for end users lacking deep RDF understanding to be able to correctly annotate their data

13.3 Ontology-Based Semantic Search Framework

13.3.1.1 Scenario for Ontology-Based Semantic Search Framework

p-medicine platform could contain an Ontology-Based Semantic Framework (OBSF) able to connects highly heterogeneous data and textual information. The semantic framework could be based on gene, tissue, disease and compound ontologies (important for drugs and clinical research frames). This framework could contain information from different organisms, platforms, data types and research areas that is integrated into and correlated within a single searchable environment using search algorithms. It will provide a unified interface for all p-medicine End Users to formulate, explore and identify new information (according to specific preferences and needs) across vast collections of experimental data.

p-medicine's OBSF will combines classical keyword-based search with text-mining and ontologies to navigate large results sets (internal & external) and facilitate information and/or knowledge discovery.

End Users will be provided with an advanced ontology based (Gene Ontology (GO) and Medical Subject Headings (MeSH)) "table of contents" in order to access, explore, structure (quickly) the millions of available resources (PubMed abstracts, news, clinical trials) according to the predefined topics "of interest" (AAL, Nephroblastoma, Breast Cancer, etc.).

13.4 P-medicine portal scenario

The p-medicine infrastructure integrates various tools, services and components, from clinical trial management and virtual organization management, through a security infrastructure and data anonymization, to database integration, ontology-based semantic mediation and the exploitation of data in end-user tools, such as literature mining, GridR and the Oncosimulator as made available to (and reusable by) the user via the workflow environment (according to the DoW). The p-medicine portal allows searching for specific

tools, models, services and data based on their semantic annotations and user generated metadata (e.g. Data Warehouse, Oncosimulator, ObTiMA, tools for education and training). Additionally the users will be supported in extending the functionality of the p-medicine workbench by registering and publishing custom tools and services as well as in using the collaboration tools.

The access to the p-medicine framework will be regulated by a roles and rights management system via the secure p-medicine portal. Unauthorized access will therefore be avoided and the risk of misuse of data within p-medicine will be restricted to people legally bound by contracts to data providers.

14 Certification of tools, software, services and modules

Introduction

Certification (of tools, software, services and modules) related activities have to start in p-medicine and represent one step in assuring a further, wide exploitation and acceptance of the p-medicine platform. Of critical importance is to assure standards based interoperability and messaging between all p-medicine tools, software and services. This represents one of the major requirements for any further certification related activities. Two tasks in p-medicine are mainly enrolled in this process: Task 9.3 and 15.4.

14.1 Good Clinical Practice (GCP) Compliance

According to the European Medicines Agency (EMA) website **Good Clinical Practice** (**GCP**) is an international ethical and scientific quality standard for designing, recording and reporting trials that involve the participation of human subjects. Compliance with this standard provides public assurance that:

- the rights, safety and wellbeing of trial subjects are protected;
- the clinical trial data are credible.

The protection of clinical trial subjects is consistent with the principles set out in the Declaration of Helsinki with adoptions[86]. This is a statement of ethical principles developed by the World Medical Association[87]. Requirements for the conduct of clinical trials in the European Union (EU), including GCP and good manufacturing practice (GMP) and GCP or GMP inspections, are implemented in:

- the Clinical Trial Directive (Directive 2001/20/EC[88])
- the GCP Directive (Directive 2005/28/EC[89]).

p-medicine Clinical Trials related activities, tools, software and services will be in strict conformance with the above EC directives.

[86] http://www.wma.net/en/30publications/10policies/b3/ July 2011

[87] http://www.wma.net/e/ July 2011

[88] DIRECTIVE 2001/20/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 4 April 2001, Official Journal of the European Communities, 2001

[89] COMMISSION DIRECTIVE 2005/28/EC of 8 April 2005, Official Journal of the European Union, 2005

15 Scenarios for Education and Training

Introduction

The p-medicine project will create a set of tools that will challenge and inspire the medical community. The use of these tools in a clinical setting is what will bridge the gap between technological development and patient benefit.

In order for the p-medicine tools to be used successfully it is vitally important that end-users are properly educated and trained. The end-users will range from clinicians to patients, from basic scientists to data managers. All of the educational tools will be developed putting the needs of these extremely varied end-users first.

15.1 Help patients understand the IEmS

"Patient empowerment" is a relatively new concept and will be quite daunting for a number of patients (as well as doctors). Patients will have varied educational backgrounds, different levels pre-existing knowledge, different psychological states as well as differing levels of interest in becoming "empowered". These factors will combine to create a very challenging environment for the educational tools to function within.

15.2 Teach health care professionals when to use the p-medicine tools

As well as teaching health care professionals when to use the p-medicine tools, it is vital to teach which tool to use and how best to make each tool work to bring most benefit to the patient. Ensuring a high level of competence within the medical community will ensure patient benefit is demonstrated on a continuous basis.

15.3 Impart understanding of the p-medicine environment

The increasing pace of technological advances has resulted in the majority of physicians being unaware of the possibilities of what modern IT can achieve. Educational tools will be developed to ensure that the medical health community are aware of today's possibilities and feel comfortable with the language and interactivity. Vital importance will also be placed on users of the tools having confidence in the background technology and security elements of the p-medicine environment.

15.4 Scenario for Education and Training

Educating end-users in how to best use the tools created by p-medicine will be vital to their continued use and success. The eLearning tools will be designed with the end-users' needs in mind. Different user-groups will be using different educational tools therefore a different set of user requirements will be identified for each tool.

A different educational tool will be required for each of the tools created by p-medicine, these tools will need to be populated with fake, but realistic data to allow the end-users to practice and demonstrate competence. Each educational tool will be created in close cooperation

with WP15 to contain an inbuilt validation process. The educational tools will be hosted on ecancer.eu as well as the p-medicine website and will be annotated to the corresponding tool within p-medicine environment.

Each tool will contain an end-user data capture introduction with a short pre-test to determine pre-existing knowledge followed by the educational content. Users will then have a practice environment with a final competence and validation requirement. An automatic reminder will be sent out after completion to help ensure retention of knowledge and competence leading to patient benefit.

15.5 Educational tools requirements

The educational tools produced within p-medicine will be web-based and hosted on a dedicated area of the ecancer.eu website. Ecancer.eu is a completely open access site and will give the educational tools the largest possible audience. The tools will be linked to from within the p-medicine environment so that users are able to find the required educational tools easily.

The educational tools will encapsulate a blended approach to learning and will include video content, narrated animations as well as a “mentoring service”. It is the aim for all of the tools to be SCORM compliant as well as EACCME accredited, however user needs will not be hampered in order to achieve these goals. It is our aim to host “lite” versions of elements of the completed p-medicine tools in order to allow users to test their competencies on dummy data within the larger educational environment.

16 Evaluation and Validation of Scenarios

Introduction

The evaluation and validation of the *p-medicine* infrastructure will be implemented in accordance with GCP, ISO and IEEE standards and criteria. The process of testing involves both users and developers.

The *p-medicine* platform has many goals that are different for each category of end-user (clinicians, data miners, bioinformaticians, statisticians, etc.), thus the achievement of these objectives will be evaluated based on realistic scenarios (previously listed in this document).

Due to the high complexity of the *p-medicine* platform, we provide several evaluation/validation examples that should be used as templates to be adapted at the level of each module:

1. Anonymize local clinical databases and upload them in *p-medicine* environment (example of workflow scenario)
2. Statistical Analysis of cancer samples with associated gene expression data and clinical features (example of VPH Toolbox scenario)
3. eCRF Developer for Prospective Clinical Trials (example of ObTiMA scenario)

We want to highlight that the evaluation of the *p-medicine* platform will be an iterative process where scenarios and evaluation procedures will evolve as new components get integrated in the environment or as some others are removed.

At this stage the validation process can be seen as:

- Reproducibility of published results
- Comparison with existing databases
- In house experiments
- Software/tools will be tested for checking the correct operation of each planned feature

At the present phase of the project the scenarios and the evaluation criteria can also be used as guidelines for developers to focus towards actual and immediate end-users' needs.

16.1 Workflow Scenario: Anonymize local clinical databases and upload them in *p-medicine* environment

The goal of this scenario is to evaluate the capability of the pseudonymization/anonymization tool in the *p-medicine* infrastructure to upload and store a clinical database in the *p-medicine* environment.

16.1.1 Evaluation process

Input: Nephroblastoma database

Required tools:

- VO/authorized user with account having write access
- *P-medicine* mirror database ready to accept data
- Interface for uploading data
- Anonymization tool, available through the *p-medicine* platform

Expected Output/Results: the Nephroblastoma database should be anonymized/pseudonymized, uploaded to the *p-medicine* platform and will be visible through the interface.

The evaluation process has to verify that:

- Personal data will be anonymized
- Clinical database is stored in the *p-medicine* environment
- Information is visible only by authorized people (with proper access rights)
- The authorized user can access anonymization/pseudonymization tool
- The authorized user can select the fields in the database to be anonymized/pseudonymized
- The authorized user can successfully upload files to database after anonymization of personal data
- The files are visible after the upload process

The authorized user sets the access rights to data (for other members of the same virtual organization)

16.1.2 Validation Process

Test procedures will be designed and documentation will be produced to formally describe these procedures in accordance with the established practices of software quality assurance. The anonymization tool will be tested to verify its correctness.

16.2 VPH Toolbox Scenario: statistical analysis of cancer samples with associated gene expression and clinical features

This scenario is an example of basic research conducted with mRNA expression and associated clinical data. It will be used to test the ability to use R to conduct a statistical analysis in the VPH Toolbox *p-medicine* environment as described in section 7.1.4.

16.2.1 Evaluation Process

Input data

- CEL files downloaded from http://www.ncbi.nlm.nih.gov/geo/
- Survival and Clinical data downloaded from http://www.ncbi.nlm.nih.gov/geo/ and manually curated
- Set of R commands to be executed in the R workflow

Required tools

Microarray database connected to the *p-medicine* environment
Interface for the workflow editor with "R-template" workflow
Web-service accessible R server (the download process can be also part of the R commands set)

Expected Output/Results:

- Reproduce the figures and results of the scenario description available on the p-medicine.eu web server, like in the example reported below:

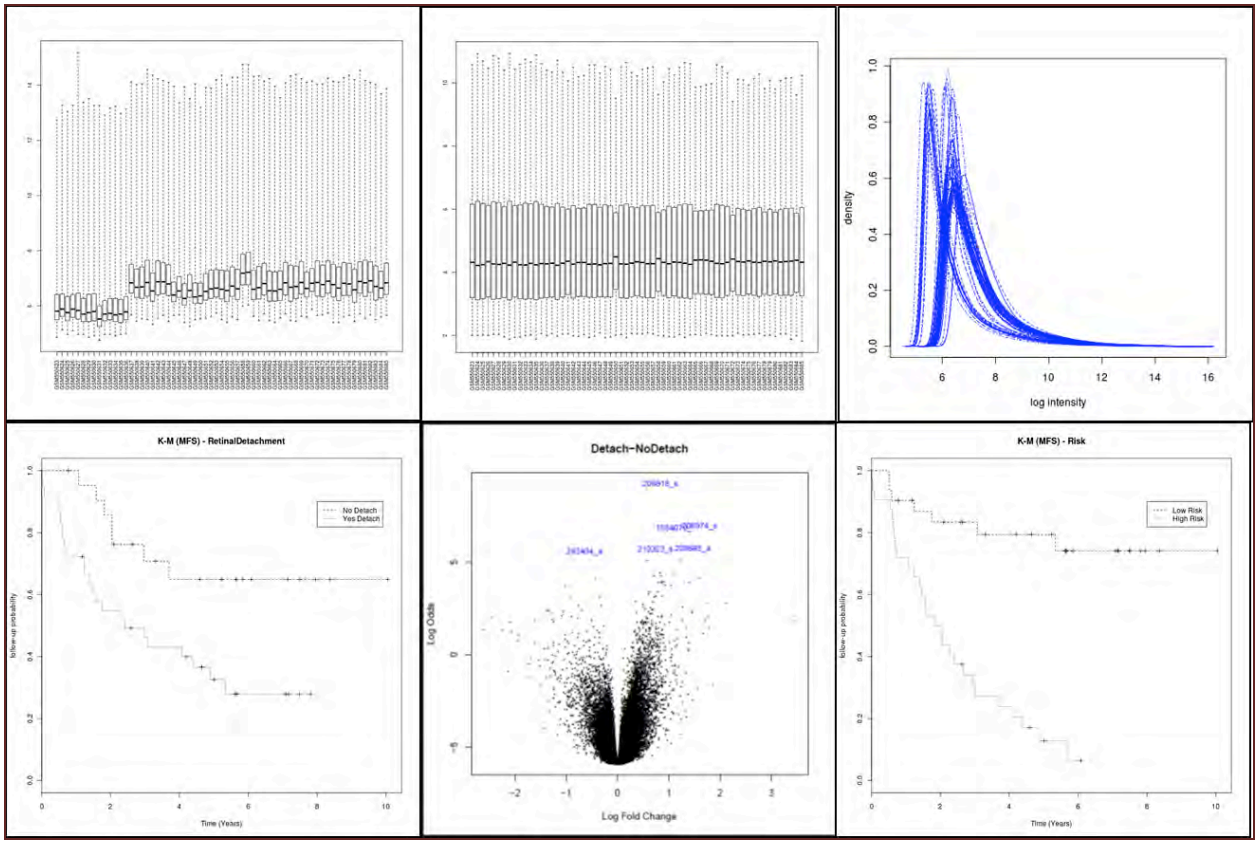


Fig.11.1: The analysis of metastatic and non-metastatic melanoma primary tumors. Starting from the left upper panel: un-normalized samples, normalized samples, intensity/density plot, survival analysis (MFS), volcano plot, survival analysis based on a risk index computed by combining the expression of the top differentially expressed genes between metastatic and non metastatic samples.

The evaluation process has to verify that:

- It is possible to connect to the *p-medicine* portal successfully
- The authorized user can open the workflow editor with R-workflow template
- The authorized user can fill in the commands to be executed
- It is possible to execute the workflow and visualize the results (e.g. survival curves)

16.2.2 Validation process

Compare the outcome of the workflow with the expected result as described in section 7.1.4.

16.3 Obtima Scenario: eCRF developer for prospective clinical trials

The use of eCRFs instead of paper based CRF would enormously increase the efficiency of the infrastructure and the treatment of data in general. The goal of this scenario is to evaluate the capability of the *p-medicine* environment to create ontology based eCRFs in a user-friendly manner.

16.3.1 Evaluation Process

Required tools:

- Authorized user with account having write access
- CRF database with form input as described in Scenario 7.2.4 for an existing clinical database
- *P-medicine* graphical interface to database

Input data

- Items needed for a CRF for a given trial
- Possible re-use of a previous created forms
- Thesaurus for the item - controlled vocabularies available from the Enterprise Vocabulary Services (EVS) (https://cabig.nci.nih.gov/concepts/EVS/) providing a semantic integration of the many diverse medical terminologies
- Input from external tools: CTC, MedDRA, TNM staging system

Expected Output/Results:

- The item will be mapped to the ontology
- An user-interface to access fields in a clinical database linked to *p-medicine* is available

The evaluation process has to verify that is possible to:

- Connect and login into *p-medicine* as authorized user account
- Login into the eCRF designer
- Upload files to database
  (Master Ontology (MO), Enterprise Vocabulary Services (EVS))
- Create a new item on the eCRF designer, that is connected to the Ontology
- Set metadata to the item on the eCRF designer
- Open data entry form preparation tool
- Enter data
- Save the design of the data entry form
- Associate the data entry form to the related clinical database
- Open clinical records

16.3.2 Validation Process

Change some fields and verify that the changes are permanently recorded to the database and logged in log file.

17 Conclusion

Introduction

The main goals of this deliverable were to identify, to elaborate and to specify the end-user needs and requirements for the proposed p-medicine project technological and clinical research infrastructure/frames. One of the core and successfully realised activities was the requirements analysis - a continuous process due to complex technological interfaces and modular infrastructure of the p-medicine platform. In this context, the main focus of all enrolled project partners was to record requirements in the form of '**use cases/scenario**' with process specifications insights.

Additionally, this deliverable has reflected in details the general architectural and technological vision of the proposed for implementation p-medicine platform. In this context we would like to mention in special the bellow sections related to Architecture and IT-Components Scenarios.

Despite the initial exclusive focus on end-user needs and requirements one of our finding was the need to describe in details the p-medicine platform architecture, workflows and interoperability specifications. It represents one of the important achievements of this document, which would serve as a 'guideline' for further project activities. As a result, the section named "Introduction and Project Background" could be easily aligned to other project Work Packages as a "template" for further project activities.

17.1.1 Use Case Scenario Success

The active enrolment of all responsible project partners was as well one of the major achievements of this deliverable, all submitted use-case scenarios have been published in the frames of this document and all received use case scenarios serve as a valuable background for the identified and underlined end-user needs and requirements.

All received use-case scenarios have been actively discussed and only the mutually agreed versions have been published. Project partners demonstrate a strong commitment and an active enrolment. The p-medicine project is an interdisciplinary and collaborative research activity and one of the learnt lessons is – we are in the right direction with high skilled and experienced partners.

17.1.1.1 Clinical use case scenarios

p-medicine is a clinically driven research project and clinical scenarios are playing the central role. Despite the specific focus on ALL, Breast Cancer and Nephroblastoma the published clinical use-case scenarios should be treated as expandable to other healthcare domains.

Clinical use case scenarios could be identified by taking into account the Use Case Scenario template (Appendix 4). Currently all use-cases are aligned to:

- Acute Lymphoblastic Leukaemia
- Breast Cancer
- Nephroblastoma
- other Cancer (with related specifications)
- Non-Cancer Domain (with related specifications)

17.1.1.2 Research use case scenarios

Research use case scenarios are represented by interdisciplinary use-cases and encompass such for implementation modules and all kinds of IT solutions including such identified in the frames of other European research projects. A special attention has been paid to VPH (and VPH Toolbox) Scenarios, Oncosimulator Scenarios, DoctorEye Scenarios and Biobanking Scenarios. Here we would like to emphasise once more the support and professional contribution received from all project partners.

The presented research use-case scenarios assure us that we have established a sound and state-of-the-art background for further project activities.

17.1.1.3 Other use case scenarios

Some presented use case scenarios are related to technological and/or ICT domain. As it is has been mentioned in 'p-medicine IT-Components Scenarios' chapter: "The technical infrastructure and security framework of the p-medicine platform will be built in accordance with legal and ethical regulations to guarantee an infrastructure that will be able to serve other VPH projects." Scenarios from this section have been presented in detail in order to guide further IT related activities.

Of high importance for further project's activities are the use cases related to Security, Education, Evaluation and Patient Empowerment scenarios. All will be continuously updated over time. In this context, the current document is a first and ambitious attempt to underline our activities.

17.1.1.4 ObTiMA and related Scenarios

ObTiMA scenarios have been presented in detail in the Section "ObTiMA Scenarios". At this stage we would like to conclude that all project partners have high expectations and ambitious requirements related to the current functionalities and possibilities of ObTiMA as a software. As a result we concluded that ObTiMA will be further developed in strong linkage with the received use case scenarios. It will allow us to revise the current functionalities of ObTiMA, to refresh the software development workflows, requirements and to identify the frames for enrolment of p-medicine technological partners in ObTiMA software developmental process. ObTiMA could serve as a central and state-of-the-art main component of p-medicine platform and further deliverables and activities can be aligned to this important conclusion.

Due to restrictions in time and available resources we have to conclude that some other ObTiMA related use case scenarios have not been included in the final version of this deliverable. Nevertheless, we would like to remind all project partners and reviewers that further ObTiMA related use case scenario will be published in the frames of next versions of this deliverable according to the agreed DoW (D2.6 - Regular update of the user needs and requirements based on evaluation and validation: Updates in M24 and M36).

The identified, elaborated and analysed ObTiMA use case scenarios are in details presented in the related chapter. All will serve as a background for further p-medicine platform functionalities.

The above use case scenarios are proposed for elaboration in close partnership with technological partners of p-medicine and will be published in the frames of the regular updates of the user needs and requirements based on evaluation and validation.

17.1.2 End-User Needs Challenge

The process of identification the end user needs was the most challenging activity of this deliverable. It required full use of the abilities, resources and professionalism of our partners. We could conclude it as being an "ambitious task", but as result of the analysis of the received use case scenario and active/continuous scientific literature review we had concluded that it was impossible to present from the very beginning an advanced and “complete” description of “all” End User Needs (EUNs). As a solution we proposed versioning control of the identified EUNs. This document presents in details only the first version of EUNs and, additionally, a flexible workflow for establishing and identifying further EUNs is proposed for implementation.

17.1.2.1 EUNs Workflow

The process of identification and elaboration of the EUNs is described in details in the related chapter ‘The identification of User Needs’. It is important to mention that we proposed two distinct versions for different end users and in special:

- ‘Flexible approach’ – applicable for Patients EUNs, and
- ‘Versioning approach’ – applicable for healthcare providers, researchers, Clinical Research Organisations EUNs.

This will assure us to design a robust, flexible and state-of-the-art platform oriented on EUNs and, as result, it will widely accepted by the targeted end users.

17.1.3 Linkage to other Deliverables and Work Packages

There are linkages to other Work Packages, where the identified and presented use case scenarios will serve as an approved and mutually agreed “starting point” for developments and research activities. We are proud of being able to present from the very beginning ‘a general view’ of research activities This was only possible by a close and successful collaboration with all project partners.

Appendices

Appendix 1 – Abbreviations and acronyms

AE	Adverse Event
AERS	Adverse Event Reporting System
AHIC	American Health Information Community
ALL	Acute Lymphoblastic Leukaemia
ARI	Access to Radiology Information
ASTM	American Society for Testing and Materials
BA	BaselineArea
BBMRI	Biobanking and Biomolecular Resources Research Infrastructure
CA	CriticalArea
CCD	Continuity of Care Document
CDA	Clinical Document Architecture
CDA R2	Clinical Document Architecture Release 2
CDASH	Clinical Data Acquisition Standards Harmonization
CDF	Cumulative Frequency Distributions
CDISC	Clinical Data Interchange Standards Consortium
CDS WG	Clinical Decision Support Work Group
CDSR	Cochrane Database of Systematic Reviews
CHG	Charge Posting
CIS	Clinical Information System
CPI	Consistent Presentation of Images
CRO	Clinical Research Organisation
CS	Communication Server
CSF	Cerebrospinal Fluid

DEISA	Distributed European Infrastructure for Supercomputing Applications
DICOM	Digital Imaging and Communications in Medicine
DSS	Decision Support Service
DSTU	Draft Standard for Trial Use
ECRIN	European Clinical Infrastructure Network
ED	Evidence Documents
EDC	Electronic Data Capture
EEA	European Economic Area
EGEE	Enabling Grids for E-sciencE
EM	Expectation Maximization
EMA	European Medicines Agency
EMD	Earth Mover's Distance
EMR	Electronic Medical Records
ENCCA	European Network for Cancer in Children and Adolescents
ESFRI	European Strategy Forum on Research Infrastructures
EHR	Electronic Health Record
EMA	European Medicines Agency
EUN	End User Need
EVCTM	EudraVigilance Clinical Trial Module
EVPM	EudraVigilance Post-Authorisation Module
FDA	Food and Drug Administration
GCP	Good Clinical Practice
GMM	Gaussian Mixture Modelling
GRID	Distributed parallel computing
GUI	Graphical User Interface
HHS	Health and Human Services
HIS	Hospital Information System

HITSP	Healthcare Information Technology Standards Panel
HL7	Health Level Seven
HPC	High Performance Computing
ICSR	Individual Case Safety Report
IdP	Identity Provider
IEEE	Institute of Electrical and Electronics Engineers
IHE	Integrating Healthcare Enterprise
IHTSDO	International Health Terminology Standards Development Organisation
IRWF	Import Reconciliation Workflow
ISO	International Organization for Standardization
ITI-TF	IT Infrastructure Technical Framework
KIN	Key Image Note
KLD	Kullback-Leibler Divergence
LIMS	Laboratory Information Management System
LOINC	Logical Observation Identifiers Names and Codes
MAUDE	Manufacturer and User Facility Device Experience Database
MedDRA	Medical Dictionary for Regulatory Activities
MoU	Memorandum of Understanding
NAV	Notification of Document Availability
NCI	National Cancer Institute
ODM	Operational Data Model
PDI	Portable Data for Imaging
PDQ	Patient Demographics Query
PGP	Presentation of Grouped Procedures
PHR	Personal Health Record
PI	Principal Investigator
PIR	Patient information Reconciliation

PRACE	Partnership for Advanced Computing in Europe
PWF	Post-Processing Workflow
RIM	Reference Information Model
RWF	Reporting Workflow
SA	StudyArea
SaaS	Software as a service
SAE	Severe Adverse Event
SINR	Simple Image and Numeric Report
SLO	Single Logout
SNOMED CT®	Systematized Nomenclature of Medicine Clinical Terms
SOA	Service Oriented Architecture
SSO	Single Sign-On
SUSAR	Suspected Unexpected Severe Adverse Reaction
SWF	Radiology Scheduled Workflow
TCE	Teaching File and Clinical Trial Export
TOB	Trial Outline Builder in ObTiMA
VO	Virtual Organization
WSDL	Web Services Description Language
XDM	Cross-Enterprise Document Media Interchange
XDS	Cross-Enterprise Document Sharing
XDS-I	Cross-Enterprise Document Sharing for Imaging
XPHR	Exchange of Personal Health Record Content

Appendix 2 – Context scenarios

Context Scenario (Bioinformatician)

Context of use	Dialogue principle	System requirements
Introduction S is a bioinformatician working at the Swiss Institute of Bioinformatics. She is involved in statistical analysis in high-throughput biological experiments (microarrays, cancer related patient samples assayed by using DNA (SNPs and methylation), RNA (mRNAs and miRNAs) and relative mutation data in the context of large international projects. She participates in the integrated project **p-medicine** (From data sharing and integration via VPH models to personalized medicine) funded by the European Community's 7th Framework Program. The goal of the project p-medicine is to create a service-oriented clinical infrastructure from current medical practice to personalized medicine. It shall bridge the gap between treatment given to patients and research to find better treatment for patients. S analyses some kind of data, e.g. some genomic data from micro array. These data are related on clinical patients' data. With these data she can imagine the survival and variables describing the patients. She analyses large omics datasets with related clinical-pathological variables like stage, age, gender, follow-up, etc.	ISO 9241 Ergonomics of human-system interaction – Part 110: Dialogue principles Suitability for the task Self-descriptiveness Suitability for the task Suitability for the task	The goal of the project **p-medicine** is to create a service-oriented clinical infrastructure from current medical practice to personalized medicine. An open, modular framework of tools and services will be implemented for efficient secure sharing and handling of large personalized data sets. A workflow should support the bioinformatician to analyse biological, genomic and clinical data in an efficient and effective way. Therefore the analysis tool should provide a structure of all available functions so that the user has the possibility to search for the appropriate function without losing much time. On the other side the user has to understand the various types of data which related on clinical patients' data. The bioinformaticians have to know the appropriate statistical methods for the analysis. The algorithms should be self-descriptive to know which functionality is

		needed for the analysed data.
The analysis concerns the following five steps: 1. read the raw data from „ncbi/geo“ in the classical Affymetrix .cel format; 2. check the data quality with the tool R and make several plots and measures; 3. normalize the expression data, extracted by reading the .cel raw data files; 4. filter the Affymetrix probe sets based on the variance of the signal through the samples; 5. analyze the omics data in relation to the clinical-pathological variables in order e.g. to extract the genes differentially expressed between stage I and stage II samples. S's work is not mechanically, it differs from task to task. Additionally she contributes on writing project proposals concerning data analysis and quality control.	Suitability for the task Self-descriptiveness Controllability Self-descriptiveness Suitability for the task	The user should be able to import the raw data and analyses in a self-descriptive and easy way. For the user interface of the system this implies to provide the user with the corresponding functionality, to lead her with simple guided actions to reach her goal in a satisfied way. The system should support the user in her different tasks in form of self-descriptiveness and reducing misleading actions. The system should support the user in writing project proposals concerning data analysis and quality control.
Assumptions S uses R and Bioconductor as analyse tools. MatLab, Perl and C++ are also used tools. For survival analysis are nice programs like Stata or SPSS. The problem with these programs are: They are designed to do different tests but they actually want to do something particularly different. They would like to write their own codes but they restrict because of the platform they need particular request, particular problem to solve, or recreate their procedure and also the costs. R is free, it is open source. They need more flexibility.	Suitability for the task Conformity with user expectations Controllability	The user uses different tools for analysis. These tools present not the required flexibility to conduct the task in an efficient and effective way. The user wanted to write her own code to solve a particular problem or to recreate her procedure. The system should support the user in these activities efficiently.

They have to deal with custom platforms of data, so they need to work with platforms which are not commercial maybe they want to use one environment to analyse the data and not exporting, importing, exporting and so on from one tool to the other. To avoid mistakes as much as possible.	Suitability for the task Conformity with user expectations	The user expected an environment to analysis the data in the whole and not exporting, importing, etc. from one environment to another. This helps to avoid mistakes as much as possible.
To read the data with a program, to analyse the data with another program. It is better to read the data and manage it inside the same environment to avoid mistakes. It is not good to manage the data with different programs. R scripts can be reused, exchanged, modified, etc. E.g. if they write a matrix with R for the first top left label and they open the file with Excel they have to remember that they don't have the label of the first left column, because actually if they don't shift it the samples will be misclassified (example: resistant and sensitive patients can be wrongly classified, potentially a sensitive sample can be wrongly classified as resistant and vice versa). This kind of mistake must be avoided.	Conformity with user expectations Suitability for the task Error tolerance	To work efficiently it is necessary to read and manage the data with the same program to avoid mistakes, too. With the analysis tool R many activities like reuse or exchange or modify the script can be handled. When writing a matrix with R for the first top left label and open the file with Excel there must be a hint or message by the system to shift the label otherwise patients are misclassified. This kind of mistake must be avoided.
Mistakes are reported by the system. Economical steps, after reading the example, quality checks quite automated. Graphical presentations of the results in form of plots are available in R.	Error tolerance Self-descriptiveness Suitability for the task	Mistakes should be reported by the system in a comprehensible and clear way. If possible the user should get additional information about the error message and its correction on request. Error messages should be written in the language of the user and not of the designer. The graphical presentation should be self-explained and annotated, so that the user needs no more explanation.
Routine activities		

Usually S uses the tool R for analysis. When she has got the data from a particular resource, called X, e.g. a public data set from GEO she has to load it. In GEO they can find also the related publication. So, in this publication she finds the link to the clinical data and maybe in GEO itself, it depends and e.g. X is the environment, the kind of variable in R, it is called environment and so related to this variable called X she has several sub variables (sub environments), one the expression from GEO, one for the clinical data, another one is the annotation in that way she has to use much memory. But to avoid mistakes (typos) everything is linked to her environment X.	Suitability for the task Conformity with user expectations	To analyse the data the user has to load the data into her environment. The user uses several sub environments which all have to link into her environment. There must be a possibility to link all data into her environment without losing data. For large data enough space must be available. The system has to check it before.
What she need is when she has to write an extraction of the clinical data to the corresponding clinic to refer all data to the same variable to avoid errors as less as possible. That is her way to organize her work. In other words it means to type a variable not five times in the same line only one time to reduce mistakes and everything is linked to this variable. They never receive sensitive patients' data, only an ID per sample. Sometimes, users are working in parallel; they use Twiki to support documentation that need to be written by several users.	Suitability for the task	The system has to support her in avoiding much typing of the same variables in the same line and everything is linking to that variable. The problem is that they receive only IDs for patients' data. Documentation should also be available when working with several users in parallel without using a different system.
A simple workflow for analyzing Affymetrix expression arrays in R / BioConductor are described in the following steps: Step 1: loading the clinical data (load packages Affymetrix pre-processing and two-color pre-preprocessing; differential expression Step 2: import "phenotype" data, describing the experimental design Step 3: RMA normalization and expression summary	Suitability for the task	The user must be able to load the data into the system or import data and describe the experimental design, start an RMA normalization and write an expression summary.

Step 4: identifying differentially expressed probe sets Step 3 and step 4 are done in R / Bioconductor. The workflow uses RMA from the affy package to pre-process Affymetrix arrays and the limma package for assessing differential expression. Clinical data is downloaded from: http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM550623 and ftp://ftp.ncbi.nih.gov/pub/geo/DATA/SeriesMatrix/GSE22138/ to extract the complete table. Raw expression data can be done directly in R, or by downloading the file GSE22138_RAW.rar from http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE22138 Results are usually saved as an rda file. The rda file is an R format to save data and variables. S can directly load these files in R. In this way the data are ready to work with because they are already in the format needed to be input for the developed algorithms. Results are reported to third parties or internally. The report is a combination of explanation of the analysis steps, intermediate and final results, visualized through plots, tables etc.	Self-descriptiveness Suitability for the task	The system should support all R format files (rda) to save data and variables, so that the user can work with them and use them as input for the developed algorithms. The user must have the possibility to report the results to third parties or internally in an easy and understandable way. The report is a reflection of the analysis steps and their explanation. It should be possible for the user to write it in an efficient time.
The feedback can be from the project manager, colleagues and/or the collaborators. The first is much more concerned on analysis details and evaluation of the results, the latter usually is much more concerned in understanding what has been done and why and if the results are as expected, he asks about new and much more powerful results visualization method.	Conformity with user expectations	All analysis stuff should be in an understandable way that it can be controlled by third parties.

They maintain the raw data and collect them, data can be shared. They work with categories of patients.	Suitability for the task	The raw data should be shared via the system in an easy and efficient way. The anonymization of patients' data should be guaranteed.
Question: are the results managed by a tool or in your own environment? The results are managed in her environment Usually they provide two kind of report, one internal and one to be sent to the clinicians, (e.g.). The internal one, is more technical but usually doesn't contain code, it contains an executive summary, figures, tables, explanations, questions and answers. The one they share is less technical and it contains an executive summary, figures, tables, explanations, etc. like the internal one. Usually no raw data goes from the bioinformaticians to the clinicians, usually is the other way around. Usually the user receives Excel files of the clinical data and the omics data are in the raw format, for example the .CEL for Affymetrix arrays (HG-U133 Plus 2.0, SNP6.0 etc.) or .txt for Agilent arrays. To connect tasks, tools or components to make a clinical content table ready to be read in R will take few minutes till few hours depending on the special task. The user's problem is to work manually on the clinical data table before importing it in the tools she is using. So, ideally the user would like to have a tool that automatically / interactively check the input clinical data.	Suitability for the task Controllability Self-descriptiveness Suitability for the task Suitability for the task Conformity with user expectations	The user should be supported by the system to provide the various kind of report. To write all results in a representative format and to avoid more explanations the system should support the user in these activities. The system should support the user in loading the different forms of raw data into the system. The import of raw data as .CEL or .txt files should be in a clear and easy manner without waste of time. The user should be supported by the system to make a clinical content table ready to read in R. The user needs support for checking the clinical data table before importing it in the tool. This means an automatic / interactive check before input can be conducted.

Special features during the working process Organisation depends from project to project. The stage of a patient with cancer disease is reported as a TNM (the size of the Tumor, the number of Nodes and the presence of Metastases) stage. The next step is to convert the TNM stage into three separated variables. The standard for the stage is the TNM but there are clinics which provide only two variables, the clinical and the pathological stage and other clinics provide only the pathological stage (only one variable). So they have to add their own defined fields in the Excel table to use them for the analysis (see http://en.wikipedia.org/wiki/TNM_staging_system).	Suitability for the task	The user should have the possibility to distinguish between the various reports of the distinct clinics and to convert the TNM stage into three separated variables. The user should be supported in adding her own defined fields in the Excel table to use them for analysis purposes.
Usually an internal report is sent before sharing the results with the external collaborators. Every colleague reports mistakes to the person that run the analysis. Every person that received the report is in the mailing list and is aware of the mistake. Once everybody agrees, the report will be sent to the external collaborator/s.	Controllability	It must be possible to send the report for verification to colleagues who are listed in a mailing list. In case of mistakes the system should inform the sender and all other involved users. The communication and coordination with colleagues and external collaborators should be easy and clear.
She has to share the data because they work in groups and different aspects of the project, e.g. clinical data. This is common but she works on one kind of data and a colleague works on another kind of data which belong to the same list of patient. They have to combine their results. The project manager takes care of to give the direction of the project. He collects all results inside the group. S and the colleagues share their code of the same project. In the case it is organized that a colleague takes care to it and resolves the problem.	Suitability for the task Controllability	In most projects the user works in groups who analyze different aspects of clinical data. To combine all results of the various groups at the end of the analysis the project manager must be supported efficiently by the system.
They use a server for saving the code to reuse it and not losing time to do many things many times once more. They work with	Suitability for the task	The user must have the possibility to save the code in an efficient way to reuse it and

small components and merge these components at the end. The results are presented in R in form of plots. Usually she makes changes to existing workflows in using new available options or substitution of a step. When creating eScience workflows or analysis processes the following phases exist: 1) Brainstorming 2) From brainstorming to entities and relationship definition 3) From entities and relationships to logical projects 4) From logical projects to physical projects 5) Evaluation by using scenarios / benchmarks	Suitability for individualization Suitability for the task	not losing time when typing the same thing at many times. The system should provide the user in the ability to make individual adjustments to work more comfortable. In the analysis process the user should be supported by the system. An eScience workflow should represent all required steps which should be conducted to reach the physical project.
Organisational conditions The aims are: first of all the high quality of the methodology, the useful experience that the bioinformatician can benefit at the end of the working task, the collaborative environment, the complete answers to the collaborator/s and the opportunity to further develop and/or use the results and the specific methods eventually implemented for the case study. The scripts are reused. It happens pretty often that they have to exchange the code. Changes can be made by updating functions, change parameters, insert new blocks and delete no longer used blocks.	Suitability for the task Suitability for the task	To achieve the working task in an efficient and effective way with satisfaction is the highest priority which a bioinformatician can benefit. The collaboration with other colleagues and groups is necessary to get complete and correct results. To exchange the code should be possible and easy to handle.
She has a clustering of computers for paralyzing. They have an interface for a web code to face the problem with sequencing data to analyzing and be able to paralyze the code. The problem is to have so many different data from a great	Suitability for the task	The system must support the user in managing the variety of patients' data to face the problem with sequencing data and to paralyze the code.

number of patients to manage the variety of data. Usually, tools are already selected, components are defined, much time is devoted to understand the task and how the components can be combined to satisfy the project's aim/s. The dissemination is so important. E.g. S's colleagues from Vitality offer courses for launching the programs to the cluster in order of important instruments To manage the possible mistakes to go to the prevent versions. TWiki is the communication tool for the analysis group. Internal instrument to upload their files and to get important information from their colleagues.	Self-descriptiveness Controllability	If the user has understood the task it must be easy and comprehensible to combine the various components to satisfy the project's aim. The system should be self-descriptive or present the user an excellent documentation. Mistakes should be explained in a clear and understandable way. The user should have the possibility to remove mistakes efficiently without loss of time.
Question: Do you use the workflow top down to generalize the task? S uses the workflow top down to generalize the task. Usually she looks for the type of the project and thinks about the components. Then she tries to split the project into several components, e.g. tasks and faces. In her brain she thinks about a workflow in PowerPoint. The first is to write something by hand. The most time consuming task is the organization, the beginning about several steps and how to do them. A stress factor for S is the same kind of activity to read the manuscripts, a new language in a restricted time. She mainly uses R and Bioconductor. Sometimes the new released packages contain bugs that usually discover the community in few days. So, it is important to be registered on mailing lists to be updated on the package life.	Suitability for the task Error tolerance	*Is it possible to realize S's preliminary plan by hand into the system as a draft?* *How can the user be supported by the system in thinking about the organization and the next steps?* To get all relevant information from the community it is important for her to be registered on mailing lists. The user must have the possibility to use the previous version of the system when discovering bugs or errors which make the execution of the task impossible.

Other comments to critical incidents which already occurred S needs a lot of flexibility due to the kind of tools she is using. She is much more concerned to the fact that she has to work manually on the clinical data table before import it in the tools. So, ideally she would like to have a tool that automatic interactively check the input clinical data. Wishes are: - be able working with Taverna for creating workflows - function, e.g. “abc” to create her own workflow with this function, or to have an icon on the screen; to have the opportunity to open a window in which the data can be imported - repeated steps should be stored and performed again in the same way when necessary - the versions, the releases, to maintain the previous versions - to go back to the previous version	Suitability for the task Suitability for the task Conformity with user expectations	Much flexibility is required to have a tool that automatically interactively checks the input clinical data before the import will happen. To work with the tool Taverna for creating workflows. The system should support her when creating her own workflow either to offer her an icon on the screen or to open a window for importing the data. Whenever analysis steps are repeated there should be a possibility to save these steps to perform them again when necessary. E.g. to have a history for all executed steps. If there are new versions or releases the user should have the possibility to go back to the previous version without getting problems.

Context Scenario (Data Manager)

Context of use	Dialogue principle	System requirements
Introduction MT is the chief data manager at the CRUK Medical Oncology Unit, Oxford. She is responsible for the implementation of clinical trial data management systems and Good-Clinical-Practice (GCP) compliant Standard Operating Procedures. She participates in the integrated project p-medicine (From data sharing and integration via VPH models to personalized medicine) funded by the European Community's 7th Framework Program. The goal of p-medicine is to create a service-oriented clinical infrastructure from current medical practice to personalized medicine. It shall bridge the gap between treatment given to patients and research to find better treatment for patients.	ISO 9241 Ergonomics of human-system interaction – Part 110: Dialogue principles Suitability for the task Suitability for the task	The goal of the project p-medicine is to create a service-oriented clinical infrastructure from current medical practice to personalized medicine. An open, modular framework of tools and services will be implemented for efficient secure sharing and handling of large personalized data sets. A workflow should support the data manager in the implementation of clinical trial data management systems and Good-Clinical-Practice (GCP) compliant Standard Operating Procedures.
MT's responsibilities cover delivery of data management services to support clinical research, in particular • eCRFs for clinical trials of investigational medicinal projects and • clinical databases associated with tissue biobank sample collections.	Suitability for the task	To analyse biological, genomic and clinical data in an efficient and effective way. Therefore the analysis tool should provide a structure of all available functions so that the user has the possibility to search for the appropriate function without losing much time.

A certain amount of common work for each clinical trial is the following: • review protocol, • generate data requirements, • development of eCRFs, • testing, • user and investigator acceptance, • a period of up to a couple of years of prospective data capture, • data cleaning, • data lock, • data export, • linking with externally generated bio data. • summary and analysis, • data archive.	Suitability for the task Self-descriptiveness Controllability	On the other side the user has to understand the various types of data which related on clinical patients' data. Reviewing the protocol should be possible in a self-descriptive way. eCRFs are supported by the system in the way that they are self-descriptive for the user. The user must be able to clean and lock the data. To export the data should be done in a clear and understandable way. To link data with externally generated biobanks should be possible and efficiently handled. The user should have the possibility to archive the data efficiently.
Assumptions MT is a member of the Medical Oncology EarlyPhase and Translational research Trial Steering and Quality Assurance committees. Most data entry is carried out by data entry personnel. Data entry personnel are often a mid-grade administrative level, not medically, statistically, IT qualified; they learn on the job how to interpret and code clinical data specific to their studies.	Suitability for learning	The user should be supported in entering data, so that it is a learning process for him/her.

In some settings data entry may be carried out by research nurses or clinicians.	Self-descriptiveness	Data entry should be executed in a clear and understandable way, so that it can be handled by each person.
Data managers are database developers with expertise in database design and limited statistics.	Suitability for the task	The data manager should be supported in the development phase.
Currently there is limited access, so requests have to be conducted through the data managers who know the data dictionaries, etc. very well. With multiple access, training in interpretation and also data dictionaries would be very important	Suitability for the task	The user must have the opportunity to use training in interpretation and also how to find data in dictionaries.
They use different software depending on the project.	Suitability for the task	The user must be supported in handling the different software to achieve her/his goal efficiently.
Clinical databases are developed using OpenClinica, open source and clinical trial specific software. Site specific databases are developed in File Maker Pro or database software.	Suitability for the task	The import of the different data files should be handled in a self-descriptive way. It should be comprehensible for the user.
Both of these data collections have data sets exported into Excel, then pulled into Stata statistics package for analysis.	Suitability for the task	The system should support the user in handling self-made workflows for statistical analysis.
They use self-made components for statistical analysis, after the data has been (but the file of commands is not re-used) just re-created (generated).		
Repository for new clinical trials, customising designs used previously with most similarities, though there will be more customising for different clinical databases as they are for different tumour sites.	Suitability for the task	Customization for clinical databases must be guaranteed.
Common sets of instructions are not formally standardised within the data team, there are sets of work instructions.	Suitability for the task	Which kind of work instructions?
It would be useful to have a sequence of producing a set of listings and tabulations for a clinical trial report plus a set of	Conformity with user expectations	The user expected a sequence of producing a set of listings and tabulations for a clinical trial

analyses relevant to the trial endpoints. Also routine listings that are required by law, e.g. annual safety reports for clinical trials, end of study report after closure. The sequence is: • select data set, • export to Excel, • load into Excel, • carry out basis set of statistical summaries • tests. This is not standardised to run as one item. Would be useful for site-specific databases.	Suitability for the task Self-descriptiveness	report plus a set of analysis relevant to the trial endpoints. Support of routine listings that are required by law, e.g. annual safety reports. The sequence of the following actions like - select a data set, from where ??? - export (the clinical data set???) to Excel, - load (the clinical data set???) to Excel - carry out the basis set of statistical summaries and - test them must all be self-descriptive. The system should provide explanation and request confirmation before carrying out the corresponding action. This is not standardised to run as one item. Would be useful for site-specific databases.
MT uses electronic clinical care systems, for which there is read only access. Data is re-typed into research databases. Also sets of patient notes used in clinic. Patient notes also hold clinical trial data capture forms, which prompt the clinician to record details required in the eCRF, for example, a toxicity would need to be recorded with the grade of severity according to published criteria (NCI Common Toxicity Criteria).	Suitability for the task	It must be possible for the user to re-type data into research databases. Patient notes must also be prompted to the clinician to get more details, e.g. a toxicity would need to be recorded with the grade of severity according to published criteria.
MT receives the data which varies according to the project • Trial data as indicated in the clinical trial protocol,	Suitability for the task	The user must have the possibility to read the clinical trial protocol in an easy and efficient way without loss of time.

typically • Clinical trial patient demographics, previous medical history, previous treatments for the disease, informed consent, appropriate inclusion and exclusion data, trial drug treatments administered, results of physical examination, vital signs, haematology and biochemistry results, all concomitant medication, all adverse events experienced, data relevant to trial endpoints (e.g. response to treatment, duration of response, survival up to a certain time point), reason off study. Typically a lot of data items over a relatively short period of time, for a relatively small number of patients. • Clinical site-specific databases supporting the biobank: typically less detail for a larger number of patients (thousands). Demographics, details of pathology and spread of tumour at first presentation (known prognostic factors), treatments given, outcomes, relapse/recurrence, survival data.	Suitability for the task	The protocol has to be written in a comprehensible way. The user must have access to the corresponding biobank to get more information about a larger number of patients. The information should include demographics, details of pathology and spread of tumour at first presentation (known prognostic factors), treatments given, outcomes, relapse/recurrence, survival data. This access should be easy manageable.
Components depend on projects • Clinical trial modules/templates are re-used as a new separate trial eCRF is designed, export and summary steps are not currently re-used, just repeated in the new setting • Site-specific databases: selection of export fields for analysis has been automated, with option to amend; statistical analysis steps are repeated to a certain extent		To re-use clinical trial modules as well as to export and summary them must be supported in an intuitive way. Export fields for analysis in site-specific databases should be automated with option to amend. Support for statistical steps which are repeated to a certain extent.
Routine activities Special data tasks are handled as:		The support of eCRFs specific to each individual clinical trial must be guaranteed.

• development of an electronic Case Report Form (eCRF) specific to each individual clinical trial, and produced to standards of ICH GCP (good clinical practice in trials: we need audit trials, data validation, documented generation and resolution of data queries, strict limited and specific address, pseudonymisation, data lock; then statistical summary and analysis). They have 5-10 new clinical trials per year. • data input is prospective, directly typed in after review of clinical care systems and patient notes and clinical annotations. These databases continue indefinitely, and are updated all the time.	Suitability for the task	The user must have in regarding the standards of ICH GCP an audit trial, data validation, documented generation and resolution of data queries, strict limited and specific address, pseudonymisation and data lock. Support of the system must also be guaranteed in statistical summarising and analysis. It must be easy for the user to import the data after review of clinical care systems and have the possibility to make notes and clinical annotations.
The data team includes data managers who • develop and maintain databases, • interrogate the databases, • export data and • carry out relatively simple statistical analysis. Also there are data entry personnel, who are end users and review clinical care systems to interpret data, then input that data into their systems.	Suitability for the task	The data manager must be supported in interrogate the databases and to export data into the own system and to carry out relatively simple statistical analysis. There are end users who review clinical care systems to interpret data, then input that data into their used systems.
A certain amount of common work for each clinical trial: • Review protocol, • generate data requirements, • development of eCRFs,	Suitability for the task	To support the user when he/she gets a clinical trial. It must be easy and conducted in an efficient way to review the protocol and generate data requirements.

• testing, • user and investigator acceptance, • a period of up to a couple of years of prospective data capture, during which data is not available for review elsewhere, • data cleaning, • data lock, • data export, • linking with externally generated biodata, • summary and analysis, • data archive. These data sets would need data validation on input, query generation and resolution recorded electronically, preferably individual record locking, definitely locking of trial datasets. Input quality control with drop-down lists etc.	Self-descriptiveness Controllability Controllability	The development of eCRFs should be provided in ObTiMA. User and investigator acceptance must be guaranteed. The user must be supported in data cleaning and data lock. The export of data and linking with externally generated biodata must be possible and conducted in an easy way. To generate a summary of the results the user should be supported efficiently. The same should be possible for analysis. All data and results should be archived in a self-descriptive way, so that data can be retrieved easily. For data sets validation on input must be guaranteed, as well as query generation and resolution recorded electronically. The user must have the possibility to individual record locking and definitely locking of trial datasets. Therefore a quality control with drop-down lists must be essential.
Clinical databases are maintained long-term, but have a sequence of tasks carried out many times per year for research projects. To review feasibility, searches are required which allow the user to specify multiple conditions (e.g. breast cancer patients who relapsed within five years, with particular combination of adjuvant treatments plus certain tumour characteristics).	Suitability for the task Conformity with user expectations Suitability for the task	For repeating a sequence of tasks in research projects many times a year it must be possible to store such task sequences in clinical databases. The user must have the possibility and be supported to review feasibility. This means that he/she has to search for special situations e.g. that a breast cancer patient who relapsed within f years, with particular combination of

When a patient population has been identified, a routine set of statistical analysis is undertaken to check that the patient population is typical: often a series of Kaplan Meier survival curves are produced and log rank statistics calculated.		adjuvant treatments plus certain tumour characteristics. The user must be supported in identifying patient population to generate a routine set of population which is typical; a series of Kaplan Meier survival curves are produced and log rank statistics calculated.
When research results are available the clinical dataset is exported from the clinical database, to provide data at that time point, and merged with that research data. The merged dataset is analysed to look for associations between known prognostic factors and the research data. Usually several cross tabulations are produced, with chi-squared tests (of goodness of fit) of association. Also multivariate regression analysis and survival curves. Some of these tasks are carried out only depending on results of previous steps.	Suitability for the task Suitability for the task	The actions export and merging datasets should be manageable intuitively. Analysis of merged datasets to look for associations beteen known prognostic factors should be conducted in a simple and efficient way. The results cross tabulations and multivariate regression analysis with survival curves should be presented in a comprehensible way. Repeating steps should be saved to minimize work effort.
For management reports and application, accrual reports some specific searches are useful, into which the date range, disease specifics etc. can be parameters. Relatively few working steps are automated. The role seeks to produce meaningful complete and accurate datasets for collaborating research colleagues. They spend a lot of time capturing the data in a quality controlled way.	Suitability for the task Controllability	The user must have the possibility to make specific searches, in which the date range and disease specifics, etc. can be parameters. E.g. looking for the patient the last three years, what happened to the disease? The steps of generating data requirements and of development of eCRFs should be automated. Support the user in producing meaningful complete and accurate datasets for

Several colleagues could be access the same data set (e.g. trial eCRF database), but not the same record. Their working steps are fairly unautomated, therefore they have the flexibility (though possibly are therefore less efficient). She should be able to generate a workflow by herself. MT has generally good feedback from collaborators, in that requests can be met. In some cases she will need to support the data manager to produce what is required.	Self-descriptiveness	collaborating research colleagues. Several colleagues must have the flexibility to have access to the same data set(e.g. eCRF) but not on the same record in the database. To generate a workflow by herself should be self-descriptive and easy manageable. Correspondence with collaborators should be listed on the system. The user should be supported by the system to produce what is required.
Special features during the working process When there is a request from a researcher, they use a standardised request document in order to clarify and define items of interest up front! Standardised project / data request form considering the database. Can be issues with version control; would be good to manage that in the system, unique names for projects including dates.	Suitability for the task	When there is a request from a researcher, they use a standardised request document in order to clarify and define items of interest up front. This should be provided by the system. Version control would be a better support for managing projects with unique names including dates.
If there is a query regarding data, there is email correspondence to clarify. If data is changed then those individuals in possession of some copy (e.g. original database, exported spread sheet, will each change their current copy manually. If someone is identified as not eligible, they will be removed from the research set, but retained in the source original database.	Suitability for the task	If there is a query regarding data, there is email correspondence to clarify. All individuals in possession of some copy have to change their current copy manually. If someone is identified as not eligible, they will be removed from the research set, but retained in the source original database.

Corrections / exclusions do not happen very often. Quality sign-offs, with a second review, is not very 'e' but would introduce a layer of assurance The one that she might be more involved in than most p-medicine partners is clinical data capture at the start. Clear data definitions and categorisation.	Self-descriptiveness	Is there a possibility to make this process automated? For all p-medicine partners there should be a clear definition and categorisation of clinical data capture.
Organizational conditions The organisational aims are strict adherence to legislation and guidance regarding confidentiality and research principles (subject to government inspections). That concerns security, long-term storage of historical data, traceability and accountability. Mechanisms to control the efficiency of work are partly clinical trial data which is currently handled in a FDA-compliant system Site-specific databases improve with audit trial and tighter control Organisational questions are not so relevant to the data team.	Suitability for the task	Security of clinical data, long-term storage of historical data, traceability and accountability must be guaranteed. Mechanisms to control the efficiency of work are partly clinical trial data handled in a FDA compliant system and site-specific databases
Other comments to critical incidents which already occurred		

Regarding their clinical site-specific databases, which are used to provide diagnostic, treatment and outcome clinical data for use in statistical analysis of experimental data such as protein expression, gene expression. • Our Biobank requires ethical approval to collect samples and associated clinical data for those who have given informed consent. When a set of tissue samples is required for a particular project, a steering committee decides whether to approve the project. Tissues are only released if approved. Similarly the accompanying clinical data is only released if approved. Using either tissue samples or any clinical data should be strictly regulated, so there must be a balance between ease of access for bona fide uses and the risk of speculative use of a data repository that is easily available for unapproved reasons (even if they are good ideas). Hence the linking, downloading merging and different data sets should be tracked; everything should maybe somehow have a project code attached, as well knowing the user. Maybe that the process of exporting a snapshot of clinical data at a certain timepoint, for the purposes of analysis, should somehow be halted until some criteria are met, e.g. completing a form with details of approval by the steering committee? In short it is important to maintain a balance between ease of access and appropriate ethically approved use.		
• Generally clinical trials data is very tightly		

regulated, and also tissue bank related data is becoming more so. It might be prudent to apply the same standards to all types of data in the system.		
• Currently most of our data has to be manually input. I think we are a long way off from having a live feed from the hospital clinical care systems. But there are negotiations around improving our efficiency by obtaining routine electronic downloads, then running a script to match and import relevant data items. This would apply to the clinical site-specific, long term, databases. An option to allow this would be helpful, the event that such downloads come to pass.		
• When we provide clinical datasets for statistical analysis, they already have some additional coding and calculation done from the raw data, e.g. an oestrogen receptor score (value will be between 0 and 8, an integer), will be coded as positive with value>=3, negative otherwise. Maybe some standard universally accepted codings should always be include in the raw data capture modules; this would be more efficient: it would save adding that step into every task of data manipulation for analysis and, once validated once at source, will be less prone to error.		
• There may be a risk that those relatively unqualified in the tasks may misinterpret easily producible statistical summary and analysis. For example could run a task to carry out an analysis that is inappropriate for the data set they selected (may be parametric tests on data that is not normally distributed), and draw false conclusions. Maybe annotations on steps that are carried out in a routine would help? (This test assumes the data is normally distributed and		

the dataset is greater than 20)		
• There is a risk that those producing reports/listings do not fully appreciate the meaning of what they think they are asking for; for example, looking for patients who had a certain procedure versus looking for that procedure (one patient may have had many procedures)		

Context Scenario (Biostatistician)

Context of Use	Dialogue Principle	System Requirements
Introduction MZ is Head of the BFM (Berlin-Frankfurt-Münster) data centre, statistician of the ALL-BFM trial group and statistical consultant to several other clinical trial groups. For a long time he has successfully coordinated the I-BFM Committee for Information Management and Methodology. He participates in the integrated project **p-medicine** (From data sharing and integration via VPH models to personalized medicine) funded by the European Community's 7th Framework Program. The goal of the project **p-medicine** is to create a service-oriented clinical infrastructure from current medical practice to personalized medicine. It shall bridge the gap between treatment given to patients and research to find better treatment for patients. Typical tasks are: data processing, the user gets data from databases, e.g. Access database, Excel files. The source are study databases of files created for the data of biological experiments;	ISO 9241 Ergonomics of human-system interaction – Part 110: Dialogue principles Suitability for the task Self-descriptiveness	The goal of the project **p-medicine** is to create a service-oriented clinical infrastructure from current medical practice to personalized medicine. An open, modular framework of tools and services will be implemented for efficient secure sharing and handling of large personalized data sets. A workflow should support the biostatistician to analyse biological, genomic and clinical data in an efficient and effective way. Therefore the analysis tool should provide a structure of all available functions so that the user has the possibility to search for the appropriate function without losing much time. On the other side the user has to understand the associations among experimental types of data and clinical patients' data. The biostatisticians have to know the appropriate statistical methods for the analysis. The algorithms should be self-descriptive to know which functionality is

import data into his analysis system; analysis of the imported data; mistakes in the data have to be clarified. There are some difficulties which related to e.g. Excel specialities, for incompatibility of fields and data. Planning of studies and analysing of biological data.	Suitability for the task Self-descriptiveness Controllability Error tolerance Suitability for the task	needed for the analysed data. To import data from databases should be executed in an easy and comprehensible way to start the analysis without much time delay. For the analysis the user must have the possibility to use his familiar tools. Incompatibility of fields and data must be clarified before the user starts the analysis. It must be possible for him to correct all incompatible stuff to work with the data effectively and efficiently. Mistakes have to be identified. Error messages should be explained to help the user to correct them. Planning of studies should be possible as well as analysing of biological data.
Assumptions He uses the Access Database or other databases from which he gets the input of clinical data. The most used analysis system is SAS statistical software (SAS is mostly used in the pharmacy industry).	Suitability for the task Conformity with user	The import of clinical data via the different databases, e.g. the Access Database must be conducted in an efficient and easy way. There should be a possibility to use SAS statistical software or the tool R. These statistical tools should be supported by the

In some cases he uses the tool R.	expectations	system.
Routine activities He gets input from different databases from study databases and from biologists which is presented mostly in form of Excel files. This is much time consuming. After importing the data he has to transform the data into his used language. Sometimes he has to transform the data into other used systems. He analyses the imported data. Sometimes he gets the task to prepare the final analysis data for presentations. He makes raw versions for the tables. He uses many SAS statistical software, sometimes R and sometimes also self-written programs. It depends on the special kind of problem. If there is no procedure in SAS, he has to use R or generates a procedure by himself. When there are the same analysis data for different projects, he writes similar programs in SAS to use them more than one time. He can re-use them.	Suitability for the task Conformity with user expectations Controllability Suitability for the task Suitability for the task Suitability for individualization	The import of Excel files should be possible. It should not be time-consuming. The system should support the user in transforming the data from biologists or study databases into his used language. The transformation of data into other used system should also be possible and carried out in an easy and comprehensible way. The input/output data are represented, should be under the control of the user. For preparing the final version of analysis data for presentation the user has to be supported by the system. It should be easy and comprehensible to create raw data for the tables. The different SAS statistical software should be supported by the system The user should also be supported by the system to generate a procedure by himself. The user should have the possibility to save such procedure respectively the workflow for later usage. The re-use of similar programs should be supported so that the user can conduct his work efficiently. Especially to simplify the execution of a sequence of repeated

		commands.
After finishing the analysis, the results can be presented in form of a paper, an administrator report, data monitoring in communities, just only input for a meeting or a presentation.	Suitability for the task Controllability	To present the results of the analysis in different ways and formats the user should have the possibility via the dialogue system to select these different tools.
In the process of data clearing he has to ask the data manager of the trial. There must be a communication he has to understand the problem. The analysis method has to be specified in advance for new prospective trials. Interaction with the clinicians is needed for trial protocols. Biometry is the application of statistics in mathematical and biological sciences.	Suitability for the task Suitability for the task	The communication with the data manager and other colleagues must be guaranteed, especially for clarification of open questions. The system should support the user to gain a general understanding of the problem the user has with the data. To specify the analysis method in advance for new prospective trials the user should be supported by the system. The interaction with the clinicians for trial protocols must be also guaranteed.
When he gets data from different sources or from biologists, there is a lot of work to do to forward it so that he can use it. The patient identification came from the lab and the people who are in the study database. This could be done automatically. If there could be a patient identification that is stored in each of the data sources they will get. This would be a good thing to have. Patient identification is very time-consuming. A kind of standardization would be fine.	Suitability for the task Conformity of user expectations Controllability Suitability for the task	The user should be enabled to work with the imported data efficiently without much time-consuming. The dialogue system should support the user in reducing this data effort. The patient identification came from the lab and the people who are in the trial database. This process could be done automatically. For the user it would be helpful to have a patient identification that is stored in each of the data sources he will get. This would

He has to struggle with names which are the same but mentioned different things, codes for the same thing. The export routines for Access are conducted without problems.	Self-descriptiveness	facilitate the users' work. Patient identification is a very time-consuming task. A kind of standardization would be an auxiliary stuff. Consistency of names is a high demand of the user. The user has to struggle with different names which mean the same thing.
Special features during the working process In most of the analysis stuff he works with scientists together who are involved in specific projects. The reports get the trial coordinator. He writes text and asks for specific analysis which is missing.	Suitability for the task	The communication to other scientists who are involved in specific projects must be possible. The system should support the user in sending the resulting reports to the trial coordinator. For missing analysis the user has to ask the responsible person. All these functionalities should be conducted by the user in an effective, efficient and satisfied way.
Organizational conditions The most consuming time is patient identification and preparation of the data. He is involved in many projects and pipelines of jobs. This is sometimes a stress factor.	Suitability for the task	The most consuming time of patient identification and data preparation should be reduced by the support of the system. The user should be enabled to conduct these tasks in an efficient way. One possibility reducing the stress factor could be to get more support by the system in time-consuming situations.

Other comments to critical incidents which already occurred		

Context Scenario (Chairman and Clinician) taken from project ACGT[90]

Context of Use	Dialogue principle	System requirements
Introduction NG has been a paediatrician in a children's oncology clinic for more than 25 years. His tasks include supervising, managing and directing clinical trials, which involve children across Europe with malignant tumours, especially with nephroblastoma. He is the trial chairman of the SIOP 2001/GPOH trial dealing with the treatment of children with nephroblastoma. He is interested in applying appropriate software to map, manage and utilize clinical trials for everyday use and to register patient data. The tasks of a trial chairman include: 1. administrative tasks and compliance with legal regulations 2. defining new trials or applying these trials 3. designing trials, i.e. graphically combining single events such as to create a trial design. 4. verifying and validate patient data during trials 5. providing (patient) data input into trials 6. management of patient data, analysis of the data, publication	ISO 9241 Ergonomics of human-system interaction – Part 110: Dialogue principles Suitability for the task Self-descriptiveness Controllability	The software should provide the possibility to map clinical trials and to manage and use them in the daily working process. Registration of patient data should be an easy and self-descriptive task. The system should support the trial chairman in performing his administrative tasks and in complying with the legal regulations. It should be easy and efficient to define new trials or to apply them. Later on, the system should help him to design trials and visualize them in graphical form. The verification of patient data in trials should be conducted in a comfortable way. Management and analysis of patient data and publication of trial results should be supported by the system in a comfortable and self-descriptive way. All clinical trials should be handled and

[90] Context Scenario (Chairman and Clinician) from project ACGT http://www.acgt.eu

of trial results and providing advice to participating centres on any questions relating to the trial He may also act as a trial participant. In this case, only task no. 5 will apply. Previously, these trials were manually recorded on paper and stored in a library or database. In the future, however, these trials shall be developed using an appropriate platform and software to support NG and all trial participants in their daily working process and to facilitate their work. They should have the possibility to enter data into the system during the daily working process through RDE (Remote Data Entry). The CRFs (Case Report Forms) needed for this purpose should be easily retrievable and editable for the clinician. Sending the completed CRFs to the trial centre should also be an easy task.	Suitability for the task Suitability for the task Suitability for the task Suitability for the task Controllability Self-descriptiveness	stored by the system. All work handled manually should be supported by an appropriate platform in order to enable efficiency and ease of use in the daily working process for all trial participants. Entering data into the system should occur in an efficient and easy way. Managing a patient in a clinical trial should occur through RDE (Remote Data Entry). This includes entry of all data related to the administrative aspects (described in the trial protocol) for patients admitted to a trial. The CRFs should be easy to call and to use. They should display to the clinician which data he will need for the trial. Data entry should be as simple and efficient as possible.
Assumptions NG has many years of experience in paediatric oncology and haematology and in the diagnosis and treatment of malignant tumours in children and adolescents. This has particularly included		

brain tumours and kidney tumours (nephroblastoma), and also blood coagulation disorders in young people. He has been a member of an IT working group developing software for children's oncology in Germany for 20 years. The disadvantage of this software has always been its low level of acceptance and the way it was used by the users. NG has realized that it is useful to first figure out what you want and what you need and then to get into touch with the software developer, to have him do the programming work and, finally, to reflect on each step in the software. This will give him a clear understanding of what he needs for his work and what he doesn't need. The software is designed to reflect the clinical routines step by step. Based on his experience, he exactly knows what functionalities the developed software must deliver to meet the needs of a clinician. As the chairman and administrator, he is fully responsible for each trial that he created and he is in charge of. He is the only person entitled to assign and distribute the rights to the single trials. 95% of all the patients supervised by NG are included in clinical trials. The care of patients occurs in interdisciplinary cooperation with all	Suitability for the task Controllability Suitability for the task Controllability Suitability for the task	The system should reflect clinical routines step by step. The structure of the system should be self-descriptive and guide him through the system without time loss so that he can perform his work efficiently. The system should automatically recognize the role of the user. The administrator is the only person authorized to assign roles and rights for the individual trials. There must be a good information system and interdisciplinary cooperation with all other specialities, hospitals and institutions. Once pseudonymized, the available results should be sent to these institutions automatically. The encryption and decryption should be realized "on-the-fly". All required tools must be uniform in the sense that the clinician should not need to think about the different types of handling. The software should not make the clinician lose his skills but rather support him in performing his work, enabling him to do

other specialities, hospitals and institutions. What is required are uniform tools. It should be avoided that the clinician stops to reflect on what he is doing and relies entirely on the software so that in case of a software failure, for example, he would not be able to treat his patients any more. Instead of making the clinician lose his clinical skills, the software should support him such as to enable him to perform his work in an efficient way, as if done manually. The clinician must be able to efficiently perform his task, no matter if the software is working properly or not.	Controllability	his work in an efficient way, as if done manually. The clinician must be able to efficiently perform his task, no matter if the software is working properly or not.
Routine activities NG has a double role. He is the trial chairman and, at the same time, may also be a trial participant in other trials.	Suitability for the task	The software should automatically recognize the role of the user when he registers in the system and support the related functionalities.
As an administrator (trial chairman), he arranges the trial, including all the content data such as graphical elements (templates = Case report form (CRF)), and also determines what rights and roles will be assigned and who will have access to the trial. In addition, he draws up the trial protocol specifying all the details of the trial. The treating physician is the only person who has access to the trial for which he himself has provided the patient-specific information. Drawing up the trial protocol should be an easy task based on templates in the system so that the physician will not	Controllability Self-descriptiveness Conformity with user expectations Suitability for the task Controllability	For the administrator, the system should arrange a trial with the entire content data such as graphical elements (templates). This should happen in a comfortable way. All relevant data relating to the patient are stored in the trial protocol. It must be easy to use and understand for any authorized user. The system should support the administrator in defining which rights and roles are assigned and who will have access to the trial.

need to think and care about the actual state of the art regulations and standards. These tasks should be made available by the system automatically using a regularly updated master protocol. It is very important for NG that this feature be supported by the software so that he will be able to work with it efficiently. When creating a trial, the clinician has to focus on questions such as “What will be the objectives of the trial?” "What should the trial be like?” “Which is the content of the trial?” “How will it be organized?”. The software should offer him support for this functionality so that he gets a guideline in form of a master protocol and can access already existing or create CRFs. A graphical implementation of the trial would be useful. A trial contains all patient data which are necessary for the treatment process. In this trial, the treating physician gave a full description of the diagnosis of the patient (child). The treatment methods and the appropriate medication are listed as well. Side effects, Severe adverse events (SAEs) and Suspected unexpected severe adverse reactions (SUSARs) caused by the medications are listed by the treating physician. NG as a trial participant would use the new software to help him perform his task efficiently and satisfactorily. The reporting should be done automatically. Until today, these data have been mainly noted on paper. Reporting is mainly done by postal, fax, mobile and, exceptional, RDE systems. The new system will be designed such as to support NG in	Conformity with user expectations Self-descriptiveness Controllability Suitability for the task Conformity with user expectations Suitability for the task Conformity with user expectations	The treating physician is the only person who has access to the trial for which he himself has provided the patient-specific information. Drawing up the trial protocol should be an easy task based on templates in the system so that the physician will not need to think about his tasks as a trial chairman. These tasks should be made available by the system automatically. The software should support both the role of the administrator and the role of the clinician/ physician. It should automatically provide the user with the corresponding rights and roles. The system should offer the possibility to design trials graphically based on specific events. A graphical implementation of the trial would be useful. The trial protocol should be drawn up in a clear and understandable way to meet the requirements of the different user groups. The chairman would like the program to illustrate exactly what he needs for his trial without him needing to think about details. The functions required for defining his trial are important and should be available in a concise and understandable form. The rights and roles assigned for a specific trial by the administrator should

performing his administrative tasks and managing all of the data. NG would like the program to visualize exactly what he needs for his trial without much reflection. The functions necessary for defining his trial are important and should be delivered in a concise and understandable form. Once he successfully defined a given trial, he defines who shall be entitled to have access to this trial, and in which form this access shall be granted.	Conformity with user expectations Error tolerance	also be defined in a clear and precise way. The software should recognize already during the registration process which role and which functions the physician will subsequently perform
As a trial participant, he needs some functionality different from the functionality that he needs as an administrator. The software should recognize already during the registration process which role and which functions the physician will subsequently perform. The administrator can dedicate the role and rights for new users by choosing from a predefined list or manual modifications.	Suitability for the task Controllability Self-descriptiveness	The registration of all patient data must be an easy and comfortable task for both the administrator and trial participant. All patient data must be provided in a consistent form so that it can be compared with other trials in other hospitals. To ensure completeness of patient data, a checklist must be available so that no important pieces of information will be forgotten and the right decisions for further treatments can be taken.
NG as a trial participant is interested to register patient data in a trial. These patient data include: age gender affliction earlier infections (previous medical history) genetic disorders in the family etc.	Conformity with user expectations Suitability for the task Controllability	Validation of data is essential and should be easily performed by the person who enters this data. The trial interface should always be the same so that the physician can quickly locate and take the same procedure without any need to think about details.
The completeness of patient data is implicitly essential in order not to distort the assessment and evaluation of the data and to enable the right decisions for further treatments.	Controllability	The system should provide the clinician with results for the compliance of his daily work. These results must be clear and understandable for him.
Validation of data is essential and should be easily performed by	Controllability Self-descriptiveness	The results should be presented in different ways. Moreover, the software should be modular

the data manager. An important feature is the interface of the patient trial. Given the large number of trials, the interface should always be the same in order to enable the physician to find the desired trial quickly and to use the same procedure without needing to think about it. The software must deliver results to the physician in order to reduce his workload in the daily working process. Moreover, the software should be modular and extensible so that NG can attach specific modules to the existing software, for example. He builds a clinical trial containing a module of a basic data set, as is the case in other trials. This module is saved in a CRF form that can also be used in other trials. Then there is a module, for example, which sends DICOM (Digital Imaging and COmmunications in Medicine) files or a file for imaging which is used for trial A as well as for trial B and also for trial C. Therefore, the software should be designed in a modular way so that NG can select exactly what he needs in the current situation. When NG is in a clinical trial he would like to be able to extract any data, e.g. a relevant treatment graph, and then set it on a "scratchboard," to import the questions from the statistic module and collect them on the "queryboard". Questions are created	Controllability Conformity with user expectations Suitability for the task Suitability for the task Suitability for the task Controllability Self-descriptiveness	and extensible. All trials should be consistent to enable comparison, better understanding and ease of use. A module which is used for trial A as well as for trial B and also for trial C, for example. The system should support the clinician in handling the different trials. The system should present the treatment graph on the scratchboard. It should support the physician to import the questions from the statisticians and collect them on the “queryboard”. Questions should be generated automatically generated and subsequently sent to the statisticians for analysis. The result is sent back to the physician who can visualize it in the form of a life table or a descriptive analysis. The visualization tool should be implemented also as a “stand-alone” tool that may be used by the statisticians for other purposes as well. For non-statisticians, it is important to collect only those data that they are interested in. They want to work using one

automatically and subsequently sent to the statisticians for analysis. The result is sent back to the physician who can visualize it in the form of a life table or a descriptive analysis, for example. He decides if he wants to get the visualization in such a form. The visualization tool may also be a “stand-alone” tool that may be used by the statisticians for other purposes as well. The clinician or physician can use this tool to see the results of his registered data and will have a broader information base. In the past, the clinician’s only option was to admit data, which was of no use for him because he got all analysis results. For him as a non-statistician it is important to only collect the data and to get the results after the analysis that he is interested in. He wants to work using one workflow only. The results that he wants to see as a life table or as a frequency distribution that can be analysed using the R-analysis tool, or the result is delivered in the form of a bar chart or list. The system should support all different forms of result visualization.		workflow only. Therefore, the system should support all different forms of result visualization. The system should be self-descriptive and controllable.
Special features during the working process The local physician in charge of the patient registers all of his patient data gathered in a trial so that he can either display all his patient data as a whole or can make a selection. The trial chairman of a specific trial can map the trial on the system. The corresponding functionality is described in detail in D2.2. The software has several levels. The lowest level gives descriptions of events. This level can only be executed by an	Suitability for the task	The local physician must have the possibility to display all his patient data as a whole or to make a selection. The trial chairman of a specific trial can map the trial on the system. He must be supported by the system to get an overview of the different existing trials. The different levels must be supported by

administrator. The next level delivers the definition of the trial. On the third level, the physician can register the data on the level of the individual patient. Each patient has its own workflow, i.e. the clinician will be guided through the trial by the branch of the patient. While the trial consists of several branches, the patient has only one branch. This is the branch that the clinician is guided through.	Controllability Self-descriptiveness	the system in a self-descriptive way for the clinician as well as for the trial chairman so that they can perform their work in an efficiently and satisfactory manner. For each level, the user needs to know which are the important input data and how to enter them into the system.
On the third level, in the patient-specific view, the trial chairman, physician or trial doctor can also display the data of the individual patient by clicking on a specific event, for example. The empty CRF will be opened. He can register data, or the CRF has already been completed so that he can again inform himself about the data already admitted. What was the point with this patient? He knows with one click where the patient is and gets an graphical overview about the individual treatment regime for the patient. He can even generate a report that can be used as a doctor's letter specifying the entire therapy of the patient and including all data. This is helpful, as it saves the local doctor a lot of time.	Controllability Controllability	All available events must be supported and displayed to the physicians / clinicians who are interested in these events. This must be a process that is controllable and understandable for every user. It must be possible to generate a doctor's letter based on the entered data for the patient, specifying his or her particular therapy. Therefore a print button should be available. The patient data should be anonymized before being sent to the database.
It is at the highest level (4) that the data analysis is performed. The data are anonymized before being admitted into the database and before a trial will be chosen. The only person entitled to see the data is the person who registered the data, the chairman and persons with dedicated rights to see the data. Furthermore, the local doctor can only display data from his own clinic. In a given trial, all data collected for this trial were available to the trial chairman.	Suitability for the task Suitability for the task Controllability	This anonymization should occur automatically and without the need for the physician to intervene manually or to perform any procedure. The anonymized data can be found in a mirror database containing all the data of the trial database in an anonymized form. The system must make sure that data is read and edited only by the user who is responsible for the patient data. The local doctor can only display data

The trial chairman instructs the system which participating hospitals and patients will be involved in the trial. When logging in into the system the physician is automatically assigned a specific role, whereas in another trial he may have a different role and other rights. Furthermore, the local doctor can only display data from his own clinic. In a given trial, all data collected for this trial were available to the trial chairman. The trial chairman is administrator only for those trials that are managed by himself.	Controllability Suitability for the task	from his own clinic. The trial chairman is the only person entitled to display the whole collection of data in a specific trial. The trial chairman instructs the system which participating hospitals and patients are involved in the trial. When logging in into the system, the physician is automatically assigned a specific role, whereas in another trial the same doctor may have a different role and as a consequence different rights. These facilities must be enabled by the system to perform clinical trials efficiently, effectively and safely.
Organizational conditions A doctor in a clinical trial wants to use the software to be guided through the trial. This has to happen intuitively and, if possible, self-descriptively. The clinical trial should offer a visualization of the results. (Deliverable D2.2) The software should enable all administrative tasks, such as automatic reporting of SAEs or SUSARs to the European database EMEA.	Controllability Self-descriptiveness Suitability for the task Suitability for the task	Going through a trial should be a self-descriptive process involving guided information for the clinician / physician and the trial chairman. The system should provide the possibility to visualize results, particularly for the statisticians. The software should enable all administrative tasks, such as automatic reporting of SAEs or SUSARs to the European database EMEA.

Context Scenario (Senior Fellow Scientist in Clinic)

Context of use	Dialogue principle	System requirements
Introduction K. is a trained biologist with expertise in clinical pharmacology. He manages clinical trials and has since focused on information technology for clinical research using the internet for clinical trial support. He led a project for the evaluation and implementation of Electronic Data Capture (EDC) solutions in the Coordination Centre for Clinical Trials (KKS). In addition he participated in the establishment of ECRIN by analyzing the differences in the national legal frameworks for clinical trials and by development of Standard Operating Procedures (SOPs) for international clinical trials conducted by ECRIN. He is involved in IT based clinical trials support and participates in WP2, WP5, WP6, WP9 and WP17 of the integrated project p-medicine (From data sharing and integration via VPH models to personalized medicine) funded by the European Community's 7th Framework Program. The goal of the project p-medicine is to create a service-oriented clinical infrastructure from current medical practice to personalized medicine. His task in the clinic can be concentrated in IT based support of the conduct of clinical trials at the	ISO 9241 Ergonomics of human-system interaction – Part 110: Dialogue principles Suitability for the task Suitability for the task Self-descriptiveness	The integrated project **p-medicine** (From data sharing and integration via VPH models to personalized medicine) is funded by the European Community's 7th Framework Program. Its goal is to create a service-oriented clinical infrastructure from current medical practice to personalized medicine. An open, modular framework of tools and services will be implemented for efficient secure sharing and handling of large personalized data sets. It shall bridge the gap between treatment given to patients and research to find better treatment for patients. In all specified points the user should be supported by the system. He should get all relevant information necessary for conducting

KKS, especially quality aspects Enabling clinical trial system interoperability by the use of data standards (e.g., CDISC, HL7, ISO) use of patient data from HIS and other care data for clinical research (secondary use) quality management of clinical trials (site audit, SOPs) electronic archiving of clinical trials documentation clinical trial system validation ECRIN data centre group: clinical data management of ECRIN IT support of the management of clinical trials.	Controllability	his daily work efficiently and effectively.
Assumptions As a biologist he has worked with prostaglandin-biosynthesis. Prostaglandin is an active ingredient which affects many processes in the body, e.g. blood vessels, pain situations and infections. Well known is the effect of prostaglandin on the musculature of the uterus and the cervix. A part of the effect of the prostaglandins is to switch on / off certain genes. In clinical medicine prostaglandins may be used as a drug e.g. during pregnancy. Since nine years K. is working on clinical trials at the KKS (coordinating centre for clinical trials). For the Telematic Platform (TMF e.V. Berlin) he has conducted together with the FhG ISST (Bernd Troschke) a RDE-project to identify, select and evaluate clinical trials remote data entry solutions. RDE-solutions were identified to be used by the KKS. Remote Data Entry uses electronic case report forms (eCRFs) for the collection of clinical trials data. After a market analysis and the evaluation of possible software solutions, two different systems were identified and		

purchased by the KKS. In the initial phase, the system MACRO™ was employed by the KKS Cologne and KKS Heidelberg. The system eResearch Network™ was employed by the KKS Düsseldorf and KKS Leipzig. In another TMF-project (Trial Master File) he has created system validation documents and a system validation master plan based on GAMP and regulatory requirements for the validation of RDE systems and other clinical trial software.		
Routine activities His task is to support the IT management of clinical trials and particularly the quality aspects. He specializes in data- and IT-management.	Suitability for the task	To manage clinical trials efficiently the user needs support in all corresponding activities.
In EU projects he participated in the development of the structure of ECRIN network. ECRIN has conducted an examination of the regulatory requirements for clinical trials in different EU countries. Despite the existence of the EU directive 2001/20/EG national differences in the regulatory requirements for the conduct of clinical trials exist, so e.g. the required documents to be filed by the Ethic Commission (EC) can vary. The number of ECs in different countries differs in the European as well. There are e.g. more than 300 ECs in Italy, so that some ECs may have not received a submission in over two or three years.	Suitability for the task	The user needs support for the construction (design) of regulatory requirements that vary in different countries.
K. deals with the legal framework for clinical trials in the EU. This is important for international clinical studies which must consider the different regulations and ethical guidelines. There has been adopted a European unique licensing	Suitability for the task	The user should get support by the system for the legal framework for clinical trials in the EU. He has to adapt the unique licensing system for clinical trials.

system for clinical trials. The EU Directive 2001/20/EC (Clinical Trials Directive) has been implemented in all EU countries into national law, for example in Germany 2001/20/EC was implemented by the 12th amendment of the AMG (German medicine law). There has been built a database that is referenced to the various regulations in different countries.	Self-descriptiveness	The user should get all information that is necessary for the legal framework. He should be guided by the system to inform about the different regulations and directives. The user should have access to the database that is referenced to the various regulations in different countries.
K. works at the KKS which supports the physicians at the University Hospital in the task of conducting clinical tests. For example, SOPs and templates (e.g. for the trial protocol, Informed consent, AE messages, etc.) are made available. For the approval of a clinical trial, a protocol including the informed consent form, insurance confirmation, approvals by the Ethics Commission and by the competent authorities is required. Trial protocol and informed consent form must be signed (this is a requirement of GCP). In addition, the trial physicians must demonstrate to have received appropriate training.	Suitability for the task	The user supports the trial physician in the task of conducting clinical tests. For these activities he should be supported by the system to make e.g. the trial protocol, informal consent, AE messages, etc. available. Support should also be given for the approval of a clinical trial. For the sign of a trial protocol and its consent form there is no electronic support available yet.
Electronic signatures are not yet routinely used in trial documents. In America for example, a password is valid as a signature (21CFR Part11). After the CRF has been filled out with the patient data, the investigator has to confirm the completeness and correctness of the data with his/her signature. This may be done e.g. by an electronic signature with a tablet PC. In this case, the electronic signature done with a pen is only an image and not already an advanced qualified electronic	Suitability for the task	Electronic signatures are a requirement of GCP and that the trial physicians must demonstrate training. **Question** to the developer: Is it possible to make electronic signatures available?

signature.		
Often, not the trial physician (investigator) him/herself collects the patient data, but because of time pressure, an assistant enters the data into the CRF; then the data input must still be confirmed by the trial physician. Because of the necessity of a signature, it is still common practical that all data that will need a signature are archived in paper form. Many other data are already stored and archived electronically. In the future, the use of an electronic document management system to support the management and filing in the course of clinical trials with integrated electronic signature and electronic archiving would be desirable.	Suitability for the task Self-descriptiveness	Entering data into the CRF should be executed into a comprehensible and easy way. It would be desirable to have an electronic document management system for supporting clinical trials with integrated electronic signature.
At the start of a clinical trial, a trial protocol is created. The lead trial physician (lead investigator) is supported by the KKS by information, trial process support and templates. The investigator and additional specialists can make comments on the study protocol and modify the plan as required. After the trial protocol has been accepted and finalised, it will be submitted to the Ethics Committee and the competent authorities for approval.	Suitability for the task Self-descriptiveness Controllability	The system should support the user in providing information and templates for the trial. To comment the trial plan and modify if required should be supported and handled in a sufficient and easy way. To add comments and modify the plan should be handled in an efficient way before accepting and finalizing it by the trial physician. It should be also possible to submit the trial protocol to the Ethic Committee and the competent authorities without loss of time.
At the beginning suitable the trial centres have to be recruited for a trial. In the centres there are the investigators who in turn	Suitability for the task	To enrol a patient should be conducted in an easy and comprehensible way. The trial has to be performed according to GCP

recruit patients and conduct the trial. In different trial centres there may be differences in quality based on the number and expertise of available personal, state of equipment, etc.. But it must be ensured in accordance with GCP that the investigator (trial physician) is able to perform his/her trial tasks according to GCP guidelines, so that he/she has sufficient resources and time available (e.g. that study nurses can work for a trial). On the other hand, for the study the necessary equipment must exist in the trial centres. For data entry using EDC system, there should be an internet connection available. The quality control in the clinical trial is conducted by monitors. If an investigator (trial physician) participates in the clinical trial, then he/she has to be trained on the trial specific issues (e.g. data entry in form of CRF, adverse effects as well as on general issues (GCP).	Self-descriptiveness	guidelines. Also study nurses must be able to conduct the trial. The necessary equipment has to exist. Monitors have to be supported by the system considering the current quality control in the clinical trial. For data entry using EDC system, there should be an internet connection available. The training on trial specific issues of a trial physician who participates in the clinical trial must be guaranteed.
Often about approximately 20-30 clinical centres participate in a trial; but there may be also smaller and larger trials. Each centre receives one Investigator Site File (ISF), in which the site specific documents are filed, stored and retained (e.g. stored in paper form). The sponsor manages the Trial Master File (TMF). The source documents of patients are often manually recorded on paper in list form and the corresponding data is then entered into the CRF. These lists are stored in the Investigator Site File in the centre. As part of the Source Data Validation the monitor controls if the source data are conform to the study data collected by CRF.	Suitability for the task Suitability for the task Controllability	The software should support the Investigator Site File (ISF) as well as the Trial Master File (TMF) to store these documents electronically. Source documents of patients have to be entered into the CRF in an easy way. The user should be guided through the CRF without thinking what to do next.

If patient data is captured electronically without the use of a paper CRF, then the monitor must still check whether the patient has existed, whether the data are collected correctly. Work overload by physicians can result mistakes during trial data collection, which may be prevented or detected by quality control.	Suitability for the task Error tolerance	The monitor has to be supported by the system, so that he/she has the possibility to control the patient data efficiently and effectively. Mistakes should be described in the user's language. The system should help the user to detect mistakes and to avoid them.
What happens when a trial is created? A trial must be created either by a sponsor or the leading investigator (in case of an investigator-sponsor). In both cases, the basis is a research idea. The leading investigator is interested e.g. in a surgical procedure that is better than the standard procedure or the pharmaceutical company is interested in whether their new drug has a better effect or is safety than existing drugs. If a trial is created and it is an interventional trial of a medicinal product it must be conducted according to the GCP guideline. Many aspects must have been considered in the planning phase. The planning of GCP trials will be supported by the KKS. The KKS has e.g. templates for the necessary content of a trial protocol, templates for the informed consent form, cover letter to the Ethics Committee, to the authorities, etc. Since these documents have to be signed, they are in paper form. The investigators need also to confirm by signature that they adhere to GCP (e.g. trial physicians has had a GCP training).	Suitability for the task Suitability for the task	The authorisation of users must be possible. The user must be legitimated to create a trial. In the planning phase of a trial many regulations must be considered according to the GCP guidelines. The system should support the GCP trials in an effective way. It should allocate templates for the trial protocol, for consent forms, cover letter to the Ethics committee, to the authorities, etc. All these activities should be regarded by the system. **Question** to the developer: Is it possible to generate an electronic signature by the system?

After the investigators (trial physicians) have been recruited and trained, the trial can start. The patient data are entered into the CRFs by the investigator (trial physician or by a study nurse). There are projects which use data that are already available in the HIS or in the electronic patient record for clinical research. This is often basic data (e.g. age, height, diagnosis) transferred from the electronic patient record into the trial database. It would be helpful if the transfer of the patient data could take place without repeated writing effort. To ensure a sufficient recruitment of patients it may be useful for the sponsor to consider during the search of suitable patients hospitals abroad where a large patient population according to the inclusion / exclusion criteria may exists. The search for suitable patient populations could be executed by search in a data warehouse or other database with anonymised patient data.	Suitability for the task Error tolerance Controllability Self-descriptiveness	It must be easy to enrol patient data into the CRF. The user should be guided through the form. There should be a connection to the available HIS system or to the electronic patient record for clinical research. The transfer of the basic data should be executed without repeated writing effort, without mistakes and loss of data. The communication of physicians with different hospitals abroad to ensure a sufficient recruitment of patients should be handled in an easy way. A search for suitable patients should be possible in a data warehouse or other database with anonymised patient data.
Then patients are recruited. This will happen as follows: The investigator (trial physician) checks the inclusion and exclusion criteria. There are inclusion and exclusion criteria that determine whether a patient can participate in the trial. The patient has to sign an informed consent form that may be stored in paper form. He is informed from by the trial physician about the trial. This task of the investigator may not be delegated to any doctor or assistant. The physician has to provide the explanation and information in the language of the patient. After the recruitment, the patient will be randomised:	Suitability for the task Self-descriptiveness	The inclusion and exclusion criteria should be offered by the system so that the trial physician can determine whether a patient can participate in the trial. Only the trial physician has to inform the patient about the trial which is selected for the treatment. This responsibility must be guaranteed by the system. If electronic signature will be supported by the system the consent form can also be stored. All explanations and information have to be provided in the language of the patient.

He/she is given a study number and an assignment to a treatment arm.		All medical expressions should be explained in a comprehensible and easy way. The randomization should be possible in an easy way.
During informed consent the patient must have the opportunity to ask questions. He must have the opportunity to leave the trial without fear of negative consequences. He must also be informed that he always has that freedom of choice. If the inclusion and exclusion criteria are fulfilled, the patient has consented and the consent form is signed, then he can participate in the trial and he is randomised.	Controllability	The opportunity must be given for the patient to ask questions before consented to the trial. To leave the trail must also be possible without negative consequences for the patient. This information must be available for the patient and that he/she has that freedom of choice.
At randomisation, there are sometimes two or three arms, e.g. one treatment arm with the new medicinal product and one arm with the standard drug (or standard treatment). Randomisation is supported by use of a software tool. The investigator (trial physician) sends a fax that he has recruited one patient to obtain a randomisation number. With the delivered randomisation number the investigator (trial physician) knows which randomisation arm must be assigned to this number. Then the corresponding treatment of the patient can start. There may be different visits, in which the patient will be analysed and / or treated, e.g. to get a drug or is irradiated. The data of the visits are recorded. Blood samples may be examined, for example the blood of a patient is sent to a central laboratory for further analysis. The data of the analysis is sent back to the investigator for input into the	Suitability for the task Controllability	The randomisation tree should be supported by the system. The trial physician should have the possibility to get a randomisation number for the recruited patient efficiently without loss of time. The visits of the patients and all further treatments should be well documented. The coordination to a laboratory for further analysis, for example the analyses of the patient's blood should be guaranteed. The data of the analysis is sent back to the investigator for input into the CRF. This should be executed in a clear and comprehensible way. The requirement is that the laboratory is certified. The task of certification review is conducted by the monitor.

CRF. The requirement is that the laboratory is certified. The review of certification documents is also part of the monitors' task.		
One problem often is unrecognized side effects. Side effects, which the patient communicates to the investigator (trial physician) are documented by the doctor. But there are a number of adverse side effects that may be not directly associated with the drug or the treatment and therefore may be not taken into consideration.	Suitability for the task	If the biostatistician uses the same system as the trial physician he/she can directly enter the analysed data into the CRF and send it to the clinician. The communication between trial physician and patient considering the side effects should be handled in a direct way. There should be a possibility to document all relevant side effects of the treatment.
Severe adverse effects (SAE), which may result in death and other severe results, must be reported immediately. For the management of serious adverse effects, there is special software in use at the KKS, SafetyNet. The trial physician must evaluate each message of an adverse effect, if there is a connection with the treatment or not. The data of SAEs must be coded according to MedDRA. Ideally the safety management system and data management system are integrated. But these databases or systems are often not yet integrated. There is no common data dictionary. It would be desirable that SafetyNet, MedDRA-coding and data / document management can communicate together. The vision would be to have a common data dictionary as basis for an integrated network.	Suitability for the task Self-descriptiveness	Severe side effects must be reported immediately and communicated. To get a common understanding about the used terminology there should be a common data dictionary available. Requirement for the user is the communication of different tools, for example Safety Net, MedDRA-coding and data / document management. An integrated network of all used tools with a common data dictionary is desirable for the user to conduct his task efficiently, effectively and with satisfaction.
Useful would be software that supports the user in the	Suitability for the task	To support the user in the quality analysis in

quality analysis in clinical trials. For the recruitment of patients it is important to get an idea of which hospitals can recruit how many patients and what may be the number of drop-outs, etc. For example, are there clinics with a large number of patients who suffer from a specific cancer disease?		clinical trials. There should be an infrastructure so that the physician gets information from other hospitals about the number of recruited patients, the number of drop-outs, etc.
All patients undergo the treatment workflow in a trial. When the last patient has been treated; this is called: last patient in, last patient out. It would be desirable if the lead investigator could see, how many patients have been recruited in which centres and in which period were the patients recruited? In case a centre has not been able to recruit the planned number of patients, the question is why have no more patients been recruited? Where is the problem? Is there software that can support this quality analysis and monitoring?	Controllability	The user needs knowledge about the number of recruited patients in which centres and in which period it happened. Questions that the trial physician has must be answered. The system should support the user in answering these open questions. **Question**: Can the software support this quality analysis and monitoring?
When data is entered by the investigator, there is a quality check, a so-called edit check which indicated spelling errors and to some degree incorrect data. In case the physician has entered are incorrect data a query management system is turned on automatically. The system creates a query and asked the investigator for the correct data. This system should guarantee the correct entry of data.	Suitability for the task Controllability	When entering patient data the system should support the trial physician to avoid errors. If spelling errors exist they should be checked by the system. Entering patient data in an incorrect way the query management system should be turned on automatically to guarantee a correct entry of data.
After all patients have been treated, last patient out, then all data are in the database. The database is closed after a quality check has been performed (database lock).	Suitability for the task	The database should be locked after the quality check has been performed. For strange looking relations on the database

The biometrician may run a query over the database for quality control, for example to search for data relations that look strange. Only when the data is clean and correct the database is closed (database lock). After the lock no more data can be changed or added. After the database has been locked, the statistical analysis of patient data can start.	Controllability	the biometrician has to run again a query for quality control. The lock of the database guarantees that no more data can be changed or added.
During the course of the trial, the investigator has collected trial documents (informed consent forms, etc.) and filed in the ISF. All trial relevant documents, even notes and training documents are collected in the TMF. Both files must be archived in Germany for ten years together with the trial database according to GCP regulation. But there may be other relevant regulations to consider: in case radiation treatment is used, for example X-rays or MRI images. These data and images must be archived according to the German radiation law for thirty years.	Suitability for the task	All collected documents, notes must be archived for ten years and X-rays or MRI images must be available in Germany for thirty years. This must be guaranteed by the system. The system should inform the user for the relevant regulations considering X-rays or MRI images.
The analysed data help the lead investigator or the sponsor to write the final study report. The patient as a person who wants to be cured does not have a benefit from the result of a trial, the benefit is with future patients. The patient has the right to see his/her data. Currently, only the investigator has access to the patient data of a trial, but not the patient him/herself. The patient wants to become healthy again and often for many other things he does not care. One problem may be, that if a patient has a health problem years after the end of the trial, this data are normally not considered in the trial.	Suitability for the task Self-descriptiveness	To involve the patient into his/her trial, so that the patient has the possibility to inform by himself about the analysed data and the results considering his/her health. The patient should have the right to see his/her data. She needs access to her trial in a very easy and self-descriptive way. The patient wants to become healthy again and for many other things he does not care. There should be a possibility for health problems that occur after the end of the trial to document them.

Often trials are seen too much in isolation, meta-analyses of trial results of several trials are necessary. A data warehouse that contains clinical trials data would be a good source to compare independently the results of different trials. It is still so that clinical trials data is kept inaccessible in pharma companies, because of the danger that competitor companies might have insight into this trial data.	Suitability for the task Suitability for the task	The trials should be more integrated into the whole process. Meta-analyses of trial results are necessary.
On the other side trial data from different trials is not easy to compare, because each trial has been created and is set up under specific and often different conditions. Even the correct collection of patient data needs standardisation. For example, the measurement of blood pressure can be done standing, lying or sitting. In principle, only values should be compared that were measured under the same conditions. If these measurement conditions are added to the data as metadata, it is easier to compare data of different measurements with each other; but that is currently often not the case.	Suitability for the task	To compare trials is not an easy activity because each trial is established under different conditions. It could be useful to compare only trials that are established under the same condition (see the example in the context of use).
Special features during the working process Clinical trials are created in a team and carried out as teamwork. For each trial, there is a responsibility list (responsibility split), in which the responsibility of each study participant is defined. There exists different responsibilities in a trial, such as project management, data management, monitoring or contact with the trial physicians.	Suitability for the task	The various responsibilities in clinical trials should be considered by the system. The system should know for which area the individual person is responsible as project management, data management, monitoring,

Furthermore even assistants (e.g. study nurse) can take on additional trial related responsibilities.	Suitability for the task	etc. Also assistants (study nurse) can have a trial related responsibility.
The lead investigator can be an investigator-sponsor and can be responsible for the trial. According to GCP the sponsor is always responsible for the clinical trial. This applies also to the case when the trials were paid for by a scientific organisation (e.g. DFG funded trials). A sponsor can be either a pharmaceutical company or a lead investigator.	Suitability for the task	The responsibilities have to be checked when entering into the system. The user should be supported correspondingly to his/her rights. According to GCP the sponsor is always responsible for the clinical trial.
Often laboratory data in clinical trials are still transferred from print-outs into CRF tables by hand. It would be helpful to import laboratory data directly into the data management system.	Suitability for the task	There should be an electronic transfer of laboratory data into tables. It would be helpful if the user could import the laboratory data directly into the data management system.
Organizational conditions To have two separated systems, e.g. SafetyNet for safety management and an EDC system for data management is an overkill	Suitability for the task	It would be helpful for the user to use only one system for side effect management and data management.
In clinical trials, more electronic documents and the possibility of signing electronically would be useful. A possibility would be to be able to use the electronic health card for the electronic signature.	Suitability for the task	The user's problem is the missing electronic signature.
The extent of logistical problems in international clinical trials is often underestimated and often are associated	Suitability for the task	Logistical problems in international trials produce high costs. Can this fact be avoided?

with high costs. In most clinical trials there is still now no efficient software support for complete site - and trial, including the medical product logistics.		Are they monitored? Has a quality control taken place? It would be desirable to have a software support and management (site - and trial management).
Because investigators often have little time and are not accessible, monitors have to wait for a meeting to get responses or to get a missing signature.	Suitability for the task	Investigators often have little time and are not accessible, monitors have to wait for a meeting to get responses or to get a missing signature. Can this process be accelerated?
Not all investigators recruit evenly. Sometimes, during the study feasibility period investigator say that they have enough patients available for a trial, but during the trial patients may be missing. Here, the access to information about prior recruitment performances of sites and the availability of certain patient populations would be, a time saving. Considering these additional information, trials may be carried out faster, better and cheaper.	Suitability for the task	The connection (communication) with data warehouses is very imported for the sponsor respectively the trial physician to get missing information. A trial could be carried out under these conditions much faster, better and cheaper.
Other comments to critical incidents which already occurred The time for setting up a trial still takes too long and may last sometimes months. To use a more standardised way for trial design and to use pre-defined building blocks would lead to easier and faster trial development and implementation. In addition, the user interfaces of many clinical trial systems in use are still not user friendly from the investigator's point of view. It is complained of that the interfaces are too colourful, with too many blinking flags,	Suitability for the task Self-descriptiveness Conformity with user expectations	The software should support the user to set up a trial in shorter time as yet possible. The user should be able to use a more standardized way for trial design and to use predefined units to reach his aim efficiently. A software system should be self-descriptive to use it and conduct the task in an efficient and effective way. The interfaces should be developed so that the user is guided through his task efficiently and

with too many items that may confuse the investigator. Another problem are error messages that the user does not understand and will not respond to.	Error tolerance	his attention is drawn to the essentials. Error messages have to be described in the user's language. The user should get help by the system to detect and remove the occurred error without loss of time.

Appendix 3 – Patients' answers of the questionnaire presented in tables

Each question has five choices, which differ correspondingly to the kind of question:

Concerning the questions with number 1 to 14, 16 to 18 and 20 to 22:

Very important, Important, Neither important or unimportant, Unimportant, Very unimportant

Concerning the question with number 15:

All of the information, Most of the information, Some of the information, None of the information, I don't know

Concerning the question with number 19:

Very likely, Likely, Neither likely or unlikely, Unlikely, Very unlikely

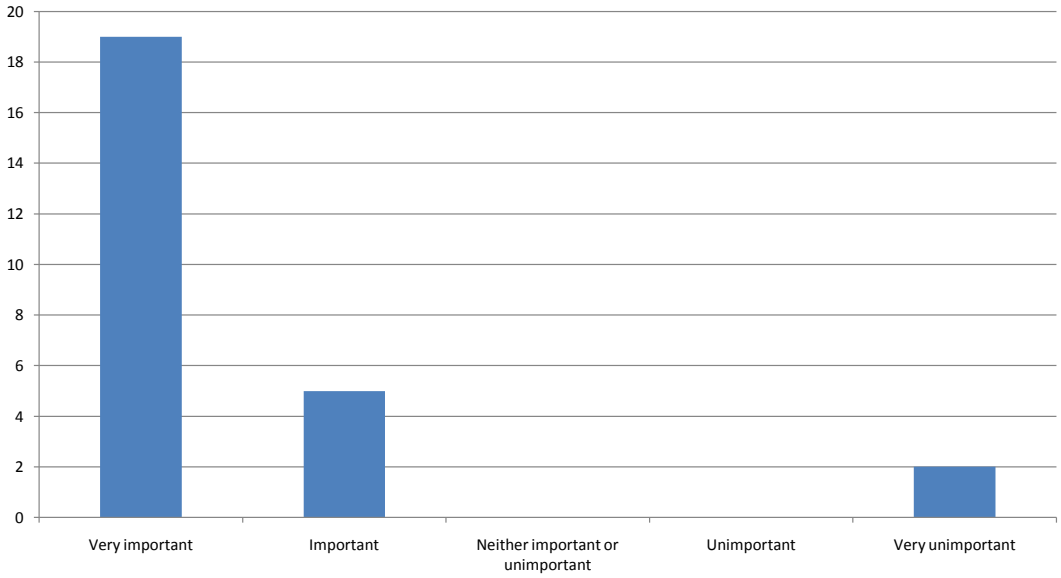
Concerning the questions with number 23 and 28:

Yes, No

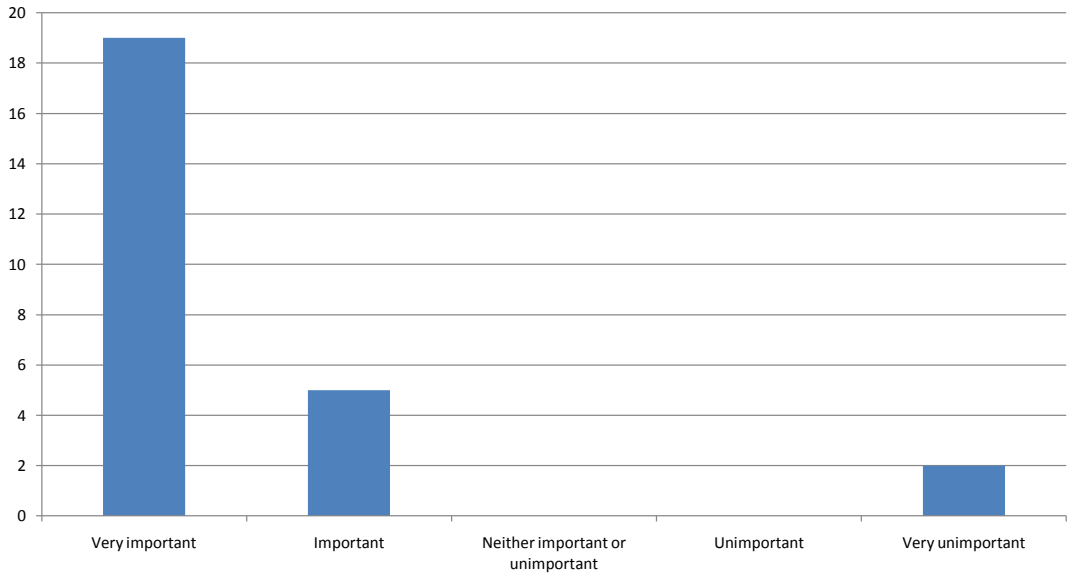
Concerning the questions with number 24 to 27 and 29:

Very often, Often, Neither often or rarely, Rarely, Very rarely

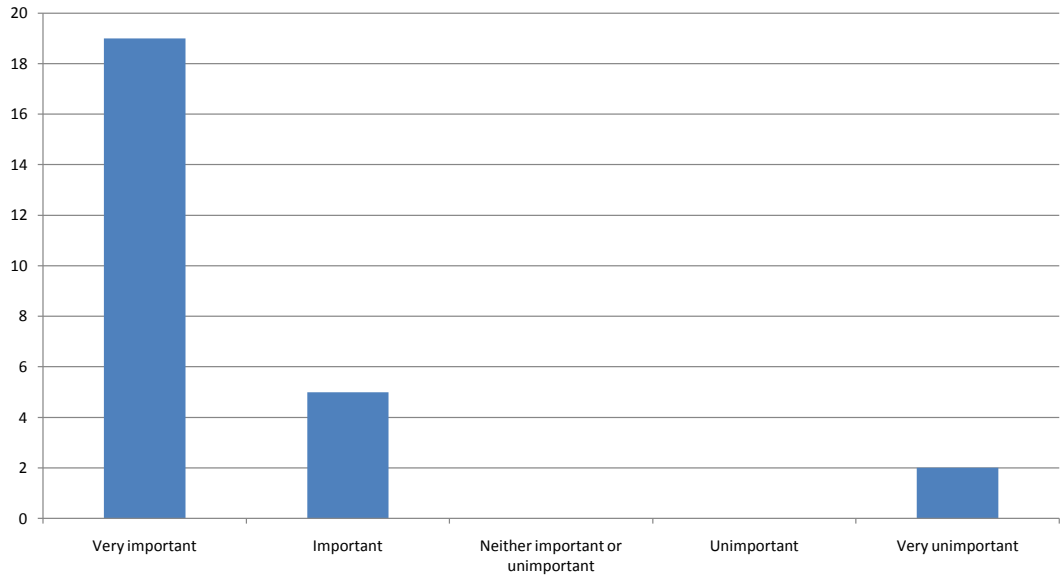
1. When discussing your treatment options with your doctor, please rate how important it is to have information on the following:
The latest new treatments



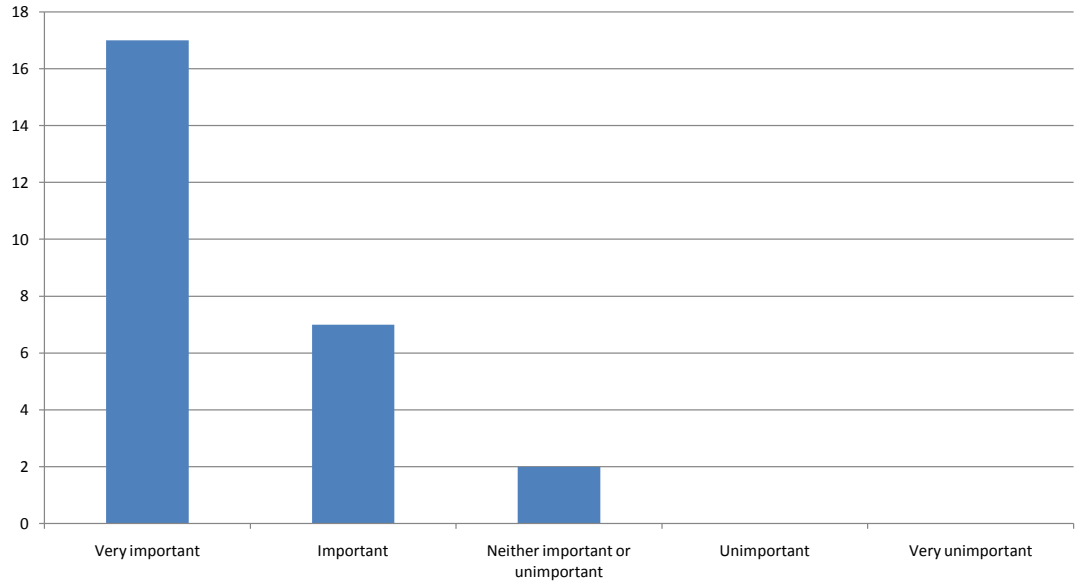
2. When discussing your treatment options with your doctor, please rate how important it is to have information on the following:
The treatment options available to me (in my particular circumstances)



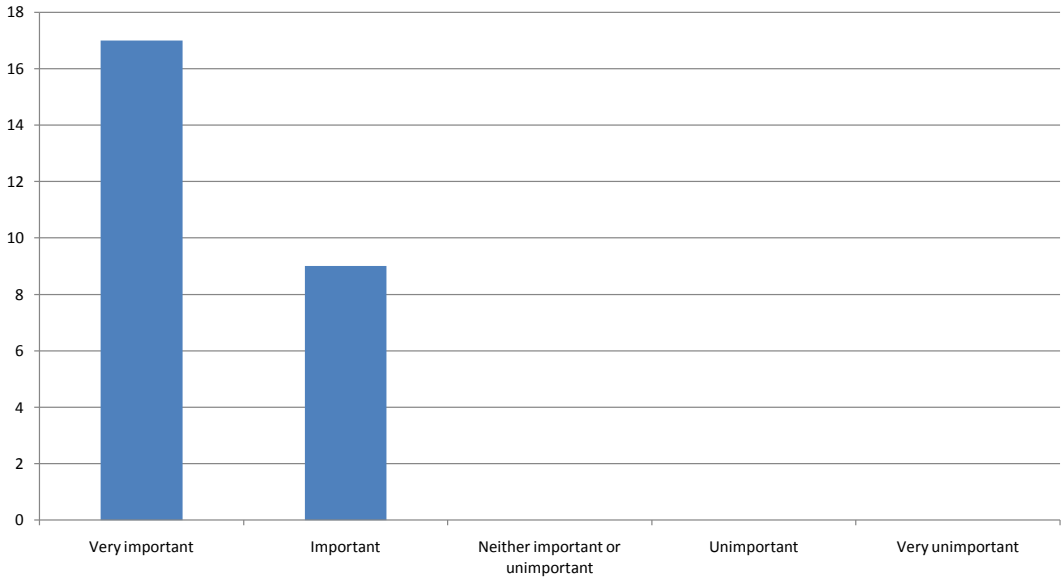
3. When discussing your treatment options with your doctor, please rate how important it is to have information on the following:
How effective the different suggested treatments are



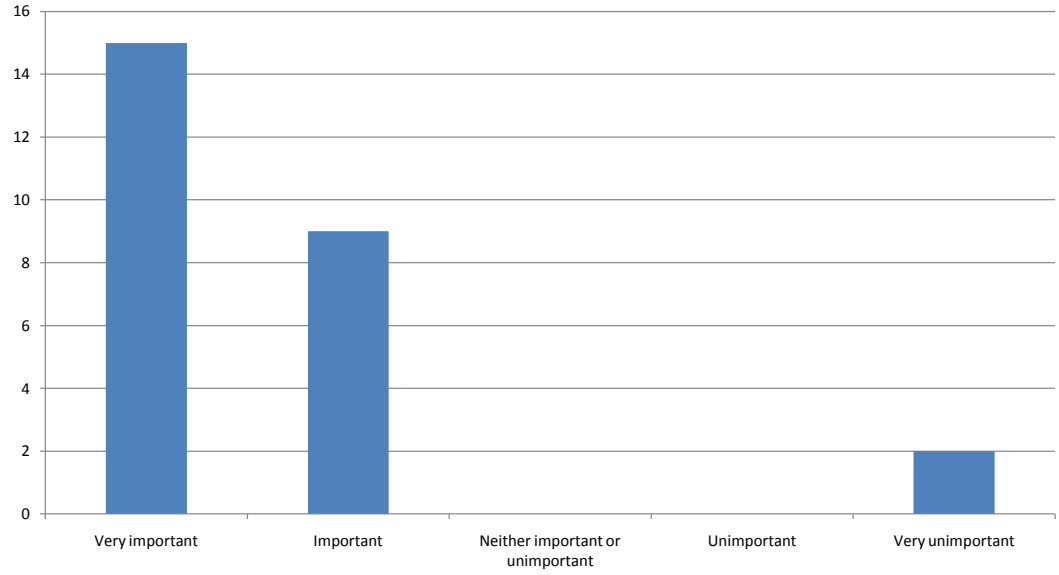
4. When discussing your treatment options with your doctor, please rate how important it is to have information on the following:
The survival rates of the suggested treatments



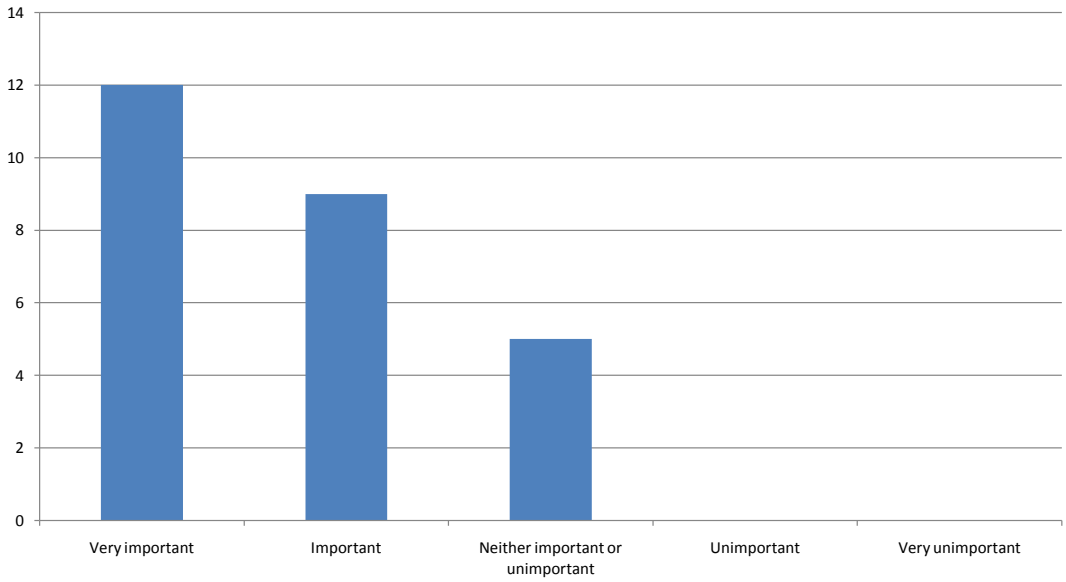
5. When discussing your treatment options with your doctor, please rate how important it is to have information on the following:
Quality of life after the suggested treatments



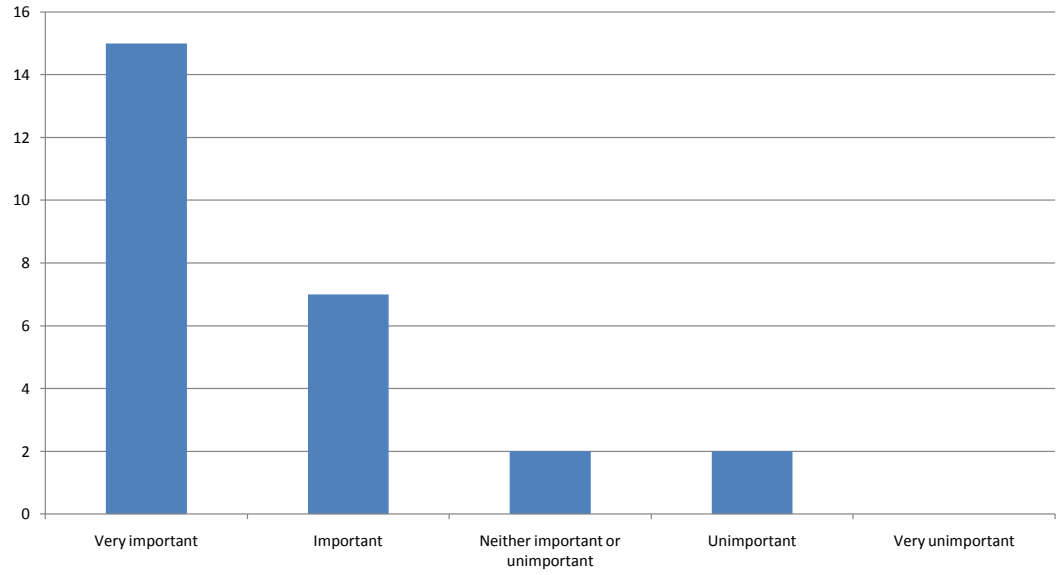
6. When discussing your treatment options with your doctor, please rate how important it is to have information on the following:
The side effects of the suggested treatments



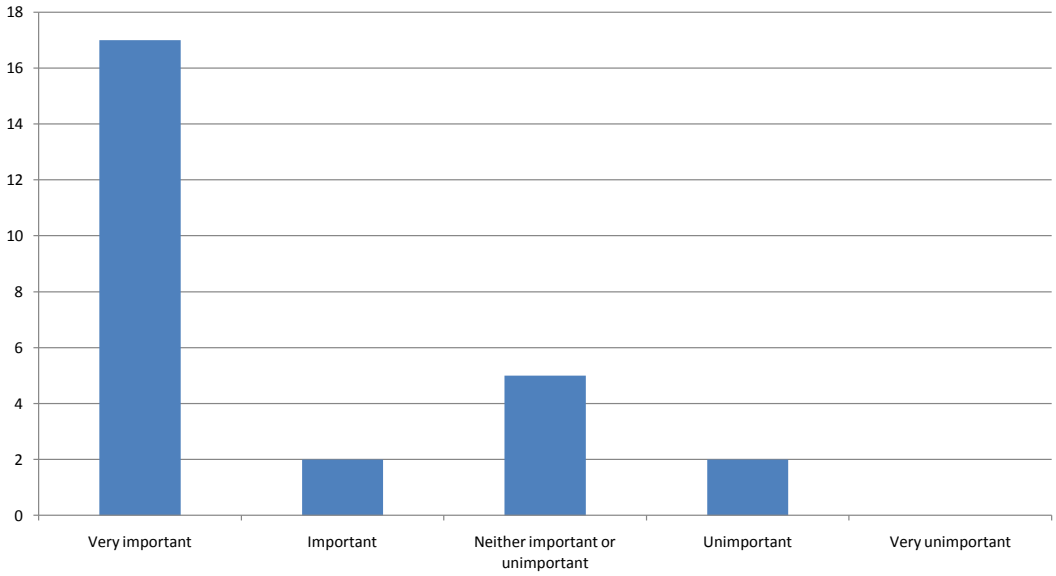
7. When discussing your treatment options with your doctor, please rate how important it is to have information on the following:
Opportunities to be involved in a clinical trial



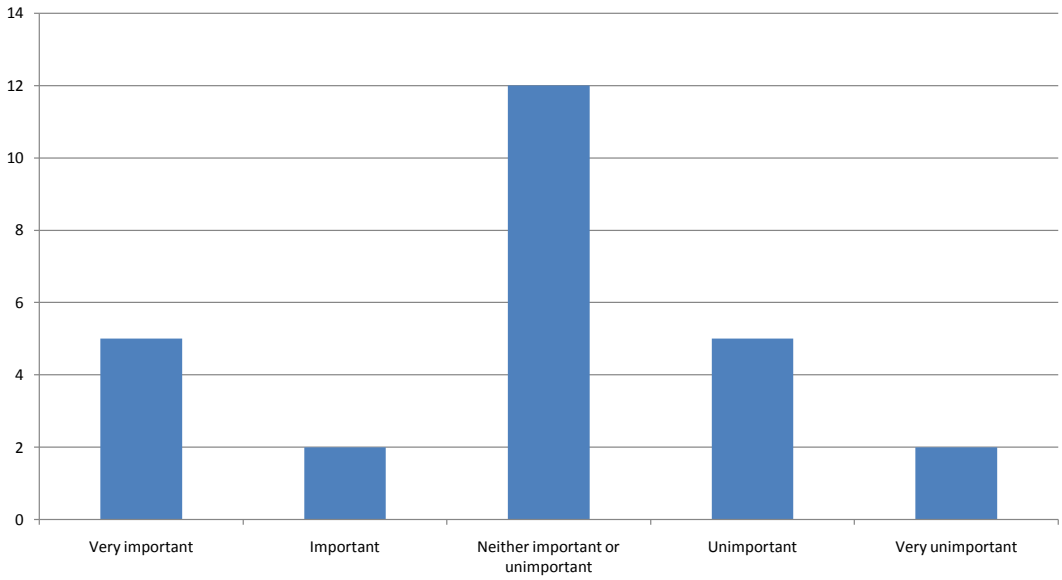
8. When discussing your treatment options with your doctor, please rate how important it is to have information on the following:
My disease



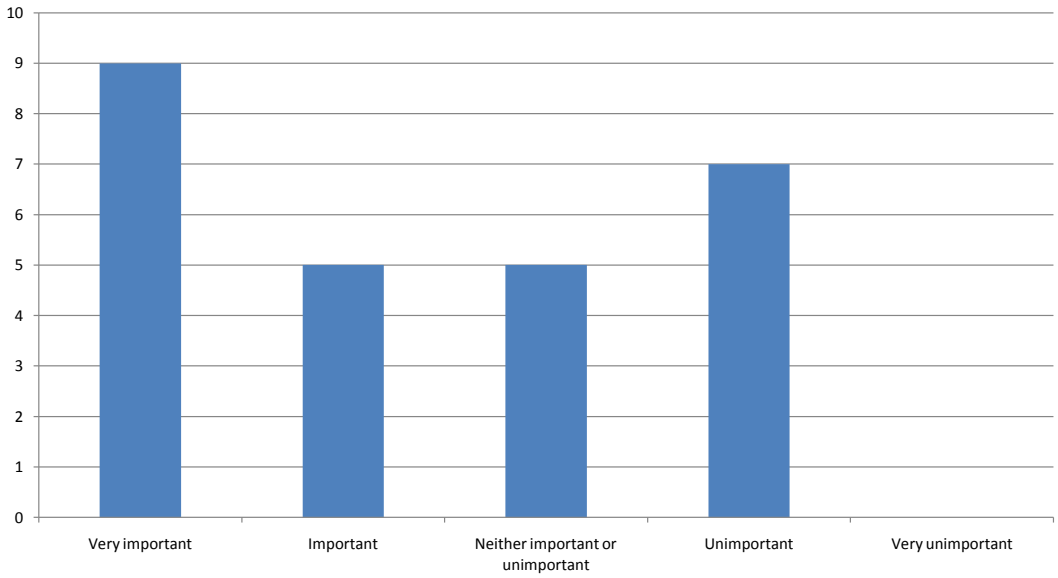
9. When discussing your treatment options with your doctor, please rate how important it is to have information on the following:
The best questions to ask the doctor for the most relevant information



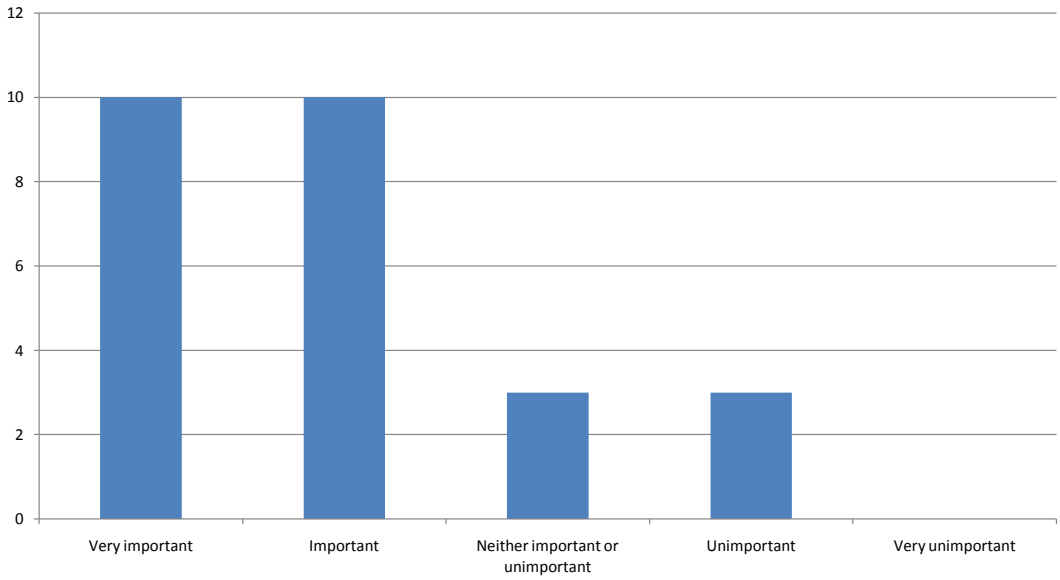
10. How important is it to be able to access all of the above information on the internet?



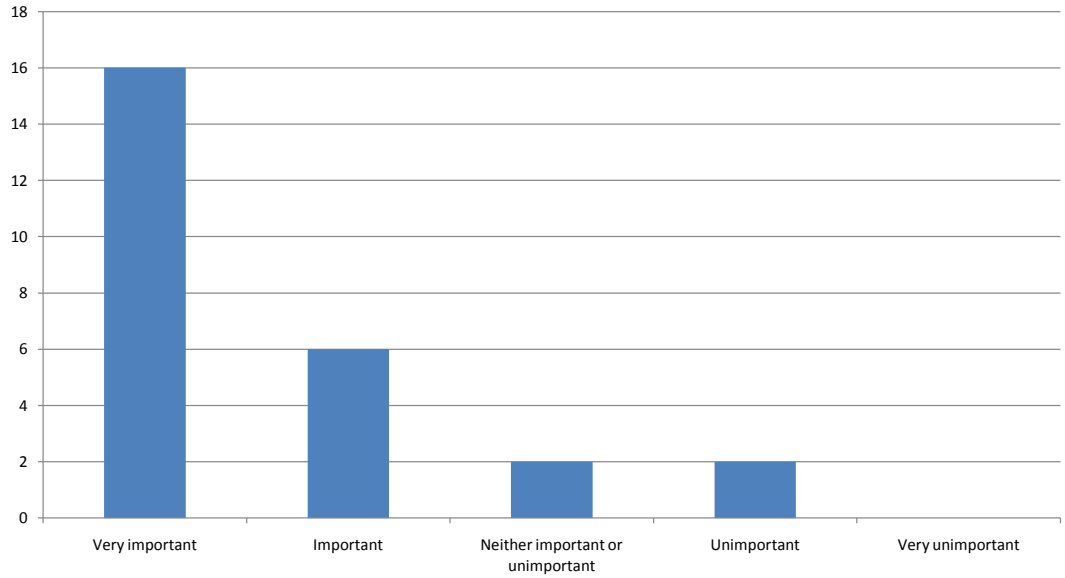
11. How important is it to have a record of what you have discussed with your doctor to refer back to at a later date?



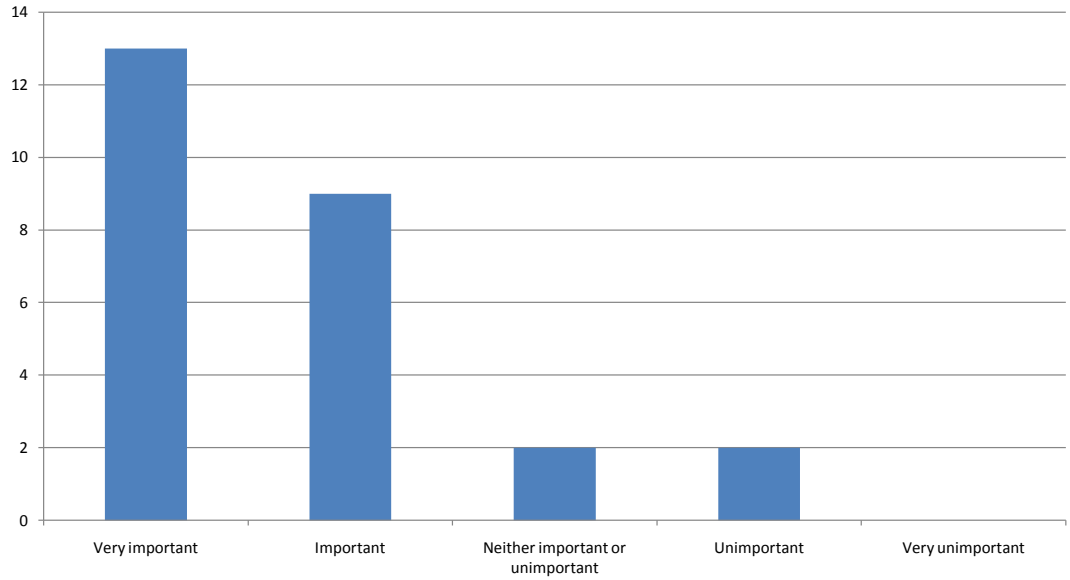
12. When discussing your treatment options with your doctor, please rate how important it is to have the following information:
How often the hospital treats this disease



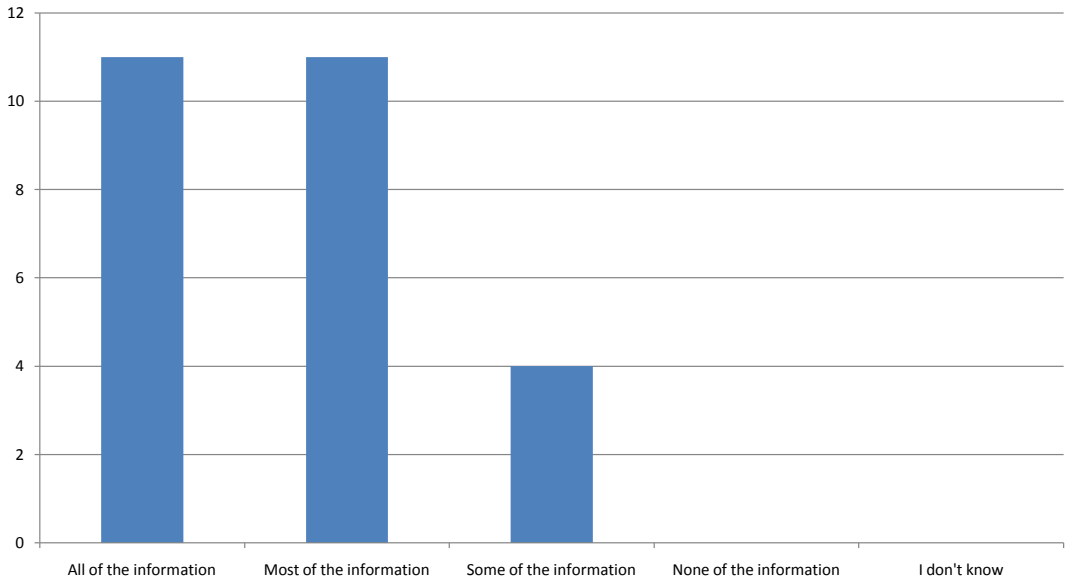
13. When discussing your treatment options with your doctor, please rate how important it is to have the following information:
How successful my hospital has been in treating patients like me



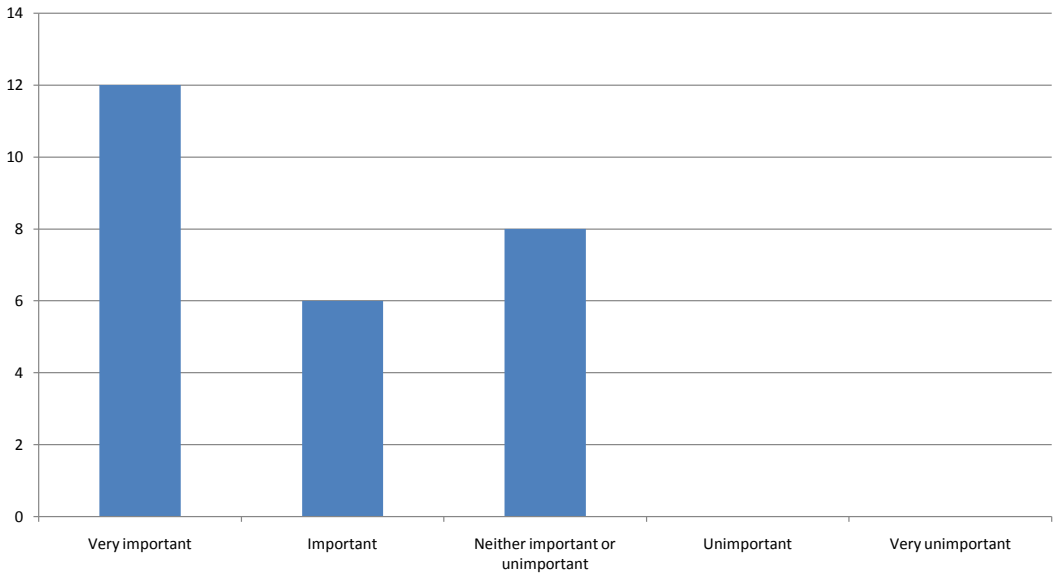
14. When discussing your treatment options with your doctor, please rate how important it is to have the following information:
How often my doctor treats this disease



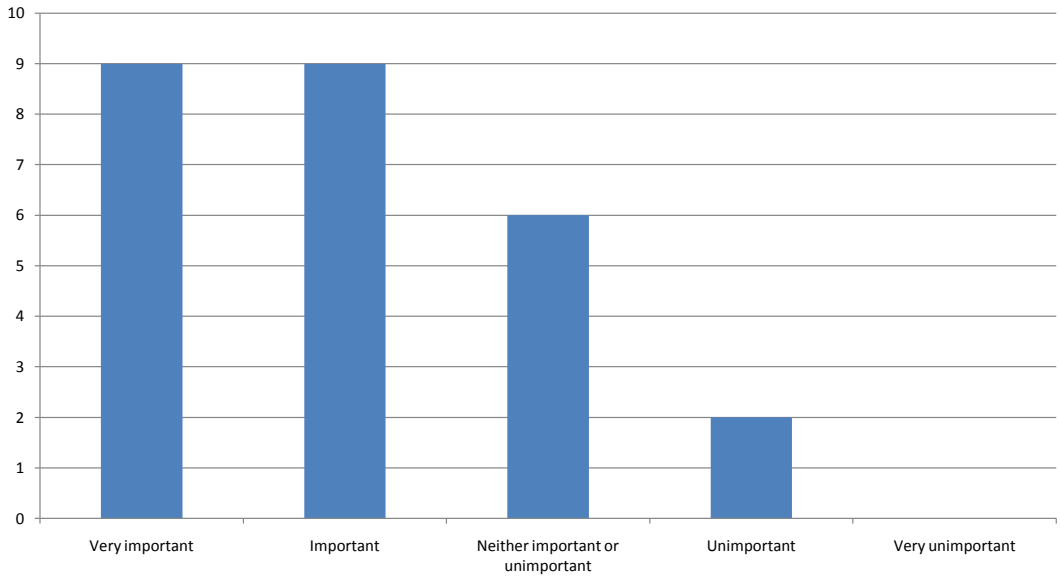
15. After speaking with your doctor, how much of the information you are given do you usually understand?



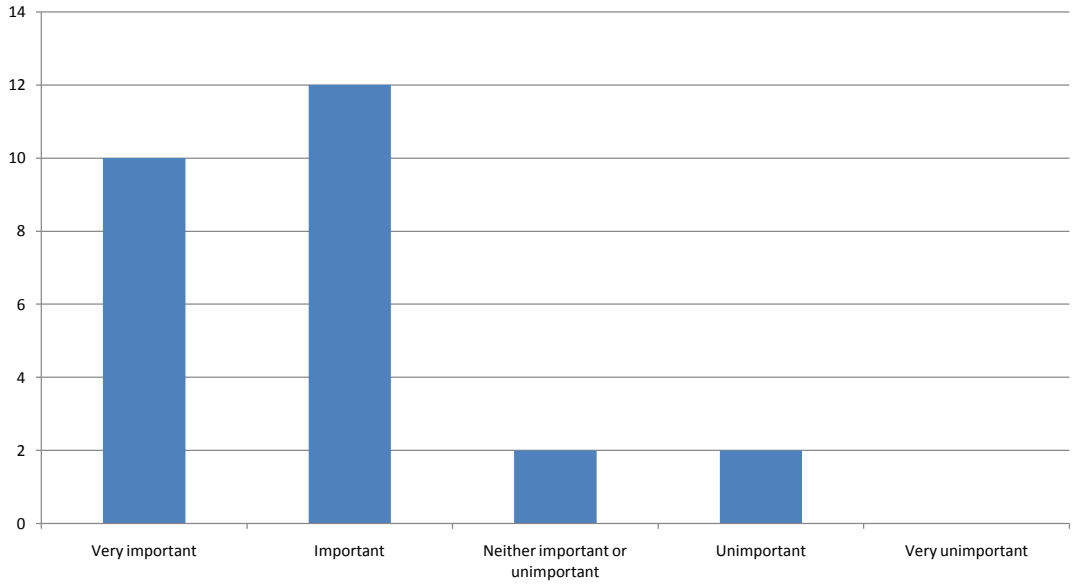
16. How important do you think it is to be given printed information explaining more detail about what the doctor has said?



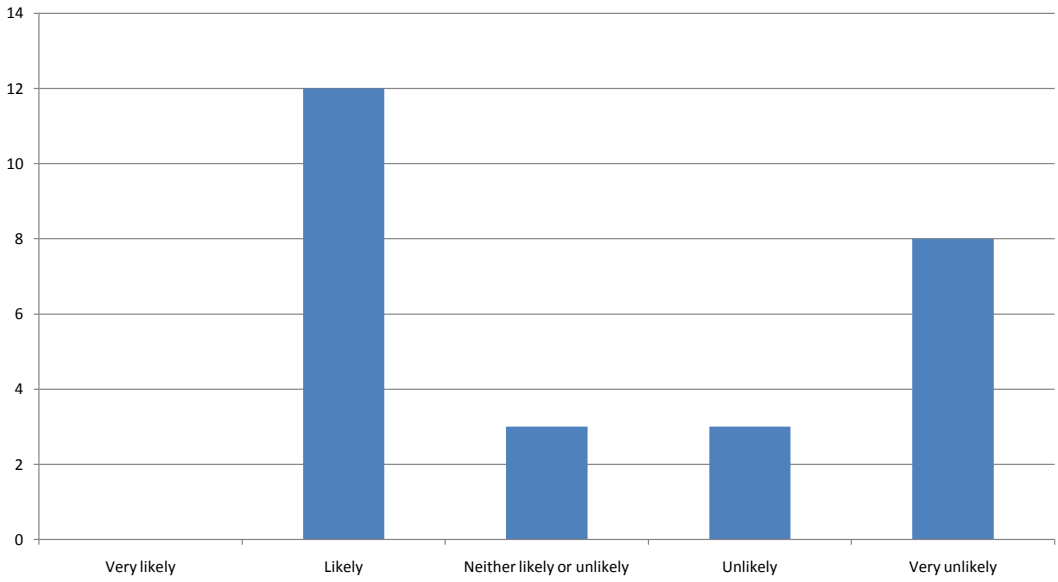
17. How important is it to be able to communicate with other patients who are affected by the same illness as you?



18. How important do you think it is that your doctor obtains information about your psychological well-being when discussing your diagnosis?

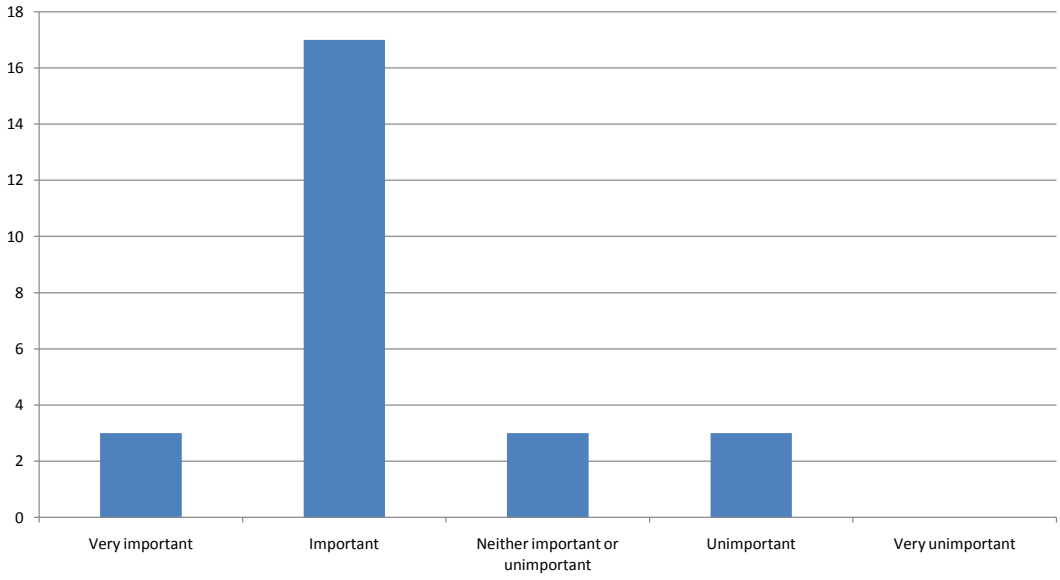


19. How likely would you be to join an internet based social media network of patients (like facebook)?

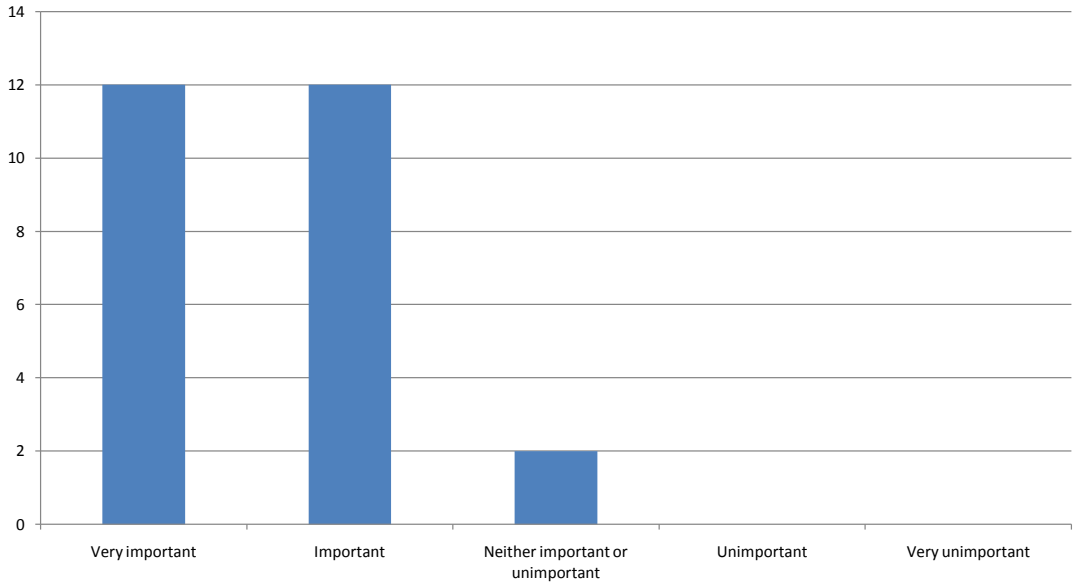


20. How important do you think it is to be given extra information at the following times:

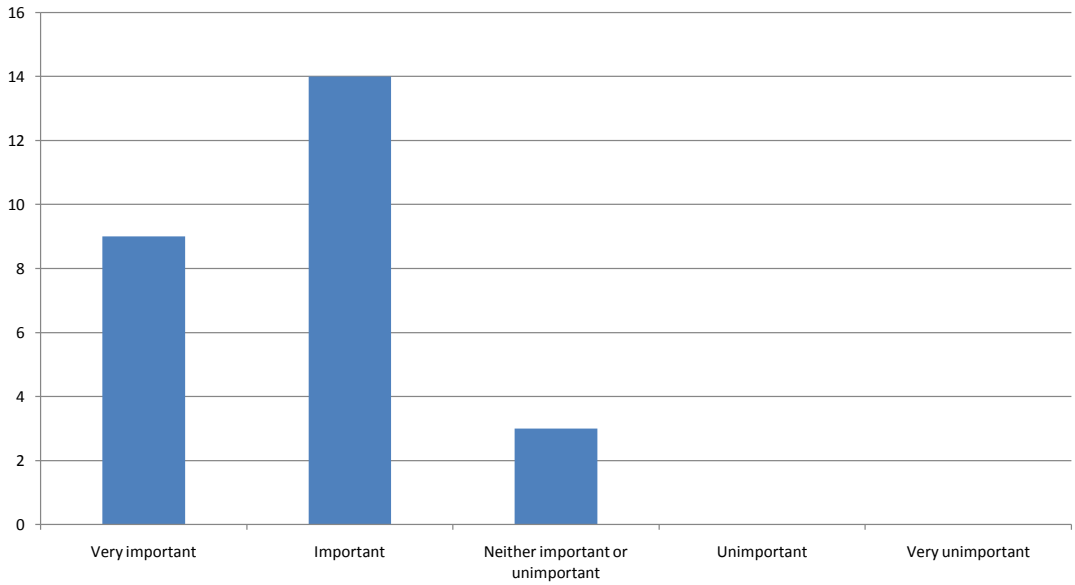
Before speaking to your doctor



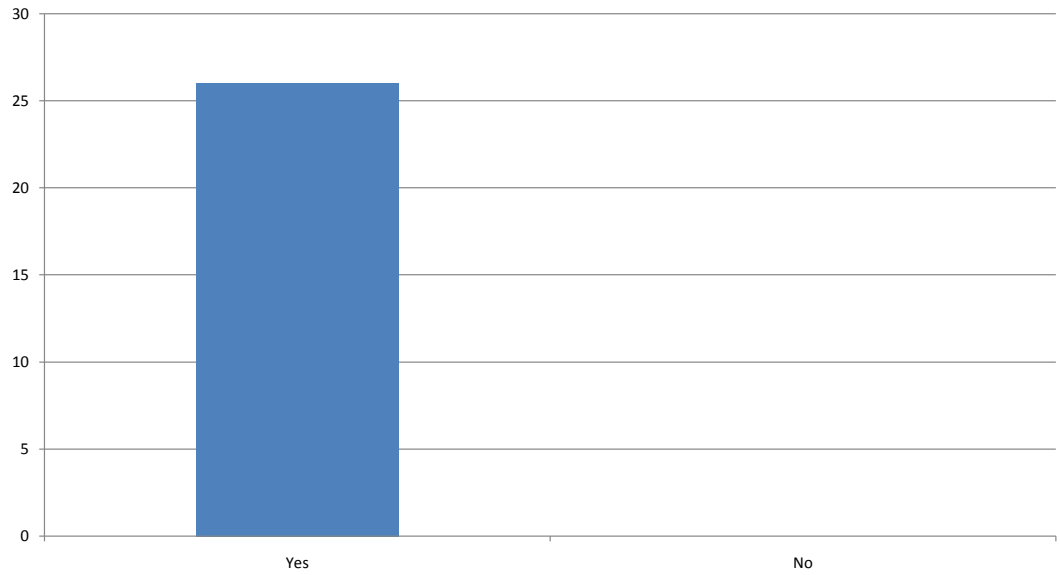
21. How important do you think it is to be given extra information at the following times:
When speaking to your doctor



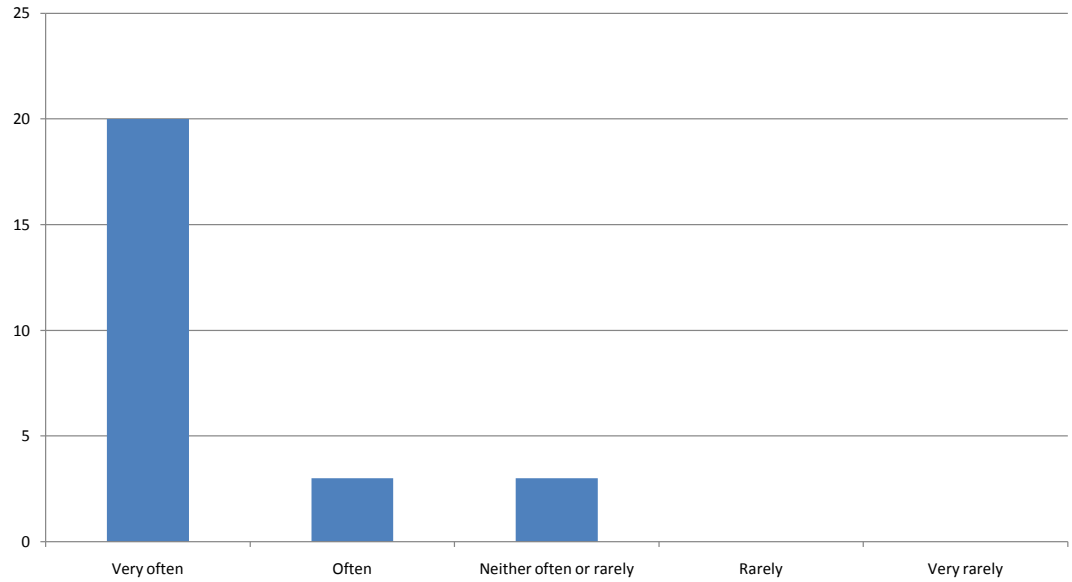
22. How important do you think it is to be given extra information at the following times:
When you return home after speaking to your doctor



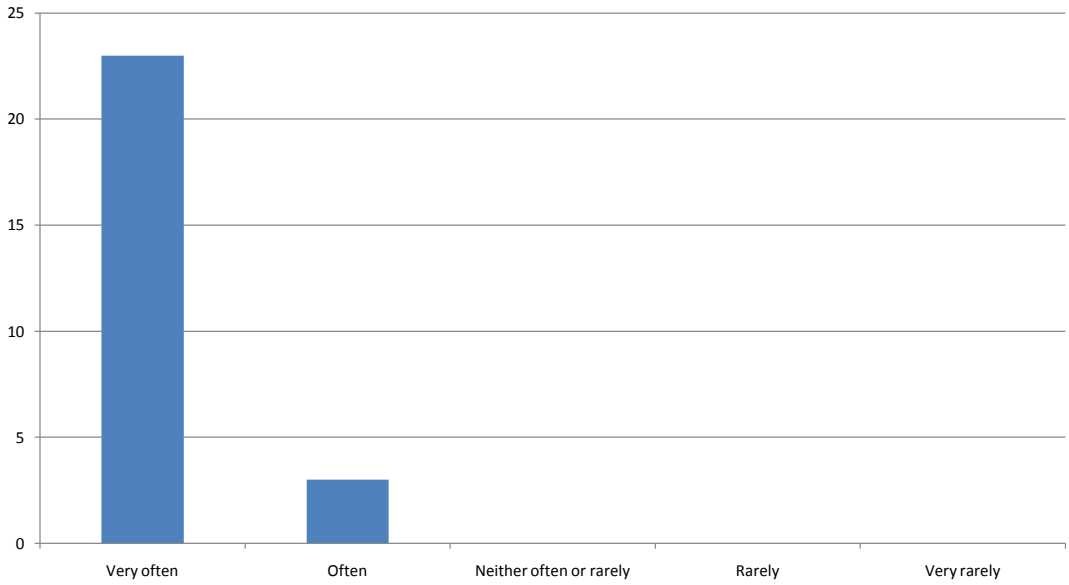
23. Do you have access to a computer at home?



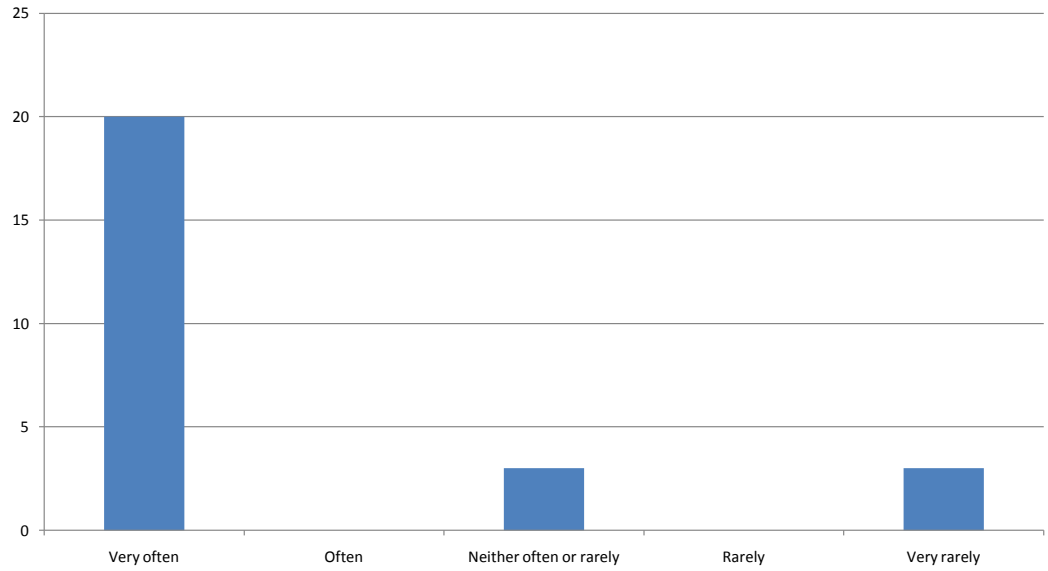
24. How often do you use the internet on any computer?



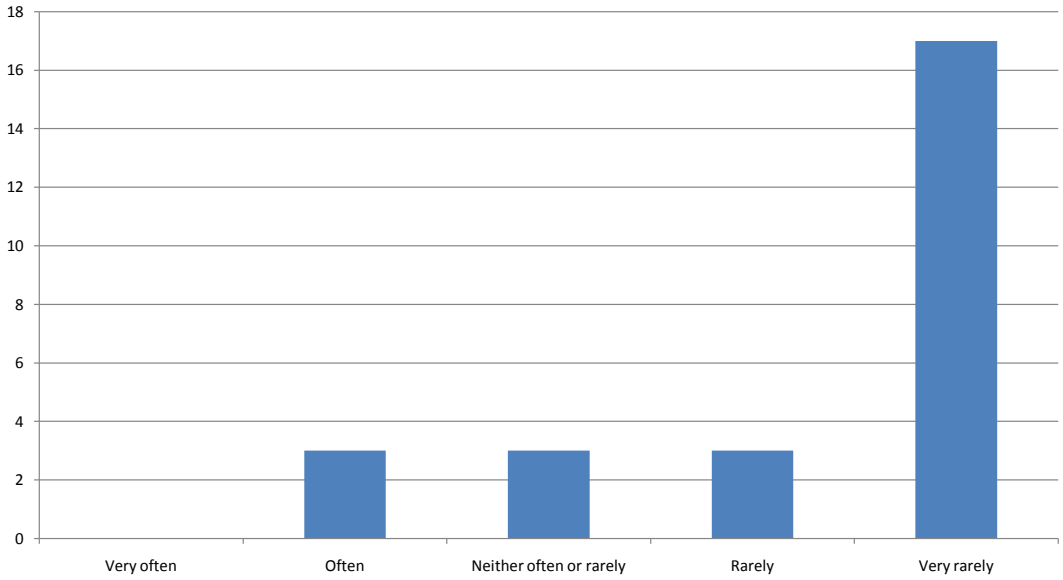
25. How often do you use email on any computer?



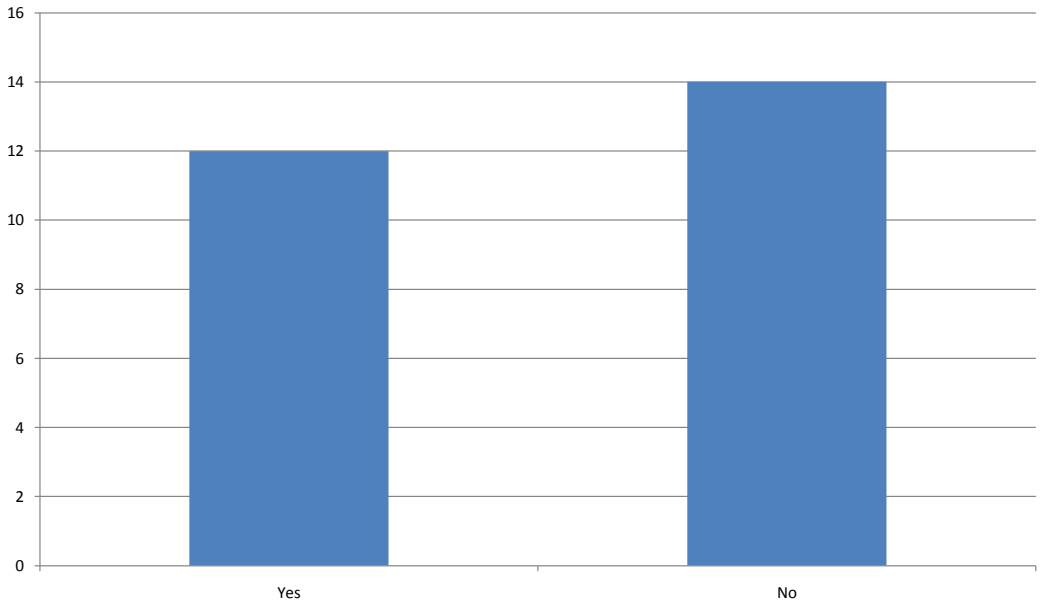
26. How often do you use Microsoft Word on any computer?



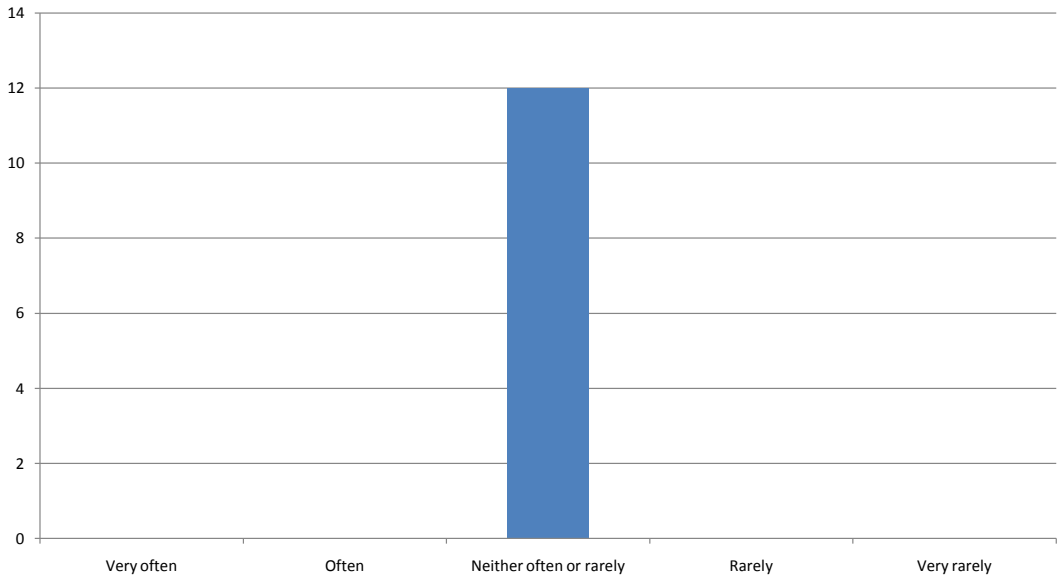
27. How often do you access the internet through your mobile phone?



28. Do you own a smart phone?



29. If you own a smart phone, how often do you download applications?



Patient comments

- **Patients do not HEAR much after being diagnosed with cancer - that word blanks out almost everything else. While speaking to a doctor it would be helpful if the conversation were actually recorded and could be played back at home in order to fully understand everything discussed and the options / treatment available. I would be most wary of using the internet to determine treatment or medication, as many sites are American and treatment/ medicines are different from here. I would advise patients to ask for a second opinion if they weren't sure of choices. Patients need to be aware that they do have some choices.**

Patient comments

- **With reference to radiotherapy treatment after having had brackytherapy, it is most important that the patient is told that he may experience rectal being after the radiotherapy course is finished. This avoids worry that bowl cancer has set it.**
- **I understand that the cause of cancer has an emotional and mental input, as well as a physical. I am addressing these myself. Some acknowledgement of this by the medical profession is sorely needed.**

Appendix 4 – Template for use cases/scenarios

Item	Description
Identifier*	
Version	
Name	
Description of the use case (end-user perspective)	
Problem(s) to solve	
Challenges	
Risks	
Expected benefits	
Characterization	○ fundamental ○ general ○ specific
If specific, please give the Domain	○ Acute lymphoblastic leukaemia ○ Breast Cancer ○ Nephroblastoma ○ other Cancer, please specify: ○ Non-Cancer Domain, please specify:

End-user	○ system
	○ person
	  ○ basic scientist
	  ○ clinician
	  ○ computer scientist
	  ○ regulatory body, lawyer, ethicist
	  ○ patient
	  ○ other, please specify:

Pre-condition(s)/pre-requisite(s)	
Requisite(s)	
Post-condition(s)/post-requisite(s)	
Constraints	

External sources needed from outside p-medicine	O data, please specify: O tools, please specify: O services, please specify: O models, please specify: O other, please specify:	
Data used	O personal O only non-personal O target population, please specify:	
Input data	O internal database, please specify: O external database, please specify: O online input	
Output data	O database, please specify: O variables for use, please specify: O structured document, please specify: O graphic, please specify:	
Data volume		
Dataflow	Please specify:	
Data storage	Please specify:	
Successful End Condition		
Fail End Condition		
Basic workflow	Actor Action	System response

Expected usage frequency		
Needed for DSS	○ yes ◉ no	
Needs HPC	○ yes ◉ no	
Needs Grid	○ yes ◉ no	
Priority for development		
Responsible for development		
Mock-up needed	○ yes ◉ no	
Responsible for Mock-up		
Who is building the tool		
Open Source tool	◉ yes ○ no, please specify why	

* The identifier should be composed of End-user (system/personal/biobanking/gene/ repository/semantic/education), Characterization (fundamental, general, specific), Domain (id specific) and consecutive numbering. E.g.: PSN_1 → First scenario used by Personal and specific for Nephroblastoma, or SG_3 → Third scenario used by the system and general

Appendix 5 – Use Cases

VPH Scenarios

VPH Toolbox Scenario

Item	Description
Identifier	SG_1
Version	1.0
Name	**Toolbox Scenario**
Description of the use case (end-user perspective)	The VPH-Toolkit serves as a source of existing tools, services, models for usage in p-medicine and as a 'toolbox' for uploading newly developed tools, services and models. This use case will cover both scenarios.
Problem(s) to solve	Sharing of tools, services and models that fulfil criteria of interoperability and user-friendliness leading to the 'Gold Standard' Toolkit Status.
Challenges	Interoperability issues and user friendliness of tools
Risks	None
Expected benefits	To increase the number of high quality tools, services and models for usage in the scientific community
Characterization	○ fundamental ◉ general ○ specific
If specific, please give the Domain	◉ Acute lymphoblastic leukaemia ◉ Breast Cancer ◉ Nephroblastoma ○ other Cancer, please specify: ○ Non-Cancer Domain, please specify:
End-user	○ system ◉ person ◉ basic scientist ◉ clinician ◉ computer scientist ○ regulatory body, lawyer, ethicist ○ patient ○ other, please specify:
Pre-condition(s)/pre-requisite(s)	None
Requisite(s)	None
Post-condition(s)/ post-requisite(s)	None

Constraints	None	
External sources needed from outside p-medicine	○ data, please specify: ○ tools, please specify: ○ services, please specify: ○ models, please specify: ○ other, please specify:	
Data used	○ personal ○ only non-personal ○ target population, please specify:	
Input data	○ internal database, please specify: ○ external database, please specify: ○ online input	
Output data	○ database, please specify: ○ variables for use, please specify: ○ structured document, please specify: ○ graphic, please specify:	
Data volume	n.a.	
Dataflow	Please specify: no	
Data storage	Please specify: no	
Successful End Condition	Download or upload of a tool, service or method	
Fail End Condition	Download or upload of a tool, service or method is not possible	
Basic workflow	Actor Action	System response
	1. Download of a tool, service, method	
	Individual user creates an account on the VPH-toolkit website (http://toolkit.vph-noe.eu/)	Credentials are send to the individual
	Login via the credentials	Individual is forwarded to the toolkit website http://toolkit.vph-noe.eu/get-involved
	Usage of the search functionality	A list of available tools is displayed with metadata and a short description of the functionality of the tool. Rating of the tool is be displayed (max. 5 stars)
	Clicking on the selected tool, service or method	More details of the tool is provided, including the website for downloading
	Download of the tool	

	via the given information	
	2. Upload of a tool, service method	
	Login via the credentials as	Individual is forwarded to the toolkit website http://toolkit.vph-noe.eu/get-involved
	Select from the User menu the submit button for a tool, or a method or a service	A new website opens http://toolkit.vph-noe.eu/home?sobi2Task=addNew&itemetype=a to enter required metadata. The last letter has the following meaning a: tools, b: method, c: service
	Metadata need to be entered as required, if finished click on the send button	The information about the tool is provide on the website of the VPH-Toolkit under the category that was chosen
Expected usage frequency	Regularly at every time someone searching for a tool, service or method, or someone wants to upload a newly developed tool, service or method	
Needed for DSS	○ yes ◉ no	
Needs HPC	○ yes ◉ no	
Needs Grid	○ yes ◉ no	
Priority for development	High	
Responsible for development	UCL	
Mock-up needed	○ yes ◉ no	
Responsible for Mock-up	n.a.	
Who is building the tool	UCL	
Open Source tool	◉ yes ○ no	

Security Scenarios

Single Sign-on Scenario

Normal Flow

Item	Description
Identifier	SG_4
Version	1.0
Name	Single sign-on (SSO) on p-medicine Platform when accessing a browser service.
Description of the use case (enduser perspective)	A p-medicine end-user will typically access multiple p-medicine sites/services. To avoid that this end-user would have to login on each site/service separately, he authenticates himself only once on a central p-medicine Identity Provider (or another federated Identity Provider). This provider will issue credentials that can be used for accessing protected P-Medicine sites/services.
Problem(s) to solve	Allow an end-user to authenticate him only once, so that he is able to access multiple sites/services (within the active browsers session) without having to login on each one of them separately (SSO).
Challenges	Hiding the complex Single Sign-On functionality for the end-user, by providing user-friendly authentication steps.
Risks	If a malicious person succeeds in stealing the credentials of the end-user, he has access to every site/service of the p-medicine Platform.
Expected benefits	p-medicine end-users will be able to authenticate themselves only once to access all p-medicine sites/services. Each site/service does not need to implement user authentication and management. The central identity provider handles this.
Characterization	○ fundamental ◉ general ○ specific
If specific, please give the Domain	○ Acute lymphoblastic leukaemia ○ Breast Cancer ○ Nephroblastoma ○ other Cancer, please specify: ○ Non-Cancer Domain, please specify:
Enduser	○ system ◉ person ◉ basic scientist ◉ clinician

	⦿ computer scientist ⦿ regulatory body, lawyer, ethicist ⦿ patient O other, please specify:	
Pre-condition(s)/pre-requisite(s)	The end-user is registered on the p-medicine Platform The end-user is not yet authenticated, does not has an active SSO session on the identity provider or any of the services he wishes to access.	
Requisite(s)		
Post-condition(s)/post-requisite(s)		
Constraints		
External sources needed from outside p-medicine	O data, please specify: O tools, please specify: O services, please specify: O models, please specify: O other, please specify:	
Data used	O personal O only non-personal O target population, please specify:	
Input data	O internal database, please specify: O external database, please specify: O online input	
Output data	O database, please specify: O variables for use, please specify: O structured document, please specify: O graphic, please specify:	
Data volume	Small	
Dataflow	Please specify:	
Data storage	Please specify:	
Successful End Condition	The end-user is signed in on the P-Medicine Platform. He can access any P-Medicine site/service (within the active browser session) without having to re-authenticate for each site/service (a SSO session is established).	
Fail End Condition	The end-user did not manage to authenticate himself. No SSO session is generated.	
Basic workflow	Actor Action	System response
	The end-user browses to a web-page that gives access to a P-Medicine service.	The web server detects that the end-user is not authenticated locally and redirects the user to the p-medicine Identity Provider (or another federated Identity Provider).

		The Identity Provider (IdP) detects whether the end-user has an active Single Sign-On session.
		If no active session is detected, the end-user is prompted to select an authentication method. Initially only one authentication method will be provided (username/password).
	The end-user tries to authenticate himself by providing his username and password.	If username and password are valid, the end-user is authenticated and the IdP redirects the end-user back to the original web-page the end-user wanted to access.
		A local service on the web server verifies the authentication token received through the redirect and creates a local session.
Expected usage frequency	High, SSO will be used every time an end-user accesses a P-Medicine site/service that has no active session running for this end-user.	
Needed for DSS	○ yes ◉ no	
Needs HPC	○ yes ◉ no	
Needs Grid	○ yes ◉ no	
Priority for development	Required	
Responsible for development	Custodix	
Mockup needed	○ yes ◉ no	
Responsible for Mockup	Not Applicable	
Who is building the tool		
Open Source tool	◉ yes ○ no, please specify why:	

Alternative Flow

Item	Description
Identifier	SG_5
Version	1.0
Name	Access a p-medicine browser servers while a p-medicine Single Sign-On (SSO) session is already active.
Description of the use case (enduser perspective)	A p-medicine end-user will typically access multiple p-medicine sites/services. To avoid that this end-user would have to login on each site/service separately, he authenticates himself only once on a central p-medicine Identity Provider (or another federated Identity Provider). This provider will issue credentials that can be used for accessing protected p-medicine sites/services.
Problem(s) to solve	Allow an end-user to authenticate himself only once, so that he is able to access multiple sites/services (within the active browsers session) without having to login on each one of them separately (SSO).
Challenges	Hiding the complex Single Sign-On functionality for the end-user, by providing user-friendly authentication steps.
Risks	If a malicious person succeeds in stealing the credentials of the end-user, he has access to every site/service of the p-medicine Platform.
Expected benefits	p-medicine end-users will be able to authenticate themselves only once to access all p-medicine sites/services. Each site/service does not need to implement user authentication and management. The central Identity Provider handles this.
Characterization	○ fundamental ◉ general ○ specific
If specific, please give the Domain	○ Acute lymphoblastic leukaemia ○ Breast Cancer ○ Nephroblastoma ○ other Cancer, please specify: ○ Non-Cancer Domain, please specify:
Enduser	○ system ◉ person ◉ basic scientist ◉ clinician ◉ computer scientist ◉ regulatory body, lawyer, ethicist ◉ patient

	○ other, please specify:	
Pre-condition(s)/pre-requisite(s)	• The end-user is registered on the p-medicine platform • The end-user has an active SSO session.	
Requisite(s)		
Post-condition(s)/post-requisite(s)		
Constraints		
External sources needed from outside p-medicine	○ data, please specify: ○ tools, please specify: ○ services, please specify: ○ models, please specify: ○ other, please specify:	
Data used	○ personal ○ only non-personal ○ target population, please specify:	
Input data	○ internal database, please specify: ○ external database, please specify: ○ online input	
Output data	○ database, please specify: ○ variables for use, please specify: ○ structured document, please specify: ○ graphic, please specify:	
Data volume	Small	
Dataflow	Please specify:	
Data storage	Please specify:	
Successful End Condition	The end-user has successfully accessed the requested service without having to re-authenticate.	
Fail End Condition	The end-user did not manage to access the requested service without having to authenticate himself on the IdP.	
Basic workflow	Actor Action	System response
	The end-user browses to a web-page giving access to a p-medicine service.	The web server detects that the end-user is not authenticated locally and redirects the end-user to the P-Medicine Identity Provider (or another federated Identity Provider).
		The Identity Provider (IdP) detects whether the end-user has an active SSO session.

		If an active SSO session is detected, the end-user's authentication token is passed back to the original web-page through a redirect of the end-user.
		A local service on the web server verifies the authentication token received through the redirect and creates a local session.
Expected usage frequency	High, SSO will be used every time an end-user accesses a P-Medicine site/service that has no active session running for this end-user.	
Needed for DSS	○ yes ◉ no	
Needs HPC	○ yes ◉ no	
Needs Grid	○ yes ◉ no	
Priority for development	Required	
Responsible for development	Custodix	
Mockup needed	○ yes ◉ no	
Responsible for Mockup	Not Applicable	
Who is building the tool		
Open Source tool	◉ yes ○ no, please specify why:	

Single Sign-out Scenario

Item	Description
Identifier	SG_6
Version	1.0
Name	Single Sign-Out from the p-medicine Platform.
Description of the use case (enduser perspective)	An end-user, who wishes to logout from the p-medicine Platform, performs one Sign-Out operation signing him out from all sites/services he is actively involved with in his current browser session.
Problem(s) to solve	An end-user should be able to sign out from all the sites/services in which he is currently authenticated, using a simple single logout action.
Challenges	Hiding the complex Single Sign-Out functionality for the end-user, by providing a user-friendly logout step.
Risks	Implementing single sign-out is very complex.
Expected benefits	A Single Sign-Out action results in the end-user being signed out from all services where he is currently authenticated.
Characterization	○ fundamental ◉ general ○ specific
If specific, please give the Domain	○ Acute lymphoblastic leukaemia ○ Breast Cancer ○ Nephroblastoma ○ other Cancer, please specify: ○ Non-Cancer Domain, please specify:
Enduser	○ system ◉ person ◉ basic scientist ◉ clinician ◉ computer scientist ◉ regulatory body, lawyer, ethicist ◉ patient ○ other, please specify:
Pre-condition(s)/pre-requisite(s)	The end-user is signed in on the p-medicine Platform. He has an active SSO browser session on the IdP and has local sessions on at least one p-medicine service.
Requisite(s)	
Post-condition(s)/post-requisite(s)	The end-user is no longer signed in on the p-medicine Platform.

	All local service sessions of the end-user are destroyed.	
Constraints		
External sources needed from outside p-medicine	O data, please specify: O tools, please specify: O services, please specify: O models, please specify: O other, please specify:	
Data used	O personal O only non-personal O target population, please specify:	
Input data	O internal database, please specify: O external database, please specify: O online input	
Output data	O database, please specify: O variables for use, please specify: O structured document, please specify: O graphic, please specify:	
Data volume	Small	
Dataflow	Please specify:	
Data storage	Please specify:	
Successful End Condition	The end-user is no longer signed in on the p-medicine Platform. The active SSO session of the end-user on the IdP is destroyed. All active local end-user sessions on p-medicine services are destroyed.	
Fail End Condition	• The IdP was not able to destroy the active SSO session of the end-user. • A p-medicine service was not able to destroy the local end-user session.	
Basic workflow	Actor Action	System response
	The end-user selects the logout link on the local p-medicine web-site/service he is currently working on.	The local service sends a Single Logout (SLO) request for the end-user to the Identity Provider (IdP).
		The IdP sends a logout request for the end-user to all connected P-Medicine services (except the one that requested logout).

		Each of the contacted services attempts to destroy their local end-user session. Upon success they send back a logout response to the IdP, indicating the end-user session was successfully destroyed.
		The IdP destroys the SSO session of the end-user.
		The IdP sends a logout request to the service that initiated the Single Logout. This service then attempts to destroy the local session of the end-user and then sends back a logout response to the IdP if the session was successfully destroyed.
		The end-user is redirected to a page on the local web-site, stating the end-user successfully logged out.
Expected usage frequency	Normal, Single Sign-Out should be used any time an end-user wishes to end his session on the P-Medicine Platform.	
Needed for DSS	○ yes ◉ no	
Needs HPC	○ yes ◉ no	
Needs Grid	○ yes ◉ no	
Priority for development	Required	
Responsible for development	Custodix	
Mockup needed	○ yes ◉ no	
Responsible for Mockup	Not Applicable	
Who is building the tool		
Open Source tool	◉ yes ○ no, please specify why:	

Access Rights Scenario

Item	Description
Identifier	SG_7
Version	1.0
Name	Set Access Rights for a P-Medicine user
Description of the use case (enduser perspective)	This use case can only be specified after further analysis of the access control model requirements and research into possible approaches. For example who provides the role attributes? A local or central service? Or a combination of both?

User Enrolment Scenario

Item	Description
Identifier	SG_8
Version	1.0
Name	Enrole a P-Medicine User
Description of the use case (enduser perspective)	An end-user wants to register himself on a P-Medicine site/service, for this two user accounts are created one for the local site/service and one for the central IdP. Both accounts are linked using a pseudonymisation service.
Problem(s) to solve	Generate a local service account and central IdP account for a end-user and link both accounts using a pseudonymisation service.
Challenges	Hiding the complex registration functionality for the end-user, by providing user-friendly registration steps.
Risks	
Expected benefits	Make it possible to use Single Sign-On on the sites/services. Enables federation for the user.
Characterization	○ fundamental ◉ general ○ specific
If specific, please give the Domain	○ Acute lymphoblastic leukaemia ○ Breast Cancer ○ Nephroblastoma ○ other Cancer, please specify: ○ Non-Cancer Domain, please specify:
Enduser	○ system ◉ person ◉ basic scientist ◉ clinician ◉ computer scientist ◉ regulatory body, lawyer, ethicist ◉ patient ○ other, please specify:
Pre-condition(s)/pre-requisite(s)	The end-user does not have an account on the site/service where he wants to create an account.
Requisite(s)	
Post-condition(s)/post-requisite(s)	

Constraints		
External sources needed from outside p-medicine	○ data, please specify: ○ tools, please specify: ○ services, please specify: ○ models, please specify: ○ other, please specify:	
Data used	○ personal ○ only non-personal ○ target population, please specify:	
Input data	○ internal database, please specify: ○ external database, please specify: ○ online input	
Output data	○ database, please specify: ○ variables for use, please specify: ○ structured document, please specify: ○ graphic, please specify:	
Data volume	Small	
Dataflow	Please specify:	
Data storage	Please specify:	
Successful End Condition	The end-user is enrolled on the p-medicine Platform, meaning he has successful created a local and central account, both linked by a pseudonymisation service.	
Fail End Condition	• The end-user is not registered on the local site/service and/or central IdP user manager • The linking of central and local accounts is not succeeded.	
Basic workflow	Actor Action	System response
	An end-user browses to a P-Medicine site/service registration page, fills in a registration form and submits the form.	The registration information is send to the site/service where a new site/service account is created and stored in a user database. The new account contains a unique user ID.
		The site/service in turn forwards the registration information to a central IdP user manager.
		The IdP user manager generates a new IdP account with the given registration information and stores it in his

		local user database. A user pseudonym is generated for the IdP user account. This pseudonym is send back to the local site/service.
		A linking request containing the local user ID and the IdP user pseudonym is send to a pseudonymisation service by the local site/service.
		The pseudonymisation service links the local user ID with the IdP user pseudonym.
		After the linking step the site/service redirects the end-user to the IdP user manager. Where a page is displayed containing two options: already registered or new IdP registration (not explained).
	The end-user selects already registered	The end-user is redirected to the IdP where he is presented an authentication form.
	The end-user provides his authentication credentials	The IdP validates the authentication credentials and if the validation was successful it redirects the user to the IdP user manager.
		The IdP user manager now links both IdP accounts (the new account created by a request from the site/service and the old account that match the authentication)
		Finally the end-user is redirected to the original local/service, which terminates the registration flow.
Expected usage frequency	High, for every site/service in the p-medicine Platform that the end-user visits, he needs to create an account, which is linked to the central IdP manager.	
Needed for DSS	○ yes ◉ no	
Needs HPC	○ yes ◉ no	
Needs Grid	○ yes	

	⦿ no
Priority for development	Required
Responsible for development	Custodix
Mockup needed	O yes ⦿ no
Responsible for Mockup	not applicable
Who is building the tool	
Open Source tool	⦿ yes O no, please specify why:

Clinical Scenarios

Pathway Scenario for Nephroblastoma

Item	Description
Identifier	PSN_1
Version	1.0
Name	**Pathway Scenario for Nephroblastoma**
Description of the use case (end-user perspective)	Gene expression data from nephroblastoma serve as the source of disrupted metabolic pathways. These data needs to be normalized and then correlated to pathway data coming from the KEEG pathway database (http://www.genome.jp/kegg/pathway.html). Another possibility is MetaCore™ (http://www.genego.com/trial) from ThomsonReuters. These tools will analyse the tumour of disrupted metabolic pathways. By correlation to clinical data of patients, individual pathway disruptions or main disruptions for a cohort of patients with nephroblastoma will be produced as a result. The tool should be made in a general way that by describing the databases and the interfaces the tool will get domain independent.
Problem(s) to solve	To find disrupted pathways in nephroblastoma
Challenges	To make the tool domain independent for usage in other cancer domains
Risks	The KEGG database will get costly, meaning it will not longer be as open source available. See their website for more info.
Expected benefits	In individual patients it will be possible to find disrupted pathways in the tumour for selecting specific drugs for treatment, like ATRA (all-trans retinoic acid) if the retinoid pathway is disrupted.
Characterization	○ fundamental ◉ general (this should be the case) ◉ specific
If specific, please give the Domain	○ Acute lymphoblastic leukaemia ○ Breast Cancer ◉ Nephroblastoma ○ other Cancer, please specify: ○ Non-Cancer Domain, please specify:
End-user	○ system ◉ person ◉ basic scientist ◉ clinician

	○ computer scientist ○ regulatory body, lawyer, ethicist ○ patient ○ other, please specify:
Pre-condition(s)/pre-requisite(s)	Availability of gene expression data, normalisation of the gene expression data, availability of the KEGG database, availability of clinical data. Anonymization of personal data is needed.
Requisite(s)	If used as clinical decision support service (DSS)
Post-condition(s)/post-requisite(s)	If used as DSS the result in individual patients needs to be on time delivered. The result of this use case might be input for a data-mining tool that searches in literature for the best drugs to normalize disrupted pathways.
Constraints	If used as DSS the data from gene expression analysis, their normalisation, as well as the clinical data needs to be available on time. These logistics have to be solved otherwise (if data are coming late) the patient will not benefit from this use case as a DSS. This risk is independent of the IT.
External sources needed from outside p-medicine	⊙ data: KEEG pathway database (http://www.genome.jp/kegg/pathway.html) ⊙ tools: If data of the gene array experiment are not normalized a toll for normalizing this data is needed ○ services, please specify: ○ models, please specify: ○ other, please specify:
Data used	⊙ personal ○ only non-personal ⊙ target population: patients enrolled in SIOP 2001
Input data	⊙ internal database: a) clinical database: description: The clinical data will be provided by ObTiMA b) gene array expression data: The gene array data will be provided as string files. They need to be further specified. The data need to be normalized. If this is not the case a further step in the workflow for normalizing the data is needed. ⊙ external database: KEGG pathway database: (http://www.genome.jp/kegg/pathway.html) ⊙ online input: for the selection of a specific patient or a cohort of patients from the clinical database
Output data	○ database, please specify:

	○ variables for use, please specify: ◉ structured document: This document should list all disrupted pathways in the tumour. In case of a analysing a cohort of patients a histogram of the frequency of disrupted pathways in the cohort is given. A heatmap of the gene expression data is provided in case of analysing a cohort of patients. In a single patient only genes are listed that are responsible for the disrupted pathways. ○ graphic, please specify:	
Data volume	Large, depending on the number of cases and the number of genes analysed in the gene array experiments	
Dataflow	The data flow needs to be specified during the development of the tool. Data should be stored in the data warehouse.	
Data storage	Data will be stored in the data warehouse after anonymization. If the tools get productive Data storage needs to be fixed.	
Successful End Condition	Delivering disrupted pathways in nephroblastoma for a single patient or a cohort of patients	
Fail End Condition	No pathways are disrupted	
Basic workflow	Actor Action	System response
	Selection of the clinical database	A view of the database will be given
	Selection of a single patient or a group of patients	Only data from the single patient or the cohort of patients will be used in running the scenario
		The system automatically finds the clinical data, the gene array expression data and the KEGG database. The workflow itself was defined by the Tool builder before
		In case of a single Patient: Results of the scenario are displayed as a structured list of disrupted pathways. Only disrupted pathways are shown In case of a cohort of patients the list of disrupted pathways is given as a structured list also displaying the percentage of patients for every pathway that is disrupted in the selected

		cohort of patients. A heatmap of the gene expression data is provided in case of analysing a cohort of patients. In a single patient only genes are listed that are responsible for the disrupted pathways.
	Download the results on the own computer	
Expected usage frequency	Regularly for every single patient entered in SIOP 2001 in whom gene expression data are available.	
Needed for DSS	⊙ yes O no	
Needs HPC	⊙ yes ? in DSS fast results needs to be available. The usage is depending on the time for running the scenario O no	
Needs Grid	⊙ yes ? in DSS fast results needs to be available. The usage is depending on the time for running the scenario O no	
Priority for development	high	
Responsible for development	Will be decided by the IT group	
Mock-up needed	⊙ yes O no	
Responsible for Mock-up	Will be decided by the IT group	
Who is building the tool	Will be decided by the IT group	
Open Source tool	⊙ yes O no, please specify why:	

Imaging Scenario

Item	Description
Identifier	PSN_2
Version	1.0
Name	**Imaging scenario for Nephroblastoma**
Description of the use case (end-user perspective)	DICOM imaging data of a patient with nephroblastoma need to be uploaded to the data warehouse for further usage of the imaging data
Problem(s) to solve	Handling of DICOM data within a clinical trial
Challenges	To make the tool independent of the domain for usage in other diseases
Risks	None
Expected benefits	Fast and safe diagnosis of nephroblastoma by involving reference radiologists for second opinion Post-processing of imaging studies for the Oncosimulator is possible This use case can be generalized for all other types of cancer
Characterization	○ fundamental ◉ general ◉ specific
If specific, please give the Domain	○ Acute lymphoblastic leukaemia ◉ Breast Cancer ◉ Nephroblastoma ○ other Cancer, please specify: ○ Non-Cancer Domain, please specify:
End-user	○ system ◉ person ◉ basic scientist ◉ clinician ○ computer scientist ○ regulatory body, lawyer, ethicist ○ patient ○ other, please specify:
Pre-condition(s)/pre-requisite(s)	Data warehouse needs to be established
Requisite(s)	Availability of DICOM data from the local hospital
Post-condition(s)/post-requisite(s)	Availability of tools for post-processing of the imaging data, like DoctorEye
Constraints	Anonymization/pseudonymization of imaging data before upload in the data warehouse

External sources needed from outside p-medicine	○ data, please specify: ○ tools, please specify: ○ services, please specify: ○ models, please specify: ○ other, please specify:
Data used	◉ personal ○ only non-personal ◉ target population, please specify: patients enrolled in SIOP 2001
Input data	◉ internal database, please specify: the clinical data will be provided by ObTiMA the imaging data will be handled in the data warehouse ○ external database, please specify: ○ online input
Output data	◉ database, please specify: post-processing imaging data for use in the Oncosimulator stored in the data warehouse ○ variables for use, please specify: ◉ structured document, please specify: report from reference radiology generated by ObTiMA after filling in a CRF for reference Radiology ◉ graphic, please specify: Histogram of signal intensities of the tumour 3d tumour volume before and after preoperative chemotherapy
Data volume	Large
Dataflow	DICOM data need to be exported from a local PACS to a communication server with pseudonymization of the DICOM files, then uploaded to the data warehouse to be used for reference radiology and post-processing. Post-processing data will be stored automatically in the data warehouse with the annotation of what segmentations, etc. done with DoctorEye. Post-processing data will be uploaded into the Oncosimulator
Data storage	DICOM data will be stored in the data warehouse after pseudonymization
Successful End Condition	DICOM files are reviewed by reference radiologists and DICOM data are post-processed for further usage
Fail End Condition	DICOM data are not available for upload to the data warehouse

Basic workflow	Actor Action	System response
	Local DICOM data need to be exported from the local PACS on a local communication server for pseudonymization or the local DICOM fields need to be uploaded from a CD to the local communication server for pseudonymization	As soon as DICOM files are stored on the communication server a notice is given to the local user. After pseudonymization a second notice is send to the local user.
	After pseudonymization the DICOM files will be automatically uploaded to the data warehouse	After upload of the data to the data warehouse the local user is notified of the successful uploading process
		After storage of the DICOM files in the data warehouse the reference radiologist is notified that new DICOM files are available for reference diagnosis The person responsible for post-processing of the data is notified that such DICOM files are available The patient is notified that his imaging data are stored in the data warehouse
	Reference radiology	
	The reference radiologist is able to select the DICOM files from the data warehouse by listing him all DICOM Files with no reference radiology.	The DICOM files are sorted according to pseudonyms of patients, modality of the DICOM files (MRI, CT, ultrasound, PET, …) and date of the study. The GUI should be according to the timeline view in ContraCancrum[91]
	After selection of the data he is able to download this pseudonymized data on his PACS system for reference radiology	At the same time ObTiMA will open and display the CRF for reference radiology for the specific patient
	Reference radiology will have access to ObTiMA and the CRF for reference	After finalizing the input of data in the CRF a standardized report of the

[91] ContraCancrum: Clinically Oriented Translational Cancer Multilevel Modelling, FP7 project: 223979

	radiology for the selected patient to input his data	reference radiologist is created
	The reference radiologists gives approval to the pseudonymized standardized report	The report is automatically send to trust centre for de-pseudonymization and then automatically send to the local physician treating the patient.
		The local physician can produce the reference radiology report at any time via ObTiMA
	User responsible for Post-processing	
	The user selects the DICOM with right mouse clicking	A menu pops up showing different tools for viewing or post-processing of the DICOM files, e.g. DoctorEye, or a DICOM viewer, etc.
	The user clicks on the tool he wants to work with	DoctorEye will open the DICOM files automatically
	The user has selected DoctorEye	The DICOM files are uploaded into DoctorEye for post-processing (segmentation of the tumour)
	The user finishes the post-processing with DoctorEye	The DICOM files are stored in the data warehouse with the annotation of finishing the segmentation and being ready for upload in the Oncosimulator. The user is guided to the DICOM files of another patient for post-processing. If all DICOM files from all patients are post-processed DoctorEye quits
	Viewing imaging files by the patient	
	The patient who wants to view his imaging files gets access to his DICOM files stored in the data warehouse after entering the p-medicine platform via the portal	The user needs to install a DICOM Viewer on his computer. The system asks if a DICOM viewer is installed
	The patient answers that he has not installed a DICOM Viewer	The system does not show any DICOM files in the timeline of the GUI and asks the user to install a DICOM viewer (DicomWorks: http://dicom.online.fr/ for

		Windows and Osirix: http://www.osirix-viewer.com/ for Mac)
	The patient answers that he has installed a DICOM Viewer	In case the patient has credentials to view his imaging files the graphical interface with the timeline of his data shows also his DICOM files.
	The patient selects his DICOM files by right mouse clicking	A menu appears where with the item DICOM Viewer and the item Download
	The patient clicks on DICOM Viewer	The DICOM files are opened in the DICOM Viewer on the client side. After closing the DICOM viewer the patient is referred back to the GUI of p-medicine
	Downloading imaging files by the patient	
	The patient who wants to view his imaging files gets access to his DICOM files stored in the data warehouse after entering the p-medicine platform via the portal	In case the patient has credentials to download his imaging files the graphical interface with the timeline of his data shows also his DICOM files.
	The patient selects his DICOM files by right mouse clicking	A menu appears where with the item DICOM Viewer and the item Download
	The patient click son Download	The patient is asked where to store the files
	The patient selects the directory to store the files	The download starts. After the end of the download the patient is referred back to the GUI of p-medicine
Expected usage frequency	high	
Needed for DSS	○ yes ◉ no	
Needs HPC	○ yes ◉ no	
Needs Grid	○ yes ◉ no	
Priority for development	high	
Responsible for development	Will be decided by the IT group	

Mock-up needed	⦿ yes O no
Responsible for Mock-up	Will be decided by the IT group
Who is building the tool	Will be decided by the IT group
Open Source tool	⦿ yes O no, please specify why

(Severe) Adverse Event ((S)AE) Prediction Scenario

Item	Description
Identifier	PSN_3
Version	1.0
Name	(S)AE prediction in nephroblastoma
Description of the use case (end-user perspective)	If the risk of a (S)AE can be predicted patients would benefit from a safer treatment. All data of a patient will be checked against data from (S)AE/SUSAR databanks, clinical trials and from literature by data mining to describe the individual risk in developing specific (S)AEs.
Problem(s) to solve	Prediction of (S)AEs
Challenges	To make the tool independent for the disease and the drug to search for
Risks	Wrong prediction and the patient will not receive a potential efficient drug for treatment
Expected benefits	Less (S)AEs in treating patients
Characterization	○ fundamental ◉ general ◉ specific
If specific, please give the Domain	○ Acute lymphoblastic leukaemia ○ Breast Cancer ◉ Nephroblastoma ○ other Cancer, please specify: ○ Non-Cancer Domain, please specify: The tool should be able to address every disease and every drug
End-user	○ system ◉ person ○ basic scientist ◉ clinician ○ computer scientist ○ regulatory body, lawyer, ethicist ○ patient ○ other, please specify:
Pre-condition(s)/pre-requisite(s)	Availability of SAE databases
Requisite(s)	Access to SAE databases, data mining tools
Post-condition(s)/post-requisite(s)	

Constraints	Only the treating physician is allowed to run the tool as personal data are needed. In case an analysis of a cohort of patients will be done as a research project personal data needs to be anonymized.
External sources needed from outside p-medicine	⊙ data, please specify: SAE databases: **EMA:** The European Medicines Agency has published its plans for granting public access to the information held in its databases of the potential side effects of human and veterinary medicines. As long as there is no access to such a database continue to be updated via EudraVigilance: http://eudravigilance.ema.europa.eu/human/index.asp **FDA:** The FDA provides a database for reporting of adverse events called the *Manufacturer and User Facility Device Experience Database* (MAUDE). The data consist of voluntary reports since June 1993, user facility reports since 1991, distributor reports since 1993, and manufacturer reports since August 1996, and is open for public view: http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfMAUDE/search.cfm?searchoptions The Adverse Event Reporting System (AERS): http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Surveillance/AdverseDrugEffects/ucm082193.htm **Canada:** The Canada Vigilance Adverse Reaction Online Database contains information about suspected adverse reactions (also known as side effects) to health products, captured from adverse reaction reports submitted to Health Canada by consumers and health professionals, who submit reports voluntarily, as well as by market authorization holders (manufacturers and distributors): http://www.hc-sc.gc.ca/dhp-mps/medeff/databasdon/index-eng.php **Clinical Trial databases** The Cochrane Central Register of Controlled Trials (Clinical Trials; CENTRAL) database contains approx. **500,000** records: http://onlinelibrary.wiley.com/o/cochrane/cochrane_clcentral_articles_fs.html

	ClinicalTrials.gov is a registry and results database of federally and privately supported clinical trials conducted in the United States and around the world. ClinicalTrials.gov gives you information about a trial's purpose, who may participate, locations, and phone numbers for more details: http://clinicaltrials.gov/ The EU Clinical Trials Register, launched on 22 March, allows to search for information on interventional clinical trials for medicines authorised in the 27 EU Member States as well as Iceland, Liechtenstein and Norway: https://www.clinicaltrialsregister.eu/ **Databases for Literature Mining** Medline/PubMed: http://www.ncbi.nlm.nih.gov/pubmed Cochrane Library: http://www.thecochranelibrary.com/view/0/ AboutTheCochraneLibrary.html Embase: http://www.embase.com/ Summary of databases can be found at: http://www.ncbi.nlm.nih.gov/guide/literature/ O tools, please specify: O services, please specify: O models, please specify: O other, please specify:
Data used	⊙ personal O only non-personal O target population, please specify:
Input data	⊙ internal database, please specify: clinical data coming from ObTiMA and research data from the p-medicine data warehouse ⊙ external database, please specify: see above O online input
Output data	O database, please specify: O variables for use, please specify: ⊙ structured document, please specify: summary report of 1.risk profile of each (S)AE of the analysed drug 2.potential risks in an individual patient related to investigated drug(s) O graphic, please specify:

Data volume	low	
Dataflow	Please specify: clinical and research data are extracted from the individual patient and with the help of data mining tools the individual risk profile of a patient is analysed related to a specific drug	
Data storage	Please specify: to be specified by IT	
Successful End Condition	Predicted risk of an (S)AE is given	
Fail End Condition	No risk profile can be given	
Basic workflow	Actor Action	System response
	Patient is selected from ObTiMA	All relevant data of a patient are selected and an individual patient profile is build
	The drug of concern is selected	Adverse events (AE) of the drug are collected from literature and clinical trials (see data sources above) by using data mining tools
		Risk factors for developing a specific (S)AE are collected for each of the detected (S)AEs of the drug by using data mining tools
		Statistical risk profiles are given for each (S)AE
		The risk profile found will be compared with the patient profile and statically analysed to define the individual risk of an (S)AE in a specific patient
		Results are given in a structured report
Expected usage frequency	High, in each individual patient	
Needed for DSS	○ yes ◉ no, only after validation of the tool t will be used for DSS	
Needs HPC	○ yes ◉ no	
Needs Grid	○ yes ◉ no	

Priority for development	high
Responsible for development	Needs to be defined
Mock-up needed	○ yes ◉ no
Responsible for Mock-up	
Who is building the tool	Needs to be defined
Open Source tool	◉ yes ○ no, please specify why

Tumour Marker Scenario for nephroblastoma

Item	Description
Identifier	PSN_4
Version	1.0
Name	Tumour Marker in Nephroblastoma
Description of the use case (end-user perspective)	There are no serum tumour markers known in nephroblastoma predicting outcome or specific subtypes. This use case will define a pattern of miRNAs, tumour specific autoantibodies and other serum proteins as specific markers for nephroblastoma.
Problem(s) to solve	There is a risk of 1% of wrong diagnosis without histological tumour diagnosis in nephroblastoma and stratified treatment for specific subtypes starts after histological diagnosis only.
Challenges	To use the tool for other cancer types by defining the specification of input data
Risks	No specific patterns will be found
Expected benefits	Better stratification of patients from the time of diagnosis
Characterization	○ fundamental ○ general ◉ specific
If specific, please give the Domain	○ Acute lymphoblastic leukaemia ○ Breast Cancer ◉ Nephroblastoma ○ other Cancer, please specify: ○ Non-Cancer Domain, please specify:
End-user	○ system ◉ person ◉ basic scientist ◉ clinician ○ computer scientist ○ regulatory body, lawyer, ethicist ○ patient ○ other, please specify:
Pre-condition(s)/pre-requisite(s)	miRNA, autoantibody and other protein data needs to be available
Requisite(s)	Pseudonymization of data is needed
Post-condition(s)/post-requisite(s)	

Constraints	none
External sources needed from outside p-medicine	○ data, please specify: ○ tools, please specify: ○ services, please specify: ○ models, please specify: ○ other, please specify:
Data used	⊙ personal ○ only non-personal ⊙ target population, please specify: patients with kidney tumours
Input data	⊙ internal database, please specify: clinical data are coming from ObTiMA Research data need to be stored in the data warehouse ○ external database, please specify: ○ online input
Output data	○ database, please specify: ○ variables for use, please specify: ⊙ structured document, please specify: A list of miRNA , autoantibodies and other protein data is given for: Best correlation with outcome, histological subtype and response to preoperative chemotherapy including a complete sensitivity/specificity report ⊙ graphic, please specify: for outcome: life tables showing the difference in outcome between patients with and without the specific pattern. This is done for all patterns that are found for histology and response to treatment: Receiver Operating Characteristic (ROC) curves
Data volume	High (research data)
Dataflow	Please specify: All data used will be anonymized, all data will be stored in the data warehouse
Data storage	Please specify: Clinical data are stored in ObTiMA and transferred to the data warehouse, research data are stored in the data warehouse
Successful End Condition	A pattern of miRNA, autoantibodies and other proteins is found that correlate with outcome, histological subtype and response to treatment

Fail End Condition	No pattern of markers is found	
Basic workflow	Actor Action	System response
	All needed data are anonymized and stored in the data warehouse	Statistical analysis of the data takes place
		Structured report and graphical output is done as specified above
Expected usage frequency	Moderate as long as it is only used for nephroblastoma. If the tool is written in a generalized way it can be used for other cancer types	
Needed for DSS	○ yes ◉ no	
Needs HPC	○ yes ◉ no	
Needs Grid	○ yes ◉ no	
Priority for development	Needs to be defined	
Responsible for development	Needs to be defined	
Mock-up needed	○ yes ◉ no	
Responsible for Mock-up		
Who is building the tool	Needs to be defined	
Open Source tool	◉ yes ○ no, please specify why	

Oncosimulator Scenario for nephroblastoma

Item	Description
Identifier	PSN_5
Version	1.0
Name	Oncosimulator for Nephroblastoma
Description of the use case (end-user perspective)	The Oncosimulator is at the same time a concept of multilevel integrative cancer biology, a complex algorithmic construct, a biomedical engineering system and eventually a clinical tool which primarily aims at supporting the clinician in the process of optimizing cancer treatment in the patient individualized context through conducting experiments *in silico* i.e. on the computer. Additionally it is a platform for simulating, investigating, better understanding and exploring the natural phenomenon of cancer, supporting the design and interpretation of clinicogenomic trials and finally training doctors, researchers and interested patients alike. The present version of the Oncosimulator refers to nephroblastoma.
Problem(s) to solve	To predict the likely response of a given patient's nephroblastoma to one or more candidate treatment schemes while toxicological limitations are taken into account.
Challenges	To clinically adapt and validate the nephroblastoma Oncosimulator in such an extent so as to allow its clinical translation.
Risks	Availability of and access to a sufficient number of sets of multiscale data which will allow both clinical adaptation and translation to be statistically reliable and trustable.
Expected benefits	Personalization of treatment, optimization of treatment outcome, increase of life expectancy and improvement of the quality of life.
Characterization	⊙ fundamental (in the sense that it contains extensive fundamental/basic science components) O general ⊙ specific
If specific, please give the Domain	O Acute lymphoblastic leukaemia O Breast Cancer ⊙ Nephroblastoma O other Cancer, please specify: O Non-Cancer Domain, please specify:
End-user	O system ⊙ person ⊙ basic scientist

	⊙ clinician O computer scientist O regulatory body, lawyer, ethicist O patient O other, please specify:
Pre-condition(s)/pre-requisite(s)	Availability of and access to sets of multiscale data (including inter alia imaging, histological, molecular and clinical data) for several patients and at several time points. Pseudo/Anonymization of personal data is needed.
Requisite(s)	If used as clinical decision support service (DSS)
Post-condition(s)/post-requisite(s)	If used as DSS the result in individual patients needs to be delivered on time.
Constraints	If used as DSS the multiscale data need to be available on time. Related logistics have to be solved, otherwise (if data is coming late) the patient will not benefit from this use case as a DSS. This risk is independent of the IT.
External sources needed from outside p-medicine	⊙ data, please specify: Population based mean or typical model parameter values (from literature) concerning pharmacokinetics, tumour biology and others O tools, please specify: O services, please specify: O models, please specify: O other, please specify:
Data used	⊙ personal O only non-personal ⊙ target population, please specify: Patients enrolled in SIOP 2001
Input data	⊙internal database, please specify: Database(s) containing the multiscale data of the patients O external database, please specify: ⊙online input: for the selection of a specific patient or a cohort of patients from the clinical database
Output data	⊙ database, please specify: in ObTiMA a table will be generated where the predicted shrinkage of the tumour in each patient will be stored and the result of the validation: correct prediction, wrong prediction O variables for use, please specify: ⊙ structured document, please specify: The document gives the probability (range of several runs) how much the tumour will shrink during preoperative chemotherapy. If there is at least 10% of shrinkage the Oncosimulator will state that preoperative chemotherapy is useful. The physician has to decide

	about the treatment. After validation with the real data a document will be produced giving as a result: correct or wrong prediction of the Oncosimulator. ⊙graphic, please specify: The Oncosimulator predictions will be provided in various forms including scalar numbers, graphs, 3D and 4D visualizations
Data volume	Large, depending on the number of cases and the number time points for which multiscale data is available.
Dataflow	Please specify: Data should be stored in the data warehouse. In the basic workflow section the data collection and processing steps are outlined. DICOM data need to be stored in the data warehouse (use case: PSN_2), the DICOM data need to be used in DrEye for pre-processing (rendering the tumour and defining the histogram of signal intensities), clinical data are coming from ObTiMA, research data from the data warehouse and ObTiMA
Data storage	Please specify: Data will be stored in the data warehouse after anonymyzation.
Successful End Condition	Predicting the response of a given patient's nephroblastoma to a candidate treatment
Fail End Condition	No response can be calculated due to incompatibility of the parameter values considered. A warning will be generated.
Basic workflow (see also fig. 5.1)	**First step**: Obtain patient's individual multiscale and inhomogeneous data. Data sets to be collected for each patient include: clinical data (age, sex, weight etc.), eventual previous anti-tumour treatment history, imaging data (e.g. MRI, CT, PET etc images) (when applicable), histopathological data (e.g. detailed identification of the tumour type, grade and stage, histopathology slide images whenever biopsy is allowed and feasible and/or haematological test data.), molecular data (DNA array data, selected molecular marker values or statuses, serum markers etc.). It is noted that the last two data categories are extracted from biopsy material and/or body fluids.
	Second step: Preprocess patient's data. The data collected are pre-processed in order to take an adequate form allowing its introduction into the "Tumour and Normal Tissue Response Simulation" module of the Oncosimulator. For example the imaging data are segmented, interpolated, eventually fused and subsequently the anatomic entity/-ies of interest is/are three-dimensionally reconstructed. This reconstruction will provide the framework for the integration of the rest of

	data and the execution of the simulation. In parallel the molecular data is processed via molecular interaction networks so as to perturb and individualize the average pharmacodynamic or radiobiological cell survival parameters.	
	Third step: Describe one or more candidate therapeutic scheme(s) and/or schedule(s). The clinician describes a number of candidate therapeutic schemes and/or schedules and/or no treatment (obviously leading to free i.e. non-inhibited tumour growth), to be simulated in silico i.e. on the computer.	
	Fourth step: Run the simulation. The computer code of tumour growth and treatment response is massively executed on distributed grid or cluster computing resources so that several candidate treatment schemes and/or schedules are simulated for numerous combinations of possible tumour parameter values in parallel. Predictions concerning the toxicological compatibility of each candidate treatment scheme are also produced or alternatively estimates of the toxicologically acceptable dosage limits are retrieved from literature.	
	Fifth step: Visualize the predictions. The expected reaction of the tumour as well as toxicologically relevant side effect estimates for all scenarios simulated are visualized using several techniques ranging from simple graph plotting to four multidimensional rendering.	
	Sixth step: Evaluate the predictions and decide on the optimal scheme or schedule to be administered to the patient. The clinician carefully evaluates the Oncosimulator's predictions by making use of their logic, medical education and even qualitative experience. If no serious discrepancies are detected, the predictions support the clinician in taking their final and expectedly optimal decision regarding the actual treatment to be administered to the patient.	
	Seventh step: Apply the theoretically optimal therapeutic scheme or schedule and further optimize the Oncosimulator. The expectedly optimal therapeutic scheme or schedule is administered to the patient. Subsequently, the predictions regarding the finally adopted and applied scheme or schedule are compared with the actual tumour course and a negative feedback signal is generated and used in order to optimize the Oncosimulator.	
	Actor Action	System response
	Patient will be selected in ObTiMA	If DICOM data are available on in the data warehouse (PSN_2) DICOM data are send to DrEye. DrEye opens

		with the DICOM data of the patient. If no DICOM data are available in the data warehouse an automated request is send to the local hospital to provide the DICOM data of this patient.
	Physician renders the tumour.	By finishing the pre-processing the data are automatically stored in the data warehouse. The stored data will be annotated that the pre-processing of the data is finished. Tumour volume will be automatically estimated by DrEye and stored in ObTiMA.
	Clinical data and research data are pseudonymized and send to the data warehouse	The Oncosimulator imports all needed data. Variables coming from literature are predefined in the Oncosimulator by specifying a range. If data are missing an automated request is send.
		If all data are available access to HPC is established and the Oncosimulator will be executed. According to the possible range of several variables several runs are needed always giving one percentage of shrinkage. All these percentages are stored together with the specific variable's data. After all runs are finished a statistic is made giving the median, mean and range of shrinkage as well as standard deviation. These will be stored in ObTiMA together with the date and time of the run.
		Structured output is given as specified above.
	At the end of the preoperative chemotherapy a new imaging study is done and the data are uploaded	After upload of the data DrEye will open with the specific data of the patient.

	after pseudonymization to the data warehouse	
	Physician renders the tumour.	After finishing the data are automatically stored in the data warehouse and annotated that the tumour is rendered. Tumour volume will be automatically estimated by DrEye and stored in ObTiMA.
		An automatic comparison between the prediction of the Oncosimulator and the real shrinkage will be done and evaluated, if the prediction was correct according to the predefined definition. An output is generated as written above. The result of the evaluation is stored in ObTiMA as given above.
Expected usage frequency	Regularly for many patients entered the SIOP 2001 clinical trial provided that the necessary multiscale data is available.	
Needed for DSS	⊙ yes O no	
Needs HPC	⊙yes (when the resolution of the predictions has to be high) O no	
Needs Grid	⊙yes (when several executions have to take place in order to offset the model parameter value expected deviations) O no	
Priority for development	high	
Responsible for development	ICCS-NTUA	
Mockup needed	O yes ⊙ no	
Responsible for Mockup		
Who is building the tool	ICCS-NTUA in collaboration with other WP12 participants	
Open Source tool	⊙ yes O no, please specify why: ⊙ Open access tool (to be also included in the European Cancer Model Repository developed by the TUMOR project)	

Oncosimulator Scenario for Breast Cancer

Item	Description
Identifier	PSB_1
Version	1.0
Name	Oncosimulator for Breast Cancer
Description of the use case (enduser perspective)	The Oncosimulator is at the same time a concept of multilevel integrative cancer biology, a complex algorithmic construct, a biomedical engineering system and eventually a clinical tool which primarily aims at supporting the clinician in the process of optimizing cancer treatment in the patient individualized context through conducting experiments *in silico* i.e. on the computer. Additionally it is a platform for simulating, investigating, better understanding and exploring the natural phenomenon of cancer, supporting the design and interpretation of clinicogenomic trials and finally training doctors, researchers and interested patients alike. The present version of the Oncosimulator refers to breast cancer.
Problem(s) to solve	To predict the likely response of a given patient's breast cancer to one or more candidate treatment schemes while toxicological limitations are taken into account.
Challenges	To clinically adapt and validate the breast cancer Oncosimulator in such an extent so as to allow its clinical translation.
Risks	Availability of and access to a sufficient number of sets of multiscale data, which will allow both clinical adaptation and translation to be statistically reliable and trustable.
Expected benefits	Personalization of treatment, optimization of treatment outcome, increase of life expectancy and improvement of the quality of life.
Characterization	⦿ fundamental (in the sense that it contains extensive fundamental/basic science components) O general ⦿specific
If specific, please give the Domain	O Acute lymphoblastic leukaemia ⦿ Breast Cancer ONephroblastoma O other Cancer, please specify: O Non-Cancer Domain, please specify:
Enduser	O system ⦿ person ⦿basic scientist

	⊙ clinician O computer scientist O regulatory body, lawyer, ethicist O patient O other, please specify:
Pre-condition(s)/pre-requisite(s)	Availability of and access to sets of multiscale data (including inter alia imaging, histological, molecular and clinical data) for several patients and at several time points. Pseudo/Anonymization of personal data is needed.
Requisite(s)	If used as clinical decision support service (DSS)
Post-condition(s)/post-requisite(s)	If used as DSS the result in individual patients needs to be delivered on time.
Constraints	If used as DSS the multiscale data need to be available on time. Related logistics have to be solved, otherwise (if data is coming late) the patient will not benefit from this use case as a DSS. This risk is independent of the IT.
External sources needed from outside p-medicine	⊙ data, please specify: Population based mean or typical model parameter values (from literature) O tools, please specify: O services, please specify: O models, please specify: O other, please specify:
Data used	⊙ personal O only non-personal ⊙ target population, please specify: Patients enrolled in SIOP 2001
Input data	⊙internal database, please specify: Database(s) containing the multiscale data of the patients O external database, please specify: ⊙online input: for the selection of a specific patient or a cohort of patients from the clinical database
Output data	O database, please specify: O variables for use, please specify: O structured document, please specify: ⊙graphic, please specify: The Oncosimulator predictions will be provided in various forms including scalar numbers, graphs, 3D and 4D visualizations
Data volume	Large, depending on the number of cases and the number time points for which multiscale data is available.
Dataflow	Data should be stored in the data warehouse. In the basic workflow section the data collection and processing steps are outlined.

Data storage	Please specify: Data will be stored in the data warehouse after anonymization.
Successful End Condition	Predicting the response of a given patient's breast cancer to a candidate treatment
Fail End Condition	No response can be calculated due to incompatibility of the parameter values considered. A warning will be generated.
Basic workflow	**First step**: Obtain patient's individual multiscale and inhomogeneous data. Data sets to be collected for each patient include: clinical data (age, s ex, weight etc.), eventual previous anti-tumour treatment history, imaging data (e.g. MRI, CT, PET etc images) (when applicable), histopathological data (e.g. detailed identification of the tumour type, grade and stage, histopathology slide images whenever biopsy is allowed and feasible and/or haematological test data.), molecular data (DNA array data, selected molecular marker values or statuses, serum markers etc.). It is noted that the last two data categories are extracted from biopsy material and/or body fluids.
	Second step: Pre-process patient's data. The data collected are pre-processed in order to take an adequate form allowing its introduction into the "Tumour and Normal Tissue Response Simulation" module of the Oncosimulator. For example the imaging data are segmented, interpolated, eventually fused and subsequently the anatomic entity/-ies of interest is/are three-dimensionally reconstructed. This reconstruction will provide the framework for the integration of the rest of data and the execution of the simulation. In parallel the molecular data is processed via molecular interaction networks so as to perturb and individualize the average pharmacodynamic or radiobiological cell survival parameters.
	Third step: Describe one or more candidate therapeutic scheme(s) and/or schedule(s). The clinician describes a number of candidate therapeutic schemes and/or schedules and/or no treatment (obviously leading to free i.e. non-inhibited tumour growth), to be simulated in silico i.e. on the computer.

	Fourth step: Run the simulation. The computer code of tumour growth and treatment response is massively executed on distributed grid or cluster computing resources so that several candidate treatment schemes and/or schedules are simulated for numerous combinations of possible tumour parameter values in parallel. Predictions concerning the toxicological compatibility of each candidate treatment scheme are also produced or alternatively estimates of the toxicologically acceptable dosage limits are retrieved from literature.
	Fifth step: Visualize the predictions. The expected reaction of the tumour as well as toxicologically relevant side effect estimates for all scenarios simulated are visualized using several techniques ranging from simple graph plotting to four multidimensional rendering.
	Sixth step: Evaluate the predictions and decide on the optimal scheme or schedule to be administered to the patient. The clinician carefully evaluates the Oncosimulator's predictions by making use of their logic, medical education and even qualitative experience. If no serious discrepancies are detected, the predictions support the clinician in taking their final and expectedly optimal decision regarding the actual treatment to be administered to the patient.
	Seventh step: Apply the theoretically optimal therapeutic scheme or schedule and further optimize the Oncosimulator. The expectedly optimal therapeutic scheme or schedule is administered to the patient. Subsequently, the predictions regarding the finally adopted and applied scheme or schedule are compared with the actual tumour course and a negative feedback signal is generated and used in order to optimize the Oncosimulator.
Expected usage frequency	Regularly for many patients entered the bevacizumab breast cancer clinical trial addressed by p-medicine provided that the necessary multiscale data is available.
Needed for DSS	⊙ yes O no
Needs HPC	⊙yes (when the resolution of the predictions has to be high) O no
Needs Grid	⊙yes (when several executions have to take place in order to offset the model parameter value expected deviations) O no
Priority for development	high
Responsible for development	ICCS-NTUA

Mockup needed	○ yes ◉ no
Responsible for Mockup	
Who is building the tool	ICCS-NTUA in collaboration with other WP12 participants
Open Source tool	◉ yes ○ no, please specify why: ◉ Open access tool (to be also included in the European Cancer Model Repository developed by the TUMOR project)

Oncosimulator Scenario for ALL

Item	Description
Identifier	PSL_1
Version	1.0
Name	Oncosimulator for Acute Lymphoblastic Leukaemia (ALL)
Description of the use case (enduser perspective)	The Oncosimulator is at the same time a concept of multilevel integrative cancer biology, a complex algorithmic construct, a biomedical engineering system and eventually a clinical tool which primarily aims at supporting the clinician in the process of optimizing cancer treatment in the patient individualized context through conducting experiments *in silico* i.e. on the computer. Additionally it is a platform for simulating, investigating, better understanding and exploring the natural phenomenon of cancer, supporting the design and interpretation of clinicogenomic trials and finally training doctors, researchers and interested patients alike. The present version of the Oncosimulator refers to acute lymphoblastic leukaemia (ALL).
Problem(s) to solve	To predict the likely response of a given patient's ALL to one or more candidate treatment schemes while toxicological limitations are taken into account.
Challenges	To clinically adapt and validate the ALL Oncosimulator in such an extent so as to allow its clinical translation.
Risks	Availability of and access to a sufficient number of sets of multiscale data, which will allow both clinical adaptation and translation to be statistically reliable and trustable.
Expected benefits	Personalization of treatment, optimization of treatment outcome, increase of life expectancy and improvement of the quality of life.
Characterization	⦿ fundamental (in the sense that it contains extensive fundamental/basic science components) O general ⦿specific
If specific, please give the Domain	⦿ Acute lymphoblastic leukaemia O Breast Cancer O Nephroblastoma O other Cancer, please specify: O Non-Cancer Domain, please specify:
Enduser	O system ⦿person ⦿basic scientist ⦿ clinician

	○ computer scientist ○ regulatory body, lawyer, ethicist ○ patient ○ other, please specify:
Pre-condition(s)/pre-requisite(s)	Availability of and access to sets of multiscale data (including inter alia histological, molecular and clinical data) for several patients and at several time points. Pseudo/Anonymization of personal data is needed.
Requisite(s)	If used as clinical decision support service (DSS)
Post-condition(s)/post-requisite(s)	If used as DSS the result in individual patients needs to be delivered on time.
Constraints	If used as DSS the multiscale data need to be available on time. Related logistics have to be solved, otherwise (if data is coming late) the patient will not benefit from this use case as a DSS. This risk is independent of the IT.
External sources needed from outside p-medicine	⊙ data, please specify: Population based mean or typical model parameter values (from literature) ○ tools, please specify: ○ services, please specify: ○ models, please specify: ○ other, please specify:
Data used	⊙ personal ○ only non-personal ⊙ target population, please specify: Patients enrolled in the ALL clinical trials included in p-medicine
Input data	⊙internal database, please specify: Database(s) containing the multiscale data of the patients ○external database, please specify: ⊙online input: for the selection of a specific patient or a cohort of patients from the clinical database
Output data	○database, please specify: ○ variables for use, please specify: ○ structured document, please specify: ⊙graphic, please specify: The Oncosimulator predictions will be provided in various forms including scalar numbers, graphs, 3D and 4D visualizations
Data volume	Large, depending on the number of cases and the number time points for which multiscale data is available.
Dataflow	Data should be stored in the data warehouse. In the basic workflow section the data collection and processing steps are outlined.

Data storage	Please specify: Data will be stored in the data warehouse after anonymyzation.
Successful End Condition	Predicting the response of a given patient's ALL to a candidate treatment
Fail End Condition	No response can be calculated due to incompatibility of the parameter values considered. A warning will be generated.
Basic workflow	**First step**: Obtain patient's individual multiscale and inhomogeneous data. Data sets to be collected for each patient include: clinical data (age, sex, weight etc.), eventual previous anti-tumour treatment history, imaging data (e.g. MRI, CT, PET etc images) (if and when applicable), histopathological data (e.g. detailed identification of the tumour type, grade and stage, histopathology slide images whenever biopsy is allowed and feasible and/or haematological test data.), molecular data (DNA array data, selected molecular marker values or statuses, serum markers etc.). It is noted that the last two data categories are extracted from biopsy material and/or body fluids.
	Second step: Pre-process patient's data. The data collected are pre-processed in order to take an adequate form allowing its introduction into the "Tumour and Normal Tissue Response Simulation" module of the Oncosimulator. For example the imaging data are segmented, interpolated, eventually fused and subsequently the anatomic entity/-ies of interest is/are three-dimensionally reconstructed. This reconstruction will provide the framework for the integration of the rest of data and the execution of the simulation. In parallel the molecular data is processed via molecular interaction networks so as to perturb and individualize the average pharmacodynamic or radiobiological cell survival parameters.
	Third step: Describe one or more candidate therapeutic scheme(s) and/or schedule(s). The clinician describes a number of candidate therapeutic schemes and/or schedules and/or no treatment (obviously leading to free i.e. non-inhibited tumour growth), to be simulated in silico i.e. on the computer.

	Fourth step: Run the simulation. The computer code of tumour growth and treatment response is massively executed on distributed grid or cluster computing resources so that several candidate treatment schemes and/or schedules are simulated for numerous combinations of possible tumour parameter values in parallel. Predictions concerning the toxicological compatibility of each candidate treatment scheme are also produced or alternatively estimates of the toxicologically acceptable dosage limits are retrieved from literature.
	Fifth step: Visualize the predictions. The expected reaction of the tumour as well as toxicologically relevant side effect estimates for all scenarios simulated are visualized using several techniques ranging from simple graph plotting to four multidimensional rendering.
	Sixth step: Evaluate the predictions and decide on the optimal scheme or schedule to be administered to the patient. The clinician carefully evaluates the Oncosimulator's predictions by making use of their logic, medical education and even qualitative experience. If no serious discrepancies are detected, the predictions support the clinician in taking their final and expectedly optimal decision regarding the actual treatment to be administered to the patient.
	Seventh step: Apply the theoretically optimal therapeutic scheme or schedule and further optimize the Oncosimulator. The expectedly optimal therapeutic scheme or schedule is administered to the patient. Subsequently, the predictions regarding the finally adopted and applied scheme or schedule are compared with the actual tumour course and a negative feedback signal is generated and used in order to optimize the Oncosimulator.
Expected usage frequency	Regularly for many patients entered the ALL clinical trials addressed by p-medicine provided that the necessary multiscale data is available.
Needed for DSS	⊙ yes O no
Needs HPC	⊙yes (when the resolution of the predictions has to be high) O no
Needs Grid	⊙yes (when several executions have to take place in order to offset the model parameter value expected deviations) O no
Priority for development	high
Responsible for development	ICCS-NTUA

Mockup needed	○ yes ◉ no
Responsible for Mockup	
Who is building the tool	ICCS-NTUA in collaboration with other WP12 participants
Open Source tool	○ yes ○ no, please specify why: ◉ Open access tool (to be also included in the European Cancer Model Repository developed by the TUMOR project)

Patient Empowerment Scenarios

Search for running clinical trials in Europe

Item	Description
Identifier	PG_1 (IEmS_1)
Version	1.0
Name	Pathway scenario for patient empowerment: Clinical trials search
Description of the use case (end-user perspective)	Users will be able to search a database of clinical trials to determine which are available and whether they are eligible
Problem(s) to solve	The ability to search available clinical trials databases
Challenges	To display information on eligibility with possible autocomplete from patient records, compatible with all clinical trials databases. Eligibility criteria can change from trial to trial
Risks	Clinical trial databases could restrict access or change format.
Expected benefits	Increase the number of patients enrolling in clinical trials through increased awareness of availability of trials
Characterization	○ fundamental ◉ general ○ specific
If specific, please give the Domain	○ Acute lymphoblastic leukaemia ○ Breast Cancer ○ Nephroblastoma ○ other Cancer, please specify: ○ Non-Cancer Domain, please specify:
End-user	○ system ○ person ○ basic scientist ◉ clinician ○ computer scientist ○ regulatory body, lawyer, ethicist ◉ patient ○ other, please specify:
Pre-condition(s)/pre-requisite(s)	Access available to clinical trial databases
Requisite(s)	Compatible with the Clinical Decision Support tools to ensure access for clinicians as well as patients. This will be regulated via the p-medicine portal.

Post-condition(s)/post-requisite(s)	When used in conjunction with the Clinical Decision Support tools patient information on the available trials must be available for the patient to access at a later date.	
Constraints		
External sources needed from outside p-medicine	⊙ data, please specify: Eudract clinical trials database as a minimum (possibly other global clinical trials databases e.g. clinical trials.gov, WHO trials registry etc.) O tools, please specify: O services, please specify: O models, please specify: O other, please specify:	
Data used	⊙ personal O only non-personal O target population, please specify:	
Input data	⊙ internal database, please specify: personal health record system O external database, please specify: O online input	
Output data	O database, please specify: O variables for use, please specify: ⊙ structured document, please specify: Should list all available clinical trials with their eligibility criteria and rank them according to the best analogy to the patients individual clinical data O graphic, please specify:	
Data volume	As needed	
Dataflow	Please specify: From the database(s) to the users records, data should be stored in the data warehouse	
Data storage	Please specify: Pseudonymized personal data should be stored in the data warehouse	
Successful End Condition	Correct information on trials displayed in an understandable way to assist in decision making	
Fail End Condition	No trial data available	
Basic workflow	Actor Action	System response
	Basic search parameters set from patient data (possibly automatically)	Searching trials that fulfil this criteria
		Available trials displayed along with additional useful information. This information is different between patients and physicians in the

		amount and detail of content, references, and the language used.
	Import function used	Details of the available clinical trials imported into the personal health record if existing
Expected usage frequency	Regularly	
Needed for DSS	⦿ yes O no	
Needs HPC	O yes ⦿ no	
Needs Grid	O yes ⦿ no	
Priority for development	Moderate	
Responsible for development	Will be decided by the IT group	
Mock-up needed	O yes ⦿ no	
Responsible for Mock-up		
Who is building the tool	Will be decided by the IT group	
Open Source tool	⦿ yes O no, please specify why	

Consent and Re-consent Scenario

Informed Consent (Patient's Perspective)

Item	Description
Identifier	PG_2 (IEmS_2)
Version	1.0
Name	Pathway scenario for patient empowerment: Informed consent
Description of the use case (end-user perspective)	Patients will be able to provide, withdraw and manage consent for clinical trials online.
Problem(s) to solve	Management of informed consent
Challenges	Communicating with the trial management system for different trials and providing the correct informed consent information for each trial will be a challenge
Risks	Patient data is handled insecurely or inaccurately or the wrong information is given at the point of consent
Expected benefits	Increased transparency for clinical trials leading to great trust, understanding and involvement
Characterization	○ fundamental ◉ general ○ specific
If specific, please give the Domain	○ Acute lymphoblastic leukaemia ○ Breast Cancer ○ Nephroblastoma ○ other Cancer, please specify: ○ Non-Cancer Domain, please specify:
End-user	○ system ○ person ○ basic scientist ○ clinician ○ computer scientist ○ regulatory body, lawyer, ethicist ◉ patient ○ other, please specify:
Pre-condition(s)/pre-requisite(s)	Clinical trial identified using the clinical trial search tool and the relevant informed consent information is provided
Requisite(s)	
Post-condition(s)/post-requisite(s)	Patient must be able to re-access and alter their information

Constraints	Informed consent information varies from trial to trial. The correct informed consent must be identified for each trial	
External sources needed from outside p-medicine	⦿ data, please specify: Trial management information O tools, please specify: O services, please specify: O models, please specify: O other, please specify:	
Data used	⦿ personal O only non-personal O target population, please specify:	
Input data	⦿ internal database, please specify: Personal health record O external database, please specify: O online input	
Output data	O database, please specify: O variables for use, please specify: ⦿ structured document, please specify: Status of informed consent O graphic, please specify:	
Data volume	Mild	
Dataflow	Please specify: The data flow needs to be specified during development.	
Data storage	Please specify: Data should be stored in the data warehouse	
Successful End Condition	Patients are able understand “informed consent” and manage their status	
Fail End Condition	No access to informed consent status	
Basic workflow	Actor Action	Actor Action
	Clinical trial is identified from a list	Information and questions relevant to this trial are displayed
	User moves through the information and the questions providing the answers	The system checks that the user is eligible for the trial and is providing the correct consent. Once the final pieces of information are gathered and the user has shown to understand what they are agreeing to an electronic signature is required
	Electronic signature provided	

	User login	Access to current status of consent which can then be modified
Expected usage frequency	Moderate	
Needed for DSS	○ yes ◉ no	
Needs HPC	○ yes ◉ no	
Needs Grid	○ yes ◉ no	
Priority for development	Moderate	
Responsible for development	Will be decided by the IT group	
Mock-up needed	○ yes ◉ no	
Responsible for Mock-up		
Who is building the tool	Will be decided by the IT group	
Open Source tool	◉ yes ○ no, please specify why	

Own Data Scenario

Item	Description
Identifier	PG_3 (IEmS_3)
Version	1.0
Name	Pathway scenario for patient empowerment: Own data
Description of the use case (end-user perspective)	Patients will be able to access the data stored on them with the data "translated" into a patient friendly format and language
Problem(s) to solve	Access to patient records within the p-medicine platform
Challenges	Displaying the information in a way that is suitable for all patients with differing levels of understanding and education, easy data transfer from existing patient records in p-medicine
Risks	Patient data stored in an insecure way, access granted to the wrong individual or hacked into
Expected benefits	Self-validation of data, greater transparency and patient empowerment
Characterization	○ fundamental ◉ general ○ specific
If specific, please give the Domain	○ Acute lymphoblastic leukaemia ○ Breast Cancer ○ Nephroblastoma ○ other Cancer, please specify: ○ Non-Cancer Domain, please specify:
End-user	○ system ○ person ○ basic scientist ◉ clinician ○ computer scientist ○ regulatory body, lawyer, ethicist ◉ patient ○ other, please specify:
Pre-condition(s)/pre-requisite(s)	Compatible with existing personal health record systems and existing patient records
Requisite(s)	
Post-condition(s)/post-requisite(s)	
Constraints	

External sources needed from outside p-medicine	○ data, please specify: ○ tools, please specify: ○ services, please specify: ○ models, please specify: ○ other, please specify:	
Data used	◉ personal ○ only non-personal ○ target population, please specify:	
Input data	○ internal database, please specify: ◉ external database, please specify: Import from existing patient records ○ online input	
Output data	○ database, please specify: ○ variables for use, please specify: ○ structured document, please specify: ○ graphic, please specify:	
Data volume	High	
Dataflow	Please specify: The data flow needs to be specified during development.	
Data storage	Please specify: In the data warehouse	
Successful End Condition	Patients have access to their own data online in an easily understandable format	
Fail End Condition	Patients don't have access to their data	
Basic workflow	Actor Action	Actor Action
	Patient login	Available health data is displayed in an understandable format. If further information or information validation is required, a message is displayed with this information
	Patient inputs data	Patient record updated
Expected usage frequency	High	
Needed for DSS	◉ yes ○ no	
Needs HPC	○ yes For IT team to decide ○ no	
Needs Grid	○ yes For IT team to decide ○ no	
Priority for development	High	

Responsible for development	For IT team to decide
Mock-up needed	⦿ yes O no
Responsible for Mock-up	For IT team to decide
Who is building the tool	For IT team to decide
Open Source tool	⦿ yes O no, please specify why

Access to Biobanks Scenario

Item	Description
Identifier	PG_4 (IEmS_4)
Version	1.0
Name	Pathway scenario for patient empowerment: Access to biobanks
Description of the use case (end-user perspective)	Patients will be able to access the biobank data stored on them with the data "translated" into a patient friendly format and language
Problem(s) to solve	Giving appropriate meaning to the biobank data for patients
Challenges	Displaying the information in a way that is suitable for all patients with differing levels of understanding and education. Access to each of the biobank repositories
Risks	Patient data stored in an insecure way, access granted to the wrong individual or hacked into
Expected benefits	Greater transparency and patient empowerment
Characterization	○ fundamental ◉ general ○ specific
If specific, please give the Domain	○ Acute lymphoblastic leukaemia ○ Breast Cancer ○ Nephroblastoma ○ other Cancer, please specify: ○ Non-Cancer Domain, please specify:
End-user	○ system ○ person ○ basic scientist ◉ clinician ○ computer scientist ○ regulatory body, lawyer, ethicist ◉ patient ○ other, please specify:
Pre-condition(s)/pre-requisite(s)	Availability of biobank data and anonymization of personal data
Requisite(s)	If used in clinical decision support
Post-condition(s)/post-requisite(s)	If used as part of the clinical decision support, needs to be delivered promptly
Constraints	

External sources needed from outside p-medicine	⦿ data, please specify: Biobank data needed O tools, please specify: O services, please specify: O models, please specify: O other, please specify:	
Data used	⦿ personal O only non-personal O target population, please specify:	
Input data	O internal database, please specify: O external database, please specify: O online input	
Output data	O database, please specify: O variables for use, please specify: O structured document, please specify: O graphic, please specify: To be determined in development depending on patient understanding	
Data volume	Moderate	
Dataflow	Please specify: To be determined in development	
Data storage	Please specify: Data could be transferred to the warehouse if appropriate	
Successful End Condition	Patients able to see biobank data in a meaningful way	
Fail End Condition	No biobank data available	
Basic workflow	Actor Action	Actor Action
	Patient logs in	User recognised and biobank data found and converted into meaningful information for the patient
	Import function	If desired biobank data imported into the personal health record
Expected usage frequency	Moderate	
Needed for DSS	⦿ yes O no	
Needs HPC	O yes For IT team to decide O no	
Needs Grid	O yes For IT team to decide O no	
Priority for development	Medium	
Responsible for development	For IT team to decide	

Mock-up needed	○ yes ◉ no
Responsible for Mock-up	
Who is building the tool	For IT team to decide
Open Source tool	◉ yes ○ no, please specify why

Summarize the history of the disease in an understandable way and increase patient-doctor understanding

Item	Description
Identifier	PG_5 (IEmS_5)
Version	1.0
Name	Pathway scenario for patient empowerment: Patient understanding
Description of the use case (end-user perspective)	A personalized medicine includes the analysis of the psychological and cognitive characteristics of each single patient. The analysis of the individual profile of the patient might serve to help physicians to evaluate how to inform the patients and which is the treatment that best fits with the personal profile of each patient. Such an approach will lead to an individualized treatment choice adjusted to the patient's needs. After a preliminary study (3 month – September/December 2011) in which ecancer will test the instruments (ipad or laptop-based questionnaires) to validate them and to verify their usability with patients, the IEmS tool will be developed. The tool will analyze the patient's answers in real time in order to provide an immediate visual feedback to the physician who will use this information to better understand the patient's needs and to propose him/her the treatment that best fits with the patient's profile. This first assessment will be followed by other periodical internet-based evaluations whose results will be accessible on-line both by physicians and patients. In this way physicians can monitor the psychological status of the patients as well as their perceived quality of life during the treatment, while patients can increase their level of empowerment having a more active role in the therapeutic process.
Problem(s) to solve	1. To help physicians to better understand the psychological and cognitive aspects of the patients so that they can find the best therapeutic approach giving them information and treatments personalized on their needs and values finding. 2. To increase the power of patients during the therapeutic process.
Challenges	1. To create a fast, easy-to-use tool to collect data from patients that can be easily interpreted by physicians. 2. To give patients the possibility to monitor their feelings and quality of life through the use of internet-based questionnaires.
Risks	Creating a personal psychological and cognitive profile through a relative small number of questions (no more

	than 50) can be very difficult. That's why it is necessary to conduct a preliminary study to test and validate the questionnaire.
Expected benefits	1. Obtaining a personal patient's profile will help physicians to better understand the patients and their needs. 2. Asking patients to answer the questionnaires will serve to increase their participation and their level of empowerment.
Characterization	○ fundamental ◉ general ○ specific
If specific, please give the Domain	○ Acute lymphoblastic leukaemia ○ Breast Cancer ○ Nephroblastoma ○ other Cancer, please specify: ○ Non-Cancer Domain, please specify:
End-user	○ system ○ person ◉ basic scientist ◉ clinician ○ computer scientist ○ regulatory body, lawyer, ethicist ◉ patient ○ other, please specify:
Pre-condition(s)/pre-requisite(s)	Availability of patients' to answer the questionnaires before the first clinical encounter and from home during the therapeutic process.
Requisite(s)	On time analysis and delivering of data obtained from the first administration of the questionnaire to the physicians.
Post-condition(s)/post-requisite(s)	Possibility for the physician and the patient to access data obtained with the internet-based questionnaires.
Constraints	If used as DSS the patient's personal profile need to be available on time. These logistics have to be solved otherwise (if data are coming late) the patient and the physician will not benefit from this use case as a DSS.
External sources needed from outside p-medicine	○ data, please specify: ○ tools, please specify: ○ services, please specify: ○ models, please specify: ○ other, please specify:
Data used	◉ personal

	○ only non-personal ○ target population, please specify:	
Input data	○ internal database, please specify: ○ external database, please specify: ◉ online input (To be defined with IT experts)	
Output data	◉ database, please specify: Excel or SPSS ○ variables for use, please specify: ○ structured document, please specify: ◉ graphic, please specify: Graphic should clearly represent the values obtained by the patient on each evaluated dimension and the range of the minimum and maximum values for that dimension.	
Data volume	Large, depending on the number of cases	
Dataflow	Please specify: The data flow needs to be specified during the development of the tool. Data should be stored in the data warehouse.	
Data storage	Please specify: Data will be stored in the data warehouse.	
Successful End Condition	1. Helping physicians to understand the personal characteristics of each patient in a very short amount of time. 2. Delivering personalized information and treatments that are compatible with the personal profile of the patient. 3. Increasing the patient's participation in the therapeutic process. 4. Eventually defining subgroups of patients with similar psychological and cognitive characteristics to identify sets of intervention strategies.	
Fail End Condition	No personal profiles identified.	
Basic workflow	Actor Action	System response
		The registration mask appears
	The patient registers him/herself in the system	
		The mask for the epidemiological variables appears
	The patient records his/her epidemiological variables	

		The questionnaire appears
	The patient answers the questions	The system elaborates the answers and produces the output graphs
		Data are stored in an online database
	The physicians selects the patient	A view of the graphs indicating the patient's profile will be given
Expected usage frequency	Several times per week	
Needed for DSS	⊙ yes O no	
Needs HPC	⊙ yes To be determined by the IT team O no	
Needs Grid	⊙ yes To be determined by the IT team O no	
Priority for development	The first part of the tool (iPad or laptop-based application) should be available for the beginning of 2012, at least in a beta version.	
Responsible for development	To be determined by the IT team	
Mock-up needed	O yes ⊙ no	
Responsible for Mock-up		
Who is building the tool	To be determined by the IT team	
Open Source tool	⊙ yes O no, please specify why The software can be open source, while the questionnaires, for scientific reasons, will be proprietary as the majority of the existing psychological validated tests.	

Biobanking Scenarios

Linking the own biomaterial data repository to the p-medicine biobank access framework for the collaboration with specific user groups

Item	Description
Identifier	BG_1
Version	0.7
Name	**Integration of biomaterial data repositories**
Description of the use case (end user perspective)	A user wants to link his own biomaterial data repository to the p-medicine biobank access framework in order to share data and material with his research community as further described in BA_3 to BA_4.
Problem(s) to solve	Biomaterial data repositories represent heterogeneous data sources and information systems to be integrated in a biobank access framework under a homogeneous search interface for biomaterial and related data. More concrete for p-medicine end users: Nephroblastoma: Excel tables and others Leukaemia: Scopeland LIMS and others Data harmonization: Agreeing on a specific biobank data set for a specific community or a specific project or on a larger general purpose biomaterial data set or biomaterial ontology. Differentiation between different research communities each with own biomaterial resources and respective access interface Legal implications on sharing biomaterial and related data within communities and over borders for research.
Challenges	a) Flexibility and general usability of the framework for the integration of heterogeneous biomaterial data sources beyond the end user scenarios for nephroblastoma and leukaemia. b) Dynamics in data definitions for a biomaterial data set. c) Provision of applicable legal guidelines/framework for biomaterial exchange across borders
Risks	a) Flexibility and general applicability of the framework for third parties. b) Collaboration of partners outside p-medicine is required.
Expected benefits	Use case is a prerequisite for the sharing of biomaterial and related data within a community.

Characterization	⊙ fundamental ⊙ general O specific
If specific, please give the Domain	O Acute lymphoblastic leukaemia O Breast Cancer O Nephroblastoma O other Cancer, please specify: O Non-Cancer Domain, please specify:
End user	⊙ system ⊙ person ⊙ basic scientist ⊙ clinician O computer scientist O regulatory body, lawyer, ethicist O patient O other, please specify:
Pre-condition(s)/pre-requisite(s)	Legal aspects must have been solved before data can be shared. Access to the biomaterial data sources is required. Descriptions of the data sets are required. The user's information system for biomaterial management must provide a respective (export) interface. Interface specification must be available or direct co-operation with the supplier of the information system may be required to develop a respective interface. The owner of the biomaterial data repository can decide which samples and which data will be made available for research via p-medicine. Informed consent from the patient does allow the use of data and material for the planned purpose.
Requisite(s)	The user has an account in p-medicine and is a member of a specific p-medicine user group
Post-condition(s)/post-requisite(s)	
Constraints	Integration of biomaterial data sources may require software adaptations and will most likely not be a plug&play like procedure
External sources needed from outside p-medicine	⊙ data, please specify: Sample biomaterial data from nephroblastoma and leukaemia use case owners O tools, please specify: O services, please specify: O models, please specify:

	⦿ other, please specify: Access to Scopeland LIMS, collaboration with Scopeland Active collaboration of Prof Gessler, Biozentrum Würzburg, as a later user of Biobank Access Framework	
Data used	○ personal ⦿ only non-personal ⦿ target population, please specify: SIOP Wilms tumor study group; ALL-leukaemia trials study groups	
Input data	○ internal database, please specify: ⦿ external database, please specify: Scopeland LIMS from Charité Berlin and University Hospital Schleswig-Holstein in Kiel Structured documents (Excel spreadsheets or CRFs) from Biozentrum Würzburg ○ online input	
Output data	⦿ database, please specify: Wrapper databases that harmonizes data sets (i.e. CRIP Inhouse Research Data Base (IRDB) or ObTiMA with biobank module) ○ variables for use, please specify: ○ structured document, please specify: ○ graphic, please specify:	
Data volume	Up to 100 data points per sample	
Dataflow	Data from biomaterial data source is filtered and sent to a wrapper database after any update of biomaterial data. Structured documents (csf, xls) are filtered and imported to a wrapper database after any update of biomaterial data.	
Data storage	Wrapper databases which can deal with a harmonized and pseudomized data set and that can eventually manage biomaterial.	
Successful End Condition	Biomaterial data source is constantly available in p-medicine biobank access interface	
Fail End Condition	Import of biomaterial data failed or is not possible	
Basic workflow	Actor Action	System response
	p-medicine administrator and biobank manager specify which data of the biobank repository will be integrated in p-medicine Biobank Access Framework, who will get access to it, and how this data is mapped to items of the standard biobank data set.	

	If necessary, the p-medicine administrator extends the standard biobank data set and annotates the extensions with HDOT. P-medicine WP10 partners implement the mapping and the interface to the local biomaterial data source. In the context of ObTiMA (use case BA_2) this means that the biomaterial data source is mapped to specific biomaterial related CRFs that represent the standard biobank data set.	
Expected usage frequency	Occasionally: Whenever a biomaterial resource shall be connected to p-medicine Biobank Access Framework	
Needed for DSS	○ yes ◉ no	
Needs HPC	○ yes ◉ no	
Needs Grid	○ yes ◉ no	
Priority for development	High; a pre-requisite for any other biobank access use case; due in Month 36	
Responsible for development	Fraunhofer IBMT	
Mockup needed	○ yes ◉ no	
Responsible for Mockup		
Who is building the tool	Fraunhofer IBMT	
Open Source tool	○ yes ◉ no, please specify why: Partly based on non-disclosed Fraunhofer ICT technology.	

Managing patient's biomaterial and related data in clinical trials with ObTiMA

Item	Description
Identifier	BG_2
Version	0.7
Name	**Managing biomaterial data in ObTiMA**
Description of the use case (end-user perspective)	A user collects biomaterial in a clinical trial, conducted with ObTiMA within p-medicine environment. The user wants to manage biomaterial and related data with ObTiMA that will enable him to link the biomaterial data directly to the clinical data of the patients and facilitates the sharing of the data and material within the trial community.
Problem(s) to solve	a) Providing pre-defined but adjustable case record forms for patient's biomaterial. The data items correspond to the standard biomaterial data set. b) Providing user interface functionality to get an overview about available biomaterial, quantity, etc. c) Providing basic functionality to administrate the use of biomaterial for research d) Providing a search interface that links clinical data and biomaterial data within a specific clinical trial e) Providing a search interface for biomaterial data for multiple clinical trials (cross-study-analyses) f) Integration in a general p-medicine biobank access framework
Challenges	Usability for biomaterial management Biomaterial management across trials requires semantic searches on the annotations. In addition roles and rights management may be challenging.
Risks	a) Efforts: Implementation within available resources in WP10
Expected benefits	a) Biomaterial data is linked to clinical data of patients b) Biomaterial data is annotated with H.dot c) Additional ObTiMA functionality is of general use in ObTiMA
Characterization	⊙ fundamental ⊙ general ⊙ specific
If specific, please give the Domain	○ Acute lymphoblastic leukaemia ○ Breast Cancer ⊙ Nephroblastoma

	O other Cancer, please specify: O Non-Cancer Domain, please specify:
End-user	O system ⊙ person ⊙ basic scientist ⊙ clinician O computer scientist O regulatory body, lawyer, ethicist O patient O other, please specify:
Pre-condition(s)/pre-requisite(s)	Legal aspects must have been solved before data can be shared. ObTiMA contains a repository with biobank specific CRFs and data items that are already annotated with a biobank ontology
Requisite(s)	The user, who is responsible for the biomaterial repository, has an account in p-medicine and is a member of a specific p-medicine user group that carries out a trial with ObTiMA. The trial chairman must plan the trial with biomaterial management.
Post-condition(s)/post-requisite(s)	
Constraints	Biomaterial management across trials requires a query interface based on the semantics. Roles and rights management may be challenging.
External sources needed from outside p-medicine	⊙ data, please specify: Sample biomaterial data from nephroblastoma; import of sample data and real data (SIOP trial) O tools, please specify: O services, please specify: O models, please specify: ⊙ other, please specify: Active collaboration of Prof Gessler, Biozentrum Würzburg, as a later user of Biobank Access Framework
Data used	O personal ⊙ only non-personal ⊙ target population, please specify: SIOP Wilms Tumor Study Group
Input data	⊙ internal database, please specify: ObTiMA database. A test trial with SIOP data and Prof. Gesslers tumor data would be required for validation of the system ⊙ external database, please specify: Structured documents (Excel spreadsheets or CRFs) from Biozentrum Würzburg and corresponding SIOP trial

	data from Norbert Graf. O online input	
Output data	⊙ database, please specify: ObTiMA trial database with integrated biomaterial data Information on available data and biomaterial O variables for use, please specify: O structured document, please specify: O graphic, please specify:	
Data volume	Up to 100 data points per sample	
Dataflow	Biomaterial data from patients is entered during the execution of the trial.	
Data storage	Biomaterial data is stored in ObTiMA as additional data of the patient	
Successful End Condition	Available biomaterial of a trial is listed in a search interface.	
Fail End Condition		
Basic workflow	Actor Action	System response
	Trial chairman creates biomaterial related CRFs during study design. For this purpose he re-uses pre-annotated biomaterial CRFs or data items from a CRF repository. Annotations and pre-defined data items represent the p-medicine standard biobank data set.	System creates CRF instances and presents them dynamically in user interface for remote data entry during trial execution.
	Clinical user enters data in biomaterial CRFs upon reception of biomaterial.	System verifies and confirms entry.
Expected usage frequency	Regularly, at least whenever biomaterial is acquired.	
Needed for DSS	O yes ⊙ no	
Needs HPC	O yes ⊙ no	
Needs Grid	O yes ⊙ no	
Priority for development	High; a user need for SIOP nephroblastoma trial; due in Month 36	
Responsible for development	Fraunhofer IBMT	

Mock-up needed	⊙ yes O no
Responsible for Mock-up	Fraunhofer IBMT
Who is building the tool	Fraunhofer IBMT
Open Source tool	⊙ yes ⊙ no, please specify why: Should follow the ObTiMA licence strategy

Offering human biomaterial to a closed and/or open clinical research community for research

Item	Description
Identifier	BG_3
Version	0.7
Name	Offering human biomaterial to a closed and/or open clinical research community for research
Description of the use case (end user perspective)	User within a research community wants to offer biomaterial for research. This use case is an extension of BA_4. The search engine includes an indicator whether and how much material is available for research and allows placing requests (use case BA_4).
Problem(s) to solve	a) A search interface on harmonized data sets or by using ontology-based annotations is required. b) A general harmonized data set versus research specific harmonized data sets c) Correct roles and rights management d) Anonymization of the data
Challenges	Roles and rights management may be challenging. Anonymization may be challenging
Risks	
Expected benefits	Biomaterial data from different sources is integrated under one search interface
Characterization	⦿ fundamental ⦿ general ⦿ specific
If specific, please give the Domain	⦿ Acute lymphoblastic leukaemia O Breast Cancer ⦿ Nephroblastoma O other Cancer, please specify: O Non-Cancer Domain, please specify:
End user	O system ⦿ person ⦿ basic scientist ⦿ clinician O computer scientist O regulatory body, lawyer, ethicist O patient O other, please specify:

Pre-condition(s)/pre-requisite(s)	Legal aspects must have been solved before data can be integrated Data sources are integrated in p-medicine biobank access framework as described in use case BA_1
Requisite(s)	The user has an account in p-medicine and is a member of a specific p-medicine user group that has access rights to biomaterial data.
Post-condition(s)/post-requisite(s)	
Constraints	
External sources needed from outside p-medicine	⦿ data, please specify: Sample biomaterial data from nephroblastoma; import of sample data and real data (SIOP trial). Sample biomaterial data from leukaemia biorepositories in Kiel and from their European partners. O tools, please specify: O services, please specify: O models, please specify: ⦿ other, please specify: Active collaboration with University hospitals Kiel, their European partners and LIMS providers. Active collaboration with Biozentrum Würzburg
Data used	O personal ⦿ only non-personal O target population, please specify: European leukaemia study groups SIOP Wilms tumor study group
Input data	⦿ internal database, please specify: ObTiMA database. A test trial with SIOP data and Prof. Gesslers tumor data would be required for validation of the system ⦿ external database, please specify: LIMS of the European partners and Scopeland O online input
Output data	⦿ database, please specify: Community specific search engine for biomaterial (CRIP) O variables for use, please specify: O structured document, please specify: O graphic, please specify:
Data volume	Up to 100 data points per sample
Dataflow	
Data storage	Anonymized and harmonized biomaterial data is stored in a central search repository.

Successful End Condition	Available biomaterial listed in a search interface and can be queried in combination with patient data. Or results of queries of patient data in another system can be further detailed with biomaterial data attributes.	
Fail End Condition	No biomaterial data available; Link between biomaterial data and clinical data is incorrect or not available.	
Basic workflow	Actor Action	System response
	User exports via wrapper service anonymized biomaterial related data to the central search engine	System removes all old entries from the providing institute and stores the anonymized data of the providing institute.
Expected usage frequency	Regularly	
Needed for DSS	○ yes ◉ no	
Needs HPC	○ yes ◉ no	
Needs Grid	○ yes ◉ no	
Priority for development	High; a user need for SIOP nephroblastoma trial and leukemia scenario; due in Month 36	
Responsible for development	Fraunhofer IBMT	
Mockup needed	◉ yes ○ no	
Responsible for Mockup	Fraunhofer IBMT	
Who is building the tool	Fraunhofer IBMT	
Open Source tool	○ yes ◉ no, please specify why: IRDB, CRIP are background of Fraunhofer	

Requesting specific human biomaterial within a closed and/or an open clinical research community for research purposes

Item	Description
Identifier	BG_4
Version	0.7
Name	Requesting specific human biomaterial within a closed and/or open clinical research community for research
Description of the use case (end-user perspective)	User within a research community needs specific biomaterial for research. This use case complements BA_3 from the perspective of the researcher, who wants to get biomaterial. It describes the request process. After selection of the required biomaterial according to use case BA_3 the user provides details about the planned research with the material. His request will then be forwarded by the system to the corresponding biomaterial owners. Legal aspects will be presented by the system (i.e. template of a material transfer agreement, privacy protection guidelines, responsibility to report about research outcome, etc). The biomaterial owners will then get in contact with the "customer" and agree on the details for the material provision.
Problem(s) to solve	a) A request process needs to be defined and implemented. b) Material deliveries need to be tracked and fed back to biomaterial data repositories.
Challenges	Provision of functionality with available resources
Risks	Low
Expected benefits	Biomaterial can be requested; research purposes can be traced?, biomaterial requests (and eventually also provisions) can be tracked. (s.o.)
Characterization	○ fundamental ◉ general ◉ specific
If specific, please give the Domain	◉ Acute lymphoblastic leukaemia ○ Breast Cancer ◉ Nephroblastoma ○ other Cancer, please specify: ○ Non-Cancer Domain, please specify:
End-user	○ system ◉ person ◉ basic scientist

	⦿ clinician O computer scientist O regulatory body, lawyer, ethicist O patient O other, please specify:
Pre-condition(s)/pre-requisite(s)	Legal aspects must have been solved before data can be integrated Biomaterial data sources are integrated in p-medicine biobank access framework as described in use case BA_1
Requisite(s)	The user has an account in p-medicine and is a member of a specific p-medicine user group that has access rights to biomaterial data.
Post-condition(s)/post-requisite(s)	The biomaterial owners accept the request.
Constraints	
External sources needed from outside p-medicine	⦿ data, please specify: Sample biomaterial data from nephroblastoma; import of sample data and real data (SIOP trial) Sample biomaterial data from leukaemia biorepositories in Kiel and from their European partners. O tools, please specify: O services, please specify: O models, please specify: ⦿ other, please specify: Active collaboration with University hospitals Kiel, their European partners and LIMS providers Active collaboration with Biocenter Würzburg
Data used	O personal ⦿ only non-personal ⦿ target population, please specify: European leukaemia study groups
Input data	O internal database, please specify: O external database, please specify: ⦿ online input A biomaterial search profile submitted through the biomaterial search engine together with a description of the research purpose
Output data	O database, please specify: O variables for use, please specify: ⦿ structured document, please specify: A list of pools with available biomaterial matching the search profile. Requests will be forwarded to the appropriate biomaterial owners for decision-making. O graphic, please specify:

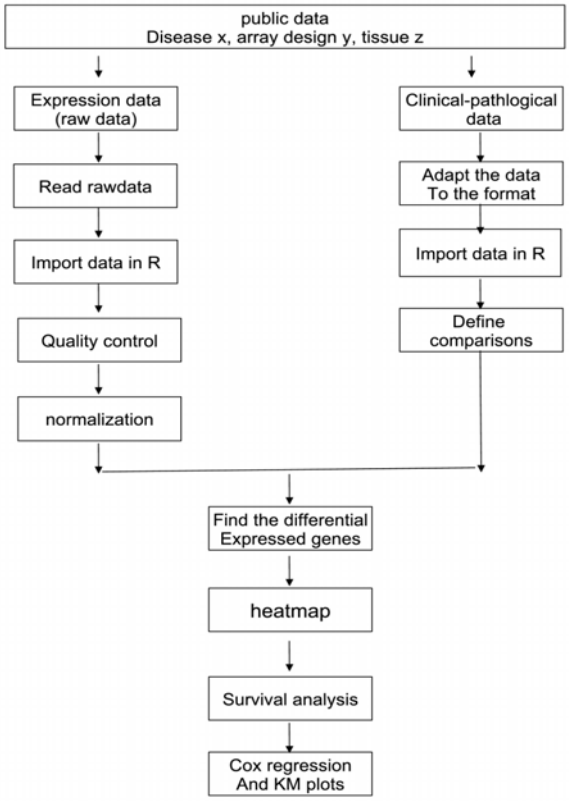
Data volume	low	
Dataflow	1) From the user to the system and back. 2) From the biomaterial search engine to the biomaterial data repositories and/or their owners	
Data storage	Requests are stored within biobank access framework	
Successful End Condition	Requests was successfully placed in biomaterial search engine	
Fail End Condition	No biomaterial available; User is not allowed to receive requested biomaterial due to access restrictions.	
Basic workflow	Actor Action	System response
	Researcher enters search criteria for biomaterial data.	System presents result list with means to any additional data and information about availability and quantity of the material for research.
	Researcher selects required biomaterial from search list as well as required quantity and submit this as an request	System stores biomaterial requests and presents a template for research purpose. Only registered users should have access! If user's identity is unknown system asks for details about his identity (address, affiliation, position, etc)
	User enters research purpose	System stores research purpose.
Expected usage frequency	Regularly	
Needed for DSS	○ yes ◉ no	
Needs HPC	○ yes ◉ no	
Needs Grid	○ yes ◉ no	
Priority for development	High; a user need for SIOP nephroblastoma trial and leukaemia scenario; due in Month 36	
Responsible for development	Fraunhofer IBMT	
Mock-up needed	◉ yes ○ no	
Responsible for Mock-up	Fraunhofer IBMT	
Who is building the tool	Fraunhofer IBMT	
Open Source tool	○ yes ◉ no, please specify why: IRDB, CRIP are background of Fraunhofer	

Clinical Trials Scenarios

Statistical Analysis of cancer samples with associated gene expression data and clinical data

Item	Description
Identifier	GEC_1
Version	0.1
Name	Statistical analysis of cancer samples with associated gene expression data and clinical features
Description of the use case (end-user perspective)	Uveal melanoma cancer samples: Affymetrix HG-U133 Plus 2 expression arrays have been extracted. The following clinical and personal features are available: tissue, age, gender, eye (right, left), tumor location, tumor diameter (mm), tumor thickness (mm), tumor cell type, retinal detachment, extrascleral extension, chromosome 3 status, months to endpoint, metastasis
Problem(s) to solve	To find genes differentially expressed between metastatic and non-metastatic tumors. To extract a prognostic gene signature
Challenges	the analysis per se doesn't represent a challenge, the real challenge is how to present the output to the several categories of end-users
Risks	implement and evaluate the scenario from less prospective than the ones the project needs to answer the P5 requirements
Expected benefits	support to clinical decisions; to a certain level, involves the patient in the decision process (i.e. the provided output can give the patient more knowledge about his disease)
Characterization	○ fundamental ◉ general - (the same steps can be applied to another type of cancer, by using others clinical variables) ◉ specific - (the platform is specific; this scenario is meant for microarray expression data)
If specific, please give the Domain	○ Acute lymphoblastic leukaemia ○ Breast Cancer ○ Nephroblastoma ○ Acute lymphoblastic leukaemia ◉ other Cancer, please specify: Uveal Melanoma primary tumours ○ Non-Cancer Domain, please specify:
End-user	◉ system

	⦿ person ⦿ basic scientist ⦿ clinician ⦿ computer scientist O regulatory body, lawyer, ethicist ⦿ patient O other, please specify:
Pre-condition(s)/pre-requisite(s)	1) R and Bioconductor packages installed; 2)clinical data need to be manually checked and re-labelled
Requisite(s)	
Post-condition(s)/post-requisite(s)	
Constraints	
External sources needed from outside p-medicine	⦿ data, please specify: GSE22138 (ncbi/GEO) ⦿ tools, please specify: preferably R/Bioconductor, but other similar script languages, statistically oriented, can be used. O services, please specify: O models, please specify: O other, please specify:
Data used	O personal ⦿ only non-personal ⦿ target population, please specify: http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE22138
Input data	O internal database, please specify: O external database, please specify: ⦿ online input: http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE22138
Output data	O database, please specify: O variables for use, please specify: O structured document, please specify: ⦿ graphic, please specify: box plots, intensity/density plot, volcano plot, K-M plots
Data volume	~300 Mbs (12 Gbs RAM to smoothly run the analysis)

<table>
<tr><td>Dataflow</td><td> 

Please specify: See flowchart</td></tr>
<tr><td>Data storage</td><td>Please specify: once the arrays have been red, the raw data can be stored “a part” and re-called in case they need to be re-normalized, merged with other data etc...</td></tr>
<tr><td>Successful End Condition</td><td>all plots have been created and no warning appears</td></tr>
<tr><td>Fail End Condition</td><td>the analysis stops with error messages</td></tr>
<tr><td>Basic workflow*</td><td>
<table>
<tr><td>Actor Action</td><td>System response</td></tr>
<tr><td>select expression data to read</td><td>data uploaded on the workspace and message “clinical data available, upload?”</td></tr>
<tr><td>normalize expression data</td><td>data normalized, new file name appears</td></tr>
<tr><td>click, find differentially expressed genes</td><td>create new file containing the genes differentially expressed, linked to ncbi/ensembl etc., and heatmap and a volcano plot to visualize the results</td></tr>
</table>
</td></tr>
<tr><td>Expected usage frequency</td><td>Moderate to high</td></tr>
</table>

Needed for DSS	○ yes ◉ no, I think it represents more a day-by-day resource for clinicians and patients mainly
Needs HPC	◉ yes, the RAM memory required on a workstation is at least 12 Gbs ○ no
Needs Grid	○ yes ◉ no
Priority for development	
Responsible for development	FhG-IAIS id developing a work flow starting from this scenario and the code that has been shared
Mock-up needed	◉ yes ○ no
Responsible for Mock-up	FhG-IAIS
Who is building the tool	FhG-IAIS in collaboration with SIB
Open Source tool	◉ yes ○ no, please specify why:

* **Description in more detail:**

A) Download of data in a tab delimited .txt file from to import as a table in R:

1) clinical data downloaded from: http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM550623 (just 1 sample); top of the file ftp://ftp.ncbi.nih.gov/pub/geo/DATA/SeriesMatrix/GSE22138/ to extract the complete table
2) raw expression data, it can be done directly in R, or by downloading the file GSE22138_RAW.rar from http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE22138
3) everything else is done in R/Bioconductor, please see the code that I shared with Dennis and Axel

B) Four steps for analysing Affymetrix expression arrays in BioConductor:

Step 1: loading the data
Step 2: import data, describing the experimental design
Step 3: RMA normalization and expression summary
Step 4: identifying differentially expressed probe sets

C) Data Import

Import data means that they are written in a tab delimited .txt file and from R the data are read, by reading them they are stored in a variable that can further be used for analysis. If the data are in an Excel format, then they are saved before as a tab delimited .txt file.

D) Data processing and format

- Raw data are from ncbi/geo, in the classical Affymetrix .cel format
- Data quality is checked by making plots and measures in R

- Expression data are normalized and extracted by reading the .cel raw data files
- The Affymetrix probe sets is based on the variance of the signal through the samples
- The omics data are analysed in relation to the clinical pathological variables, e.g. extraction of genes differentially expressed between stage I and stage II samples
- 3 variables are provided for stage: the size of the tumor, the number of nodes and the presence of metastases; starting from them, if not provided by the pathologist, stage can be deduced to only one variable: (http://en.wikipedia.org/wiki/TNM_staging_system)

Data management in international clinical trials by ECRIN

Item	Description
Identifier	PGE_1
Version	1
Name	Data management in international clinical trials by ECRIN
Description of the use case (end user perspective)	Data management in international clinical trials is especially challenging. During protocol implementation, data entry and trials conduct specific requirements regarding countries involved, user training and languages; as well differences in time zones must be considered. The CDMS (Clinical Data Management System) will be installed at an ECRIN data centre, guaranteeing that clinical trials can be performed according to GCP. User training and first level user support will be conducted by national ECRIN representatives using their native languages. The CDMS must not only support data entry including data checks during input, but also being able to support data querying, Adverse Events (AE) collection, simple data analysis and reporting (e.g. number of queries per site, increase in enrolled patients) to guarantee data quality and efficient data collection in clinical trials.
Problem(s) to solve	Training and support should be done in native languages. National specifics in data protection and the implementation of the GCP directive must be considered.
Challenges	Harmonisation of international data management processes, including difference in time zones during data entry and differences in the date of implementation of amendments
Risks	Inefficient training and data management, error prone software use, bad translation of user guides/SOPs
Expected benefits	High quality data and efficient conduct of international trials with large number of patients enrolled
Characterization	○ fundamental ● general ○ specific
If specific, please give the Domain	○ Acute lymphoblastic leukemia ○ Breast Cancer ○ Nephroblastoma ○ other Cancer, please specify: ○ Non-Cancer Domain, please specify:

Enduser	○ system ● person ● basic scientist ● clinician ○ computer scientist ● patient ● other, please specify: investigator, study nurse
Pre-condition(s)/pre-requisite(s)	Availability of a CDMS incorporated in an international support and clinical trials infrastructure, internet connection at study sites, study protocol
Requisite(s)	
Post-condition(s)/post-requisite(s)	Successfully conducted clinical trial
Constraints	
External sources needed from outside p-medicine	○ data, please specify: ○ tools, please specify: ● services, please specify: training and user support ○ models, please specify: ● other, please specify: internet connection
Data used	○ personal ● only non-personal ○ target population, please specify:
Input data	● internal database, please specify: Clinical trial database ○ external database, please specify: ● online input by eCRF
Output data	● database, please specify: Clinical trial database ○ variables for use, please specify: ● structured document, please specify: generated reports, e.g. patient recruitment rate per site, number of queries per investigator,… ● graphic, please specify: completed and signed CRFs, queries still unresolved, plan of visits
Data volume	

Dataflow	data collected with eCRF, data cleaning by querying process, all data is collected in a single clinical trial database	
Data storage	Please specify: clinical trial database	
Successful End Condition	Completed and locked clinical trials database, ready for archiving	
Fail End Condition		
Basic workflow	Actor Action	System response
	investigator	Log-in
	investigator	Displays visit
	system	Displays CRF
	investigator	Inputs data into CRF
	system	
	investigator	Signature of completed CRF
	Data manager	Initiates query
	investigator	Inputs data in query
	system	Displays completed CRFs
Expected usage frequency	High, because of use in different cancer and non-cancer studies in ECRIN	
Needed for DSS	O yes O no	
Needs HPC	O yes O no	
Needs Grid	O yes O no	
Priority for development		
Responsible for development		
Mockup needed	O yes O no	
Responsible for Mockup		
Who is building the tool		
Open Source tool	O yes O no, please specify why:	

Use of data mining to improve study feasibility

Item	Description
Identifier	PGE_2
Version	1
Name	Use of data mining to improve study feasibility
Description of the use case (end user perspective)	The data mining functionality of the p-medicine platform can be used to improved protocol feasibility for planned clinical trials. Data warehouses containing data from hospital information systems, registers, biobanks, study databases are part of the p-medicine platform and are searched to identify possible patient populations, number of eligible patients, efficiency of defined inclusion / exclusion criteria, availability of special surgical or therapeutic procedures, cancer treatment options, etc. In this way potential study populations, effects of changes in inclusion / exclusion criteria on recruitment, availability of medical treatments are determined and modelled. Results are used to improved study protocol and study planning.
Problem(s) to solve	Access to data from heterogeneous resources (different standards), establishment of an easy to use querying interface, enabling of searches in free text
Challenges	Effective way of broadening and narrowing search pattern
Risks	Insufficient data available, insufficient data quality
Expected benefits	Better study protocols can be created enabling more efficient trial conduct and more efficient patient recruitment
Characterization	○ fundamental ● general ○ specific
If specific, please give the Domain	○ Acute lymphoblastic leukemia ○ Breast Cancer ○ Nephroblastoma ○ other Cancer, please specify: ○ Non-Cancer Domain, please specify:
Enduser	○ system ● person ● basic scientist ● clinician

	● other, please specify: Trial feasibility expert
Pre-condition(s)/pre-requisite(s)	Anonymised, searchable datapool from heterogenous resources, including study databases, HIS, data warehouses, cancer registers
Requisite(s)	
Post-condition(s)/post-requisite(s)	Feasibility information is displayed in a useful way: for example, number of eligible patients for a specific combination of inclusion criteria, or information about the do-ability of a study (availability of special surgical procedures, novel cancer treatments, etc.)
Constraints	Data anonymisation
External sources needed from outside p-medicine	● data, please specify: data from registries, HIS, study databases combined in a data warehouse ● tools, please specify: data anonymisation and linking ● services, please specify: TTP services ○ models, please specify: ○ other, please specify:
Data used	○ personal ● only non-personal ○ target population, please specify:
Input data	● internal database, please specify: p-medicine data warehouse ● external database, please specify: registers, external data warehouses ● online input into query forms
Output data	● database, please specify: query database ○ variables for use, please specify: ● structured document, please specify: generated query results report: for example: eligible patient number per tumor form, hospital, country,… ● graphic, please specify: for example the display of maps; map of countries, regions with number of eligible patients (e.g. color coded)
Data volume	
Dataflow	Please specify:
Data storage	Please specify: data stored in data warehouse

Successful End Condition	Display of meaningful query results	
Fail End Condition	Displayed query results are not useful for feasibility	
Basic workflow	Actor Action	System response
	Researcher	Log-in into query system
	system	Displays query form
	Researcher	Input and selection of criteria for query
	Researcher	Selection of logical connection between criteria
	Researcher	Sending of query to data warehouse
	System	Search procedure in data warehouse
	System	Displays results in a way meaningful way
	Researcher	Formulates new fine-tuned query
	Researcher	Sending new query to data warehouse
	System	Search procedure in data warehouse
	System	Displays improved results
Expected usage frequency	High, usage for the improvement of many clinical trial protocols	
Needed for DSS	○ yes ○ no	
Needs HPC	○ yes ○ no	
Needs Grid	○ yes ○ no	
Priority for development		
Responsible for development		
Mockup needed	● yes, for query interface design ○ no	
Responsible for Mockup		
Who is building the tool		
Open Source tool	○ yes ○ no, please specify why:	

Improved Patient Recruitment in oncological clinical trials

Item	Description
Identifier	PGE_3
Version	1
Name	Improved patient recruitment in oncological clinical trials
Description of the use case (enduser perspective)	The p-medicine platform delivers a unique combination of data warehouse with data mining tool, biobank access, import of data from HIS, laboratories and clinical trials databases and an integrated patient empowerment tool. This novel combination of components can be used to improve patient recruitment in oncological clinical trials. Only a small number of patients suffering cancer have the possibility to profit from innovative therapies in clinical trials. Therefore, the improvement of patient recruitment is of special importance. The process covers the aspects of advertising the trial, identifying and contacting patients, pre-screening of patients, information of patient and informed consent, monitoring patient flow throughout the enrolment process. P-medicine's tools can be used to identify possible candidates and conduct some pre-screening to increase patient quality and help investigator sites. Because successful recruitment is determined by the patient's understanding and acceptance of the trial the Patient Empowerment Tool is used to enable information exchange with the patient.
Problem(s) to solve	Searching diverse data for patient identification.
Challenges	Data privacy and confidentiality. Integration of Patient Empowerment into clinical trial enrolment.
Risks	Inefficient patient identification, disturbing contribution of patient empowerment
Expected benefits	Enabling better patient recruitment; most patients that want to participate in a trial are considered
Characterization	○ fundamental ● general ○ specific
If specific, please give the Domain	○ Acute lymphoblastic leukemia ○ Breast Cancer ○ Nephroblastoma ○ other Cancer, please specify: ○ Non-Cancer Domain, please specify:
Enduser	○ system ● person

	● basic scientist ● clinician ○ computer scientist ○ regulatory body, lawyer, ethicist ● patient ○ other, please specify:
Pre-condition(s)/pre-requisite(s)	System enables the finding and identification of possible patients
Requisite(s)	
Post-condition(s)/post-requisite(s)	High number of patients are enrolled. Patients are properly informed about a trial
Constraints	Identification of patients can only be done by the treating physician
External sources needed from outside p-medicine	● data, please specify: HIS data, registry data, biobank data, patient empowerment tool data ● tools, please specify: pseudonymisation and re-identification tool ● services, please specify: TTP ○ models, please specify: ○ other, please specify:
Data used	● personal ○ only non-personal ○ target population, please specify:
Input data	● internal database, please specify: Clinical trial database, data warehouse, patient empowerment tool data ● external database, please specify: HIS and EHR data, data given by patients ○ online input
Output data	● database, please specify: patient contact information ○ variables for use, please specify: ○ structured document, please specify: ○ graphic, please specify:
Data volume	
Dataflow	Please specify: pseudonymous CIS data, Patient Empowerment Tool data and data warehouse data are analysed for the pre-screening and the identification of

	study participants; the treating physician recruits the patient if inclusion / exclusion criteria are met. For this the treating physician receives re-identified information only visible to him to contact the patient. In addition, information exchange between investigator and the patient enabled by the Patient Empowerment tool to inform and engage the patient.	
Data storage	Please specify:	
Successful End Condition	Patient recruited and patient is informed about the trial	
Fail End Condition	Patient is not recruited	
Basic workflow	Actor Action	System response
Expected usage frequency	High: used by investigators and patients. Recruitment is necessary for all trials	
Needed for DSS	○ yes ○ no	
Needs HPC	○ yes ● no	
Needs Grid	○ yes ● no	
Priority for development		
Responsible for development		
Mockup needed	○ yes ○ no	
Responsible for Mockup		
Who is building the tool		
Open Source tool	○ yes ○ no, please specify why:	

Use of p-medicine platform with decision support to conduct oncological clinical trials by ECRIN

Item	Description
Identifier	PGE_4
Version	1
Name	Use of p-medicine platform with decision support to conduct oncological clinical trials by ECRIN
Description of the use case (enduser perspective)	The p-medicine platform offers a set of tools like data warehouse, biobank access, decision support and a CDMS to increase the efficiency of oncological clinical trials and enable translational research. Used in ECRIN the p-medicine platform can be used in large international trials. Decision support can be evaluated as part of the intervention. Based on the prediction of the decision support tool, patients in clinical trials obtain different treatments. Training and support of the application of the decision support tool in the environment of an international clinical trial will be of special importance.
Problem(s) to solve	Integrative access and use of the decision support tool
Challenges	Integration of decision support in clinical trial process flow
Risks	Inefficient training and error prone use of the tool
Expected benefits	Evaluation of the usefulness of the results of the decision support. Application of a tool to enable translational medicine in international oncological clinical trials
Characterization	○ fundamental ● general ○ specific
If specific, please give the Domain	○ Acute lymphoblastic leukemia ○ Breast Cancer ○ Nephroblastoma ○ other Cancer, please specify: ○ Non-Cancer Domain, please specify:
Enduser	○ system ● person ○ basic scientist ● clinician ○ patient ● other, please specify: investigator

Pre-condition(s)/pre-requisite(s)	Integration of decision support tool into an international clinical trial infrastructure, access to the tool from different study sites	
Requisite(s)	Decision support tool gives correct decisions for treatment options to the investigator	
Post-condition(s)/post-requisite(s)		
Constraints		
External sources needed from outside p-medicine	O data, please specify: O tools, please specify: ● services, please specify: training and user support for investigators O models, please specify: O other, please specify:	
Data used	O personal ● only non-personal ● target population, please specify: Participants in clinical trials	
Input data	● internal database, please specify: Clinical trial database O external database, please specify:	
Output data	● database, please specify: Clinical trial database O variables for use, please specify: ● structured document, please specify: Report of results of decision support system O graphic, please specify:	
Data volume		
Dataflow	Please specify: enrolled patients are treated according to the decision support system or according to standard therapy. Investigator inputs patient data into decision support system and receives a treatment decision.	
Data storage	Please specify:	
Successful End Condition	Patient undergoes treatment according to the results of the decision support system	
Fail End Condition		
Basic workflow	Actor Action	System response
	investigator	Input of relevant patient data
	system	System develops a treatment decision

	system	System displays a treatment decision
	investigator	Evaluates decision
	investigator	Treatment of patient accordingly
Expected usage frequency	Middle, only in trials where decision support is possible and the outcome is evaluated	
Needed for DSS	○ yes ○ no	
Needs HPC	○ yes ○ no	
Needs Grid	○ yes ○ no	
Priority for development		
Responsible for development		
Mockup needed	○ yes ○ no	
Responsible for Mockup		
Who is building the tool		
Open Source tool	○ yes ○ no, please specify why:	

Increased Subject Retention rates in oncological clinical trials

Item	Description
Identifier	PGE_5
Version	1
Name	Increased Subject Retention rates in oncological clinical trials
Description of the use case (end user perspective)	The p-medicine platform delivers a unique combination of tools, covering data warehouse biobanking, HIS and Lab data import, with an integrated Patient Empowerment Tool. Therefore, p-medicine tools can be used to improve patient retention in oncological clinical trials. Especially in oncological trials, visit reminders, compliance reminders, assistance, self-monitoring and educational support can improve patient retention by for example offering the possibility to intercept potential drop-outs. The Patient Empowerment Tool is used to enable such information exchange with the patient during the clinical trial flow.
Problem(s) to solve	Searching diverse data in clinical study data base and Patient Empowerment Tool.
Challenges	Data privacy and confidentiality. Integration of Patient Empowerment into clinical trial conduct (e.g. visit schedule).
Risks	Patient is discouraged to us the Patient Empowerment Tool
Expected benefits	Better patient retention in clinical trial; the patient is engaged and satisfied with participating in the clinical trial
Characterization	○ fundamental ● general ○ specific
If specific, please give the Domain	○ Acute lymphoblastic leukemia ○ Breast Cancer ○ Nephroblastoma ○ other Cancer, please specify: ○ Non-Cancer Domain, please specify:
Enduser	○ system ● person ○ basic scientist ● clinician ○ computer scientist ○ regulatory body, lawyer, ethicist

	● patient ○ other, please specify:
Pre-condition(s)/pre-requisite(s)	System enables two-way communication with the patient
Requisite(s)	
Post-condition(s)/post-requisite(s)	
Constraints	Exchange of identifiable and personal information of the patient can only be done with the treating physician
External sources needed from outside p-medicine	● data, please specify: Clinical trials data, Patient Empowerment Tool data ● tools, please specify: pseudonymisation and re-identification tool ● services, please specify: TTP ○ models, please specify: ○ other, please specify:
Data used	● personal ○ only non-personal ○ target population, please specify:
Input data	● internal database, please specify: Clinical trial database, Patient Empowerment Tool data ● external database, please specify: EHR data ○ online input
Output data	● database, please specify: patient contact information, patient reported data, ○ variables for use, please specify: ○ structured document, please specify: ○ graphic, please specify:
Data volume	
Dataflow	Please specify: pseudonymous Patient Empowerment Tool data, clinical trials data and self-reported patient data are used for communication and information exchange with study participant. Data is analysed to identify potential drop-outs.
Data storage	Please specify:
Successful End Condition	Patient finishes clinical trial and is satisfied with participating in the trial
Fail End Condition	

Basic workflow	Actor Action	System response
Expected usage frequency	High: used by investigator and patient	
Needed for DSS	○ yes ○ no	
Needs HPC	○ yes ○ no	
Needs Grid	○ yes ○ no	
Priority for development		
Responsible for development		
Mockup needed	○ yes ○ no	
Responsible for Mockup		
Who is building the tool		
Open Source tool	● yes ○ no, please specify why:	

ObTiMA Scenarios

Pseudonymization Scenario

Item	Description
Identifier	SF_1
Version	1.0
Name	**Pseudonymization Scenario**
Description of the use case (end-user perspective)	This sis a fundamental use case dealing with the pseudonymization of personal data.
Problem(s) to solve	To use personal data is only possible after anonymization/pseudonymization.
Challenges	Developing of a tool, that can be used in a general way in all disease domains
Risks	The need for a Trust Centre and consecutive costs
Expected benefits	The use of personal data in a safe and secure way
Characterization	⦿ fundamental ⦿ general O specific
If specific, please give the Domain	⦿ Acute lymphoblastic leukaemia ⦿ Breast Cancer ⦿ Nephroblastoma ⦿ other Cancer, please specify: This is a fundamental tool and independent from the disease ⦿ Non-Cancer Domain, please specify: This is a fundamental tool and independent from the Cancer Domain
End-user	O system ⦿ person ⦿ basic scientist ⦿ clinician O computer scientist O regulatory body, lawyer, ethicist O patient O other, please specify:
Pre-condition(s)/pre-requisite(s)	Use of ObTiMA
Requisite(s)	
Post-condition(s)/post-requisite(s)	

Constraints	The need for Trust Centre	
External sources needed from outside p-medicine	○ data, please specify: ○ tools, please specify: ◉ services, please specify: pseudonymization service ○ models, please specify: ○ other, please specify:	
Data used	◉ personal ○ only non-personal ○ target population, please specify:	
Input data	◉ internal database, please specify: clinical data, research data, imaging data, etc. ○ external database, please specify: ○ online input	
Output data	○ database, please specify: ◉ variables for use, please specify: predefined personal data will be pseudonymized ○ structured document, please specify: ○ graphic, please specify:	
Data volume		
Dataflow	Please specify: personal data will be send to the trust centre, after pseudonymization the pseudonymized database can be stored in the data warehouse	
Data storage	Please specify: Only pseudonymized data will be stored in the data warehouse	
Successful End Condition	Personal data are pseudonymized	
Fail End Condition	Pseudonymization is not possible	
Basic workflow	Actor Action	System response
	Pseudonymization on the fly	
	e.g.: treating physician needs to send biomaterial to a laboratory, or to send DICOM files to reference radiology,	
	Physician selects patient in ObTiMA and selects form for sending biomaterial to a laboratory	The form is automatically filled with the personal data of the patient and with needed and existing data from ObTiMA
	Physician fills in the rest of needed information for the lab on the form	After all information is written in the form the form is send to the trust centre to pseudonymized the personal data. The form is send back

		with a barcode representing the personal data and an additional barcodes with which the biomaterial can be labelled.
	The physician attaches the label with the barcode on the biomaterial and sends it with the pseudonymized request from per mail to the laboratory.	
	People in the laboratory receive only pseudonymized requests with pseudonymized biomaterial. With the help of a barcode scanner the pseudonym can be store in the database of the laboratory. Analytical results of the biomaterial will be store in the database of the lab as pseudonymized data. The result will be send electronically from the laboratory via the trust centre to the hospital.	In the trust centre the pseudonymized data will be de-pseudonymized and send as personal data to the treating physician, so that he will be able to work only with personal data and no pseudonyms.
	Pseudonymization of databases for storage in the data warehouse	
	Physician or researcher selects the database and annotates all personal data fields.	
	Physician or researcher sends the selected and annotated database to the data warehouse	The selected database will automatically sent via the trust centre to the data warehouse. In the trust centre a service starts to pseudonymized all personal data in the database. After finishing the pseudonymization the database is send to the data warehouse for storage.
Expected usage frequency	Frequently, very high	
Needed for DSS	⊙ yes O no	

Needs HPC	○ yes ◉ no
Needs Grid	○ yes ◉ no
Priority for development	Very high
Responsible for development	Custodix
Mock-up needed	○ yes ◉ no
Responsible for Mock-up	
Who is building the tool	Custodix
Open Source tool	◉ yes ○ no, please specify why

Data Entry of Prospective Clinical Trial Data

Item	Description
Identifier	PG_6
Version	1.0
Name	Data Entry of Prospective Clinical Trial Data
Description of the use case (end-user perspective)	Clinical trials of investigational medicinal products must be carried out in accordance with ICH GCP and national legislation. Requirements for the data capture include quality control, quality assurance. The end user requires clear instructions and prompts, drop-down lists etc. to help with speed and accuracy of data input
Problem(s) to solve	To input manually, throughout the course of the clinical trial, data captured from review the subjects' medical records
Challenges	Unambiguous questions; ease of navigating through different screens; quality control to reduce error/omission/incompatibility in data submitted
Risks	Poor data will jeopardise the results produced in analysis.
Expected benefits	Well-designed electronic CRFs will ensure efficient data capture and help to produce complete and accurate data.
Characterization	○ fundamental ◉ general ○ specific
If specific, please give the Domain	○ Acute lymphoblastic leukaemia ○ Breast Cancer ○ Nephroblastoma ○ other Cancer, please specify: ○ Non-Cancer Domain, please specify:
End-user	○ system ◉ person ○ basic scientist ○ clinician ○ computer scientist ○ regulatory body, lawyer, ethicist ○ patient ◉ other, please specify: data entry clerk
Pre-condition(s)/pre-requisite(s)	Availability of trial-specific CRFs developed within the system
Requisite(s)	
Post-condition(s)/post-requisite(s)	(NA: Other users)

Constraints		
External sources needed from outside p-medicine	○ data, please specify: ○ tools, please specify: ○ services, please specify: ○ models, please specify: ○ other, please specify:	
Data used	○ personal ◉ only non-personal : pseudonymized ○ target population, please specify:	
Input data	◉ internal database, please specify: Trial-specific CRFs developed in ObTiMA Direct data entry is the rule of the user in this scenario ○ external database, please specify: ◉ online input is one of the features to be developed within ObTiMA	
Output data	○ database, please specify: ○ variables for use, please specify: ○ structured document, please specify: ◉ graphic, please specify: summary of data status for trial / subject under review, e.g. which items are complete, which are outstanding, whether there are data queries to be addressed	
Data volume	As per protocol, the total dataset will be input in relatively small amounts over a period, typically 1-3 years, maybe for as few as 50 patients in a study, possibly also thousands (depending on phase and protocol). Several clinical trials could be on-going simultaneously.	
Dataflow	Please specify: eCRFs developed and version-controlled in advance. Regular data input and both automated and personalised electronic queries.	
Data storage	Please specify: As per design of ObTiMA	
Successful End Condition	Set of protocol-specific clinical trial data, quality controlled for accuracy and completeness, which is ready for export in order to carry out statistical analysis according to the statistical analysis plan in the clinical trial protocol.	
Fail End Condition	Difficulty for data entry person to navigate between screens; slowness in system connection or response time.	
Basic workflow	Actor Action	System response
	Data entry person logs in to ObTiMA	A listing of user's clinical trials is presented

	User selects project	Summary of project is produced: whether there are new messages, how many subjects have had data started / completed
	User enters data items	Auto-checks of some data items (e.g. alerts if value is out of normal range)
	User responds to queries	Replies are retained within ObTiMA for review by data managers etc.
Expected usage frequency	At least one session per week per trial for the duration of the trial, daily for large trials, from many different remote investigating centres	
Needed for DSS	○ yes ◉ no not at this stage of the research	
Needs HPC	○ yes ◉ no	
Needs Grid	○ yes ◉ no	
Priority for development	Collaborative trials within p-medicine would use this software if available and validated at start of recruitment (otherwise alternative data capture systems would need to be implemented, ready as trial opens, and the data capture would remain in the other system throughout the trial)	
Responsible for development	ObTiMA developers	
Mock-up needed	◉ yes, needs user testing ○ no	
Responsible for Mock-up	Suggested collaboration between ObTiMA developers and clinical end-users (Marian Taylor is willing); to provide typical / retrospective set of data fields for building a mock trial	
Who is building the tool	ObTiMA developers	
Open Source tool	◉ yes ○ no, please specify why:	

Data Manager of Prospective Clinical Trials

Item	Description
Identifier	PG_7
Version	1.0
Name	Data Manager of Prospective Clinical Trial
Description of the use case (end-user perspective)	Clinical trials of investigational medicinal products must be carried out in accordance with ICH GCP and national legislation. Requirements for the data capture include quality control, quality assurance. This end user needs the facility to raise data clarification queries within the ObTiMA software, and allocate status to queries (e.g. close them when satisfied); the role plays an important part in demonstrable quality assurance.
Problem(s) to solve	The data manager will need to be able to manage user accounts for the trials allocated to them. The data manager needs be able to generate electronic clarifications requests, and allow the data entry person at the clinical site to provide a response. When satisfied, the data manager can close the query. The data manager will need to be assured that an audit trail is in place, to track changes as per GCP requirements. The data manager will need privilege to lock data once assured that it is clean. The data manager will to be able to export data in a format that is compatible with the statistical package used for analysis of data as per protocol
Challenges	
Risks	Poor data will jeopardise the results produced in analysis.
Expected benefits	Well-designed electronic CRFs will ensure efficient data capture and help to produce complete and accurate data.
Characterization	○ fundamental ◉ general ○ specific
If specific, please give the Domain	○ Acute lymphoblastic leukaemia ○ Breast Cancer ○ Nephroblastoma ○ other Cancer, please specify: ○ Non-Cancer Domain, please specify:
End-user	○ system ◉ person ○ basic scientist ○ clinician

	○ computer scientist ○ regulatory body, lawyer, ethicist ○ patient ◉ other, please specify: data manager
Pre-condition(s)/pre-requisite(s)	Availability of trial-specific CRFs developed within the system
Requisite(s)	
Post-condition(s)/post-requisite(s)	(NA: Other users)
Constraints	
External sources needed from outside p-medicine	○ data, please specify: ○ tools, please specify: ○ services, please specify: ○ models, please specify: ○ other, please specify:
Data used	○ personal ◉ only non-personal : pseudonymized ○ target population, please specify:
Input data	◉ internal database, please specify: Trial-specific CRFs in ObTiMA ○ external database, please specify: ◉ online input and exchange with users at clinical centres
Output data	◉ data output for analysis in a statistical package, in order to address research questions of trial protocol ◉ variables for use, please specify: simple data items for summary and analysis ◉ structured document, please specify: routine safety reporting for independent trials review board and/or research/ethics committees; end of study report, with many lists and tabulations ◉ graphic, please specify: summary of data status for trial / subject under review, e.g. which items are complete, which are outstanding, whether there are data queries to be addressed
Data volume	As per protocol, the total dataset will be input in relatively small amounts over a period, typically 1-3 years, maybe for as few as 50 patients in a study, possibly also thousands (depending on phase and protocol). Several clinical trials could be on-going simultaneously. The data manager will review data input for completeness, chronology and compatibility, in addition to any automated checked in the OpenClinica program (more programmed as automatic, less for the data manager, and vice versa)

Dataflow	Please specify: eCRFs developed and version-controlled in advance. Regular data input and both automated and personalised electronic queries.	
Data storage	Please specify: As per design of ObTiMA	
Successful End Condition	Set of protocol-specific clinical trial data, quality controlled for accuracy and completeness, which is ready for export in order to carry out statistical analysis according to the statistical analysis plan in the clinical trial protocol.	
Fail End Condition	Difficulty for data entry person to navigate between screens; slowness in system connection or response time.	
Basic workflow	Actor Action	System response
	Data manager logs in to ObTiMA	A listing of user's clinical trials is presented
	Data Manager selects project	Summary of project is produced: whether there are new messages, how many subjects have had data started / completed
	Data manager reviews response to data clarification requests,	Allows close of queries into resolved status (no longer on the to-do list)
	Data manager reviews new data	Allows for generation of queries for the attention of personnel at the investigating centres
Expected usage frequency	At least one session per week per trial for the duration of the trial, daily for large trials, from many different remote investigating centres	
Needed for DSS	○ yes ◉ no not at this stage of the research	
Needs HPC	◉ yes? Instance response required, ○ no	
Needs Grid	○ yes ○ no ◉ don't know technical requirements to achieve standards needed	
Priority for development	Collaborative trials within p-medicine would use this software if available and validated at start of recruitment (otherwise alternative data capture systems would need to be implemented, ready as trial opens, and the data capture would remain in the other system throughout the trial)	

Responsible for development	ObTiMA developers
Mock-up needed	⦿ yes, needs user testing O no
Responsible for Mock-up	Suggested collaboration between ObTiMA developers and clinical end-users (Marian Taylor UOXF is willing); to provide typical / retrospective set of data fields for building a mock trial
Who is building the tool	ObTiMA developers
Open Source tool	⦿ yes O no, please specify why:

eCRF Developer for Prospective Clinical Trials

Item	Description
Identifier*	PG_8
Version	1.0
Name	eCRF Developer for Prospective Clinical Trials
Description of the use case (end-user perspective)	Clinical trials of investigational medicinal products must be carried out in accordance with ICH GCP and national legislation. Requirements for the data capture include quality control, quality assurance. This user need to design electronic case report forms (eCRFs)
Problem(s) to solve	The suite of eCRFs for a given trial needs to capture all clinical data generated by participating on the trial. Re-use and customisation of previously used forms would increase efficiency
Challenges	Potentially: mapping of very specific data requirements for a given trial into data items already existing in the ObTiMA ontology pool. How quickly would a new ontology item be incorporated? A couple of months would be available to produce the full set of eCRFs in a timely manner.
Risks	Poor data capture tools will jeopardise data quality and therefore the results produced in analysis. ObTiMA software would need to be available for the duration of a trial employing it: possibly several years. ICH GCP requires long-term archiving of data (typically 15 years) so if this were not be possible, there would huge logistic and quality control issues for the trial coordinating centre
Expected benefits	Efficient development of eCRFs, improving as the pool of trials grows. Use of an electronic data management system is more efficient than paper then data input centrally. Computer system validation will be available, audit trail will be incorporated This type of clinical trials software has been largely unavailable to academic units as it is prohibitively expensive, particularly units undertaking many small (early phase) studies, who have struggled to meet the GCP standards
Characterization	○ fundamental ◉ general ○ specific
If specific, please give the Domain	◉ Acute lymphoblastic leukaemia

	⊙ Breast Cancer ⊙ Nephroblastoma ⊙ other Cancer, please specify: ⊙ Non-Cancer Domain, please specify:
End-user	O system ⊙ person O basic scientist O clinician ⊙ computer scientist/database developer (nearest role) O regulatory body, lawyer, ethicist O patient O other, please specify: data manager
Pre-condition(s)/pre-requisite(s)	ObTiMA system with user-friendly eCRF building, easy finding and mapping of data items in the ontology. Would there be a development and tester environment separate from the 'live' system, where eCRF designers can design/test/amend/re-test before producing a version for formal end-user testing?
Requisite(s)	Test environment as above
Post-condition(s)/post-requisite(s)	Long-term operation of the ObTiMA system, since some clinical trials will run for years
Constraints	
External sources needed from outside p-medicine	O data, please specify: ⊙ tools: public, widely used reference for coding, e.g. CTC, MedDRA, TNM staging O services, please specify: O models, please specify: O other, please specify:
Data used	O personal ⊙ only non-personal : pseudonymized O target population, please specify:
Input data	O internal database, please specify: O external database, please specify: ⊙ online input: The eCRF developer produced a specialised data capture tool specific to the individual trial BUT the input FOR THIS TASK is reference to the ethically approved clinical trial protocol
Output data	O database, please specify ⊙ variables for use, please specify: as specified in the trial protocol, for the trial participants ⊙ structured document, please specify: an entity diagram , a data dictionary, for the individual trial, as part of the data management documentation (along with evidence that the ObTiMA system has been validated)

Data volume	As per protocol, the total dataset will be input in relatively small amounts over a period, typically 1-3 years, maybe for as few as 50 patients in a study, possibly also thousands (depending on phase and protocol).	
Dataflow	Please specify: eCRFs developed and version-controlled in advance, then user tested (and documented) Some sort of sign-off of the eCRFs.	
Data storage	Please specify: not available for this particular role	
Successful End Condition	Set of protocol-specific clinical trial data capture forms	
Fail End Condition	Incomplete or ambiguous capture forms, trial research question cannot be adequately addressed	
Basic workflow	Actor Action	System response
	eCRF developer logs in	A list of items available is produced
	eCRF developer creates and tests (themselves), building up the eCRF bit by bit	Success or otherwise of the design will become apparent with self-testing during development
Expected usage frequency	Intensively but intermittently: Probably used full-time for a period of a few weeks at the setting up of each new trial.	
Needed for DSS	○ yes ◉ no not at this stage of the research	
Needs HPC	◉ yes? Would need to avoid time lags as they make development very slow, with all the trying, testing, amending, re-testing rounds. ○ no	
Needs Grid	○ yes ○ no ◉ don't know technical requirements to achieve standards needed	
Priority for development	Collaborative trials within p-medicine would use this software if available and validated at start of recruitment (otherwise alternative data capture systems would need to be implemented, ready as trial opens, and the data capture would remain in the other system throughout the trial)	
Responsible for development	ObTiMA developers	
Mock-up needed	◉ yes, needs user testing ○ no	
Responsible for Mock-up	Suggested collaboration between ObTiMA developers and clinical end-users (Marian Taylor UOXF is willing); to provide typical / retrospective set of data fields for building a mock trial	
Who is building the tool	ObTiMA developers	

Open Source tool	⊙ yes O no, please specify why:

Data Synchronization with HIS during running trial in ObTiMA

Item	Description
Identifier	PG_9
Version	0.5
Name	Data Synchronization with HIS during running trial in ObTiMA
Description of the use case (end-user perspective)	During a running trial, a clinician or data clerk can import data from a hospital information system (HIS) to fill patient CRFs in ObTiMA.
Problem(s) to solve	
Challenges	Identify data that can be accessed from HIS and can be reused in CTMS.
Risks	Data synchronization is uncomfortable, needs too much time, manual data entry would be more time efficient, incorrect data is transferred from HIS into ObTiMA,
Expected benefits	o In HIS systems only unstructured data is stored that cannot be reused to fill CRFs. o Data synchronization is uncomfortable, needs too much time, manual data entry would be more time efficient, incorrect data is transferred from HIS into ObTiMA.
Characterization	○ fundamental ◉ general ○ specific
If specific, please give the Domain	○ Acute lymphoblastic leukaemia ○ Breast Cancer ○ Nephroblastoma ○ other Cancer, please specify: ○ Non-Cancer Domain, please specify:
End-user	○ system ○ person ○ basic scientist ◉ clinician ○ computer scientist ○ regulatory body, lawyer, ethicist ○ patient ◉ other, please specify: data entry clerk
Pre-condition(s)/pre-requisite(s)	Sync services to retrieve data from data repositories in which current data of HISs is stored (e.g. communication server, data warehouse, which are in the following called HIS data repositories) have been set up and are running.

	ObTiMA has been configured in order that it is able to access the sync services.	
Requisite(s)		
Post-condition(s)/post-requisite(s)		
Constraints		
External sources needed from outside p-medicine	⊙ data, please specify: data from HIS data repositories O tools, please specify: O services, please specify: O models, please specify: O other, please specify:	
Data used	O personal ⊙ only non-personal, but for HIS data in data repositories same pseudonyms need to be provided as in ObTiMA in order that HIS data can be associated to the appropriate patient in ObTiMA O target population, please specify:	
Input data	O internal database, please specify: ⊙ external database, please specify: database of HIS / communication server of HIS O online input	
Output data	⊙ database, please specify: ObTiMA database O variables for use, please specify: O structured document, please specify: O graphic, please specify:	
Data volume		
Dataflow	Please specify: Data from HIS data repositories is transferred into the ObTiMA database.	
Data storage	Please specify: Transferred data is stored into the ObTiMA database.	
Successful End Condition	Transferred data is stored into the ObTiMA database	
Fail End Condition	Data cannot be transferred due to technical problems / user aborts data transfer.	
Basic workflow	Actor Action	System response
	User enrols or selects a patient.	Main patient view is shown.
	User requests to transfer data for the selected patient from HISs to CRFs.	The data that can be retrieved from the HIS data repositories for the patient is shown to the user for verifying. Suggestions are shown into

		which items on the CRFs the data may be filled in. Data that is already stored for the patient in ObTiMA is not shown.
	From the shown data, the user selects the data that he wants to store for the patient in ObTiMA and the according items into which the data should be filled in.	Data is stored in the ObTiMA database and the main patient view is shown.
Expected usage frequency	unknown	
Needed for DSS	○ yes ◉ no	
Needs HPC	○ yes ◉ no	
Needs Grid	○ yes ◉ no	
Priority for development	Needs to be developed until month 30 prototypically according to the DoW.	
Responsible for development	IBMT	
Mock-up needed	◉ yes ○ no	
Responsible for Mock-up	Software developers from IBMT together with clinicians from USAAR	
Who is building the tool	IBMT	
Open Source tool	○ yes ○ no, please specify why: ◉ not clear yet	

SAE/SUSAR Scenario

Item	Description
Identifier	PG_10
Version	1.0
Name	SAE/SUSAR Reporting
Description of the use case (end-user perspective)	The tool will be used for reporting of SAEs and SUSARs. This includes reporting of SAEs and SUSARs from a local hospital to the trial centre and the needed action regarding GCP criteria to be done in the trial centre.
Problem(s) to solve	Handling of SAEs and SUSARs in clinical trials
Challenges	To build a tool that can be used in every clinical trial
Risks	
Expected benefits	Faster and better reporting of SAEs and SUSARS according to GCP criteria
Characterization	○ fundamental ◉ general ○ specific
If specific, please give the Domain	◉ Acute lymphoblastic leukaemia ◉ Breast Cancer ◉ Nephroblastoma ◉ other Cancer, please specify: ◉ Non-Cancer Domain, please specify:
End-user	○ system ◉ person ○ basic scientist ◉ clinician ○ computer scientist ○ regulatory body, lawyer, ethicist ○ patient ◉ other, please specify: Data Manager
Pre-condition(s)/pre-requisite(s)	Access to: EudraVigilance Clinical Trial Module (EVCTM): http://eudravigilance.ema.europa.eu/human/docs/Directives/Dir2001-20_en.pdf EudraVigilance Post-Authorisation Module (EVPM): http://ec.europa.eu/health/documents/eudralex/vol-9/index_en.htm
Requisite(s)	License to MedDRA Database

Post-condition(s)/post-requisite(s)	
Constraints	
External sources needed from outside p-medicine	⊙ data, please specify: MedDRA O tools, please specify: ⊙ services, please specify: EudraVigilance Clinical Trial Module (EVCTM): http://eudravigilance.ema.europa.eu/human/docs/Directives/Dir2001-20_en.pdf EudraVigilance Post-Authorisation Module (EVPM): http://ec.europa.eu/health/documents/eudralex/vol-9/index_en.htm O models, please specify: O other, please specify:
Data used	⊙ personal O only non-personal O target population, please specify:
Input data	O internal database, please specify: ObTiMA database O external database, please specify: O online input
Output data	⊙ database, please specify: Database for SAEs and SUSARs at EMA, needs to be discussed with EMA how to get access. See also EVCTM and EVPM ⊙ variables for use, please specify: structured data are given below** O structured document, please specify: O graphic, please specify:
Data volume	Low
Dataflow	Please specify: SAEs/SUSARs are reported by the local hospital using ObTiMA. The SAEs/SUSARs needs to be checked by the PI of the trial and then send to EMA and all regulatory bodies
Data storage	Please specify: All SAEs/SUSARs need to be stored in ObTiMA, as well the decision that are taken from those SAEs/SUSARs
Successful End Condition	SAEs/SUSARs are reported at the EMA and regulatory bodies and all reports are send around according to GCP
Fail End Condition	No SAEs/SUSARs are reported, No routine reports are generated

Basic workflow	Actor Action	System response
	Local hospital	
	Physician needs to report a SAE/SUSAR	ObTiMA will open the CRF for SAE/SUSAR of the patient
	Physician fills in the data of the SAE/SUSAR	After finishing data input the data are stored in the CRF. And a notation is given to the PI via email and/or in ObTiMA
	PI of a trial/study centre	
	PI is informed about a new SAE/SUSAR via email or by opening ObTiMA a notation is found that a new SAE/SUSAR needs to be checked	
	Physician clicks on the notation in ObTiMA	The CRF of the SAE/SUSAR opens
	Physician checks the data in the CRF and validates the data and annotates the CRF as valid SAE/SUSAR or no SAE/SUSAR	In case of valid SAE/SUSAR the SAE/SUSAR is reported to EMA, regulatory bodies and local hospitals automatically. The reporting to EMA should be a direct import into their SAE database if possible/allowed. Information about the electronic exchange of SUSARs with EMA is given at: **Error! Hyperlink reference not valid.**. The reporting of the SAE/SUSAR is notified on the CRF of this specific SAE/SUSAR. After 4 weeks a request is send to the local hospital asking for update information on the status of the patient and the treatment. This is done with a specific CRF for this purpose.
	Local physician gives an update on the status of the patient after receiving an announcement via email or at the time he logs into ObTiMA again.	If the data are completed the SAE is closed, if there are still open data the process of asking for updating info at the local hospital is repeated as described above
		Further information is given at (technical documents): http://eudravigilance.ema.europ

		a.eu/human/EudraVigilanceRelatedDocs.asp and: http://www.ich.org/products/guidelines/efficacy/article/efficacy-guidelines.html
	Reporting of SAEs/SUSARs	
	At specific time-points reports of all SAEs/SUSARs need to be reported to regulatory bodies and ethical committees according to GCP criteria. The PI of a trial can do so by using the report facility of the SAE/SUSAR use case.	By clicking on the report button the system will ask the PI for the time period to report.
	The PI enters the time period	The system lists all SAEs in one list and all SUSARs in another list. These lists can be stored as Excel files, text files or XML files and as a PDF file. The PDF file will be send by email to the regulatory bodies and the ethical committees.
Expected usage frequency	Regularly, often	
Needed for DSS	○ yes ◉ no	
Needs HPC	○ yes ◉ no	
Needs Grid	○ yes ◉ no	
Priority for development	high	
Responsible for development	Needs to be decided by IT people	
Mock-up needed	◉ yes ○ no	
Responsible for Mock-up	Needs to be decided by IT people	
Who is building the tool	Needs to be decided by IT people	
Open Source tool	◉ yes ○ no, please specify why	

Drug interaction Scenario

Item	Description
Identifier	PG_11
Version	1.0
Name	Drug Interaction checker
Description of the use case (end-user perspective)	The tool should help to find dangerous interaction between two drugs that are prescribed to a patient. A physician should do this check always before subscribing drugs. If all the drugs a patient gets are stored in CRFs in ObTiMA then such a service can automatically check for interaction and send a warning to the treating physician, announcing that there is incompatibility between drugs. In addition this service names the drugs and gives information about what are the risk for the patient. This use case can be combined with the use case for the prediction of an SAE (see chapter 7.1.3, use case: PSN_3)
Problem(s) to solve	Interaction between drugs causing severe side effects in patients
Challenges	Integration into ObTiMA and the IEmS
Risks	None
Expected benefits	Improve safety of treatment of patients
Characterization	○ fundamental ◉ general ○ specific
If specific, please give the Domain	◉ Acute lymphoblastic leukaemia ◉ Breast Cancer ◉ Nephroblastoma ◉ other Cancer, please specify: every cancer ◉ Non-Cancer Domain, please specify: every disease
End-user	○ system ◉ person ○ basic scientist ◉ clinician ○ computer scientist ○ regulatory body, lawyer, ethicist ◉ patient ○ other, please specify:
Pre-condition(s)/pre-requisite(s)	Access to databases with the information about interactions between drugs and their incompatibilities.

Requisite(s)	Having access to external databases for drug-drug interactions	
Post-condition(s)/post-requisite(s)		
Constraints		
External sources needed from outside p-medicine	⊙ data, please specify: drug-drug interaction database, licensing policy, not open source. Alternative data mining tool searching for interaction between drugs O tools, please specify: ⊙ services, please specify: e.g. Medscape Interaction service: http://reference.medscape.com/drug-interactionchecker O models, please specify: O other, please specify:	
Data used	O personal ⊙ only non-personal O target population, please specify:	
Input data	⊙ internal database, please specify: Data from ObTiMA: name of drugs given together in one patient O external database, please specify: ⊙ online input: name of drugs, if used as a service outside of ObTiMA	
Output data	O database, please specify: O variables for use, please specify: ⊙ structured document, please specify: list of drugs and their interaction: none, or yes, if yes describing the interaction and the severity of the risk for the patient. O graphic, please specify:	
Data volume	low	
Dataflow	Please specify: depending from the way the tool will be developed; usage of existing databases or developing an own data mining tool	
Data storage	Please specify: in ObTiMA or the data warehouse	
Successful End Condition	Answer is given, if there is an interaction and the explanation of the interaction.	
Fail End Condition	No result is given	
Basic workflow	Actor Action	System response
	Within ObTiMA	
	A new drug given to a patient is entered in the	As soon as the data are entered into the CRF the tool

	treatment CRF of the patient.	will check, if there is an interaction between the given drugs to the patient. If there is no interaction nothing happens, meaning there is no incompatibility. If there is an interaction, a message will appear immediately telling there is an interaction between the following drugs. The severity of the incompatibility and the risk for the patient is given with an explanation of the interaction.
	The physician decides what to do.	
	Service outside of ObTiMA	
	Physicians and patients can use this service. As soon as the service opens the person will enter the drugs that will be checked. After entering the all drugs a check box will be pressed.	After pressing the check box the service will check for interactions and incompatibilities of the drugs. If there is an interaction, a message will appear immediately telling there is an interaction between the following drugs. The severity of the incompatibility and the risk for the patient is given with an explanation of the interaction.
Expected usage frequency	high	
Needed for DSS	⊙ yes, can be used ⊙ no, is not always the case	
Needs HPC	O yes ⊙ no	
Needs Grid	O yes ⊙ no	
Priority for development	Moderate	
Responsible for development	Needs to be decided by IT people	
Mock-up needed	O yes ⊙ no	
Responsible for Mock-up		
Who is building the tool	Needs to be decided by IT people	
Open Source tool	⊙ yes, if no licensing is needed O no, please specify why: if licensing is needed	

DICOM Scenario

Item	Description
Identifier	PG_12
Version	1.0
Name	DICOM Scenario
Description of the use case (end-user perspective)	This use-case describes how DICOM data can be send from a local hospital to the data warehouse after automatic pseudonymization of the data. In a second step it describes how DICOM data can be downloaded for reviewing or post-processing.
Problem(s) to solve	In time availability of DICOM data for reference radiology, reference surgeons and for post-processing
Challenges	On fly pseudonymization of the data
Risks	Failure of pseudonymization
Expected benefits	More reference radiology and better diagnosis. Enhancing of research with imaging studies
Characterization	○ fundamental ◉ general ○ specific
If specific, please give the Domain	◉ Acute lymphoblastic leukaemia ◉ Breast Cancer ◉ Nephroblastoma ○ other Cancer, please specify: later in the project ○ Non-Cancer Domain, please specify: later in the project
End-user	○ system ◉ person ◉ basic scientist ◉ clinician ◉ computer scientist ○ regulatory body, lawyer, ethicist ○ patient ○ other, please specify:
Pre-condition(s)/pre-requisite(s)	Contract between data provider and data user
Requisite(s)	
Post-condition(s)/post-requisite(s)	
Constraints	

External sources needed from outside p-medicine	⦿ data, please specify: local DICOM data O tools, please specify: ⦿ services, please specify: pseudonymization service O models, please specify: O other, please specify:	
Data used	⦿ personal O only non-personal O target population, please specify:	
Input data	O internal database, please specify: O external database, please specify: ⦿ online input: data are send from local hospitals directly to the data warehouse after pseudonymization	
Output data	⦿ database, please specify: DICOM data are stored in the data warehouse O variables for use, please specify: O structured document, please specify: ⦿ graphic, please specify: presentation of the data in a timeline (according to: http://www.simile-widgets.org/timeline/)	
Data volume	High	
Dataflow	Please specify: from local hospital to the data warehouse. On the way the DICOM files are automatically pseudonymized.	
Data storage	Please specify: in the data warehouse	
Successful End Condition	Successful storage of pseudonymized DICOM files in the data warehouse	
Fail End Condition	No storage of DICOM files in the data warehouse	
Basic workflow	Actor Action	System response
	Sending of DICOM files	
	Local physician wants to send DICOM files of a patient. He logs into ObTiMA and selects via the menu sending of DICOM files. He selects the DICOM files from his local computer or the PACS system of the hospital. An eCRF opens asking for the following information: 1. Ultrasound, x-ray, CT, MRI, PET, scintigraphy with further specification (e.g.: bone scan, etc.)	

	2. DICOM files at diagnosis, during follow-up (e.g.: after pre-operative chemotherapy) or at relapse 3. Report of the local radiologists to be uploaded after pseudonymization	
	Press button: send	The DICOM files are send to the data warehouse. On the fly a pseudonymization service will automatically pseudonymize the DICOM files. In the data warehouse the DICOM files are listed according to a timeline. This is graphically visualized according to: http://www.simile-widgets.org/timeline/
		The sender will be notified that the DICOM files are uploaded successfully. This is also stored on the eCRF. Further notifications are going to reference radiologists that new imaging studies are available.
	Usage of DICOM files: reference radiology	
	A reference radiologist having the role and right to use DICOM files from the data warehouse logs into ObTiMA, he selects a patient from the list of patients awaiting reference radiology. He can only see the pseudonym but not the patient's personal data. This list is automatically updated after new DICOM files are stored in the data warehouse and when the reference radiologist has done his report.	After selecting the patient the reference radiologist can view the DICOM files in his DICOM viewer to do his report. For that purpose a structured eCRF opens, where the reference radiologist can write his report. The files can also be opened in DrEye for pre-processing, etc. If he wants to download the files he needs to accept to destroy them after review.
	Usage of DICOM files: for consultation or second opinion	
	If DICOM files are needed for consultation or second	After annotation the consultants will be notified

	opinion the local physician needs to get informed consent by the patient/parents that the DICOM files can be used for that purpose. After getting this informed consent he will annotate the patient in ObTiMA that one or more specific person(s) are allowed to see the personal data of the patient as they are put in the role of a consultant. The annotation is specifically addressed with the names of these persons and can be restricted to a specific time frame. (see also the consultation use case)	that their opinion is asked regarding the imaging studies. This is also needed for surgical consultations etc.
	The consultant will find a list of patients where he is a consultant. He selects the specific patient in ObTiMA and clicks on DICOM files.	The consultant can view the DICOM files in his DICOM viewer to do his report. For that purpose a structured eCRF opens, where he can write his report in ObTiMA.
Expected usage frequency	high	
Needed for DSS	○ yes ◉ no	
Needs HPC	○ yes ◉ no	
Needs Grid	○ yes ◉ no	
Priority for development	high	
Responsible for development	USAAR (ObTiMA), UCL (data warehouse), Custodix (pseudonymization)	
Mock-up needed	○ yes ◉ no	
Responsible for Mock-up		
Who is building the tool	Needs to be decided by IT people	
Open Source tool	◉ yes ○ no, please specify why	

Consultation Scenario

Item	Description
Identifier	PG_13
Version	1.0
Name	Consultation Scenario
Description of the use case (end-user perspective)	Local physicians can ask for consultation of a patient treated within a clinical trial
Problem(s) to solve	Store consultations in a systematic way
Challenges	
Risks	
Expected benefits	Research of the stored data can be done to improve clinical trials
Characterization	○ fundamental ◉ general ○ specific
If specific, please give the Domain	◉ Acute lymphoblastic leukaemia ◉ Breast Cancer ◉ Nephroblastoma ○ other Cancer, please specify: alter in the project ○ Non-Cancer Domain, please specify: later in the project
End-user	○ system ◉ person ○ basic scientist ◉ clinician ○ computer scientist ○ regulatory body, lawyer, ethicist ○ patient ○ other, please specify:
Pre-condition(s)/pre-requisite(s)	
Requisite(s)	
Post-condition(s)/post-requisite(s)	
Constraints	
External sources needed from outside p-medicine	◉ data, please specify: data of a specific patient need to be available to the consultant ○ tools, please specify: ○ services, please specify:

	○ models, please specify: ○ other, please specify:	
Data used	◉ personal ○ only non-personal ○ target population, please specify:	
Input data	◉ internal database, please specify: ObTiMA, request is stored in an eCRF in ObTiMA ○ external database, please specify: ○ online input	
Output data	◉ database, please specify: ObTiMA, response is stored on an eCRF in ObTiMA ○ variables for use, please specify: ◉ structured document, please specify: a summary of the consultation will be automatically produced and sent by email to the person requesting the consultation ○ graphic, please specify:	
Data volume	low	
Dataflow	Please specify: from participating centre in a trial to the trial centre	
Data storage	Please specify: in OBTIMA	
Successful End Condition	Successful consultation is given and stored in ObTiMA	
Fail End Condition	No consultation is given or stored in ObTiMA	
Basic workflow	Actor Action	System response
	Local physician need to fill in an eCRF in ObTiMA asking for a consultation, giving needed information and asking the question to be answered by the consultant. On the CRF he has to annotate that he has the informed consent from the patient or the parents allowing him to send personal data to the consultant.	The consultant gets a notification via email and within ObTiMA that a new consultation is needed.
	The consultant logs into ObTiMA, where he/she immediately sees all requests for consultations listed according to date of request and patients. The consultant chooses a patient.	The CRF with the request is opened.
	The consultant can select all data including research	An structured eCRF is opened in ObTiMA, where

	data, DICOM data to see and analyse.	the consultant can enter his recommendations.
	The consultant finishes the input in the structured CRF	The name is deleted in the list of requested consultations. The recommendation is stored in the eCRF in ObTiMA and an email with the content of the eCRF is sent to the person requesting consultation.
Expected usage frequency	high	
Needed for DSS	⊙ yes, not always, but will be very helpful in complex situations ⊙ no, mostly	
Needs HPC	O yes ⊙ no	
Needs Grid	O yes ⊙ no	
Priority for development	high	
Responsible for development	USAAR	
Mock-up needed	O yes ⊙ no	
Responsible for Mock-up		
Who is building the tool	USAAR	
Open Source tool	⊙ yes O no, please specify why	

Trial Development Scenario

Item	Description
Identifier	PG_14
Version	0.5
Name	Trial Development Scenario
Description of the use case (end-user perspective)	Templates will guide the trial chairman or people responsible for writing a new trial through all needed tasks according to legal, ethical and GCP regulations. There are also templates available for writing a standardized trial protocol.
Problem(s) to solve	To reduce time efforts in creating a new trial.
Challenges	Integrate the scientific text of a trial protocol, documents from regulatory bodies, CRFs, consent forms, list of participants etc. in one document.
Risks	None
Expected benefits	Faster writing of a trial protocol
Characterization	○ fundamental ◉ general ○ specific
If specific, please give the Domain	◉ Acute lymphoblastic leukaemia ◉ Breast Cancer ◉ Nephroblastoma ◉ other Cancer, please specify: ◉ Non-Cancer Domain, please specify:
End-user	○ system ◉ person ○ basic scientist ◉ clinician ○ computer scientist ○ regulatory body, lawyer, ethicist ○ patient ○ other, please specify:
Pre-condition(s)/pre-requisite(s)	
Requisite(s)	
Post-condition(s)/post-requisite(s)	
Constraints	

External sources needed from outside p-medicine	○ data, please specify: ○ tools, please specify: ◉ services, please specify: to apply for the EUDRACT number ○ models, please specify: ◉ other, please specify: templates for different parts of the trial protocol, for consent, for ethical approval, for approval at regulatory bodies, etc.	
Data used	○ personal ○ only non-personal ○ target population, please specify:	
Input data	○ internal database, please specify: ○ external database, please specify: ○ online input	
Output data	○ database, please specify: ○ variables for use, please specify: ◉ structured document, please specify: Trial protocol ○ graphic, please specify:	
Data volume	Low	
Dataflow		
Data storage		
Successful End Condition	A complete trial protocol is available as a PDF document and stored in ObTiMA	
Fail End Condition	No trial protocol could be created.	
Basic workflow	Actor Action	System response
	User logs in ObTiMA and selects Request Trial, fills in requested data and push button send	User will be checked if he is a trial chairman. In case he is, the system will allow him to built a new trial and a new menu will appears on the left site of ObTiMA guiding him through all regulatory steps to be taken, including registering the trial and receiving an EUDRACT number, developing consent forms, developing eCRFs, developing the trial outline in a graphical way, using templates for writing the scientific background, primary and secondary objectives, inclusion and exclusion criteria, the treatment plan, etc.

	User is guided through the whole process and finishes.	A file will be created as a draft the trial protocol as a Word document by integrating all templates etc. The filled in templates will be stored in an ObTiMA database.
	Formatting and customizing the Word document by the user will finalize the document.	A PDF file is generated and the Word document and the PDF file are stored in the trial repository of ObTiMA.

Procedures to be done

1. List of procedures to be done

1.1. Registering of the trial and applying for an EUDRACT number
1.2. Sponsorship documents
1.3. Funding issues
1.4. Statistical procedures (calculating trial number, etc.)
1.5. Safety Desk and Pharmacovigilance
1.6. Pseudonymization and Trust Centre
1.6.1. Templates
1.6.2. Consent forms
1.6.3. Ethical approval
1.6.4. National Regulatory Bodies approval
1.6.5. Contract Forms (Sponsor, hospitals, etc)
1.6.6. Trial investigator approval
1.6.7. Trial investigator file
1.7. Creating eCRFs within ObTiMA
1.8. Creating the Trial Outline (graphical schema)
1.9. Selection of Participating Centres
1.10. Insurance issues
1.11. Selection of ObTiMA modules

Structure of the Trial Master Protocol

1. Introduction

1.1. Study Abstract
1.2. Primary Hypothesis
1.3. Purpose of the Study Protocol

2. Background

2.1. Prior Literature and Studies
2.2. Rationale for this Study

3. Study Objectives

	3.1. Primary Aim 3.2. Secondary Aim 3.3. Rationale for the Selection of Outcome Measures **4. Investigational Agent** 4.1. Preclinical Data 4.2. Clinical Data to Date 4.3. Dose Rationale and Risk/Benefits **5. Study Design** 5.1. Overview or Design Summary 5.2. Subject Selection and Withdrawal 5.2.1. Inclusion Criteria 5.2.2. Exclusion Criteria 5.2.3. Ethical Considerations 5.2.4. Subject Recruitment Plans and Consent Process 5.2.5. Randomization Method and Blinding 5.2.6. Risks and Benefits 5.2.7. Early Withdrawal of Subjects 5.2.8. When and How to Withdraw Subjects 5.2.9. Data Collection and Follow-up for Withdrawn Subjects 5.3. Study Drug 5.3.1. Description 5.3.2. Treatment Regimen 5.3.3. Method for Assigning Subjects to Treatment Groups 5.3.4. Preparation and Administration of Study Drug 5.3.5. Subject Compliance Monitoring 5.3.6. Prior and Concomitant Therapy 5.3.7. Packaging 5.3.8. Blinding of Study Drug 5.3.9. Receiving, Storage, Dispensing and Return **6. Study Procedures** 6.1. Screening for Eligibility 6.2. Schedule of Measurements 6.3. Visit 1 6.4. Visit 2 etc. 6.5. Safety and Adverse Events

	6.5.1. Safety and Compliance Monitoring 6.5.2. Medical Monitoring 6.5.2.1. Investigator only 6.5.2.2. expert to monitor 6.5.2.3. Institutional Data and Safety Monitoring Board 6.5.2.4. Data and Safety Monitoring Board 6.5.3. Definitions of Adverse Events 6.5.4. Classification of Events 6.5.4.1. Relationship 6.5.4.2. Severity 6.5.4.3. Expectedness 6.5.5. Data Collection Procedures for Adverse Events 6.5.6. Reporting Procedures 6.5.7. Adverse Event Reporting Period 6.5.8. Post-study Adverse Event 6.6. Study Outcome Measurements and Ascertainment **7. Statistical Plan** 7.1. Sample Size Determination and Power 7.2. Interim Monitoring and Early Stopping 7.3. Analysis Plan 7.4. Statistical Methods 7.5. Missing Outcome Data 7.6. Unblinding Procedures **8. Data Handling and Record Keeping** 8.1. Confidentiality and Security 8.2. Training 8.3. Case Report Forms and Source Documents 8.4. Records Retention 8.5. Performance Monitoring **9. Study Monitoring, Auditing, and Inspecting** 9.1. Study Monitoring Plan 9.2. Auditing and Inspecting **10. Study Administration** 10.1. Organization and Participating Centres 10.2. Funding Source and Conflicts of Interest 10.3. Committees 10.4. Subject Stipends or Payments 10.5. Study Timetable

	11. Publication Plan **12. Attachments** 12.1. Tables 12.2. Informed consent documents 12.3. Patient education brochures 12.4. Special procedures protocols 12.5. Questionnaires or surveys **13. References**
Expected usage frequency	frequently
Needed for DSS	○ yes ◉ no
Needs HPC	○ yes ◉ no
Needs Grid	○ yes ◉ no
Priority for development	high
Responsible for development	USAAR
Mock-up needed	◉ yes ○ no
Responsible for Mock-up	USAAR
Who is building the tool	USAAR
Open Source tool	◉ yes ○ no, please specify why

Trial Outline Builder Scenarios

Statistical Toolbox

Item	Description
Identifier	PG_15
Version	0.5
Name	Statistical toolbox
Description of the use case (end-user perspective)	This use case describes how clinical data from a clinical trial can be statistically analysed within ObTiMA
Problem(s) to solve	Statistical analysis of trial data
Challenges	To analyse trial data across different trials
Risks	
Expected benefits	A trial chairman is able to analyse data very fast and easy. This is important, when he needs to give consultations and needs statistical analyses of concrete questions.
Characterization	○ fundamental ◉ general ○ specific
If specific, please give the Domain	◉ Acute lymphoblastic leukaemia ◉ Breast Cancer ◉ Nephroblastoma ◉ other Cancer, please specify: ◉ Non-Cancer Domain, please specify: as it is a general tool it can be used independent of the trial
End-user	○ system ◉ person ◉ basic scientist ◉ clinician ○ computer scientist ○ regulatory body, lawyer, ethicist ○ patient ○ other, please specify:
Pre-condition(s)/pre-requisite(s)	Annotation of trial data with an ontology
Requisite(s)	
Post-condition(s)/post-requisite(s)	
Constraints	End- user must have the role and right to use the statistical toolbox

External sources needed from outside p-medicine	⊙ data, please specify: data of clinical trial(s) within ObTiMA that will be analyzed O tools, please specify: O services, please specify: O models, please specify: O other, please specify:	
Data used	O personal ⊙ only non-personal O target population, please specify:	
Input data	O internal database, please specify: ObTiMA O external database, please specify: O online input	
Output data	O database, please specify: O variables for use, please specify: ⊙ structured document, please specify: The document should be structured according to the questions to be answered giving the statistical results. ⊙ graphic, please specify: The structured report should contain basic graphics, like histograms, box-plots, life tables, regression curves, etc.	
Data volume	Depending on the amount of trial data	
Dataflow	Please specify: working with the data from the data warehouse	
Data storage	Please specify: data warehouse	
Successful End Condition	Statistical analysis is performed an a structured document available for print or download	
Fail End Condition	No statistical report can be generated	
Basic workflow	Actor Action	System response
	Search of trial(s) by the end-user within ObTiMA	ObTiMA lists all trials that the end-user has the right to analyse
	The end-user selects the trial(s) he wants to analyse	The trials are shown on in level 4 of the trial outline builder in a graphical mode (trial schema).
	End-user selects one or more trial elements and can select from a list of the corresponding CRFs all items he wants to analyse	All selected items will be displayed as parallel coordinates. A list of statistical tools is shown.

	The end-user can select from a list of statistical tools the tool he needs for analysis and selects the cohort of patients to analyse by narrowing parallel coordinates of those items/variables that describe the cohort	The analysis is immediately shown in a new window in a graphical way. A document is created describing the trial(s) analysed, the cohort of patients and the statistical method used, followed by the results including the graphical output.
	The end-user can select an other cohort of patients by using the parallel coordinates	Immediately a new structured output is given
	The end-user selects the structured output for downloading and printing	
Expected usage frequency	High	
Needed for DSS	○ yes ◉ no	
Needs HPC	○ yes ◉ no	
Needs Grid	○ yes ◉ no	
Priority for development	high	
Responsible for development	UHok	
Mock-up needed	○ yes ◉ no	
Responsible for Mock-up		
Who is building the tool	UHok	
Open Source tool	◉ yes ○ no, please specify why	

Gene expression parallel coordinates

Item	Description
Identifier	PG_16
Version	0.5
Name	Gene expression parallel coordinates
Description of the use case (end-user perspective)	This use case describes how clinical data from a clinical trial can be statistically analysed together with molecular data within ObTiMA
Problem(s) to solve	Statistical analysis of trial data together with molecular genetic data
Challenges	To upload gene expression data or any other molecular data as parallel coordinates
Risks	
Expected benefits	A trial chairman is able to analyse molecular genetic data together with clinical data very fast and easy.
Characterization	○ fundamental ◉ general ○ specific
If specific, please give the Domain	◉ Acute lymphoblastic leukaemia ◉ Breast Cancer ◉ Nephroblastoma ◉ other Cancer, please specify: ◉ Non-Cancer Domain, please specify: as it is a general tool it can be used independent of the trial
End-user	○ system ◉ person ◉ basic scientist ◉ clinician ○ computer scientist ○ regulatory body, lawyer, ethicist ○ patient ○ other, please specify:
Pre-condition(s)/pre-requisite(s)	Annotation of trial data
Requisite(s)	
Post-condition(s)/post-requisite(s)	
Constraints	The user needs to have the right to perform this analysis
External sources needed from outside p-medicine	○ data, please specify: ○ tools, please specify: ○ services, please specify:

	O models, please specify: O other, please specify:	
Data used	O personal ⊙ only non-personal O target population, please specify:	
Input data	O internal database, please specify: O external database, please specify: O online input	
Output data	O database, please specify: O variables for use, please specify: ⊙ structured document, please specify: The document should be structured according to the questions to be answered giving the statistical results. ⊙ graphic, please specify: The structured report should contain basic graphics, like histograms, box-plots, life tables, regression curves, etc.	
Data volume	Depending on the amount of trial data	
Dataflow	Please specify: working with the data from the data warehouse	
Data storage	Please specify: data warehouse	
Successful End Condition	Statistical analysis is performed an a structured document available for print or download	
Fail End Condition	No statistical report can be generated	
Basic workflow	Actor Action	System response
	Search of trial(s) by the end-user within ObTiMA	ObTiMA lists all trials that the end-user has the right to analyse
	The end-user selects the trial(s) he wants to analyse	The trials are shown on in level 4 of the trial outline builder in a graphical mode (trial schema).
	End-user selects one or more trial elements and can select from a list of the corresponding CRFs all items he wants to analyse	All selected items will be displayed as parallel coordinates. A list of available gene data / molecular data corresponding to the selected trial(s) is shown. A list of statistical tools is shown.
	End-user selects the genes or molecular data he wants to analyse	
	The end-user can select from a list of statistical tools	The analysis is immediately shown in a

	the tool he needs for analysis and selects the cohort of patients to analyse by narrowing parallel coordinates of those items/variables that describe the cohort	new window in a graphical way. A document is created describing the trial(s) analysed, the cohort of patients and the statistical method used, followed by the results including the graphical output.
	The end-user can select an other cohort of patients by using the parallel coordinates	Immediately a new structured output is given
	The end-user selects the structured output for downloading and printing	
Expected usage frequency	high	
Needed for DSS	○ yes ◉ no	
Needs HPC	○ yes ◉ no	
Needs Grid	○ yes ◉ no	
Priority for development	high	
Responsible for development	UHok	
Mock-up needed	○ yes ◉ no	
Responsible for Mock-up		
Who is building the tool	UHok	
Open Source tool	◉ yes ○ no, please specify why	

Participating Centres Scenario

Item	Description
Identifier	PG_17
Version	0.5
Name	Graphical View of participating centres
Description of the use case (end-user perspective)	A trial chairman needs to select participating centres and trial investigators from specific centres. Researchers to include research institutes can also use this tool.
Problem(s) to solve	To find participating centres for a trial
Challenges	
Risks	
Expected benefits	To get GCP conform centres enrolled in clinical trials
Characterization	○ fundamental ◉ general ○ specific
If specific, please give the Domain	◉ Acute lymphoblastic leukaemia ◉ Breast Cancer ◉ Nephroblastoma ◉ other Cancer, please specify: ◉ Non-Cancer Domain, please specify:
End-user	○ system ◉ person ◉ basic scientist ◉ clinician ○ computer scientist ○ regulatory body, lawyer, ethicist ○ patient ○ other, please specify:
Pre-condition(s)/pre-requisite(s)	
Requisite(s)	
Post-condition(s)/post-requisite(s)	
Constraints	
External sources needed from outside p-medicine	○ data, please specify: ○ tools, please specify: ○ services, please specify: ○ models, please specify: ◉ other, please specify: maps for the graphical view
Data used	◉ personal

	○ only non-personal ○ target population, please specify:	
Input data	○ internal database, please specify: ○ external database, please specify: ◉ online input: the specific data for the centres and for the trial investigator are given in eCRFs, they will be stored in ObTiMA for editing at any time	
Output data	◉ database, please specify: list of all centres containing all infos for the centre as well as for all ○ variables for use, please specify: ○ structured document, please specify: ◉ graphic, please specify: representation of the centres on a map. Information of Centres can be seen by clicking on the centre listing also all trial investigators of the centre	
Data volume	low	
Dataflow	Please specify:	
Data storage	Please specify:	
Successful End Condition	Database and map with the above mentioned information is established	
Fail End Condition	No database and no map is established	
Basic workflow	Actor Action	System response
	A new centre applies for participation	
	The user logs into ObTiMA. He selects from the menu: Applying for as a new centre.	A eCRF is shown
	The user fills in all data and uploads files if needed. The user selects on a map his city.	The CRF is stored and the centre, will be listed on a map graphically. CRF is shown for adding data Trial investigators
	User enters for each trial investigator the data and uploads needed files.	
	A new centre updates his data	
	User selects his centre from the map	The CRF is shown and data can be updated for the centre and for all trial investigators of the centre
	Trial chairman selects a centre	
	User logs in ObTiMA and selects the map representing known trial	The map is shown with all centres

	centres	
	User selects a centre	All data of the centre are shown to the user
	User can decide if he wants to have the centre as a participant of the trial. If he selects yes	An automatic email is sent to the centre asking for participation in the trial
	User receives feedback from the centre that they want to participate	Contracts between the centre and the study centre are automatically generated and send to the centre for signatures
	Signed contract comes back	The centre will be automatically annotated in the trial as a participating centre
	If the user selects no after having seen the data of the centre the user can select another centre	After selection of all centres user gets back to the menu.
Expected usage frequency	frequently	
Needed for DSS	○ yes ◉ no	
Needs HPC	○ yes ◉ no	
Needs Grid	○ yes ◉ no	
Priority for development	high	
Responsible for development	USAAR	
Mock-up needed	◉ yes ○ no	
Responsible for Mock-up	USAAR	
Who is building the tool	USAAR	
Open Source tool	◉ yes ○ no, please specify why	

Patient Access to his/her trial data and Diary Scenario

Item	Description
Identifier	PG_18
Version	1.0
Name	Patient access to his/her trial data and diary scenario
Description of the use case (end-user perspective)	If patients are enrolled in clinical trials, they are allowed to see there stored data and might be able to write data into a specific diary CRF. This will allow to check and validate data of patients as well as enhance data curation. The patient is not allowed to change data in the database, but he is allowed to comment to data. He can only write in the diary CRF.
Problem(s) to solve	To increase data validity and enhance curation of data
Challenges	To make data understandable for patients. This means that the database needs to be translated in a language a patient will understand.
Risks	
Expected benefits	Better validated and curated data within clinical trials. This transparency will increase patient empowerment
Characterization	○ fundamental ◉ general ○ specific
If specific, please give the Domain	◉ Acute lymphoblastic leukaemia ◉ Breast Cancer ◉ Nephroblastoma ◉ other Cancer, please specify: ◉ Non-Cancer Domain, please specify: as this is a general tool, it can be applied for any kind of trial within ObTiMA
End-user	○ system ◉ person ○ basic scientist ○ clinician ○ computer scientist ○ regulatory body, lawyer, ethicist ◉ patient ○ other, please specify:
Pre-condition(s)/pre-requisite(s)	Patient needs to get the right to use this tools. A contract needs to be signed between the patient and the PI of the trial.
Requisite(s)	

Post-condition(s)/post-requisite(s)		
Constraints		
External sources needed from outside p-medicine	○ data, please specify: ○ tools, please specify: ○ services, please specify: ○ models, please specify: ○ other, please specify:	
Data used	◉ personal ○ only non-personal ○ target population, please specify:	
Input data	◉ internal database, please specify: ObTiMA ○ external database, please specify: ○ online input	
Output data	○ database, please specify: ○ variables for use, please specify: ○ structured document, please specify: ○ graphic, please specify:	
Data volume		
Dataflow	Please specify: only within ObTiMA, no output data will be generated	
Data storage	Please specify: in ObTiMA	
Successful End Condition	Patient is able to use the tool	
Fail End Condition	The tool will not be used by patients	
.Basic workflow	Actor Action	System response
	Patient logs into ObTiMA with his/her credentials	ObTiMA recognized the patient and displays all data of the patient in an understandable language for the patient and a CRF for the patient diary opens if such a CRF was created within the trial
	The patient can comment on every single item of his data	
	The patient can use the CRF diary and enters new data into the CRF	
	The patient logs out of ObTiMA	
Expected usage frequency	high	

Needed for DSS	○ yes ◉ no
Needs HPC	○ yes ◉ no
Needs Grid	○ yes ◉ no
Priority for development	high
Responsible for development	USAAR
Mock-up needed	○ yes ◉ no
Responsible for Mock-up	
Who is building the tool	USAAR
Open Source tool	◉ yes ○ no, please specify why

Repository Scenario

Item	Description
Identifier	REP_1
Version	1.0
Name	Repository Scenario
Description of the use case (enduser perspective)	The end user can store both parts as well as the entirety of CRFs into a (centralized) repository. This end user or others can subsequently retrieve, (re)assemble and reuse those full or partials CRFs in other new trials or studies.
Problem(s) to solve	Design and develop the actual repository together with the necessary interfaces to the core ObTiMA. Create an intuitive user interface both for selecting partial and full CRFs when storing them into the repository (best with some user-specifiable meta-data or tags). Such an interface is also needed for an easy-to-perform search to find full or partial CRFs fitting the specified end-user criteria within the repository.
Challenges	The search for fitting CRFs or parts of them is not trivial at all. It has to be experimented what criteria can best be used to make such a search both quick and easy to use and how this can be implemented in a suitable user interface. Otherwise this feature will not be used.
Risks	If the search interface is not efficient then this feature might not be used at all but users will recreate their CRFs scratch anyway. Also, it is not clear how often parts of CRFs are actually recurring and can therefore be reused and thus
Expected benefits	Large time savings for the end user can be the result because CRFs can be composed of recurring, existing pieces (or even be reused in their entirety) instead of tediously recreating them from scratch again and again for each study or trial.
Characterization	○ fundamental ◉ general ○ specific
If specific, please give the Domain	◉ Acute lymphoblastic leukaemia ◉ Breast Cancer ◉ Nephroblastoma ◉ other Cancer, please specify: ◉ Non-Cancer Domain, please specify: as this is a general tool, it can be applied for any kind of trial within ObTiMA

Enduser	○ system ◉ person ◉ basic scientist ◉ clinician ◉ computer scientist ○ regulatory body, lawyer, ethicist ○ patient ○ other, please specify:
Pre-condition(s)/pre-requisite(s)	
Requisite(s)	
Post-condition(s)/post-requisite(s)	
Constraints	
External sources needed from outside p-medicine	○ data, please specify: ○ tools, please specify: ○ services, please specify: ○ models, please specify: ○ other, please specify:
Data used	◉ personal ◉ only non-personal ○ target population, please specify:
Input data	◉ internal database, please specify: CRFs and their parts created by the current user who wants to put them into the repository ○ external database, please specify: ○ online input
Output data	○ database, please specify: Repository holding the CRFs and their parts ○ variables for use, please specify: ○ structured document, please specify: ○ graphic, please specify:
Data volume	
Dataflow	CRFs in their entirety and/or parts thereof are sent to the repository and later retrieved from it using a search mechanism (probably added with some additional metadata to improve retrieval)
Data storage	CRFs in their entirety and/or parts
Successful End Condition	1) Storage of the CRFs and their parts was successful 2) Search for and retrieval of the appropriate CRFs or part was successful
Fail End Condition	1) Storage was not working because of technical

<table>
<tr><td></td><td colspan="2">difficulties
2) Search and retrieval was not working because of…
a. technical difficulties
b. the partial or entire CRF fitting the needs of the end user was not in the repository
c. an entire or partial CRF fitting the end user's needs was in the repository but could not be found (e.g. invalid search criteria, inappropriate metadata)</td></tr>
<tr><td rowspan="4">Basic workflow</td><td>Actor Action</td><td>System response</td></tr>
<tr><td>The end user selects a one or more CRFs or one ore more parts within a CRF in ObTiMA's section for developing CRFs.
Then he/she can add some metadata in a special dialogue window (creator and creation date of the CRF are added automatically).
Then he/she presses a button to submit the storage.</td><td>The system acknowledges the storage of the data within the repository (and perhaps shows again a window with the names of the CRFs or their parts stored and also the metadata submitted).</td></tr>
<tr><td>1) The end user opens a special search window where he/she can specify some metadata of the CRFs he/she is looking for (e.g. all CRFs edited by a specific author).
He/she can also specify some keywords that should be contained in the content of the stored CRFs or their parts.
Then he/she pressed a button to submit the search.</td><td>The system returns a list of CRFs or parts thereof that fit the search criteria.</td></tr>
<tr><td>3) From the returned list, the user selects…
a) one or more entire CRFs and submits
b) one or more parts of CRFs</td><td>For a) the entire CRFs are added to the list of CRFs for the current trial
For b) the selected parts are added to the currently opened/edited CRF</td></tr>
<tr><td>Expected usage frequency</td><td colspan="2">medium</td></tr>
<tr><td>Needed for DSS</td><td colspan="2">O yes
⊙ no</td></tr>
<tr><td>Needs HPC</td><td colspan="2">O yes</td></tr>
</table>

	⦿ no
Needs Grid	O yes ⦿ no
Priority for development	high
Responsible for development	USAAR, FhG-IBMT
Mockup needed	⦿ yes O no
Responsible for Mockup	USAAR
Who is building the tool	USAAR, FhG-IBMT
Open Source tool	⦿ yes O no, please specify why:

Semantic interoperability Scenario

Item	Description
Identifier	SEM_1
Version	1.0
Name	Semantic Interoperability Scenario
Description of the use case (enduser perspective)	Data from both external as well as internal data sources should be integrated and used along with the data collected using the CRFs within ObTiMA.
Problem(s) to solve	How can external data be accessed? (e.g. What technologies are needed to connect to external databases over the Internet? Are there licensing issues?) How can external data sources be enriched semantically in order for the semantic interoperability to happen? …and who does that? How should a user interface look like to include (and map) external data sources and how can items stemming from them (visually) defined on CRFs?
Challenges	Persuade database curators to enrich their databases semantically with the ontologies as used within p-medicine and provide open and standardized interfaces to make them accessible.
Risks	Unavailability of enough data sources that fulfil the requirements mentioned above.
Expected benefits	Data from various internal and external sources can be used within ObTiMA transparently along with the data collected using the CRFs.
Characterization	○ fundamental ⊙ general ○ specific
If specific, please give the Domain	⊙ Acute lymphoblastic leukaemia ⊙ Breast Cancer ⊙ Nephroblastoma ⊙ other Cancer, please specify: ⊙ Non-Cancer Domain, please specify: as this is a general tool, it can be applied for any kind of trial within ObTiMA
Enduser	○ system ⊙ person ⊙ basic scientist ○ clinician ⊙ computer scientist

	○ regulatory body, lawyer, ethicist ○ patient ○ other, please specify:
Pre-condition(s)/pre-requisite(s)	Data sources which are semantically enriched with annotations from ontologies used in the p-medicine environment (as proposed by the WP4).
Requisite(s)	
Post-condition(s)/post-requisite(s)	
Constraints	
External sources needed from outside p-medicine	◉ data, please specify: Depends heavily on the specific trial or study at hand which (type of) data sources should (and can) be integrated into that trial or study ○ tools, please specify: ○ services, please specify: ○ models, please specify: ○ other, please specify:
Data used	◉ personal ◉ only non-personal ○ target population, please specify:
Input data	◉ internal database, please specify: Various types of data that have been collected or generated by or incorporated into p-medicine's own tools and which are to be used as complement to the data collected within ObTiMA with the CRFs ◉ external database, please specify: Various data that contain relevant data (e.g. KEGG) which are to be used as complement to the data collected within ObTiMA with the CRFs ◉ online input (see "external database" above – those databases are probably accessed via the internet mostly)
Output data	○ database, please specify: ○ variables for use, please specify: ○ structured document, please specify: ○ graphic, please specify:
Data volume	medium to high
Dataflow	External and internal databases have to offer standardized interfaces that can be queried from ObTiMA with semantically enriched queries based on the ontologies used within p-medicine (as proposed by the WP4). The result set of the query is subsequently sent back to ObTiMA where it is integrated on-the-fly with the data collected with the CRFs.

Data storage	No additional data storage because the original data is to be kept at its place of origin and only "virtually" integrated on-the-fly.	
Successful End Condition	Successful integration of other data (sources) with the data collected via the CRFs in ObTiMA.	
Fail End Condition	External data cannot be "made compatible" (i.e. sensibly annotated) following structure and content of the data internally collected via the CRFs.	
Basic workflow	Actor Action	System response
	User selects the external data source to connect to, i.e. he/she specifies the needed connection parameters like URL, username, password, etc. Additionally, it would be nice to offer a list of default data sources and of "last visited" data sources.	The system shows an identifier of the selected data source to use in ObTiMA.
	In the editor for CRFs in development, the editor can create links from CRF questions/ items to fields in the data source. For this, links to the available data sources are shown. When clicked upon, then a window pops up where the user can specify a SPARQL query to query the given data source and whose result shows up in the CRF for a given patient.	In the CRF item, the link to the linked data source field is shown along with the information about its ontology annotation. If clicked upon then the SPARQL query is shown.
	The user opens a CRF in use for a patient.	For the CRF item linked to a data source, the result of the SPARQL query is shown. (Possibly marked with a sign or a different colour to show the external origin. Also possibly, the name of data source could be shown when hovering the item.)
Expected usage frequency	Medium	

Needed for DSS	⦿ yes O no
Needs HPC	O yes ⦿ no
Needs Grid	O yes ⦿ no
Priority for development	high
Responsible for development	USAAR, FhG-IBMT, UPM, UCL
Mockup needed	⦿ yes O no
Responsible for Mockup	USAAR
Who is building the tool	USAAR, FhG-IBMT, UPM, UCL
Open Source tool	⦿ yes O no, please specify why:

Reporting Scenario

Item	Description
Identifier	REP_2
Version	1.0
Name	Reporting Scenario
Description of the use case (enduser perspective)	The end user receives a summary report of the data collected of a patient. The end user can be a physician but also the patient him/herself. Therefore the look and content of the report should be adaptable in relation to the end user.
Problem(s) to solve	What are the details that should be shown on the report according to the actual end user group? How can those details be shown in a sensible way?
Challenges	The report has to be understandable and fitting for the different end user groups.
Risks	The report might not be structured or its content might not be suitable for the end user at hand.
Expected benefits	The end user receives a succinct overview over the data of a given patient.
Characterization	○ fundamental ◉ general ○ specific
If specific, please give the Domain	◉ Acute lymphoblastic leukaemia ◉ Breast Cancer ◉ Nephroblastoma ◉ other Cancer, please specify: ◉ Non-Cancer Domain, please specify: as this is a general tool, it can be applied for any kind of trial within ObTiMA
Enduser	○ system ◉ person ○ basic scientist ◉ clinician ○ computer scientist ○ regulatory body, lawyer, ethicist ◉ patient ○ other, please specify:
Pre-condition(s)/pre-requisite(s)	
Requisite(s)	
Post-condition(s)/post-requisite(s)	

Constraints		
External sources needed from outside p-medicine	○ data, please specify: ○ tools, please specify: ○ services, please specify: ○ models, please specify: ○ other, please specify:	
Data used	○ personal ○ only non-personal ○ target population, please specify:	
Input data	◉ internal database, please specify: Database that contains the data of patients collected by using the CRFs ○ external database, please specify: ○ online input	
Output data	○ database, please specify: ○ variables for use, please specify: ○ structured document, please specify: ○ graphic, please specify:	
Data volume		
Dataflow	Please specify:	
Data storage	Please specify:	
Successful End Condition	Production of the report containing all necessary details fitting for the end user group in an understandable fashion.	
Fail End Condition	Production of a report that is not suitable or does not contain all necessary data.	
Basic workflow	Actor Action	System response
	In the case that a clinicians is logged in, then there are two menu items: 1) Select “Create Report for me” from the menu to create a report that is suitable for a clinician as end user 2) “Create Report for Patient” to create a report suitable for a patients. If a patient is logged in then there is only the item “Create Report” which creates a patient suitable report.	The corresponding report is created on-the-fly on the screen (either as HTML or PDF) and is ready to be stored or printed.

Expected usage frequency	medium
Needed for DSS	○ yes ◉ no
Needs HPC	○ yes ◉ no
Needs Grid	○ yes ◉ no
Priority for development	high
Responsible for development	USAAR
Mockup needed	◉ yes ○ no
Responsible for Mockup	USAAR
Who is building the tool	USAAR
Open Source tool	◉ yes ○ no, please specify why:

p-medicine IT-Components Scenarios

Push Scenario – Usage of clinical data from hospital information systems (see 11.1.5)

Data translation for PUSH services

Item	Description
Identifier	PG_19
Version	1.0
Name	Data translation for PUSH services
Description of the use case (end-user perspective)	When a user pushes his data into the p-medicine data warehouse (DW), this needs to translate it into HDOT format. The DW invokes the translation services in the semantic layer, providing the data received and an ontology annotation that permits to translate that data. The semantic layer returns the data in HDOT format
Problem(s) to solve	Translate data to HDOT format using HDOT annotations
Challenges	The tool must be aware of previously translated data in order to support real data integration (avoid duplicate instance creation, support related data linkage)
Risks	Performing too many requests to the DW to identify common instances could provoke excessive performance loss. Requests should b kept to a minimum
Expected benefits	The translation of raw data into HDOT format will allow the DW to offer automatic integration of biomedical data
Characterization	○ fundamental ◉ general ○ specific
If specific, please give the Domain	○ Acute lymphoblastic leukaemia ○ Breast Cancer ○ Nephroblastoma ○ other Cancer, please specify: ○ Non-Cancer Domain, please specify:
End-user	◉ system: the p-medicine Data Warehouse ○ person ○ basic scientist ○ clinician ○ computer scientist ○ regulatory body, lawyer, ethicist ○ patient ○ other, please specify: Database manager

Pre-condition(s)/pre-requisite(s)	The Data Warehouse has a valid ontology annotation for the data he wants to translate	
Requisite(s)		
Post-condition(s)/post-requisite(s)		
Constraints		
External sources needed from outside p-medicine	⦿ data, please specify: Data pushed by end-users to the Data Warehouse O tools, please specify: O services, please specify: O models, please specify: O other, please specify:	
Data used	⦿ personal O only non-personal O target population, please specify:	
Input data	O internal database, please specify: ¤ external database, please specify: The Data warehouse receives from end users raw data from their biomedical databases. These data undergo a pseudonymization process and are translated into RDF form. The result is the data sent to the tool for translation into HDOT format. In addition, the tool received an ontology annotation document O online input	
Output data	⦿ database, please specify: RDF-based data in terms of the HDOT ontology O variables for use, please specify: ⦿ structured document, please specify: An XML document with a specific format containing the HDOT-based annotation of the database O graphic, please specify:	
Data volume	Data size will depend on the amount of data pushed by the users (we can expect up to several megabytes)	
Dataflow	Please specify:	
Data storage	Please specify:	
Successful End Condition	The provided ontology annotation contains all necessary information for translating the received data	
Fail End Condition	Some data cannot be translated due to missing information in the provided ontology annotation	
Basic workflow	Actor Action	System response
	The Data Warehouse (DW) access the tool and	The system begins translating the data by

	submits data and an ontology annotation	means of the provided ontology annotation. The systems produces the translated data and returns it to the DW
Expected usage frequency	High (once every time data at the hospital-side is updated, and user wants to flush it to the DW)	
Needed for DSS	⦿ yes O no	
Needs HPC	O yes ⦿ no	
Needs Grid	O yes ⦿ no	
Priority for development	Medium	
Responsible for development	UPM	
Mock-up needed	⦿ yes O no	
Responsible for Mock-up	UPM	
Who is building the tool	UPM	
Open Source tool	⦿ yes O no, please specify why:	

Ontology annotation of external databases

Item	Description
Identifier	PG_20
Version	1.0
Name	Ontology annotation of external databases
Description of the use case (end-user perspective)	Annotation of external databases in terms of the HDOT ontology is necessary for data to be stored and integrated in the p-medicine Data Warehouse. The tool will offer data managers a graphical interface to perform this annotation. The interface should be intuitive enough for end users lacking deep RDF understanding to be able to correctly annotate their data
Problem(s) to solve	Define an HDOT-based annotation of an external database
Challenges	The tool must support different database formats (excel, access, SQL, RDF). Deep knowledge on RDF/Ontologies should not be a requirement for users
Risks	Unwillingness of end users to provide their database schema, or to perform the annotation process
Expected benefits	The annotation of external databases will allow automatic data access in terms of the HDOT ontology. In addition, integration of heterogeneous databases will be enabled
Characterization	○ fundamental ◉ general ○ specific
If specific, please give the Domain	○ Acute lymphoblastic leukaemia ○ Breast Cancer ○ Nephroblastoma ○ other Cancer, please specify: ○ Non-Cancer Domain, please specify:
End-user	○ system ◉ person ○ basic scientist ○ clinician ○ computer scientist ○ regulatory body, lawyer, ethicist ○ patient ◉ other, please specify: Database manager
Pre-condition(s)/pre-requisite(s)	The tool is able to access the metadata (schema description) of the database to annotate

Requisite(s)		
Post-condition(s)/post-requisite(s)		
Constraints		
External sources needed from outside p-medicine	○ data, please specify: ○ tools, please specify: ○ services, please specify: ○ models, please specify: ◉ other, please specify: Schemas of the databases to be annotated	
Data used	○ personal ◉ only non-personal ○ target population, please specify:	
Input data	○ internal database, please specify: ◉ external database, please specify: Any database to be accessible through the p-medicine data warehouse should undergo the annotation process. The actual data contained in the database is not needed, only its schema ○ online input	
Output data	○ database, please specify: ○ variables for use, please specify: ◉ structured document, please specify: An XML document with a specific format containing the HDOT-based annotation of the database ○ graphic, please specify:	
Data volume	Data involved will have a size of a few KBs	
Dataflow	Please specify:	
Data storage	Please specify:	
Successful End Condition	The user completes the annotation of all database fields that he wants to annotate	
Fail End Condition	The user cannot find an appropriate path in HDOT to map to any path in her/his database.	
Basic workflow	Actor Action	System response
	User logs into the system	The system displays the form for the user to input his database metadata
	User provides access to the database he wants to annotate	The system displays graphical representations of HDOT and the user's database. On this screen the user can select and drag items to perform the annotation

	The system annotates fields of his database	The system marks the annotated fields to give visual feedback to the user
	The user informs that the annotation process is complete	The system stores the XML document containing the performed annotations
Expected usage frequency	Medium (at least once for each database to be integrated in the DW)	
Needed for DSS	⊙ yes O no	
Needs HPC	O yes ⊙ no	
Needs Grid	O yes ⊙ no	
Priority for development	medium	
Responsible for development	UPM	
Mock-up needed	⊙ yes O no	
Responsible for Mock-up	UPM	
Who is building the tool	UPM	
Open Source tool	⊙ yes O no, please specify why:	

Ontology-Based Semantic Search Framework

Item	Description
Identifier	SG_2
Version	1.0
Name	Scenario for Ontology-Based Semantic Search Framework
Description of the use case (end-user perspective)	p-medicine platform could contain an Ontology-Based Semantic Framework (OBSF) able to connects highly heterogeneous data and textual information. The semantic framework could be based on gene, tissue, disease and compound ontologies (important for drugs and clinical research frames). This framework could contain information from different organisms, platforms, data types and research areas that is integrated into and correlated within a single searchable environment using search algorithms. It will provide a unified interface for all p-medicine End Users to formulate, explore and identify new information (according to specific preferences and needs) across vast collections of experimental data. p-medicine's OBSF will combines classical keyword-based search with text-mining and ontologies to navigate large results sets (internal & external) and facilitate information and/or knowledge discovery. End Users will be provided with an advanced ontology based (Gene Ontology (GO) and Medical Subject Headings (MeSH)) "table of contents" in order to access, explore, structure (quickly) the millions of available resources (PubMed abstracts, news, clinical trials) according to the predefined topics "of interest" (AAL, Nephroblastoma, Breast Cancer, etc.). Examples: - GoPubMed, http://www.gopubmed.com - NextBio, http://www.nextbio.com - ResearchGate, http://www.researchgate.net
Problem(s) to solve	Unified and semantic based p-medicine platform search engine
Challenges	Insufficient practical experience on data-mining and ontology-based search solutions. Term extraction from external data (PubMed abstract, Clinical Trial, News article) and semantic benchmarking with GO and MeSH.
Risks	Risks are associated with the needs on advanced Search Algorithms and data mining technics/approaches.
Expected benefits	All End-users will be able to use and explore the p-medicine's OBSF (wide usage and acceptance)
Characterization	○ fundamental ◉ general ○ specific

If specific, please give the Domain	⊙ Acute lymphoblastic leukaemia ⊙ Breast Cancer ⊙ Nephroblastoma ⊙ other Cancer, please specify: *(applicable to all domains related to biomedical research)* ⊙ Non-Cancer Domain, please specify: *(applicable to all domains related to biomedical research)*
End-user	⊙ system ⊙ person ⊙ basic scientist ⊙ clinician ⊙ computer scientist ⊙ regulatory body, lawyer, ethicist ⊙ patient ⊙ other, please specify: *(Applicable to all end users)*
Pre-condition(s)/pre-requisite(s)	p-medicine platform with access to external and/or local databases with publically available data (PubMed, Clinical Trials, News, etc.)
Requisite(s)	Gene Ontology (GO) and Medical Subject Headings (MeSH)
Post-condition(s)/post-requisite(s)	End Users Feedback and usability suggestions tracking
Constraints	Insufficient experience with similar, publically available, on-line projects
External sources needed from outside p-medicine	⊙ data, please specify: *PubMed Repository, Clinical Trials information, news articles, etc.* *Example:* - *Acute Lymphoblastic Leukaemia - 27190 PubMed search Results (July 2011)* - *Breast Cancer - 232072 PubMed search Results (July 2011)* - *Nephroblastoma - 10039 PubMed search Results (July 2011)* ⊙ tools, please specify: - *Text mining applications* - *Apache Lucene(TM) is a high-performance, full-featured text search engine* - *GATE: a full-lifecycle open source solution for text processing* ⊙ services, please specify: *OpenCalais Web Service allows to automatically annotate the content with rich semantic metadata* ⊙ models, please specify:

	Semantic data model, http://en.wikipedia.org/wiki/Semantic_data_model (July, 2011) ⊙ other, please specify: *Ontologies/Vocabularies* *Gene Ontology (GO), Medical Subject Headings (MeSH)*	
Data used	⊙ personal (should be decided by the End Users) ⊙ only non-personal O target population, please specify:	
Input data	⊙ internal database, please specify: ⊙ external database, please specify: ⊙ online input	
Output data	⊙ database, please specify: ⊙ variables for use, please specify: ⊙ structured document, please specify: ⊙ graphic, please specify:	
Data volume	High	
Dataflow	Please specify:	
Data storage	Please specify: *Data Warehouse and/or Cloud Computing/Network*	
Successful End Condition	Contribution and efforts from all p-medicine project partners (is not excluded the need for external contribution)	
Fail End Condition		
Basic workflow	Actor Action	System response
	Automatic (on-the-fly) semantic analysis of End user data and prefilled sections with content (i.e. Latest PubMed Articles, News, Clincal Trials, etc.)	Click on the provide resources with access to Ontology/Semantic based “table of contents” and/or “Tag Cloud”. For an on-line examples please explore: GoPubMed, http://www.gopubmed.com NextBio, http://www.nextbio.com ResearchGate, http://www.researchgate.net
	Access to advanced search frames	p-medicine’s OBSF provide with an advanced, user friendly and powerful ontology based “table of content” similar to GoPubMed or “Tag Cloud” similar to NextBio
	Subscription frames	End user has the possibility to subscribe and receive regular alert messages in case of new content

Expected usage frequency	Advance usage is expected
Needed for DSS	⦿ yes O no
Needs HPC	⦿ yes O no
Needs Grid	⦿ yes O no
Priority for development	High
Responsible for development	All p-medicine project partners
Mock-up needed	O yes ⦿ no (please explore the above examples)
Responsible for Mock-up	
Who is building the tool	All p-medicine project partners
Open Source tool	O yes ⦿ no, please specify why: Needs for proprietary search algorithms and the enrolment of high skilled and experienced semantic and/or data mining partners is not excluded.

p-medicine portal scenario

Item	Description
Identifier	SG_3
Version	0.1
Name	p-medicine portal
Description of the use case (enduser perspective)	The p-medicine infrastructure integrates various tools, services and components, from clinical trial management and virtual organization management, through a security infrastructure and data anonymization, to database integration, ontology-based semantic mediation and the exploitation of data in end-user tools, such as literature mining, GridR and the Oncosimulator as made available to (and reusable by) the user via the workflow environment (according to the DoW). The p-medicine portal allows searching for specific tools, models, services and data based on their semantic annotations and user generated metadata (e.g. Data Warehouse, Oncosimulator, ObTiMA, tools for education and training). Additionally the users will be supported in extending the functionality of the p-medicine workbench by registering and publishing custom tools and services as well as in using the collaboration tools. The access to the p-medicine framework will be regulated by a roles and rights management system via the secure p-medicine portal. Unauthorized access will therefore be avoided and the risk of misuse of data within p-medicine will be restricted to people legally bound by contracts to data providers.
Problem(s) to solve	
Challenges	Identify tools, services, data sources that can be accessed from p-medicine portal Integration of new custom tools and services on the running portal. Select an appropriate portal framework.
Risks	Heterogeneity of tools, services and data sources could lead to problems during integration in the portal.
Expected benefits	
Characterization	○ fundamental ◉ general ○ specific
If specific, please give the Domain	◉ Acute lymphoblastic leukaemia ◉ Breast Cancer ◉ Nephroblastoma

	⊙ other Cancer, please specify: ⊙ Non-Cancer Domain, please specify:
Enduser	O system ⊙ person ⊙ basic scientist ⊙ clinician O computer scientist O regulatory body, lawyer, ethicist ⊙ patient ⊙ other, please specify: technical portal administrator
Pre-condition(s)/pre-requisite(s)	The p-medicine portal framework should - have a robust functionality, - have a comfortable user Interface - have a flexible service oriented architecture - have robust user management and security features including password policies - have high availability and high performance - needs to be fully compatible with all major databases, operating systems, and application servers. - offer a customizable single sign-on (SSO)
Requisite(s)	
Post-condition(s)/post-requisite(s)	
Constraints	
External sources needed from outside p-medicine	⊙ data, please specify: tools, services and data sources that should be accessed from the portal O tools, please specify: O services, please specify: O models, please specify: O other, please specify:
Data used	O personal O only non-personal O target population, please specify:
Input data	O internal database, please specify: O external database, please specify: O online input
Output data	⊙ database, please specify: portal database O variables for use, please specify: O structured document, please specify: O graphic, please specify:

Data volume		
Dataflow	Please specify:	
Data storage	Please specify:	
Successful End Condition		
Fail End Condition		
Basic workflow	Actor Action	System response
	The portal user logs in the portal. The login data are valid for the whole content of the portal (single-sign-on commitment)	
		The portal framework check credentials of the user and allows or rejects access to the portal content.
	The portal user performs different asynchronous actions on the portal according to a set of assigned roles and permissions: - administration activities for different user groups: management of communities, organisations, user groups, teams in the portal; management of roles and permissions; definition and management of GUI; management of pages, contents and available resources; configuration of the portal framework; - using the tools, services and data sources developed by the p-medicine workbench; sharing documents and resources for different user groups; - using collaboration tools.	
		The portal framework will save all modifications of the user in the database and will provide access to different resources available in the

		portal to users according the last modifications.
Expected usage frequency	?	
Needed for DSS	○ yes ◉ no	
Needs HPC	○ yes ◉ no	
Needs Grid	○ yes ◉ no	
Priority for development	Needs to be developed until month 18 prototypically according to the DoW.	
Responsible for development	IBMT, FORTH, CUSTODIX	
Mockup needed	◉ yes ○ no	
Responsible for Mockup	IBMT and USAAR	
Who is building the tool	IBMT, FORTH, CUSTODIX	
Open Source tool	○ yes ○ no, please specify why: ◉ not clear yet	

Scenario for Education and Training

Item	Description
Identifier	EdTr_1
Version	1.0
Name	Scenario for Education and Training
Description of the use case (end-user perspective)	Educating end-users in how to best use the tools created by p-medicine will be vital to their continued use and success. The eLearning tools will be designed with the end-users' needs in mind. Different user-groups will be using different educational tools therefore a different set of user requirements will be identified for each tool. A different educational tool will be required for each of the tools created by p-medicine, these tools will need to be populated with fake, but realistic data to allow the end-users to practice and demonstrate competence. Each educational tool will be created in close cooperation with WP15 to contain an inbuilt validation process. The educational tools will be hosted on ecancer.eu as well as the p-medicine website and will be annotated to the corresponding tool within p-medicine environment. Each tool will contain an end-user data capture introduction with a short pre-test to determine pre-existing knowledge followed by the educational content. Users will then have a practice environment with a final competence and validation requirement. An automatic reminder will be sent out after completion to help ensure retention of knowledge and competence leading to patient benefit.
Problem(s) to solve	Creation of all of the required educational tools suitable for each end-user group
Challenges	Low user satisfaction, low knowledge retention or low knowledge to performance conversion.
Risks	Any of the challenges resulting in low patient benefit or the incorrect use of p-medicine tools
Expected benefits	End-user education resulting in continued and competent use of p-medicine tools
Characterization	○ fundamental ◉ general ○ specific
If specific, please give the Domain	○ Acute lymphoblastic leukaemia ○ Breast Cancer ○ Nephroblastoma

	○ other Cancer, please specify: ○ Non-Cancer Domain, please specify:	
End-user	○ system ◉ person ◉ basic scientist ◉ clinician ◉ computer scientist ◉ regulatory body, lawyer, ethicist ◉ patient ○ other, please specify:	
Pre-condition(s)/pre-requisite(s)		
Requisite(s)	Realistic dummy data	
Post-condition(s)/post-requisite(s)	Reminder for knowledge reinforcement activity	
Constraints		
External sources needed from outside p-medicine	○ data, please specify: ○ tools, please specify: ○ services, please specify: ○ models, please specify: ○ other, please specify:	
Data used	○ personal ◉ only non-personal ○ target population, please specify:	
Input data	○ internal database, please specify: ○ external database, please specify: ◉ online input	
Output data	○ database, please specify: ◉ variables for use, please specify: End-user contact details ○ structured document, please specify: ○ graphic, please specify:	
Data volume	Low	
Dataflow	Please specify:	
Data storage	Please specify: In the data warehouse?	
Successful End Condition	Required educational tools available to end-users	
Fail End Condition		
Basic workflow	Actor Action	System response
	Selection of correct educational tool	Tool opens

	Sign in	System retains end-user details
	User views educational content	
	User enters the practice area	Dummy data is available
	Competence and validation area	System verification needed
	User completes educational tool	Reminder for knowledge reinforcement activity send after a set period
Expected usage frequency	Each time a new user uses the p-medicine tools or an old user needs knowledge reinforcement	
Needed for DSS	○ yes ◉ no	
Needs HPC	○ yes ◉ no	
Needs Grid	○ yes ◉ no	
Priority for development		
Responsible for development	Ecancer	
Mock-up needed	○ yes ◉ no	
Responsible for Mock-up		
Who is building the tool	Ecancer	
Open Source tool	◉ yes ○ no, please specify why:	

Appendix 6 – Ethical Approval

Der Vorsitzende

Ärztekammer
des Saarlandes
Körperschaft
des öffentlichen Rechts

Ärztekammer des Saarlandes · Postfach 100262 · 66002 Saarbrücken
Ethik-Kommission

Herrn Professor
Dr. med. N. Graf
Kliniken für Kinder- und Jugendmedizin
Klinik für Pädiatrische Onkologie und Hämatologie
Universitätsklinikum des Saarlandes
66421 Homburg

Ethik-Kommission
Geschäftsstelle

Faktoreistraße 4
66111 Saarbrücken
Telefon-Durchwahl (0681) 4003-378
Telefax (0681) 4003-394
E-Mail: ethikkommission@aeksaar.de
Internet: www.aerztekammer-saarland.de

Unsere Kenn-Nr.:	Ihr Schreiben vom:	Ihr Zeichen:	Datum:
Prof. Schie./Ha			16. März 2011

Europäische Forschungsprojekte:

1. **ACGT (Advancing Clinico Genomic Trials)**
2. **ContraCancrum (Clinically Oriented Translational Cancer Multilevel Modelling)**
3. **TUMOR (Transatlantic Tumor Model Repositories)**

Jetzt: p-medicine (from Data Sharing and Integration via VPH-Models to Personalized Medecine)

Unsere Kenn-Nr.: 104/10 ***(bitte stets angeben!)***

Sehr geehrter Herr Graf!

Wir bestätigen den Erhalt Ihres Schreibens vom 02.03.2011, Eingang hier: 10.03.2011, mit folgender Anlage:

- *Unser Votum vom 20. Juli 2010.*

Gegen die Benutzung der im Rahmen des Forschungsprojektes ACGT erhobenen Daten auch im Forschungsprojekt p-medicine bestehen keine Bedenken.

Wir wünschen Ihnen weiterhin viel Erfolg.

Mit freundlichen Grüßen

San.-Rat Prof. Dr. Schieffer

Die Ethik-Kommission bei der Ärztekammer des Saarlandes ist unter Beachtung der internationalen Richtlinien der ICH, GCP-V und der 12. Novelle AMG tätig, nach Landesrecht (Saarländisches Heilberufekammergesetz, § 5 Abs. 1) anerkannt und beim Bundesinstitut für Arzneimittel und Medizinprodukte gem. § 17 Abs. 7 des Medizinproduktegesetzes sowie beim Bundesamt für Strahlenschutz nach § 92 der Strahlenschutzverordnung und nach § 28g der Röntgenverordnung registriert.