Deliverable No. 10.3

Biobank Access Services fully integrated into the p-medicine infrastructure

Grant Agreement No.:	270089
Deliverable No.:	D10.3
Deliverable Name:	Biobank Access Services fully integrated into the p-medicine infrastructure
Contractual Submission Date:	31/01/2014
Actual Submission Date:	31/01/2014

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 10.3
Document name:	Biobank Access Services fully integrated into the p-medicine infrastructure
Nature (R, P, D, O)[1]	O
Dissemination Level (PU, PP, RE, CO)[2]	PU
Version:	1
Actual Submission Date:	31/01/2014
Editor: Institution: E-Mail:	Gabriele Weiler Fraunhofer Institute for Biomedical Engineering Gabriele.Weiler@ibmt.fraunhofer.de

ABSTRACT:

Biobanks represent key resources for clinico-genomic research and advances in personalized medicine. With its ICT infrastructure and contractual framework, p-medicine will support researchers' growing demand to access and share high quality biomaterial and related data for their research projects. For this purpose a Biobank Access Framework called p-BioSPRE has been developed to enable and simplify access to existing biobanks, but also to offer own biomaterial collections to research communities and manage biobank specimens over the ObTiMA Trial Biomaterial Manager. This deliverable describes the architecture of p-BioSPRE, the functionality of its different components and its integration into the p-medicine infrastructure.

KEYWORD LIST: Biobank Access, specimen, p-BioSPRE, ObTiMA Trial Biomaterial Manager, IDB, p-Biobank Wrapper

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement nº 270089.

The author is solely responsible for its content, it does not represent the opinion of the European Community and the Community is not responsible for any use that might be made of data appearing therein.

[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	01/12/13	Draft	Gabriele Weiler
0.2	19/12/13	Added Chapter 4	Christina Schröder
0.3	10/01/14	Added Chapter 2, 3, 6 and 7	Gabriele Weiler
0.5	17/01/14	Draft for internal review	Gabriele Weiler, Christina Schröder, Fatima Schera
1.0	24/01/14	Final version	Gabriele Weiler, Christina Schröder, Norbert Graf, Stefanie Hänold, Nikolaus Forgó, Fatima Schera, Iheanyi Nwankwo

List of contributors

- Gabriele Weiler, Fraunhofer IBMT
- Fatima Schera, Fraunhofer IBMT
- Dr. Christina Schröder, Fraunhofer IBMT, Branch Potsdam-Golm
- Matthias Dobkowicz, Dipl.-Inf. (FH), Fraunhofer IBMT, Branch Potsdam-Golm
- Nikolaus Forgó, LHU
- Stefanie Hänold, LHU
- Iheanyi Nwankwo, LHU
- Norbert Graf, USAAR

Contents

1 Executive Summary

Biobanks represent key resources for clinico-genomic research and are needed to pave the way to personalised medicine. To achieve this goal it is crucial that scientists can securely access and share high quality biomaterial and related data. Therefore, there is a growing interest in integrating biobanks into larger biomedical information and communication technology (ICT) infrastructures.

Within the p-medicine platform, we have developed a Biobank Access Framework called p-BioSPRE. This generic framework not only enables and simplifies access to existing biobanks, but also assists users in offering own biomaterial collections to research communities and in managing biobank specimens and related clinical data over the ObTiMA Trial Biomaterial Manager.

P-BioSPRE takes into consideration all relevant ethical and legal standards, e.g. safeguarding donors' personal rights and enabling biobanks to keep control over the donated material and related data. The framework thus enables secure sharing of biomaterial within research communities, while flexibly integrating related clinical and omics data.

Although the development of the framework is mainly driven by user scenarios from the cancer domain, in this case Acute Lymphoblastic Leukemia (ALL) and Wilms tumor, it can be extended to further disease entities.

In this deliverable we describe the architecture and functionality of the Biobank Access Framework and its integration into the p-medicine environment.

2 Introduction

A research-integrated biobanking solution is crucial for the advent of personalised medicine for cancer patients. Technological progress in cancer-related molecular biology and molecular imaging has led to an exponential increase in the possibilities for individualised cancer care. Moreover, the accompanying complexity of the resulting biomarkers and their application in, for instance, radiotherapy planning and targeted drug choice and dosing has made the knowledge base too complex for physicians and certainly incomprehensible for patients. Many more new potential biomarkers will be discovered in the next years, and their validation as useful biomarkers in diagnosis and therapy will require fast reference to the global collection of well-characterised and annotated human biospecimens. Beyond access to high-quality biological samples, research will also require integrated access to sample-related clinical and omics data of increasing granularity. In order to address this need, a biobank access framework called p-BioSPRE ("Biospecimen Search and Project Request Engine") has been designed and integrated into the p-medicine platform.

Within its overall ICT and contractual architecture, the Biobank Access Framework and the p-medicine Data Warehouse are foreseen to complement and synergise with each other to generate and contribute critical VPH knowledge.

The aim of p-BioSPRE is to provide access to different kind of human biomaterials and related data for research purposes. A biobank operator is supported in offering his biomaterial and related data to research communities, compliant with ethical and legal standards. A researcher is enabled to search the biomaterial that is offered within his communities. It is furthermore possible for him to request biomaterial for a research project and his request is forwarded on-line to the biobank operator.

Furthermore, within p-BioSPRE the ObTiMA Trial Biomaterial Manager is developed that enables users of ObTiMA, the p-medicine's ontology based trial management system, to manage their biomaterial data within clinical trials. For this purpose a pre-defined but adjustable case report form for patient's biomaterial according to a standard biobank dataset is provided in ObTiMA. The biomaterial data can be integrated with clinical data.

In D10.2 we have described the implementation of the first prototype of p-BioSPRE. In this deliverable we describe the functionality and architecture of the first version of p-BioSPRE and its integration into the p-medicine platform. It can therefore be seen as an update and amendment of D10.2. Please note that hence several parts will inevitably overlap in the two deliverables, since the aim of this deliverable is to provide a complete and consistent description of the functionality and the architecture.

The deliverable is structured as follows:

In Chapter 3 an overview of the architecture of p-BioSPRE and its integration into the p-medicine IT environment is given.

Chapter 4 describes the main component of the Biobank Access Framework, namely the p-BioSPRE metabiobank. P-BioSPRE provides researchers the possibility to search for and request biomaterial that fits their research purposes.

In Chapter 5 the p-Biobank Wrappers and the underlying Inhouse Database (IDB) is described. The P-Biobank Wrappers enable biobank operators to share their biomaterial and related data in p-BioSPRE.

Chapter 6 describes the functionality of the ObTiMA Trial Biomaterial Manager that enables management of biobanks and associated specimen data in clinical trials and sharing selected specimen data.

Chapter 7 concludes the deliverable by outlining the future work.

3 Overview of Architecture and Integration into the p-medicine environment

In order to build an integrated biobanking solution for p-medicine, two typical biobanking user scenarios from the cancer domain (Acute Lymphoblastic Leukemia (ALL) and Wilms tumor) have been analysed and user requirements identified. Furthermore, different tools and projects for integrated biobanking have been evaluated with the aim to find models, guidelines or software that can be adapted to the biobanking user scenarios and constitute the base of an integrated biobanking framework. The user scenarios and results of this analysis are described in Deliverable 10.1. Based on this research a biobank access framework has been developed in close interaction with end users and legal experts. In the following, we summarize the main functionality and describe its architecture and the integration of the different components into the p-medicine platform.

3.1 Main Functionality

The aim of p-medicine's biobank access framework is to provide access to different kind of human biomaterials and related data for research purposes. Starting with a minimum data set (Appendix 2), the framework provides means to harmonise data and facilitates data import by help of a metadata repository. In particular, the p-medicine biobank framework supports the following main functionalities:

a. Offering human biomaterial for research

A biobank operator is supported in providing data on his biomaterial and related clinical data to research communities, as defined in p-BioSPRE. Data provision is regulated in p-BioSPRE's legal framework and by p-medicine's Biobank Data Transfer Agreement (see Deliverable 5.3 for a detailed description). Any biobank management system can be used to import data in p-BioSPRE. The biobank operator decides which data will be disclosed or not, thus maintaining full control of his material and data.

b. Requesting specific human biomaterial for research purposes

A registered researcher is enabled to search for and request biomaterial that is available within his research communities. He can access information about anonymized data linked to biomaterial and the number of available cases (a case ("psn" in Appendix 2) being defined as the data belonging to a specific disease of a donor/patient, backed by specimens). The p-BioSPRE web portal enables researchers to define a search profile according to the classification and annotations stored in the underlying database. The result of searches is presented as statistical groups only. The project portal allows selection of statistical groups from the search result. In addition, researchers are requested to describe the scope of the envisioned project and can request additional services (e.g. PCR, IHC, FISH; cf. Figure 5, "Requested Methods"). Upon submission of this project request, the providing hospital or consortium will be shown.

c. Managing Biomaterial Data in ObTiMA

Users of ObTiMA, the p-medicine's ontology based trial management system, can manage their biomaterial data within clinical trials. For this purpose, a pre-defined but adjustable case report form (CRF) for patient's biomaterial is provided in ObTiMA, the so-called biobanking specimen CRF. The biomaterial data can be integrated with clinical data within a trial or across several trials for further analysis. Legacy biomaterial data can be imported into ObTiMA from excel files.

3.2 Main Components and their Interaction

We have designed the biobank access framework as a set of coupled components, which are depicted in Figure 1.

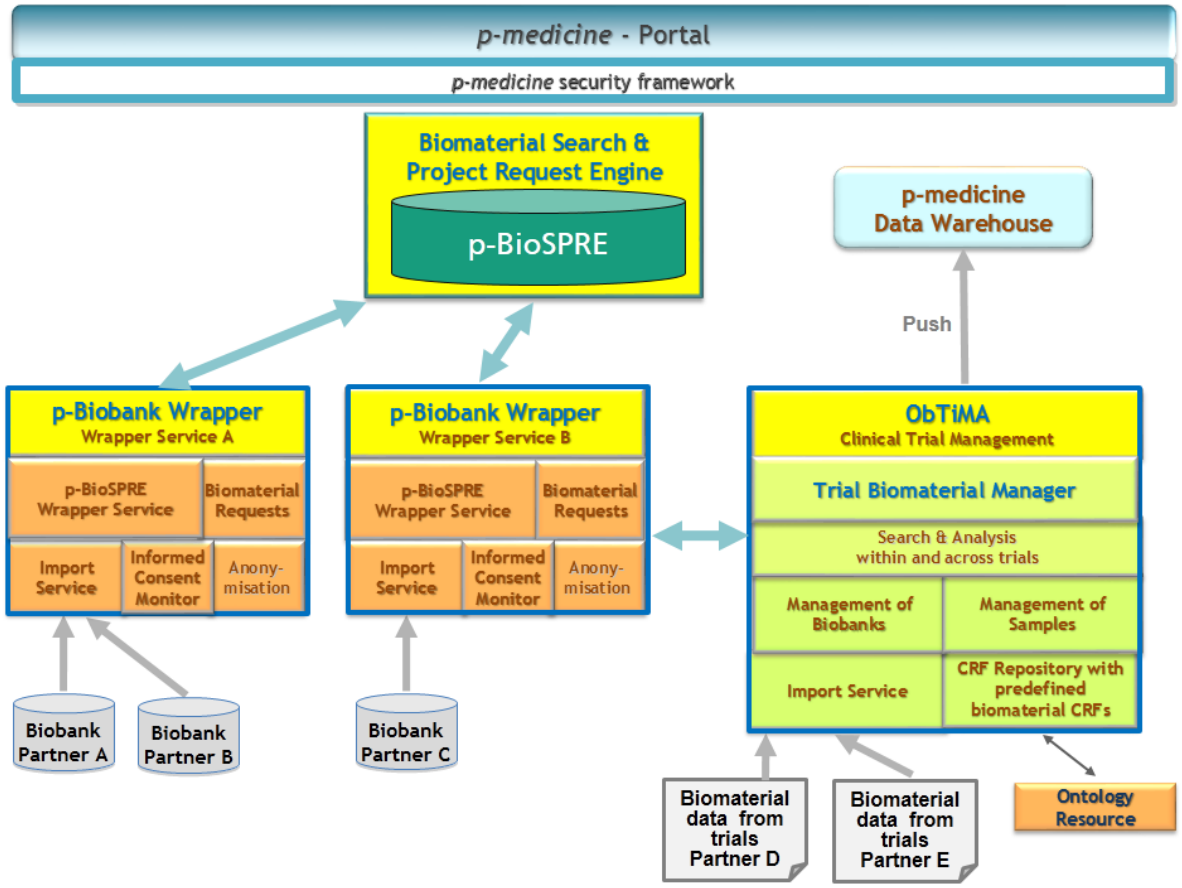


Figure 1 Architecture of the p-medicine Biobank Access Framework.

The main component of the framework is ***p-BioSPRE***, the ***p****-medicine* **Bio**material **S**earch and **P**roject **R**equest **E**ngine, which is a meta biobank to share biomaterial for research purposes. P-BioSPRE is based on the CRIP meta biobank concept[3] and toolbox.

Furthermore, the framework comprises the ***p-Biobank Wrappers,*** which are tools to support biobank operators to make their biomaterial and related data available in p-BioSPRE and to manage associated requests, independent from any biobank management system they may have in place. This is achieved by the core of each p-Biobank Wrapper, ***the local In-house Database (IDB)***, which is also a component of the CRIP Toolbox.

In order to enable users of the *p-medicine* trial management system ObTiMA to integrate biomaterial data in clinical trials and to offer them in p-BioSPRE a ***Trial Biomaterial Manager*** is provided.

[3] Schröder C, Heidtke KR, Zacherl N, Zatloukal K, Taupitz J (2010) Safeguarding donors' personal rights and biobank autonomy in biobank networks: the CRIP privacy regime. Cell Tissue Bank doi: 10.1007/s10561-010-9190-8; 12(3): 233 – 240

3.3 Integration into the p-medicine environment

The components of the biobank access framework are seamlessly integrated into the p-medicine platform. The aim of the p-medicine platform is to provide an integrated platform to foster research and clinical decision support for personalised therapy. This platform integrates tools and services to support all kind of research for personalized medicine. Beside tools and services to semantically integrate and analyse heterogeneous biomedical data in a data warehouse, it also comprises the ontology-based trial management system ObTiMA to design and conduct clinical trials. The integration of p-BioSPRE into the p-medicine platform shall ensure that researchers' needs regarding biobanking are met and supported in line with other services needed for research in personalized medicine in integrated workflows.

The Biobank Access framework is integrated with the portal, the security infrastructure, the Data Warehouse and ObTiMA. In the following we will describe the components in more detail.

Portal

The main access point to the p-medicine platform is the *p-medicine* portal that provides clinicians, patients and researchers a platform to collaborate, share data and expertise, and use tools and services to improve personalized treatments of patients. The *p-medicine* portal (cf. Figure 2) is a web-based environment providing a single access point from which most of the *p-medicine* applications can run. These applications are integrated in a consistent and systematic way (see Deliverable 8.1.2).

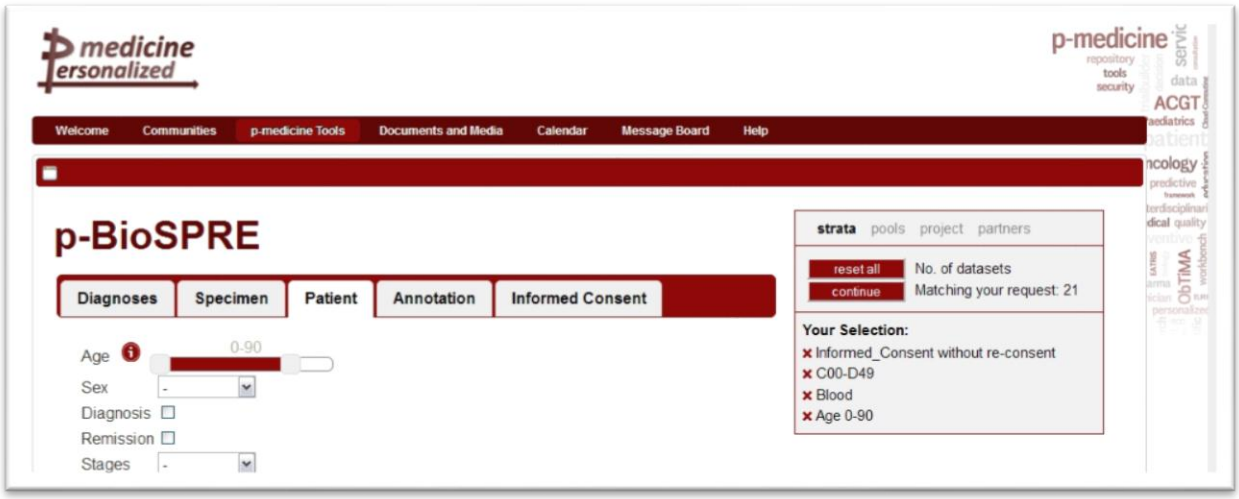


Figure 2: p-medicine portal with p-BioSPRE selected from the menu "p-medicine Tools". To access these tools, user authentication is required.

The *p-medicine* portal users do not need to sign in for every application separately. With Single Sign-On functionality the login procedures for different applications are centralized in one system, accessible with only one password. After entering the right password, only the applications that the user is authorized to see become available.

The p-BioSPRE metabiobank and the Trial Biomaterial Manager can both be accessed via the portal.

Security Infrastructure

The *p-medicine* platform provides a lightweight dynamic security architecture, which is integrated in the portal. It consists of modular re-usable components, which deal with security issues such as authentication, authorisation, auditing and de-identification. The p-medicine

security framework[4] is based on open commonly used stable standards such as SAML 2.0, SSL, X.509[5], WS-*[6] (e.g. WS-Security[7], WS-Trust[8]) and XACML[9].

For authentication and authorisation p-BioSPRE relies on the p-medicine security framework. This framework guarantees confidentiality in p-BioSPRE by restricting access only to authenticated users who have sufficient access rights through a security gateway and encrypting all communication through SSL.

Data Warehouse and Semantic Integration Layer

The p-medicine Data Warehouse is the central research repository of *p-medicine* with respective services for collecting, sharing and further elaborating 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. Data is semantically integrated in terms of the HDOT ontology (see Deliverable 4.3).

From the Trial Biomaterial Manager biomaterial data can be pushed into the DWH via push services that automatically translate the data into an HDOT compliant format that can be stored in the DWH (see Section 6.7).

ObTiMA

ObTiMA is an ontology-based clinical trial management web application intended to support clinicians in both designing and conducting clinical trials. The Trial Builder in which all aspects of a clinical trial can be specified facilitates the design phase. A graphical user interface allows defining content, navigation, and layout of ontology-based Clinical Report Forms (CRFs) developed by a trial chairman to capture all patient data during a trial, e.g., medical findings or diagnostic data.

One of the user needs that we identified for an integrated biobanking solution in personalized medicine is that trial data and biomaterial data can be managed in one system in order to enable the seamless integration of these data. Therefore, the Trial Biomaterial Manager was developed as a module of ObTiMA assuring that clinical and biomaterial data can be captured and managed similarly and retrieved with integrated queries.

Legal Framework

Alongside the ICT framework, a legal framework has been provided for p-BioSPRE to ensure compliance with applicable regulations. This framework deals with data protection and data security issues within the p-BioSPRE meta biobank, and regulates in this respect the relationship between the actors: meta biobank operator, biobank operators and researchers. A detailed description of this framework can be found in Deliverable 5.3.

Moreover, p-BioSPRE is compliant with the BBMRI list of requirements for biobank data integration systems (Deliverable 10.1, chapter 5.2.2), including (amongst others)

- compliance with local database policies, national ELSI regulations and EU data protection regulations (R 1)
- user authentication and authorization (R 9)

[4] D5.1: Setting up of the data protection and data security framework.

[5] X.509 is a standard for public key infrastructure (http://www.ietf.org/html.charters/pkix-charter.html).

[6] WS – Web Service

[7] https://www.oasis-open.org/committees/tc_home.php?wg_abbrev=wss

[8] http://docs.oasis-open.org/ws-sx/ws-trust/200512/ws-trust-1.3-os.html

[9] XSAML is an XML-based declarative access control policy language defining both a policy, decision request and decision response language. It is based on the Attribute Based Access Control (ABAC) model which incorporates Role Based Access Control (RBAC): http://docs.oasis-open.org/xacml/3.0/xacml-3.0-core-spec-os-en.pdf

- full control of local biobanks on the data they expose (R 14)
- anonymization / k-anonymization of data and metadata (R 15, R 16)
- metadata driven query and analysis tool, not custom-written against a fixed data model (R 27)

In the following chapters we describe the functionality of the different components of p-BioSPRE and their integration into the p-medicine platform in more detail.

4 p-BioSPRE - The p-medicine Biomaterial Search and Project Request Engine

P-BioSPRE is a meta biobank that provides researchers the possibility to search for and request biomaterial that fits their research purposes.

Technically, p-BioSPRE is based on the CRIP meta biobank. It is a web application and database architecture and can be accessed via the p-medicine portal, the main access point of the p-medicine platform.

Database Architecture

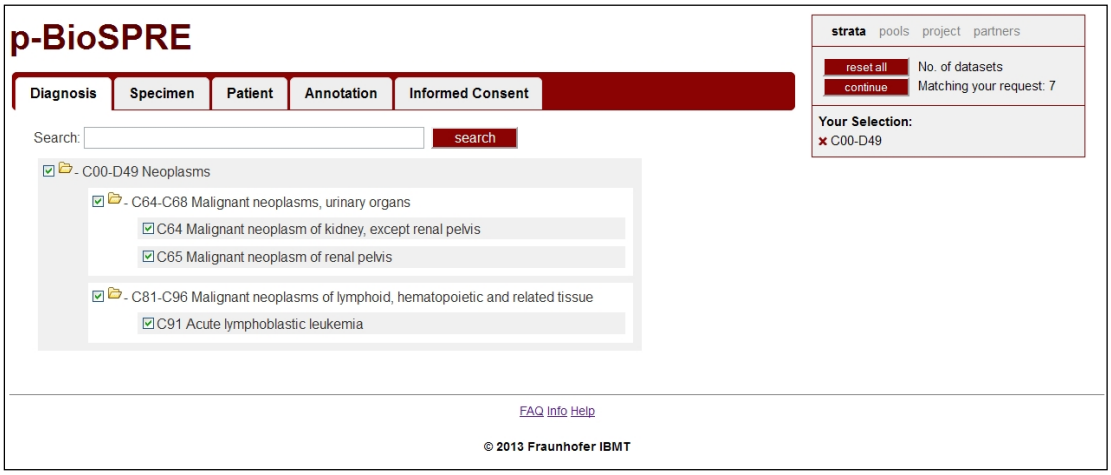
From the biobank partners local In-house databases (see Chapter 5) data is uploaded to the central database. The anonymized data are encrypted via ssl and then uploaded to the p-BioSPRE central database, which is hosted at Fraunhofer IBMT. Before upload, biobank partners need to authenticate. In the central database, each subsequent upload will completely replace the previous one of this biobank partner ensuring that single cases (cf. Appendix 2) cannot be traced back. Although search in the central database is performed on anonymized cases, the result is displayed as pools of cases/statistical data only (see text below and Figure 4) ensuring a k-anonymized view of the data. This design delivers, through a differentiated query of high granularity, yet anonymized data, a comprehensive and satisfying result (Figure 4 andFigure 5) to the p-BioSPRE user, though compliant with all privacy and data protection requirements.

p-BioSPRE Search Tool

P-BioSPRE provides a search interface that enables authorised users to search for biomaterial. After authentication, users access an interactive search tool (Figure 3) allowing to select:

- **diagnosis** (based on ICD 10 and ICD-O))
- **specimen** (group and type of specimen, e.g. whole blood, serum, tissue (FFPE or cryo-preserved), DNA, RNA, etc.)
- **patient data** (age, sex, body mass index (BMI) etc.)
- **annotation** (including information on clinical parameters, genetic subtypes and omics data)
- **consent** (information on patient's informed consent given for the underlying biobank/trial)

a)



b)

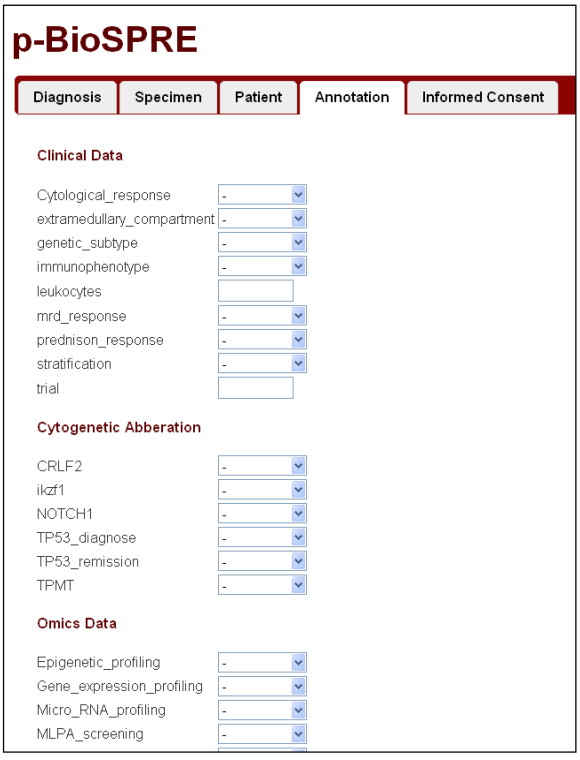


Figure 3: p-BioSPRE Search Tool

a) Search criteria can be entered under five tabs representing multiple options each. Selected criteria are shown in the box "Your selection", overall number of available matching cases just above.

b) Annotation of specimens includes clinical, cytogenetic, and omics data. The selection of parameters shown here is representing the ALL data set and can be extended for any other datasets if required.

Although initially only cases of Wilms tumor and ALL will be processed in p-medicine, the system is open to integrate and display all diagnoses, as technically provided by the CRIP concept.

Once the user has selected appropriate criteria ("strata", cf. Figure 3 a) for his envisioned project, the "continue" button will open up the list of "pools" (Figure 4) or, in other words, of (statistical) groups of cases matching the user's request. Although providing the user with the sought-after information how many cases and specimens would be available for his/her project, tracking of single cases or patients will obviously not be possible over the p-BioSPRE Search Tool. Hence the statistical (or "aggregated") data shown on the p-BioSPRE pool list are fully in line with data and privacy protection requirements. This design allows a maximum protection of privacy-related interests of donors, in particular anonymity of donors within the p-BioSPRE infrastructure, whereas at the same time access to information on available specimen is provided.

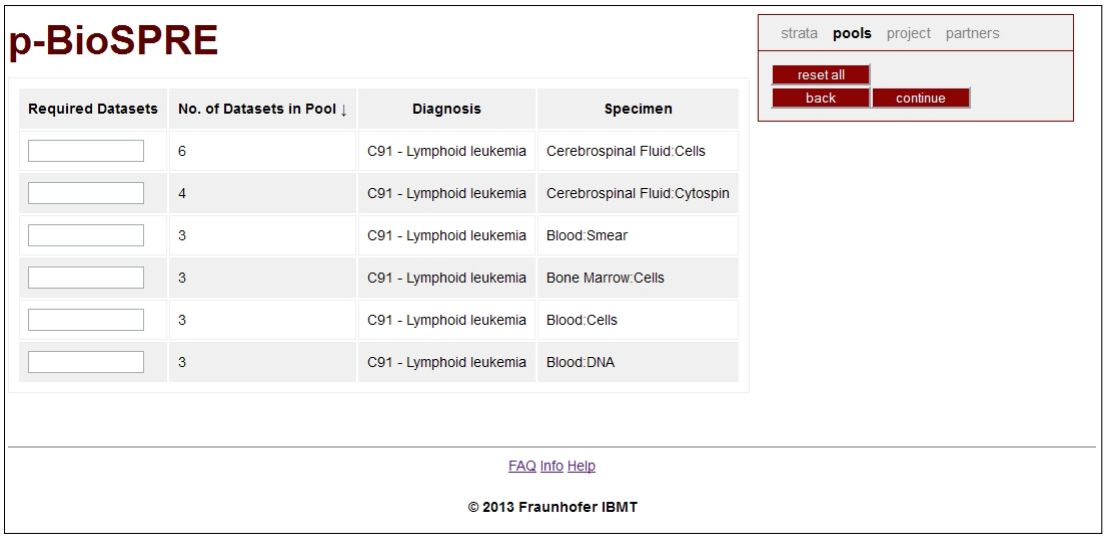


Figure 4: p-BioSPRE Pool List.
Upon "continue", the pools selected by requesting a number of cases ("Required Datasets", left column) are automatically inserted into the Project Request Form (Figure 5). Note that pools shown in this screenshot are small since to date only 50 datasets have been imported into p-BioSPRE. Display of minimum no. of datasets in a pool can be set by default (see Chapter 5, Elements of k-anonymization).

When a user has found appropriate biomaterial for his research, he can request the biomaterial. For this purpose, p-BioSPRE provides an interactive request form (Figure 5) that enables the user to specify the amount of biomaterial he needs and to outline his research project in some detail (i.e. as far as relevant for the biobank). Having submitted his project request, a receipt window will open up to the user indicating the biobank operator(s) to whom his request has been forwarded. He is then free to contact them directly or await their response within ten workdays, as agreed over the Biobank Data Transfer Agreement (see Deliverable 5.3).

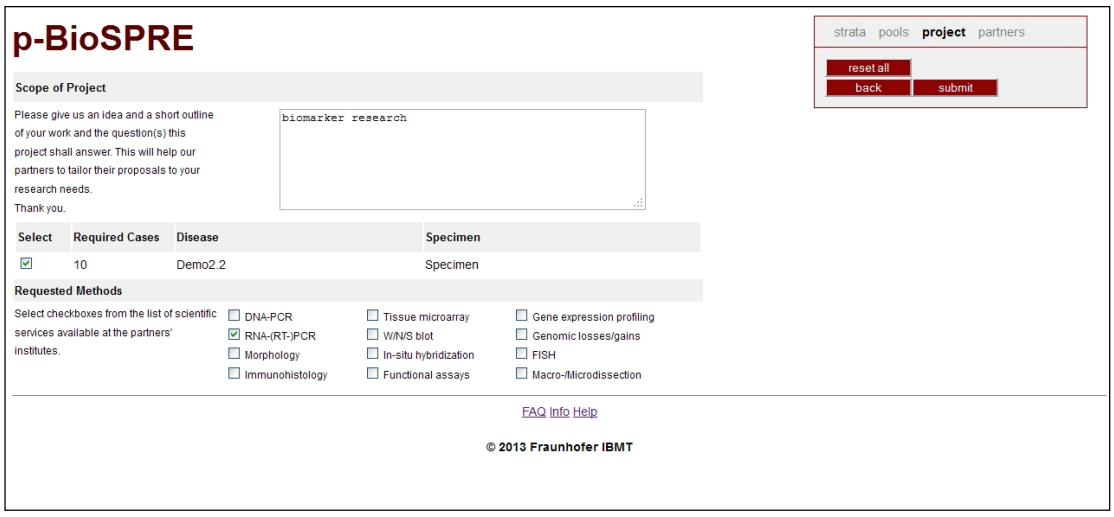


Figure 5: p-BioSPRE Project Request Form:
Submission of the "Project Request Form" generates a file with the search profile that is automatically conveyed to the participating biobank partners. Requested biobanks will run this „input file" on their local In-house Database (Figure 7: button "Project request") and retrieve a list of pseudonymized cases matching the project request.

P-BioSPRE forwards the request as an xml file to the biobank operator enabling him to quickly retrieve the requested material over his/her In-house Database. In other words: The result of the project request tool is a *search profile* that can be applied on the data stored in the In-house Databases, delivering the list of locally available cases matching the user's project request/search criteria. The biobank shall then contact the researcher and decide about the request.

Integration into the p-medicine environment

p-BioSPRE is accessible via the p-medicine portal. p-BioSPRE is integrated in the p-medicine portal in an iframe container (see Deliverable 8.1.2). The application's management of user roles and rights is compliant to the p-medicine security framework. For authentication p-BioSPRE relies on this framework. The security framework provides brokered authentication to services. To access p-BioSPRE a user is redirected to the p-medicine central identity provider, which is responsible for authenticating the user. On successful authentication, the user is redirected to p-BioSPRE with a SAML 2.0[10] identity token, which proves the user's identity[11].

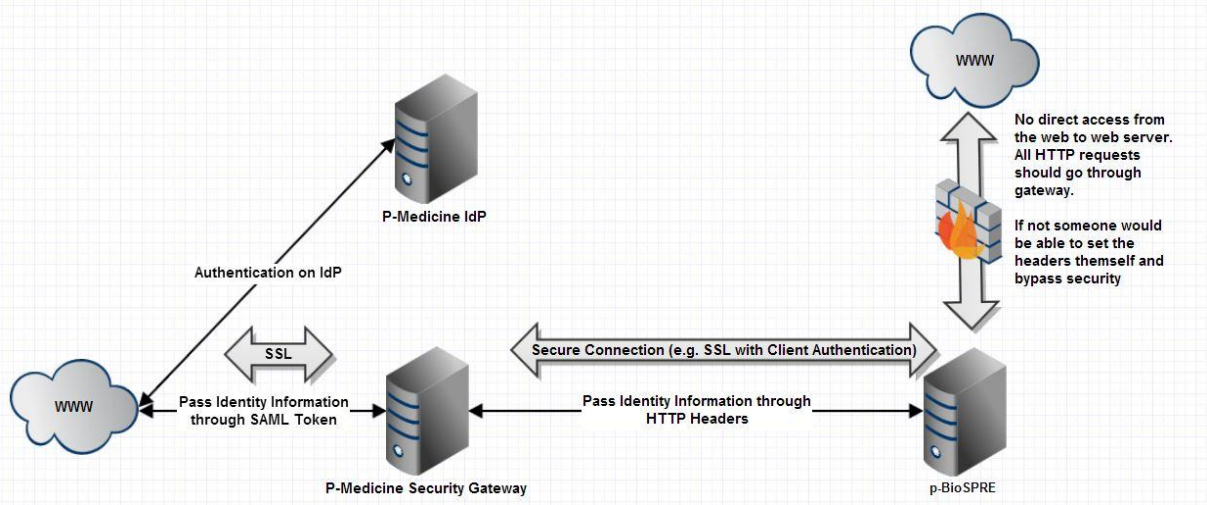


Figure 6: Access to p-BioSPRE via p-medicine security gateway

Integration into the p-medicine security framework implies that p-BioSPRE needs to support the SAML 2.0 standard as defined in Deliverable 3.4[12]. For easy integration and to avoid a tight coupling between p-BioSPRE and SAML, p-medicine provides a security gateway (Figure 6). It is the responsibility of the gateway to handle the SAML protocol. Once a user is successfully authenticated, the gateway will pass all identity related information (e.g. user's name, identifier, roles) to the shielded web application (p-BioSPRE). It is hereby very important that p-BioSPRE is only accessible through the security gateway. To ensure this, p-BioSPRE should only allow connections from the gateway by using SSL client authentication.

[10] Security Assertion Markup Language 2.0 (http://docs.oasis-open.org/security/saml/v2.0/saml-core-2.0-os.pdf).

[11] SAML 2.0 Web Browsers SSO Profile (http://docs.oasis-open.org/security/saml/v2.0/saml-profiles-2.0-os.pdf).

[12] D3.4: Service Integration Guidelines.

5 In-house Database / p-BioBank Wrappers

Biomaterial data is uploaded into p-BioSPRE over so-called p-Biobank Wrappers. Technically, a p-Biobank Wrapper is based on the In-house Database, a component of the CRIP toolbox that comprises a local database, software and web services that can be installed at the biobank's site.

A p-Biobank Wrapper provides biobank operators a comfortable user interface to share their biomaterial and related data in p-BioSPRE (cf. Figure 7).

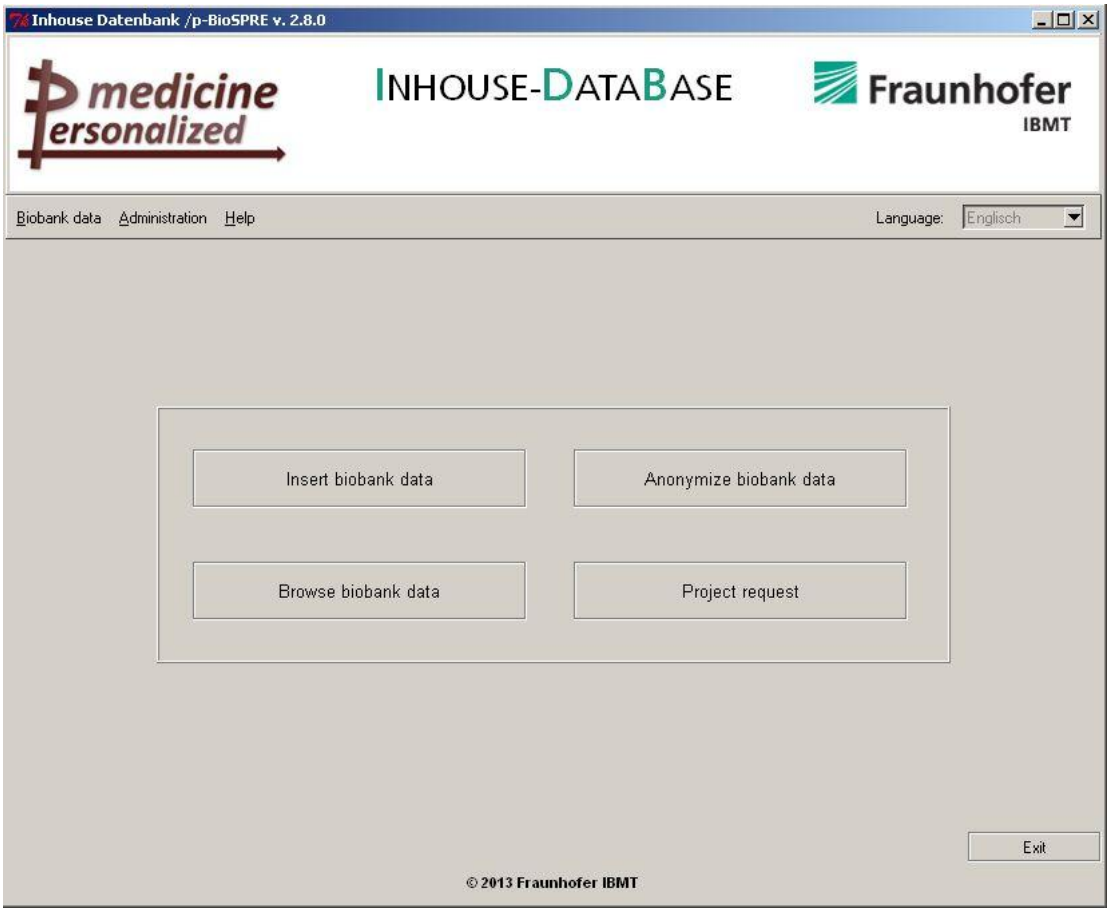


Figure 7: p-Biobank Wrapper User Interface:
Management of In-house Database
Functionalities of buttons are as follows: Insert biobank data: Select and import a data file; Anonymize biobank data: De-identify data and create a file ready for export to p-BioSPRE; Browse biobank data: View cases (e.g. in order to restrict export); Project request: Execute a project request (file) and retrieve cases matching the request.

It imports pseudonymized data on biomaterial and provides a selected, anonymized export to p-BioSPRE for on-line queries (Figure 7, buttons "Insert biobank data" and "Anonymize biobank data"). Upon import, data is harmonised (e.g. by unifying different formats for time stamps or for sample donor's sex (male/female; m/f etc.), and de-identified, e.g. by stripping off identifiers or pseudonyms and by converting a patient's exact age into full years. Further features of the the p-Biobank Wrappers are

- elements of k-anonymization (see below)
- calculation of values (e.g. age of patient from dates of birth and sample preservation, or of body mass index from patient's height and weight)
- metadata repository

Thanks to their meta datarepository (MDR), p-Biobank Wrappers can easily be adapted to the export functionality of the biobank information management system: The MDR just needs to be configured accordingly without any programming efforts.

The user model of the p-Biobank Wrappers provides roles for

- one or more biobank(s)/user(s) (import / export / search data) and
- an administrator (delete data from database / add users / add biobanks).

Data will not be exported automatically to p-BioSPRE, export can only be triggered by the participating biobank operators (cf. Figure 7, button “Anonymize biobank data”). Thus before data will be sent to the p-BioSPRE database, the biobank operator exerts control over and takes responsibility for the data he is going to expose. Upon export, all identifying data will be stripped off, and no personal information will be transferred anyway.

Elements of k-anonymisation

K-anonymity means that attributes are suppressed or generalized until each dataset is identical with at least k-1 other datasets.[13]

Some p-BioSPRE features towards an approach for k-anonymity, such as de-personalizing data by widening the interval that is being processed or displayed for this data and removing personal identifiers, have already been mentioned in the text above: A patient’s exact age is always converted into full years already upon import into the p-Biobank Wrapper. In addition, over the p-BioSPRE Search Tool a minimum of 5-years intervals may be selected for sample donor’s age.

As indicated in the legend of Figure 4, a minimum number of datasets in a pool can be set by default in the Search Tool, and in the In-house Database as well. In the Search tool, this would prevent any pools below a certain limit from being displayed. In the In-house Database, this feature allows implementing specific requirements of local Ethics Review Boards saying that a pool must contain for example a minimum of five cases in order to be processed, safeguarding that projects based on a smaller number of cases cannot be performed.

[13] Sweeney L. Achieving k-anonymity privacy protection using generalization and suppression, International Journal on Uncertainty and Knowledge-based Systems, 10 (5), 2002:571-588.

6 ObTiMA Trial Biomaterial Manager

The p-BioSPRE metabiobank and the p-Biobank Wrappers ensure that biospecimens can be shared in research communities, one of the needs for personalized medicine. However, our requirements analysis revealed that it is also needed that clinical data and biomaterial data can be managed in one information system preferably in a trial management system. To meet this need the Trial Biomaterial Manager has been developed as a third component of the Biobank Access Framework.

It has been developed as a module of the web based trial management system of the p-medicine infrastructure ObTiMA. It enables management of biobanks and associated specimen data in clinical trials and sharing selected specimen data.

The Trial Biomaterial Manager provides users an interface to manage specimen data in clinical trials according to the standard SIOP biomaterial data set. For this purpose a predefined biobanking specimen CRF is provided that can be adjusted to the user's needs. This CRF is stored in the ObTiMA CRF repository.

Furthermore, the Trial Biomaterial Manager enables users to upload their legacy biomaterial data into ObTiMA. Therefore, it provides an import service that enables users to import excel files on existing biomaterial data.

An interface is provided to get an overview about the available biomaterial. It is possible to link clinical data and biomaterial data within clinical trials.

Furthermore, it is possible to select the biomaterial data that the biobank wants to share, and upload the data to p-BioSPRE. In the following, the current functionality of the Trial Biomaterial Manger is described in more detail. In Section 6.1 to 6.5 we describe the core functionality to manage biobanks and samples that is already in a stable state. In Section 6.6 we describe an extension for efficient integration of clinical and biomaterial data that is currently under development. In Section 6.7 we describe the integration of the Trial Biomaterial Manager into the p-medicine environment.

6.1 Managing Biobanks

The Trial Biomaterial Manager provides biobank operators the possibility to create biobanks, edit and view the biobank metadata and manage the samples of the biobanks. Furthermore, biobank operators can assign trials to the biobank and specify and control, which users may access the data of the biobank.

Creating Biobank / Editing Biobank Metadata

To create a new biobank or edit the metadata of an existing biobank, the biobank operator has to select the item "Manage Biobanks" from the ObTiMA main menu. After selecting this item, a list of biobanks for which the user has rights is shown (2, Figure 8).

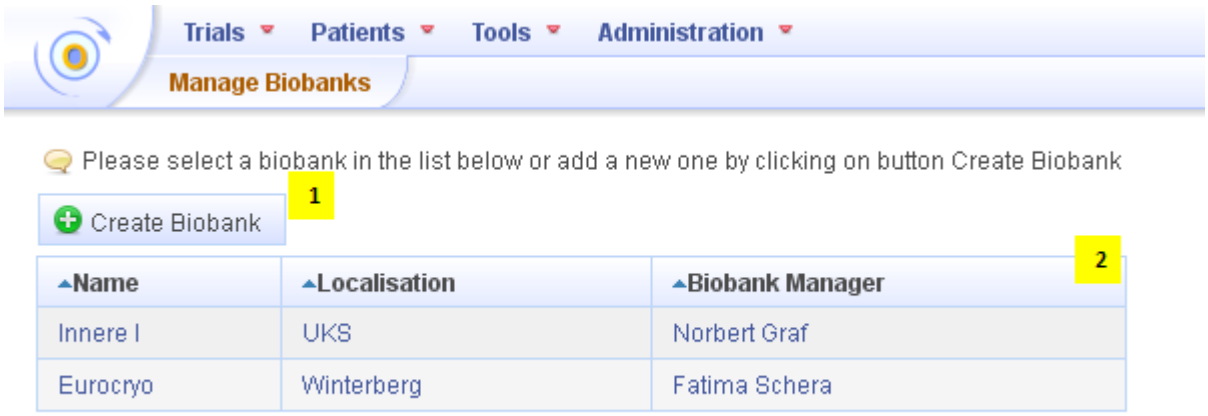


Figure 8: List of Biobanks.

The user can then create a new biobank by pressing the button “Create Biobank” (1, Figure 8) or select a biobank from the list to edit or view its metadata. The metadata of a biobank can be edited in several tabs:

In the tab “Biobank Details” (1, Figure 9) the main metadata for the biobank can be filled in by the user. The biobank details comprise:

- General Information (2, Figure 9) as the name, the biobank operator and the location of the biobank.
- Specification of the provided services as e.g. the specification of the storage cost and if the biobank is used commercially (3, Figure 9).
- Specimen Information as the number of specimens that can be stored and the number of specimens that are stored (4, Figure 9).
- Detailed information: Additional information about the biobank, as e.g. the homepage and the contact mail (5, Figure 9).
- Tissue Information: Information about the tissue types stored in the biobank (6, Figure 9).

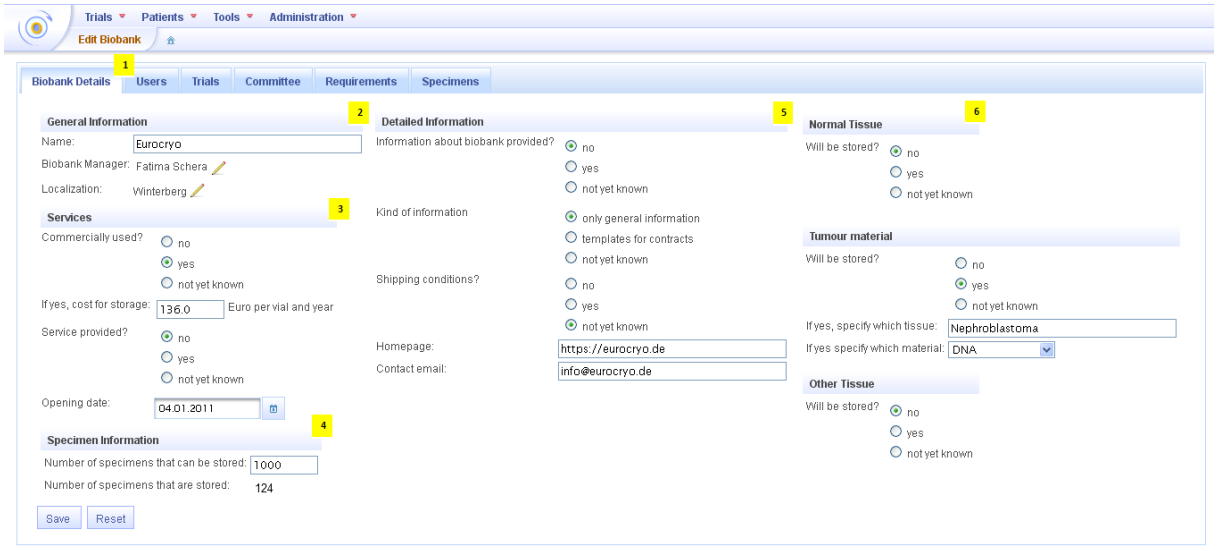


Figure 9: Biobank Details.

In the tab “Committee”, the user can list the members of the biobank committee. He can enter the name, organization, mail and phone number for each committee member (

Figure 10). The members do not need to be users of ObTiMA.

Figure 10: Biobank Committee

In the tab “Requirements”, the requirements for the biobank to participate in the *p-medicine* project can be filled in.

Assigning Trials

It is possible to assign biobanks to trials in order to enable collection of biomaterial for the biobank in the according trial.

To this aim, the biobank can be assigned to a trial in the tab “Trials” (1, Figure 11). A specimen that is created in a trial can only be assigned to a specific biobank if the biobank is assigned to the trial. For assigning a biobank to a trial, the biobank operator needs to select a trial by clicking the button “Add Trial” (2, Figure 11) and select the appropriate trial from the shown list. The trial is then shown in the table “Related Trials” with status “Pending” (3, Figure 11), meaning that the biobank still needs to be confirmed by the trial chairman.

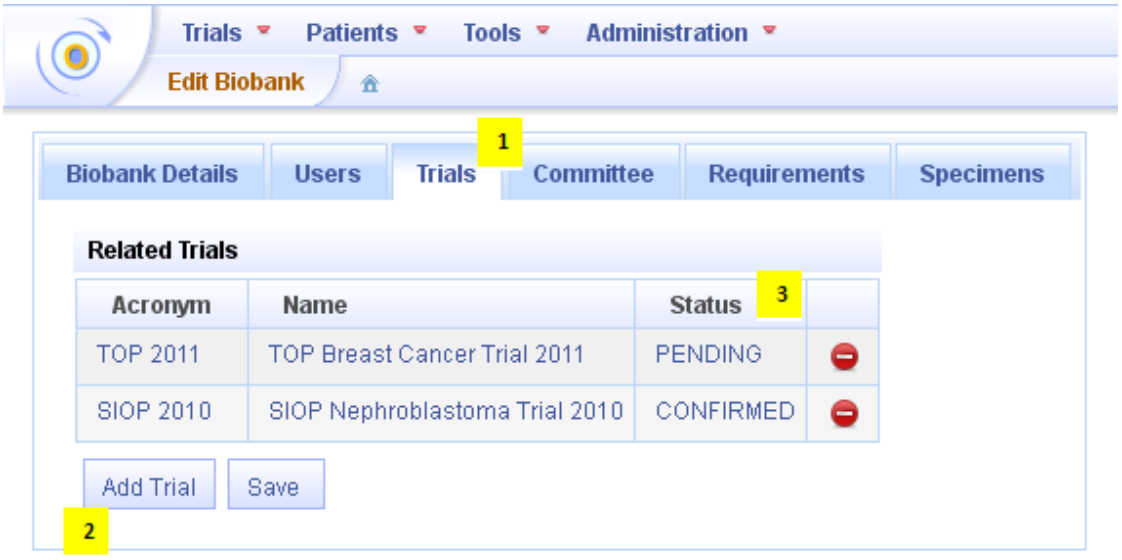


Figure 11: Trials related to a biobank

The trial chairman can confirm a biobank in the tab “Biobanks” in the trial menu after choosing the appropriate trial by pressing the 'plus'-sign (Figure 12).

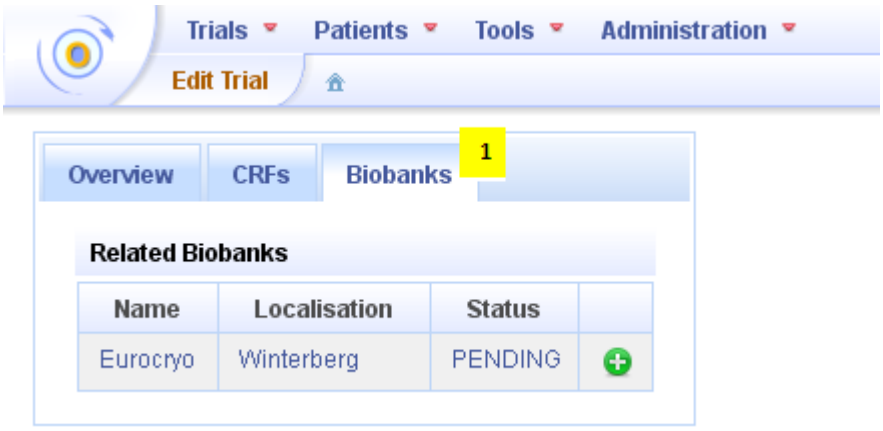


Figure 12: Biobanks related to a trial

User Rights

In the tab “Users” the biobank operator can assign rights for specific ObTiMA users to manage the biobank (1, Figure 13). Firstly, the biobank operator needs to add a user by pressing the button “Add user” (2, Figure 13) and selecting the appropriate user from a list with all ObTiMA users. Each added user has the right to view the metadata of the biobank and to add specimens to a biobank in a trial. The user can only add the specimen if he has the appropriate rights in the trial. Additionally the biobank operator can assign the following rights to the user:

- Search: The user can view the list of specimens that is added to the trial.
- Metadata: The user can change the metadata of the biobank.

Figure 13: Biobank Users

View Specimens

In the tab “Specimens” a user with the appropriate rights can view all specimens that are stored in the biobank (cf. Figure 14). By clicking on a specimen the user is directed to the CRF in the appropriate trial, where the specimen has been created and can view the clinical as well as the specimen data of the patients.

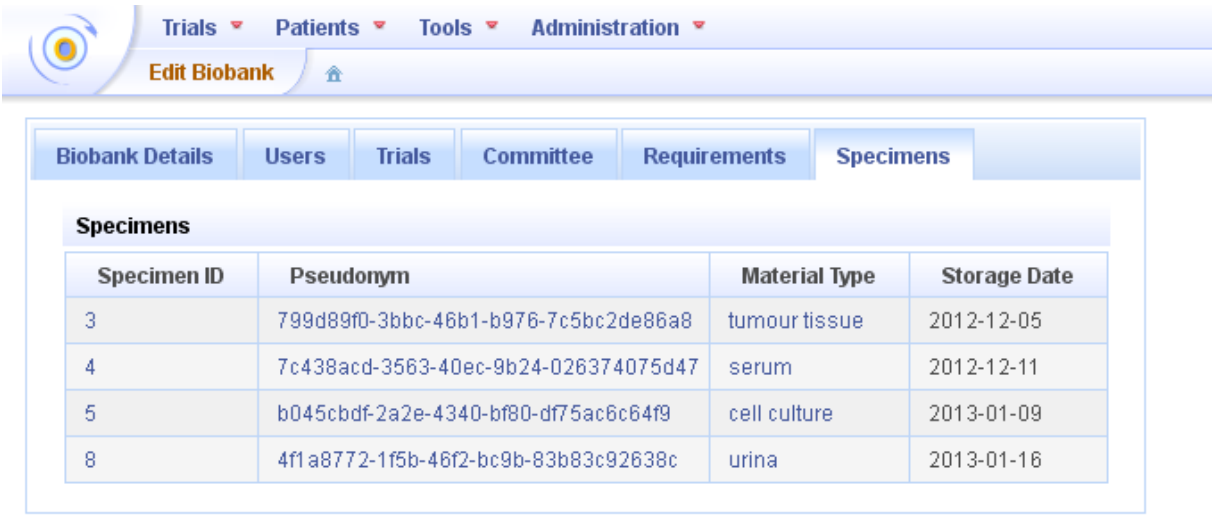


Figure 14: Biobank Specimens

6.2 Creating Specimens and Biobanking Specimen CRF

A specimen can be created by assigning the biobanking specimen CRF to a patient (Figure 15). This CRF is stored in the ObTiMA CRF Repository. On this CRF relevant information describing a biobank sample can be collected on items that are structured into the following sections and subsections:

- General: General information about the specimen as e.g. information about informed consent or storage location.
- Material: Material type as e.g. blood or tumor tissue and depending on the material type items to characterize the material, e.g. the total volume of collected blood or the available volume for material type blood.
    - Transport, Processing, Storage: Quality information about the transport, processing and storage as e.g. the processing method or the freezing temperature.
    - Vials: Information about the available vials, e.g. when the material in the vials was isolated.
        - Analyses: Information about the analyses made with the specimen, e.g. information about the normalization and the used platform.

A precondition to create a specimen in a trial is adding the biobanking specimen CRF to the trial in the design phase and adapt it to the trial's needs. For this purpose, ObTiMA provides a user-friendly interface that allows adding, changing and/or deleting items on a CRF. It is possible to add, change and delete items on the biobanking specimen CRF; however, the items specimen id and biobank id must not be deleted.

When creating a specimen for a patient a specimen id is automatically assigned and filled in the according item. The specimen is associated to a biobank by selecting the according biobank id. In the according item all biobank ids are shown for which the user has rights to add specimens.

The Biobanking Specimen CRF has been created to store samples in ObTiMA. The mechanism to store sample data on CRFs simplifies the integration with clinical data that is as well stored on patient CRFs.

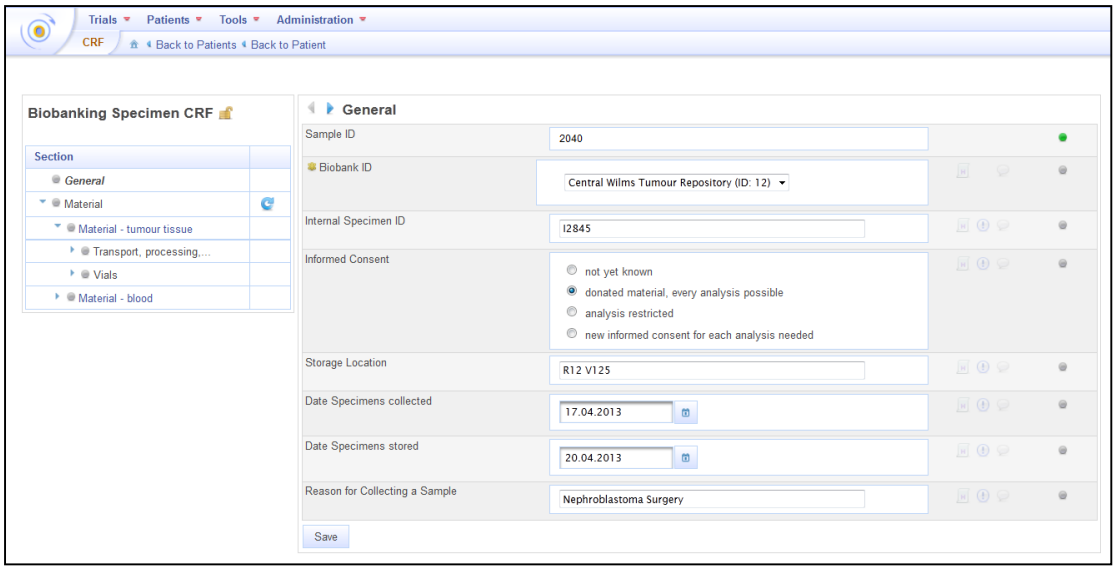


Figure 15: Biobanking Specimen CRF

6.3 Shipping Biomaterial

The Trial Biomaterial Manager supports also the process of shipping biomaterial samples from a study centre to a biobank operator. The biobank operator must not know the identifying data of the patient. Therefore, when shipping the biomaterial the study centre needs to send the pseudonym to the biobank operator to enable matching the referenced patient in ObTiMA.

Barcode labels are used to facilitate handling of the very long pseudonyms. For this purpose, a study centre can print a barcode for the patient and label the specimen for transport to the biobank operator on the page "Patient Details" (1, Figure 16). In turn, the biobank operator can find the referenced patient by simply scanning the barcode (Figure 17). The biobank operator will not be able to see the identifying data of the patient as shown in Figure 16, but only a pseudonymized view.

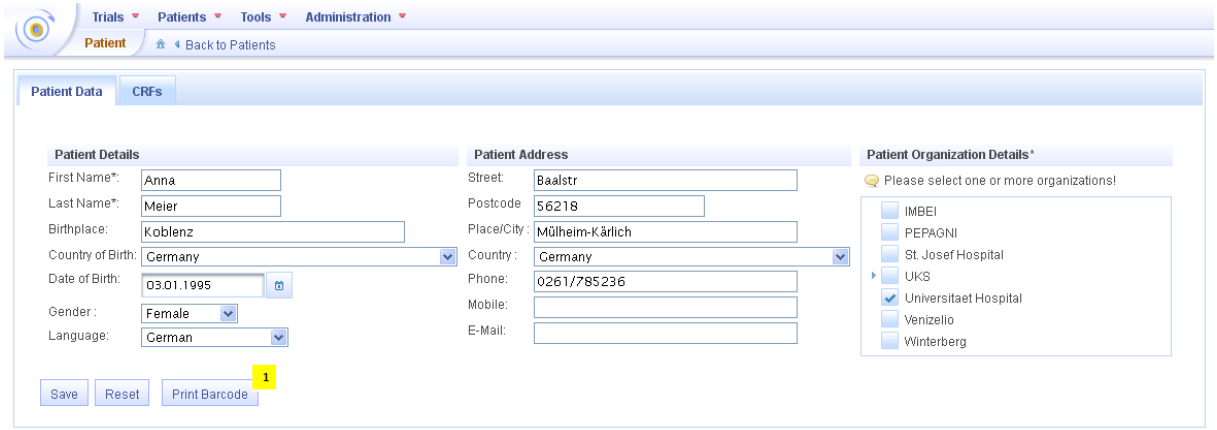


Figure 16: Patient Details

Figure 17: Patient's Barcode

6.4 Import Functionality for SIOP Scenario

For biobank operators, who have already collected biomaterial data an import interface is provided, where biobank operators can import data according to the SIOP biobanking data set. Therefore, we have defined a set of Excel tables (or csv files) with corresponding linked files with analysis results. The format and data types are based on the SIOP biobanking data set, which was also used for designing the Biobanking Specimen CRF.

6.5 Interface to p-BioSPRE

A trial chairman or biobank operator can export the specimen data of a trial to provide it in p-BioSPRE. For this purpose the data firstly needs to be loaded into his p-Biobank Wrapper installation. To this end the user can select a trial and all specimen data stored in the trial and the data is exported pseudonymized into a file in the CDISC ODM format. The user can upload the exported file into his p-Biobank Wrapper installation.

6.6 Search Interface of Trial Biomaterial Manager

For integrating clinical and biobanking data, a search interface as a component of the Trial Biomaterial Manager is currently under development.

The search interface allows to link clinical data and biomaterial data within clinical trials and across trials for further analysis. Trial chairman or biobank operators can search for patient or sample data with individual characteristics. As search criteria each item on a CRF can be used, hence all kind of clinical and biomaterial data can be queried.

Selecting Search Items

To start a search a trial needs to be selected, from which data should be retrieved. When cross trial analysis shall be performed it is possible to select other trials that shall be included in the search in a later step. After selecting the desired trial, a search interface is shown, where further search criteria can be specified. A sketch of the search interface is shown in Figure 18.

As a first step search items need to be selected from the trial that shall be shown in the result table (1, Figure 18). These search items need to be selected from the items on the CRFs in the trial. To select an item first the CRF has to be selected from the list of all CRFs of the trial that is shown in a multi selection item. Then the item group can be selected from a list of all item groups contained in the selected CRF. Last, the item can be selected from a list of all items contained in the item group. Then a name for the search item can be specified that is later shown in the result table.

Specifying Search Constraints

In a second step, constraints for each search item can be specified (2, Figure 18). For numerical items a minimum and maximum value can be specified, for date items a range and for multiple choice items values can be selected from the list of all values for the item.

This can be either done in input fields that are generated from the search items or in a parallel coordinate view[14].

Selecting trials for cross trial analysis

The search items are selected from the trial that had been chosen first. Cross trial analysis are only possible, when other trials that shall be involved into the analysis also contain these items on CRFs. The trial biomaterial manager is capable of detecting those trials automatically with the help of the HDOT ontology. Those trials that contain items with the same ontology annotation as the search items are presented to the user and can be selected to be included into a cross trial analysis (3, Figure 18). When none of the shown trials is selected, then the results are only shown for the trial selected in the first step.

Result Table

When all search items, constraints and trials that shall be included into the analysis are selected and the user presses button “Show Results”, a table with all patients fulfilling the criteria is shown (4, Figure 18). When clicking on a row the CRFs of the according patient are shown and can be inspected in more detail. The patients in this view are all shown pseudonymized, i.e. without identifying information. The table can be exported into a csv file for further use, e.g. analysis.

The search interface can be used to answer queries like: ‘How many blood samples are available from patients between 20-40 years that have tumors with diameters between 1.0-2.0 centimeters?’

[14] Parellel coordinates are a common way of visualizing high-dimensional geometry and analyzing multivariate data.

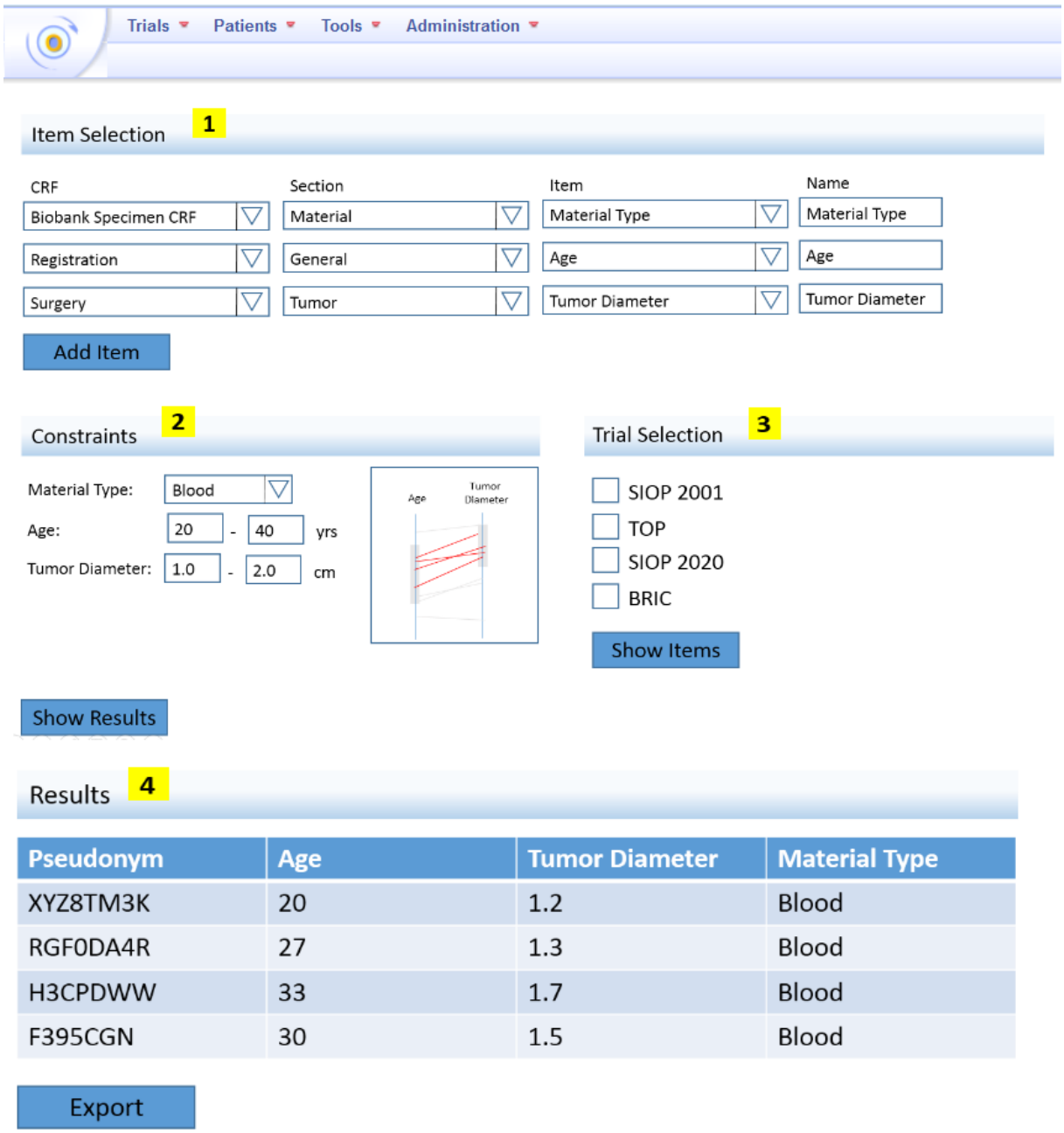


Figure 18: Sketch of search interface for Trial Biomaterial Manager.

6.7 Integration into the p-medicine environment

The Trial Biomaterial Manager is integrated into the p-medicine Portal and has an interface to the p-medicine DWH as described in the following.

Integration into Portal

ObTiMA and therefore the Trial Biomaterial Manager can be accessed via the p-medicine portal. ObTiMA is integrated in the Portal as a web link. The user management of ObTiMA and the portal are harmonized. To access the Trial Biomaterial Manager over the portal it is only once required to log in, that is assured by the single sign-on functionality of the portal.

Interface to Data Warehouse

The Trial Biomaterial Manager is seamlessly integrated with the p-medicine DWH via push services. Via the push services data collected on ontology based CRFs in ObTiMA can be pushed into the p-medicine DWH and be provided in a format compliant to the HDOT ontology. That means the push services can push as well biomaterial data from the Trial Biomaterial Manager as clinical data into the Data Warehouse.

For pushing biomaterial data from the Trial Biomaterial Manager to the DWH a trial chairman has to select the corresponding trial in which the biomaterial is stored and needs to press the button “Push to Data Warehouse” (Figure 19).

Data collected on the CRFs can only be pushed to the p-medicine DWH if the questions have been annotated with the ontology when designing the CRF. Since the biobank specimen CRF is fully annotated with the ontology all biobanking data collected on this CRF can be pushed into the DWH.

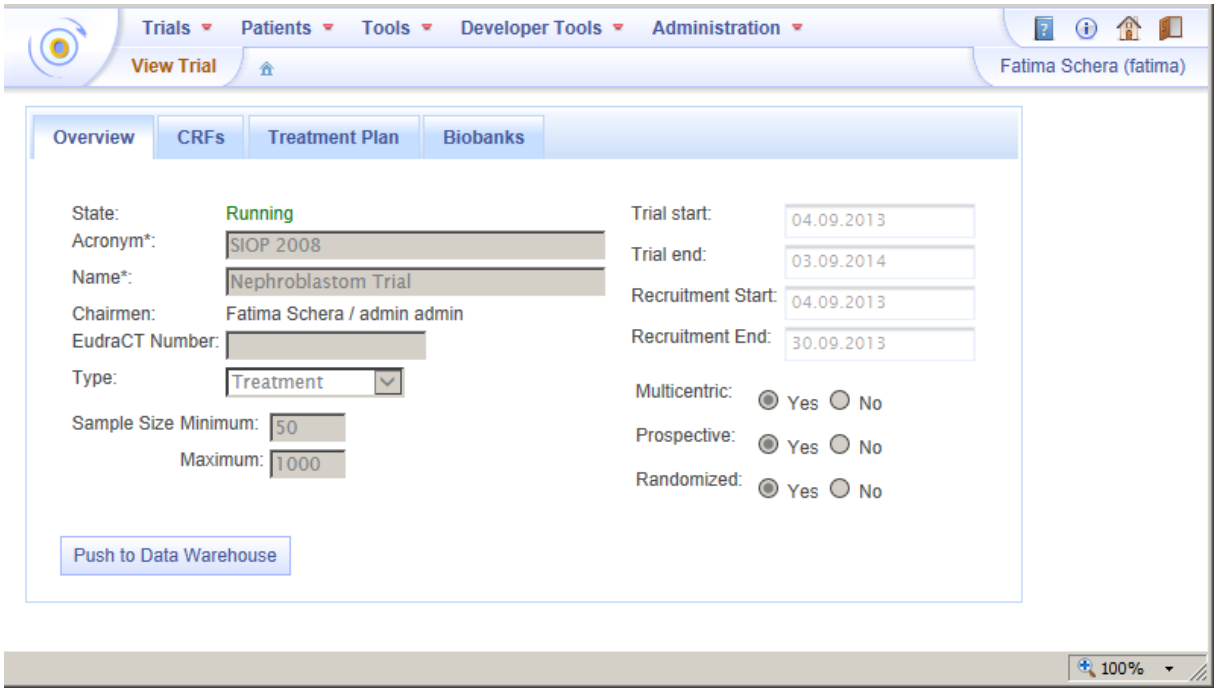


Figure 19: User interface for pushing clinical data of a trial conducted in ObTiMA to Data Warehouse.

The process of pushing data from ObTiMA into the DWH is depicted in Figure 20.

In order to translate the data from the ObTiMA database to a format compliant to HDOT that can be stored in the DWH the Data Translation Services of the p-medicine semantic layer are utilized (see Deliverable 8.5 for a description). These services require two input files as depicted in Figure 20, a “data file” and a “mapping file” for the selected trial.

The “data file” contains the biomaterial and clinical data that will be pushed in form of RDF triples. The “mapping file” contains the ontology-annotations necessary to translate the biomaterial and clinical data into an HDOT compliant format. For pushing data into the DWH, the two files are automatically generated and sent to the Data Translation Services, which translate the data from the ObTiMA database into a format that can be stored into the DWH.

The technical details of the ObTiMA push services are described in D8.5.

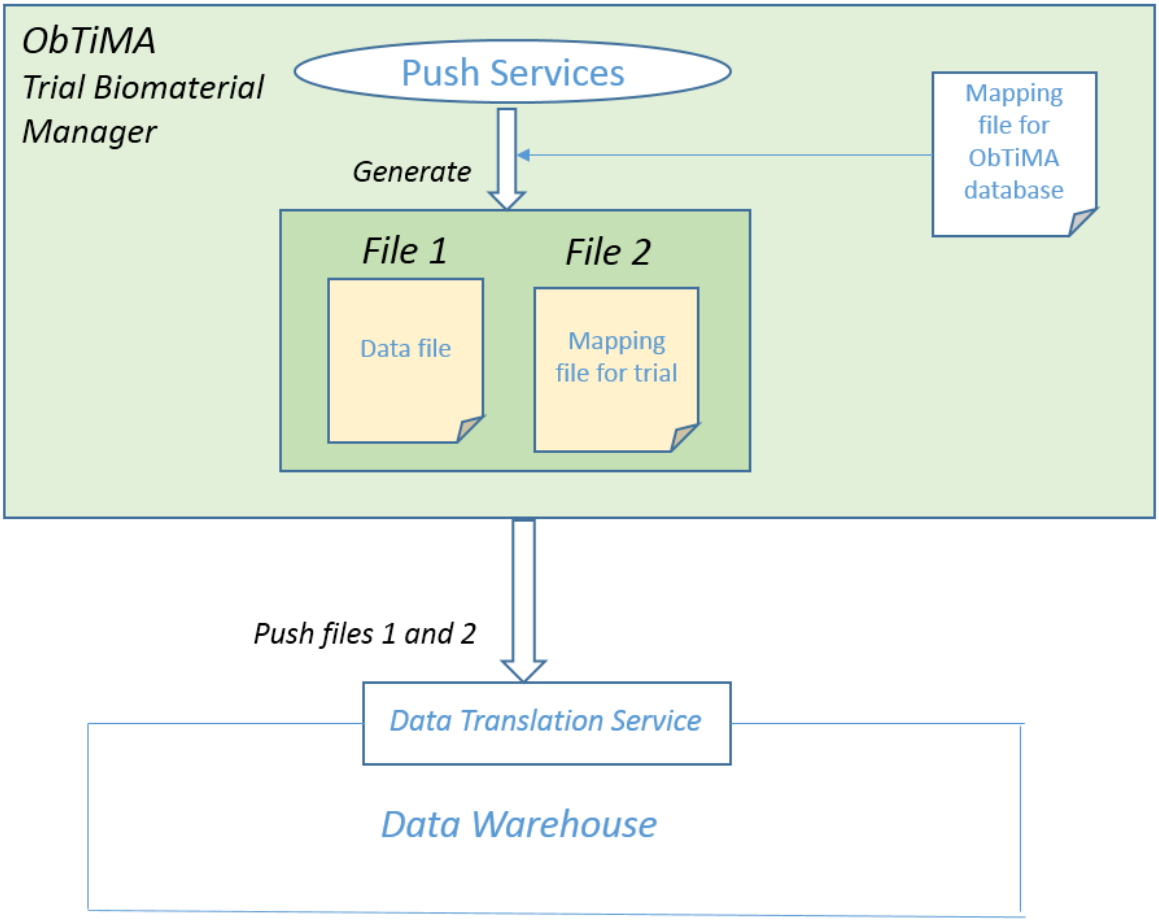


Figure 20: Pushing biomaterial and clinical data from the Trial Biomaterial Manager and ObTiMA to the Data Warehouse.

7 Conclusion

P-medicine's biobank access framework p-BioSPRE is an integrated biobank solution to support research in personalised medicine. It enables and facilitates secure trans-institutional and transnational access to existing biobanks, but also offering biomaterial collections to research communities. Biobank specimens can be managed using the ObTiMA Trial Biomaterial Manager. Along with biobank data, information on patients' informed consent is also processed and displayed. The development of p-BioSPRE is driven by two user scenarios on ALL and Wilms tumor, in order to seamlessly adapt the framework to the needs of integrated biobanking.

The Biobank Access Framework is integrated into the p-medicine platform, enabling that biobank data can be integrated with heterogeneous biomedical information, as e.g. clinical or microarray data. Biobank data from the Trial Biomaterial Manager can be pushed into the DWH in a format compliant with the p-medicine semantic integration framework for further analysis. Furthermore, the framework is secured by the p-medicine security framework. The different components of the Biobank Access Framework can be accessed in the p-medicine portal.

An initial version of the Biobank Access Framework is available on a test server of the p-medicine Portal and can be evaluated by interested and authorised research communities under the URL *https://pmedportal.ibmt.fraunhofer.de*.

Currently the Biobank Access Framework is under evaluation by the SIOP and the ALL study groups. The results of this evaluation will be reported in Deliverable 10.4. Components of the framework will then be further adapted to user needs and improved with additional features. Based on the results of the evaluation the Biobank Access Framework will be further improved.

In the following, several possible improvements for the Trial Biomaterial Manager are described. In the future, the process "Shipping Biomaterial" will be extended. With this functionality the trial chairman will be supported to ship biomaterial. He will be enabled to add a biobank sample CRF to the patient, where the trial chairman can specify sample details and select the biobank to which the sample will be shipped. From this data a letter for the receiving biobank operator including the patient's barcode is generated automatically and can be printed from the trial chairman. The biobank operator, who receives the sample, can then find the according patient by scanning the barcode and fill in more details into the biobank sample CRF that was already created by the trial chairman.

A reporting functionality will be developed for the Trial Biomaterial Manager. That means a biobank operator may generate a report about the movements in his biobank. The report will show administrative data as e.g. which samples were stored or removed or which samples were send to whom, for which research purpose.

Appendix 1 - Abbreviations and acronyms

BMI	Body Mass Index
DWH	Data Warehouse
FISH	Fluorescence in situ hybridization
ICD	International Classification of Diseases
IHC	Immunohistochemistry
IDB	In-house Database
ObTiMA	Ontology based Trial Management Application
PCR	Polymerase Chain Reaction
p-BioSPRE	p-medicine Biomaterial Search and Project Request Engine

Appendix 2 – Minimum data set

In the following table the minimum data set (in grey) and metadata (in alphabetical order) of p-BioSPRE is shown. Minimum data include: diagnosis_class (standard used, e.g. ICD-O or ICD-10); diagnosis_code (actual code, e.g.C64 for Malignant neoplasms of kidney); groupofspecimen (e.g. blood, tissue); ppsn (patient); psn (case). Note: A patient can constitute several cases, and a case can be backed by several specimen types and aliquots.

Description	**Name**	**Type**
Age of donor*	age	int
Biobank acronym	biobank	text
Name of contact person	biobank_contact	text
e-mail address of contact person	biobank_email	text
Name of biobank	biobank_name	text
	blasts_bm	text
	blasts_pb	text
	crlf2	text
	cytological_response	text
Date of birth (full years only)	dateofbirth	date
Date of sample preservation	dateofsamplepreparation	date
Diagnosis	diagnosis	text
Class of code	diagnosis_class	text
Code of diagnosis	diagnosis_code	text
	epigenetic_platform	text
	epigenetic_profiling	text
	exom_sequencing_platform	text
	extramedullary_compartment	text
	gene_expression_platform	text
	gene_expression_profiling	text
	genetic_subtype	text
	genome_sequencing_platform	text
Group of material	groupofspecimen	text
	ikzf1	text

	immunophenotype	text
	labeldateofbiopsy	text
	leukocytes	text
	localization	text
	localization_class	text
	localization_code	text
	micro_rna_platform	text
	micro_rna_profiling	text
	mlpa_kit	text
	mlpa_screening	text
	mrd_response	text
	notch1	text
Patient	ppsn	text
	prednison_response	text
Case	psn	text
Sex of donor	sex	text
	snp_diagnose	text
	snp_platform_diagnose	text
	snp_platform_remission	text
	snp_remission	text
	stage	text
	stratification	text
	tp53_diagnose	text
	tp53_remission	text
	tpmt	text
	trial	text
	trial_follow_up	text
Material type	typeofspecimen	text

	whole_exome_seq	text
	whole_genome_seq	text

*calculated from year of birth and year of sample preservation