Deliverable No. 10.1

Analysis report about existing tools, platforms, and initiatives for integrated biobanking

Grant Agreement No.:	270089
Deliverable No.:	D10.1
Deliverable Name:	Analysis report about existing tools, platforms, and initiatives for integrated biobanking
Contractual Submission Date:	31/10/2011
Actual Submission Date:	28/10/2011

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 10.1
Document name:	Analysis report about existing tools, platforms, and initiatives for integrated biobanking
Nature (R, P, D, O)[1]	R
Dissemination Level (PU, PP, RE, CO)[2]	PU
Version:	1.00
Actual Submission Date:	28/10/2011
Editor: Institution: E-Mail:	Stephan Kiefer Fraunhofer Institute Biomedical Engineering stephan.kiefer@ibmt.fraunhofer.de

ABSTRACT:

This deliverable lays the technical and legal foundation for p-medicine's Biobank Access Framework. It investigates on one hand the state-of-the art in biobank integration by describing the relevant international activities in the domain and by analysing existing tools for integrated biobanking, but also related legal, ethical and social issues. On the other hand the scenarios of the use case owners in p-medicine are further described, required data sets are listed and an initial architecture for Biobank Access Framework is proposed together with a corresponding legal framework for biomaterial sharing within the p-medicine platform.

KEYWORD LIST: **Biobanking, federated biobanks, biomaterial sharing, sample management, cross-border sample transfer, biobank access**

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 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	30/06/11	Draft - structure	Stephan Kiefer
0.2	31/08/11	Draft with chapters 3 and 6	Stephan Kiefer, Kerstin Rohm, Nikolaus Forgo, Tina Krügel ,Hartwig Geringer
0.5	16/09/11	Next draft	Stephan Kiefer, Kerstin Rohm, Christina Schröder, Gabriele Weiler, Patrick Dörr Nikolaus Forgo, Tina Krügel ,Hartwig Geringer Vangelis Kritsotakis, FORTH
0.6	30/09/11	First draft for circulation among WP10 partner	Stephan Kiefer, Christina Schröder
0.8	11/10/11	Second draft for circulation among WP10 partner	Stephan Kiefer, Nikolaus Forgo, Martin Zimmermann
0.81	12/10/11	Norbert's contribution added.	Stephan Kiefer, Norbert Graf
0.90	21/10/11	Chapter 1 and 4 added.	Giuseppina La Bonizzi (Chapter 4), Gordon McVie (Chapter 4 & 1), Christina Schröder (Chapter 1)
1.00	28/10/11	Finalised.	Final version after revision by Holger Stenzhorn

List of contributors

- Stephan Kiefer, Kerstin Rohm, Christina Schröder, Gabriele Weiler, Patrick Dörr, Fraunhofer-Institut Biomedizinische Technik
- Nikolaus Forgo, Tina Krügel, Hartwig Gerhartinger, Gottfried Wilhelm Leibniz Universität Hannover
- Vangelis Kritsotakis, FORTH
- Martin Zimmermann, Martin Stanulla, Christian-Albrechts-Universität zu Kiel
- Norbert Graf, Universitätsklinikum des Saarlandes
- Giuseppina La Bonizzi, Gordon McVie, Istituto Europeo Di Oncologia SRL

Contents

1 Executive Summary

p-medicine's mission is to foster and support the advent of personalized medicine for patients, and an integrated biobanking solution is critical to its success. The explosion of molecular medicine and molecular imaging in cancer has led to an exponential increase in the possibilities for individualised cancer care. But the accompanying complexity of the resulting cellular biomarkers, their application in radiotherapy planning and targeted drug choice has made the knowledge base too great for physicians to retain, and certainly too complex for patients to understand. Many more new molecules will be designed in the time course of this project and their validation as useful biomarker driven treatments will require fast reference to the global collection of carefully, (and molecularly) characterised stored tissue samples. An access system as envisaged in p-medicine could cut drug development time from lab to patient by at least five years.

Within its overall ICT and contractual architecture, the planned Biobank Access Framework and Data Warehouse are foreseen to complement and synergise with each other to generate and contribute critical VPH knowledge. This analytical report has initiated the design process of the Biobank Access Framework encompassing both technical and legal issues. It contains a screen of the continuously evolving and differentiating spectrum of biobank integration and standardization initiatives throughout the world. It is organised in several sections:

1) Two specific p-medicine use cases, Wilm's tumour in children and acute and recurrent lymphocytic leukaemia, as well as a formal scenario for biobank access were introduced as cornerstones for the endeavour (chapter 3).
2) As background and reference information, an overview of international and interconnecting initiatives for integrated biobanking is given (chapter 4).
3) Preparing the development and implementation of the Biobank Access Framework, existing ICT tools for integrated biobanking were screened and evaluated in terms of their eligibility and fitness for p-medicine's purpose (chapter 5).
4) As a prerequisite for the Framework's legal and contractual integrity, relevant legal and ethical rules and guidelines were analyzed in detail (chapter 6) and compiled as a set of practical recommendations (6.4).
5) Data standardization and harmonization tools as well as relevant core data sets required to start work were collated and presented (chapter 7).
6) Finally, p-medicine's approach to the Biobank Access Framework from both technical and legal perspectives was documented.(chapter 8). The basic architecture of the p-medicine Biomaterial Search and Project Request Engine or metabiobank "p-BioSPRE" is outlined. Cooperation between p-medicine biobank operators and the p-BioSPRE metabiobank operator relating to data protection issues will be regulated by contract, embedded in the overall governance of p-medicine. For access to p-medicine resources by researchers and third parties, clear access rules complying with all relevant ethical and legal standards will be set up.

Given the rapid progress of research in the biobanking field, even an elaborate analysis such as this document can only provide a snapshot of the state-of-the art in biobank integration. Therefore the field will continue to be monitored and the project's development will be adapted technically and contractually throughout the remaining three years.

2 Introduction and project background

Biobanks, sometimes called biorepositories or biomaterial banks represent key resources for clinico-genomic research and advances in personalized medicine. Through the increasing use of molecular and genetic factors in the study of disease causes and targeted individualised therapies biobanks have gained growing importance as part of the scientific infrastructure. The discovery of critical genes and pathways as well as the analysis of their impact and significance will critically depend on sufficient access to biomaterial and related information. In consequence, there is a growing interest in integrating biomaterial repositories into larger infrastructures in order to satisfy research communities' need to specific, quality-assessed samples and up-to-date sample related data with integrated (or at least the possibility to link to) clinical and epidemiological information. A prominent initiative in this direction represents the European Research Infrastructure project BBMRI (Biobanking and Biomolecular Resources Research Infrastructure) whose preparatory phase came to its end in January 2011.

p-medicine will support with its ICT infrastructure and tools researchers growing demands to access and to share high quality biomaterial and related data for their research projects. In Work Package 10 we aim at developing an integrated service framework within the p-medicine platform to enhance and simplify access to existing biobanks but also to make own biomaterial collections available to larger research communities. WP10 addresses the question on how to integrate existing human biomaterial repositories in such an ICT infrastructure and in particular those biomaterial collections where corresponding data about the clinical cases can be made available in *p-medicine's* data warehouse. If a researcher wants to test for example a new biomarker in a specific disease, he can ask what data are stored in the data warehouse and what biomaterial is available from corresponding patients. This will allow him under the legal framework to run new analyses in the lab, to correlate them with the corresponding data and to include the findings in new VPH models. Such an approach will gain new research results much faster and will drive medicine into personalized medicine.

In this context we speak about 'integrated biobanking' in this document which means the integration of heterogeneous, federated biobanks or biomaterial repository under one search engine for the centralized access to samples and sample related data within the p-medicine infrastructure and under good clinical and legal practice.

With this document we want to lay the foundation for p-medicine's biobank access framework including the legal and ethical perspective.

3 Analysis of p-medicine use cases for biobank access

The technical developments of this work package towards a biobank access framework for p-medicine are mainly driven by use scenarios provided by p-medicine partners Universität des Saarlandes (USAAR) in the context of the SIOP Wilms Tumor trials and Christian-Albrecht-Universität zu Kiel (CAU) in the context of clinical research and trials related to lymphoblastic leukaemia. The scenarios are briefly summarized in the following two subchapters. As a result four general use cases of p-medicine's biobank access framework have been defined and contributed to Deliverable D2.2 "Definition of scenarios and use cases and report on scenario based user needs and requirements". These use cases are further summarized and analysed in subchapter 3.3.

3.1 SIOP Wilms tumor biobanking scenario

During the runtime of the SIOP Wilms tumor trials biological material (tumor, normal kidney, blood samples) is collected by the participating centres. Fresh tumour material and nomal kidney is shock frozen and also fixated and embedded in paraffin, From some tumours vital material is preserved for cell culture. All the biomaterial is sent to central labs in the participating country for further processing and analysis and for long-term preservation. The Biocenter of the University Würzburg is the central facility in Germany to preserve this biomaterial analyzed and made available for further research to SIOP partners. Standardized procedures and ethical approval are in place how requests to material are handled. A committee of the SIOP Wilms tumor study group decides about material request. Among others, the approval of the committee depends on the research that the requester wants to conduct with the biomaterial, whether similar research has already been conducted or not, the availability of material and so on. Childhood Wilms tumor is a rare disease. Around 100 cases in Germany are reported annually (for paediatric leukaemia 550 cases) and, as almost all of the patients are treated in the prospective SIOP trial, their biomaterial is stored in the Wilms tumor repository by the Biocenter. The tumor tissue received from the different pathology institutes of the clinical partners is shipped on dry ice and varies in quality on arrival. A quality control system based on SOPs with observable quality variables is not implemented yet (i.e. time from surgery to asservation; cooling chain observation, etc) beside the visual test whether dry ice can still be found in the shipping box on arrival.

Data about the material is captured in several Excel files, which fulfil the needs of researchers regarding flexibility (i.e. new data fields) and further data analysis options (i.e. data export). Additional data is created in further experiments using tumor material, which include microarray analysis, LOH (loss of heterozygosity) and others. The data are of particular interest for the clinical researchers, and are not only relevant for retrospective studies but also in some cases in routine clinical care. A so-called SIOP number is the identifier of the biomaterial that also links the material to the clinical data of a patient in the SIOP trial. LOH analysis data is captured in a simple Excel file, while micro array data are more complex. There standardization is important and depends on the platform used and the SOPs of the lab.

The primary goal of the clinicians involved in SIOP Wilms tumor research is to combine the clinical data of the patient with molecular data in order to find new biomarkers or stratification markers for upcoming trials. This will allow applying better treatments with less side effects to patients. In summary there are four important issues of this biobank scenario:

1. to access data from the central Wilms tumor repository in Würzburg and link them with clinical data.

2. to check if there is biomaterial available for further research, after getting such a request from any researcher. This needs approval by the respective SIOP committee. The approval will depend on the experience of the researcher, the scientific question, the ethical approval, the availability of material, the knowledge that this kind of research will generate, and financial issues.
3. to share data and biomaterial cross borders. Prof. Norbert Graf and Prof. Manfred Gessler, responsible for the German Wilms tumor repository at the Biocenter Würzburg, consider this as relevant. As trials in North America follow a different treatment approach (primary surgery) comparison of biomaterial between these approaches would add significantly to the knowledge about Wilms tumour.
4. to collect through the p-medicine biobank access framework information about the planned research with samples and to link this information to the remaining aliquots on stock. Ultimately, information of the outcome of such research should be linked to the remaining stock of aliquots as well. According to resource limitations of p-medicine this is out of the scope of WP10.

The strategy for access to the biobank in p-medicine regarding the SIOP Wilms tumor biobank scenario is

- To demonstrate the feasibility and advantage of legally integrating biomaterial data with clinical data and sharing data and biomaterial within the German SIOP study group through the p-medicine infrastructure and its biobank access framework to be developed in WP10,
- To establish this technical approach as the standard in the SIOP Wilms tumor community,
- To transfer this approach to other SIOP countries.

In consequence the operator of the German Wilms tumor repository requests from the Biobank Access Framework of p-medicine the following, that allows

i) To upload/export existing Excel based specimen data sheets to the p-medicine infrastructure for further searching and processing by SIOP researchers

ii) To upload/export also the analytical/research data, in particular LOH data and microarray data

iii) To flexibly manage the German biomaterial repository through the WP10 Biobank access components instead of filling and exporting Excel tables.

Through the Biobank Access Framework researchers of the German Wilms tumor study group want

i) To search and view all biomaterial data and

ii) To combine the biomaterial data with clinical data from ObTiMA, the data warehouse, or from any other external data base in particular the KEGG Pathway Database,

iii) To request the availability of Wilms tumor material for a certain research question

Regarding aspects of pseudonymization/anonymization the Biobank Access Framework has to follow the privacy policy of the p-medicine infrastructure in all aspects.

3.2 Lymphoblastic Leukaemia biobanking scenario

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 of the Christian-Albrecht-Universität zu Kiel (CAU). Annually, there are about 500-550 new cases of minor ALL patients reported in Germany.

Clinical data, biobanking data and samples for German ALL-BFM 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).

Local clinics document ALL-BFM cases paper-based. The CRFs summarize approximately 600 parameters per patient. The entered document is sent to the study centre in Kiel, where all data is entered manually into the data base management system.

When minor ALL patients, which were treated in an ALL-BFM study, have a relapse, usually are treated according to ALL-REZ BFM studies, which are coordinated by the university hospital Charité in Berlin. Annually, around 60 relapses occur in Germany. Biobanking data and samples for German relapse patients are collected in the Charité study centre in Berlin. The data is collected in a web-based biobanking management system that was tailored for the Charité based on the universal database application platform Scopeland[3]. Clinical data will be collected for the next trial (IntReALL 2010) in the trial management system Marvin. The ALL-BFM and ALL-REZ BFM study groups collaborate with several other study groups in Europe. Each study group has own solutions to store clinical and biobanking data. The trial IntReALL 2010 will start at the beginning of 2012. It is intended that in all partners, who don't possess qualified solutions for biobanking, can use the customized Scopeland system.

Currently, ALL partners can only access clinical data and biomaterial data of their own patients that are stored in their own databases. In particular, it is not possible for the ALL-REZ BMF study centre in Berlin to get automated access for their patients with relapses to biomaterial and relevant analytical and clinical data from the acute phase in the study centre in Kiel. Regarding European and world wide collaborations of CAU for research on Lymphoblastic Leukaemia it is wished to access and share biomaterial and related analytical data within the respective international research collaboration. Relevant collaborations are the study AIEOP-BFM ALL 2009 with biobanks from Austria, Australia, Czech Republic, Germany, Israel, Italy and Switzerland as well as the EC funded relapse study IntReALL 2010 with partners from UK, Germany, Australia, Austria, Czechia, France, Israel, Italy, Japan, Netherlands, Sweden, Switzerland and Poland.

The strategy for biobank access in p-medicine regarding the Lymphoblastic Leukaemia biobanking scenario is

- To demonstrate feasibility of the access to biomaterial and related data of patients collected in different trials from their acute and recurrent phase for the study centres Kiel and Berlin,
- To show advantage of integrating and accessing biomaterial data provided by collaborating international partners through their local bio-repository management systems for international research collaborations on lymphoblastic leukaemia. For this purpose, initially one or two biobanks of international ALL research partners of CAU shall be integrated under p-medicine's biobank access framework.
- To demonstrate access to such samples under a model based legal framework which ensures legally correct sample and data transfer between partners.

[3] Scopeland is a universal database application platform of the German SME Scopeland Technology GmbH, Berlin, http://www.scopeland.de

3.3 Formal use cases for biobank access in p-medicine

The main purpose of p-medicine's biobank access framework is to get access to biomaterial with specific characteristics that matches the intended research purpose or from the perspective of a biobank operator to make the own biomaterial stock and characterising data available to the research community. In consequence the p-medicine biobank access framework will support this main intended use of biomaterial sharing while aspects like data sharing and linkage of different data resources is covered by the data warehouse and its corresponding data push services and data annotation resources. For the Biobank Access Framework in total four main use cases have been distinguished in Deliverable 2.2. They mainly cover different tasks and user perspectives when using p-medicine's Biobank Access Framework in addition to a general integration use case for existing biomaterial repositories. The following table gives a description of each use case.

Identifier	Name	Description
BA_1	Integration of biomaterial data repositories	A user wants to link his own biomaterial data repository to the p-medicine biobank access framework in order to share biomaterial and related data with his research community as further described in BA_3 to BA_4.
BA_2	Managing biomaterial data in ObTiMA	A user collects biomaterial in a clinical trial, conducted with ObTiMA within p-medicine environment. 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 sharing of data and material within the trial community.
BA_3	Offering human biomaterial to a closed and/or open clinical research community for research	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).
BA_4	Requesting specific human biomaterial within a closed and/or open clinical research community for research	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 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.

Table 1: Use cases for biobank access in p-medicine.

4 International initiatives for integrated biobanking

4.1 BBMRI - Biobanking and Biomolecular Resources Research Infrastructure

During the Preparatory Phase 2007 – 2010, BBMRI announced their overall mission "to prepare for the construction of a pan-European Biobanking and Biomolecular Resources Research Infrastructure (BBMRI)[4] for biomedical and biological research in Europe and worldwide, building on existing infrastructures, resources and technologies, specifically complemented with innovative components and properly embedded into European ethical, legal and societal frameworks."At present BBMRI is undergoing a transition phase: Whilst the Preparatory Phase has been terminated in February 2011, the envisioned Construction Phase will be implemented under the ERIC (European Research Infrastructure Consortium) legal entity. ERIC has been presented to Member States of the European Union for approval and funding and will start, at the earliest, by the middle of 2012. Meanwhile, multiple national hubs foreseen to compose the BBMRI network are under construction in the Member States (e.g. France, Germany, Ireland, Italy, Netherlands, Sweden).

This reflects BBMRI's broad impact and success in promoting and structuring the European biobanking landscape, community, and practices. Notwithstanding the open questions of implementation, funding and financing, particularly BBMRI's WP 5 on ICT (cf. chapter 5.3) has revealed important models and standards to coin future infrastructure. Nevertheless, although they all will strive to adhere to WP 5's recommendations, the emerging National Hubs' ICT infrastructures are currently lacking binding central European coordination of efforts. It is up to p-medicine and its biobanks to identify and define their aims, role and position in this evolving federated European biobanking system.

4.2 ISBER – International Society for Biological and Environmental Repositories

ISBER, the International Society for Biological and Environmental Repositories, based in Maryland, USA, is an international forum that addresses the technical, legal, ethical, and managerial issues relevant to biological environmental repositories. ISBER was founded in 1999 as an educational forum for discussion of repository management and dissemination of information on operational issues in order to ensure that specimen collections are available for study, as new biomarkers emerge and more sensitive measurement technologies become pertinent. ISBER has around 850 institutional and individual members worldwide, and is a Division of the American Society for Investigative Pathology (ASIP), a nonprofit educational organization.[5]

Actually 10 different working groups of ISBER contributes to international biobank issues. They comprise 'Automated Repositories', 'Biospecimen Science', 'Clinical Biobanking', 'Environmental Biospecimen', 'Informatics', 'Informed Consent Procedures for the Collection of Biospecimens', 'International Repository Locator', 'Pharma-Academia', 'Rare Diseases', and 'Rights to and Control of Human Tissue Samples'. The ISBER Informatics Working Group includes representatives from ISBER member vendors who provide custom built and commercial off-the-shelf systems and ISBER members who are involved in informatics within a repository. The group aims at developing best practices for information management systems to support biobanking including decision making tools for selection of home-grown,

[4] http://www.bbmri.eu/

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

custom built, or off the shelf system, appropriate use of an information management system in the biobank, sample labelling as well as federal guidelines as they pertain to information management systems in the biobank. In addition, the group develops self-evaluation tools for providers of information management systems that support biobanking.

ISBER also offers a market place for its vendors that provide services and products to biorepositories. Members have access to ISBER's digital library that makes documents of the working groups, strategic plans and speaker presentations from meetings available. ISBER has published best practices for repositories, a self assessment tool for repositories and the regular journal 'Biopreservation and Biobanking'. ISBER also maintains a registry of any kind of biological repositories.

ISBER has also launched a biorepository proficiency testing program in collaboration with IBBL, the Integrated BioBank of Luxembourg. The IBBL is a biorepository and technology centre that serves Luxembourg and its partners to collect, store and redistribute biospecimens and their related clinical data, and produces analytes suitable for analyses by state-of-the-art genomics and proteomics platforms. IBBL works in partnership with ISBER, and is in charge of preparing and shipping the testing materials to all proficiency testing participants. Proficiency Testing (PT) as defined in ISO/IEC 17043:2010 (the International Standard on "Conformity assessment – General requirements for proficiency testing"), is seen as a powerful tool to help laboratories/repositories demonstrate their competence in biospecimen characterization to researchers, industry, or accreditation bodies. PT enables laboratories/repositories to monitor their quality control tests over time, identify longer term trends and consider any necessary corrective actions. The scope of this program is to develop, coordinate and implement PT Programs for quality control assays and biomolecular characterization of biospecimens. The PT Programs include assays performed by repositories and/or end-users for the validation/characterization of biospecimens, and their cellular and molecular derivatives. This Program is expected to improve the quality management systems of repositories through PT of their quality control assays. PT Programs are designed to promote the quality and the economic health of the particular industry of biorepositories, by diminishing the actual "asymmetric information" gap between biospecimen providers and biospecimen end-users.

4.3 ESBB – European, Middle Eastern & African Society for Biopreservation & Biobanking

ESBB's mission is to advance the field of biobanking in support of research relating to healthcare, agriculture and the environment.

ESBB is an open society for people interested in all aspects of biobanking, including biopreservation science, biobank management, quality assurance, informatics and automation as well as ethical, legal, regulatory and social issues. Members represent the full spectrum of biobanks and biological resource centres, both human and non-human. As stated on the ESBB website[6], ESBB membership now includes 67 organisations, comprising 35 companies and 32 biobanks and institutions. Since 2011 ESBB is a regional chapter of ISBER. ESBB works by exchanging, enhancing and disseminating relevant knowledge, by identifying and solving problems, and by encouraging high professional standards in the biobanking field. Its annual meeting, the ESBB Inaugural Conference, which takes place in Marseille in this year and which represent an important meeting place for the Biobank community of Europe and beyond, provides an optimal platform for the dissemination of p-medicine's work for integrating biobanks in an e-research infrastructure.

[6] http://www.esbb.org

4.4 BioSHaRE.eu - Biobank Standardisation and Harmonisation for Research Excellence in the European Union

BioSHaRE is large scale integrating project of the FP7 Health Programme that started in December 2010 and is carried out by a consortium of leading biobanks and international researchers from all domains of biobanking science under the coordination of Academisch Ziekenhuis Groningen, Netherlands. According to the project's website[7], the overall aim of the project is to build upon tools and methods available to achieve solutions for researchers to use pooled data from different cohort and biobank studies, in order to obtain the very large sample sizes needed to investigate current questions in multifactorial diseases, notably on gene-environment interactions. This aim shall be achieved through the development of harmonization and standardization tools, implementation of these tools and demonstration of their applicability.

The mission of BioSHaRE is to ensure the development of harmonized measures and standardized computing infrastructures enabling the effective pooling of data and key measures of life-style, social circumstances and environment, as well as critical sub-components of the phenotypes associated with common complex diseases. Case studies shall guide the development of tools like harmonization, ethical solutions, laboratory standards, etc.

Participants are University Medical Centre Groningen, University of Leicester, Norwegian Institute of Public Health, University of Helsinki, Helmholtz Zentrum München, Norwegian and Technology, Institute National de la Sante et de la Recherche Médicale, University of Manchester, Legal Pathway, McGill University, Medical University of Graz, Public Population Project in Genomics, Ontario Institute For Cancer Research, University of Oxford and Imperial College.

4.5 P3G - Public Population Project in Genomics

P3G (Public Population Project in Genomics)[8] is a not-for-profit international consortium dedicated to facilitating collaboration between researchers and biobanks working in the area of human population genomics. P3G is member-based and composed of experts from the different disciplines in the areas of and related to genomics, including epidemiology, law, ethics, technology, biomolecular science, etc. P3G and its members are committed to a philosophy of information sharing with the goal of supporting researchers working in areas that will improve the health of people.

P3G has four international interdisciplinary working groups focussed on scientific development in areas of social, environmental and biochemical investigations, information curation and information technology, ethics, governance and public engagement, and epidemiology and biostatistics. These groups conduct research on their areas and deliver their results to members at P3G meetings. The results are then posted online at the P3G website.

The P3G Observatory[9] provides online access to information and scientific tools for the purposes of facilitating and promoting development and achievement in the harmonization of research.

[7] http://www.p3g.org/bioshare/
[8] http://www.p3g.org
[9] http://www.p3gobservatory.org/

The P3G Observatory site is free accessible, and all documents, web sites and tools included on the site are non-commercial and open source. Overviews of the research projects being undertaken by P3G members are posted in the cores section of the Observatory and results of their research are published online in the P3G publications section.

4.6 ENCCA - European Network for Cancer Research in Children and Adolescents

The ENCCA FP7 Network of Excellence[10] aims to bring together the existing informal clinical trials groups in paediatric and adolescent oncology towards a European virtual institute to reduce knowledge fragmentation and enhance their communication, collaboration and management of effective clinical research in Europe. ENCCA aims to establish a durable, European Virtual Institute for clinical and translational research in childhood and adolescent cancers that will define and implement an integrated research strategy and will facilitate the necessary investigator-driven clinical trials to introduce the new generation of biologically targeted drugs into standard of care for children and adolescents with cancer. A virtual biobank is seen as an important cornerstone of the initiative. This project will lead to more efficacious and less toxic therapies that will maximise the quality of life of the increasing number of survivors of cancer at a young age in Europe and allow them to assume their proper place in society. The objective is to restructure knowledge-sharing through the integration of the whole chain of stakeholders (epidemiologists, biologists, clinicians, drug developers, statisticians, industrials, imaging developers and IT partners in electronic health records, parents and ethical and regular authorities) and to support the acceleration of the development of innovative therapeutic strategies for children and adolescents with cancer. This shall facilitate access to efficient cancer services for children and adolescents across Europe and support enhanced interest in scientific and clinical careers for young European students.

ENCCA aims to structure and harmonize clinical trials throughout Europe in order to improve treatment and care as well as access to best treatment in paediatric oncology all over Europe. The ENCCA consortium comprises influential European research institutes and organisations recognised for their excellence in paediatric oncology that are dedicated to improve the treatment of children and adolescents suffering from cancer.

ENCA has been included in this chapter as a collaborative network of cancer researchers that could substantial benefit of the p-medicine platform. Some p-medicine partners are members of this network. Their joint research activities imply the need of sharing biomaterial.

4.7 GEN2PHEN

The GEN2PHEN project[11], which started 2008 as FP7 project in the Health Programme, aims to unify human and model organism genetic variation databases towards increasingly holistic views into Genotype-To-Phenotype (G2P) data, and to link this system into other biomedical knowledge sources via genome browser functionality. The project tries to establish the technological building-blocks needed for the evolution of today's diverse G2P databases into a future seamless G2P biomedical knowledge environment. This shall consist of a European-centred but globally networked hierarchy of bioinformatics GRID-linked databases, tools and standards, all tied into the Ensemble genome browser. The project has the following specific objectives:

[10] http://www.encca.eu

[11] http://www.gen2phen.org

- To analyse the G2P field and thus determine emerging needs and practices
- To develop key standards for the G2P database field
- To create generic database components, services and integration infrastructures for the G2P database domain
- To create search modalities and data presentation solutions for G2P knowledge
- To facilitate the process of populating G2P databases
- To build a major G2P internet portal
- To deploy GEN2PHEN solutions to the community
- To address system durability and long-term financing
- To undertake a whole-system utility and validation pilot study

GEN2PHEN project aims to become the key European contribution to the challenges listed above, harmonised with similar projects elsewhere, and dovetailed into many related European programmes of work. It intends to provide a solution to a current research need that was highlighted by the European Strategy Forum on Research Infrastructures (ESFRI) - Priority area: 'Upgrade of European Bio-Informatics Infrastructure. It seeks to provide European G2P research and biotech industries with support related to database technologies and data integration systems.

4.8 Genetic Alliance

Genetic Alliance[12] is the leading nonprofit health advocacy organization committed to transforming health through genetics and promoting an environment of openness centered on the health of individuals, families, and communities. According to the website Genetic Alliance's network includes more than 1,000 disease-specific advocacy organizations, as well as thousands of universities, private companies, government agencies, and public policy organizations. The network is a dynamic open space for shared resources, creative tools, and innovative programs. Genetic Alliance has developed a series of programs designed to increase the visibility of genetics and advocacy, establish strong networks and advance important campaigns, including community based family history. Genetic Alliance dissolve boundaries to foster dialogue that includes the perspectives of all stakeholders: from industry professionals, researchers, healthcare providers, and public policy leaders to individuals, families, and communities. The mission of Genetic Alliance are transform health through genetics, for that they promote an environment of openness centered on the health of individuals, families, and communities. The goal is to build capacity within the genetics community, to eliminate obstacles and limitations through novel partnerships, informed decision-making and individual, family, and community perspectives. The goal of the congenital conditions program is to collect and disseminate evidence-based information, while coordinating the availability of supportive services for parents whose child receives a diagnosis prenatally, at birth, or up to one year after birth. Among others Genetic Alliance also offers a web bulletin about developments in the field of registries and biorepositories[13]. Among others the actual bulletin covers the topics governance of genomic biobanks and provenance of biobank samples.

[12] http://www.geneticalliance.org

[13] http://www.geneticalliance.org/biobank.bulletin

4.9 BioMedBridges

BioMedBridges is a project under negotiation of the FP7-Infrastructures Programme (INFRA-2011-2.3.2 Implementation of common solutions for a cluster of ESFRI infrastructures in the field of "Life sciences"). The BioMedBridges consortium brings together the six under implementation ESFRI biomedical science infrastructures (ELIXIR, BBMRI, EATRIS, ECRIN, InfraFrontier and INSTRUCT). The project seeks to identify and implement standard interoperable services to allow the linking, exchange and deposition of large volumes of data from one infrastructure to another, across the biological and biomedical domain. In consequence BioMedBridges aims at developing and harmonising standards and ontologies across the domains represented by the ESFRI life sciences infrastructures in order to implement a federated access system to diverse data sources in these e-infrastructures and to develop and implement protocols to ensure secure and appropriate access to heterogeneous data types across them.

Public data will be freely accessible through these standard interoperable services. Also, standards for secure and restricted access will be identified and implemented where projects need to share sensitive data (medical information or data with intellectual property issues).

4.10 Related initiatives on ethical, legal and social issues

4.10.1 PrivateGen

PRIVATE Gen[14] investigates the existing privacy regimes - which encompass statutory regulation (both national and international), self-regulation, and technology-based privacy instruments - in relation to (post-) genomic research in general and more specifically in relation to the creation of large-scale life science infrastructures in Austria, Finland, and Germany. Each national case shall be first studied from a different disciplinary angle by the consortium's four subprojects (juridical, ethical, sociological, and political science) and subsequently integrated into a coherent governance framework.

Furthermore, the privacy related complexity of these endeavours is amplified by increasing scientific collaborations that operate on a transnational scale. Private Gen's selection of case studies reflects this trend, since the cases under investigation play an important role in an effort to create the pan-European Biobanking and Biomolecular Resources Research Infrastructure (BBMRI). They are therefore excellent locations to study particular dynamics of privacy regimes that occur on a transnational scale.

4.10.2 Human Sample Exchange Regulation Navigator

The Human Sample Exchange Regulation Navigator (hSERN)[15] is a Web-based service developed within the Global Allergy and Asthma European Project, GA2LEN, to import-export human biological samples, together with Material Transfer Agreements (MTA) beyond the legal requirements. The Human Sample Exchange Regulation Navigator was born from the need to exchange human biological material enforces researches to deal with procedural issues across countries with different legislations. hSERN provides users with information on theoretical and practical legal aspects for exchanging biological samples across borders.

In terms of sample exchange, international collaborations between biobanks, or between biobanks and their research partners, have two important aspects. First, the donors' consent usually implies that the scope and purpose of any sample transfer to third parties is subject

[14] http://private-gen.eu

[15] http://www.hsern.eu/

to major constraints. Since the legal, ethical and political framework of biobanking may differ substantially, even between countries of comparable jurisdictional systems, general rules for the international sharing of biomaterial are difficult, if not impossible, to define. Issues of uncertainty include the right to transfer the material, the scope of research allowed, and intellectual property rights. Since suitable means of international law enforcement may not be available in the context of biobanking, collaborators are advised to clarify any residual uncertainty by means of bilateral contracts, for example, in the form of material transfer agreements. Second, biobank partners may rightly expect that the biomaterial they receive for further analysis attains a certain level of quality. This implies that a biobank has to implement stringent quality control measures covering, in addition to the material transfer itself, the whole process of material acquisition, transport, pre-analytical handling and storage. Again, it may be advisable for biobank partners to claim contractual warranties for the type and quality of the biomaterial they wish to acquire.

hSERN allows addressing practically a series of relevant requests, for different countries, on the issue of regulatory aspects of exchanging human biological samples across borders. It will permit to get information on validated theoretical as well practical legal aspects, for exchanges of human biological samples for research purposes.

The existing solution between France and the UK is expected to be extended to other countries to reduce the obstacles for cooperation within biobank networks.

This web based tool is under construction and not yet publicly available.

4.10.3 Tiss.EU – Assessing legal and ethical aspects of human tissue research

The Tiss.EU Project (Evaluation of Legislation and Related Guidelines on the Procurement, Storage and Transfer of Human Tissues and Cells in the European Union - an Evidence-Based Impact Analysis) was funded by the European Commission as part of the 7th Framework Programme.[16] The project that commenced on the April 2008 and finished in early 2011 addressed questions of ethical and legal regulation in relation to research using human tissue.

Tiss.EU was analysed ethical and legal aspects of:

- procurement, storage and transfer of tissue and cells for research;
- rights and entitlements to tissue and cells;
- anonymisation and pseudonymisation to protect privacy rights;
- research using biobanks.

The project assessed the impact of European Union's regulatory activities on research in the Member States and Switzerland. For this purpose, the project compared national legislative instruments and guidelines with the help of external experts.

[16] For further information see http://www.tisseu.uni-hannover.de.

5 ICT tools for integrated biobanking

The purpose of the investigation of existing tools is to evaluate how far they fulfil the user requirements for the Biobank Access Framework in p-medicine. Key questions are whether they can serve as a model for our own p-medicine Biobank Access Framework or whether they can even used and adapted to p-medicine's biobanking use cases. The application focus in p-medicine lays on sharing and accessing biomaterial and related data within scientific communities through the integration of their biomaterial repositories in the p-medicine infrastructure. For this reason we excluded commercial biobank software products which are usually LIMS for the specific purpose of biobanking. We assume that users have already their Biobank LIMS or database application or spread sheets to manage their biomaterial and want to connect it now to the p-medicine infrastructure through the p-medicine Biobank Access Framework. An exception represents the product Biotracker Biobanking SaaS, which includes many features for the integration of other data sources and means for data harmonization. Biotracker serves as a promising representative for other commercial biobank information systems.

From the perspective of p-medicine the following criteria seams to be important and have been investigated as far as possible based on publically available information about the solutions or in some cases demonstrators available on the Internet or on request. The criteria comprise the purpose of the solution, main functionality and architecture, supported biomaterial set and its adaptation, the role and use of metadata, vocabulatories, terminologies and ontologies, import functionality and data harmonisation, search functionality, availability and maturity as well as the licence model. A short conclusion is given for each solution.

5.1 caBIG Tissue Banks and Pathology Tools Workspace

The **cancer Biomedical Informatics Grid (caBIG)** initiative was launched by the National Cancer Institute, aiming to create a virtual network of interconnected data, individuals and organizations that collaborate in order to redefine the way that cancer research is conducted.[17] Several tools have been developed under this initiative that assist in collecting, analyzing, integrating and disseminating data information that is related with cancer care and research. Objective of these tools is to promote data sharing in a syntactically interoperable manner.

caTissue Suite offers a comprehensive solution for biological sample inventory, tracking and annotation and has been developed in order to assist in the management of the biological specimen annotation data and the operations that have to be performed in a distributed bio-repository environment.[18] This tool is a web-based open source application that utilizes the principles of caBIG by adopting standardized vocabularies and ontologies and enabling data sharing over the caGrid infrastructure (caGrid is the underlying network architecture responsible for maintaining connectivity between caBIG® tools across cancer research institutions).

The suite consists of the following three integrated applications:

- *CaTissue Core:* provides the central biological specimen management.

[17] Rakesh Nagarajan, Poornima Govindrao, Mukesh Sharma, Amy Brink, David Mulvihill, Sachin Lale, Srikanth Adiga, Mark Watson, "caTissue Suite and caBench-to-Bedside (caB2B): Managing and querying forbiospecimens on the caGrid", http://www.cagrid.org/display/community/caTissue+and+caB2B, April 2011.

[18] Catissue suite, https://cabig.nci.nih.gov/tools/catissuesuite, caBIG® Cancer Biomedical Informatics Grid®, accessed on 22-06-2011.

- *Cancer Text Information Extraction System (caTIES):* automates the process of coding, de-identifying, storing and retrieving information from free-text pathology reports.
- *caTissue Clinical Annotation Engine:* provides a web based graphical user interface for the manual annotation of specimens with clinical information, based on standards.

The targeted users are basically bio-specimen resource staff who track and store, retrieve or distribute biological sample collection data and scientists willing to search for samples and associated data that are useful for their translational research project. Since the suite supports the operation of multiple repositories within a single instance of caTissue, it is important to regulate and control user access to operations and data across these repository sites. Regulating access to repository sites is achieved by associating collection protocols, users, and storage containers.[19]

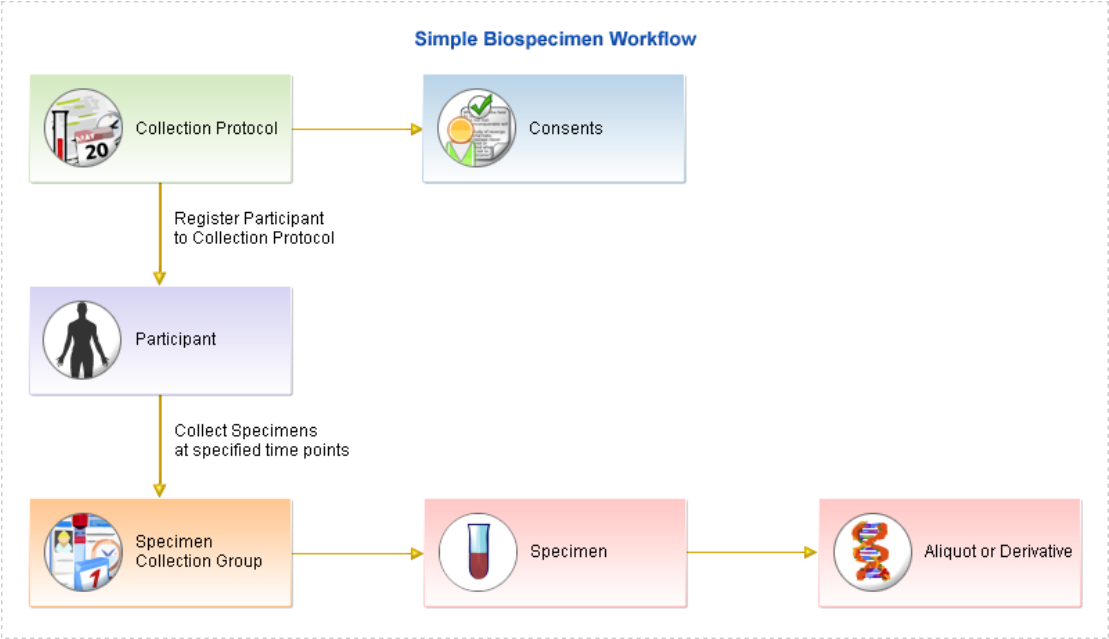


Figure 1: A simple caTissue Suite workflow diagram [20].

The core functionalities/modules are:

- *Administration*: Creating and editing users, protocols and storage systems associated with a bio-specimen inventory.
- *Bio-specimen*: creating and editing data concerning participants and their corresponding bio-specimens.
- *Query*: Identifying bio-specimens and their data based upon one or more selection criteria.

The key features are:

- Web-based, open source
- Role-based security
- Configurable storage container hierarchy
- Supports protocols, patient registry, and specimens
- Supports pathology reports upload

[19] "caTissue Core v1.2.2 (LSD) Overview", caTissue elearning Materials, May 2008.
[20] "caTissue Suite v1.0 Overview" presentation accessible under cabigtrainingdocs.nci.nih.gov/caTissue/pps/catissuesuite_overview.ppt

- Supports clinical and pathology annotations
- Facilitates ordering and distributing specimens
- Provision to define local extensions
- Facilitates querying, exporting results, and saving query

Certified for:

- JBoss Application Server
- Oracle and MySQL databases
- Windows and Linux servers
- Internet Explorer and Mozilla, and Safari for Mac
- Well defined caBIG compatible API
- Ability to interface with external IT systems
- Facilitates label & barcode printing and generation

The architecture of caTissue Suite consists of n-tiers. In general applications on the client layer request operations on the server.[21] The Web browser and API are two client programs in the Suite. The presentation layer provides a Web interface as well as support for the API. The object layer is the core part of the application and consists of domain objects, model classes, and the data access layer. The object layer is responsible for all requirements related to tissue banking operations. Finally the data storage of the application is a local database that holds all the tissue banking information.

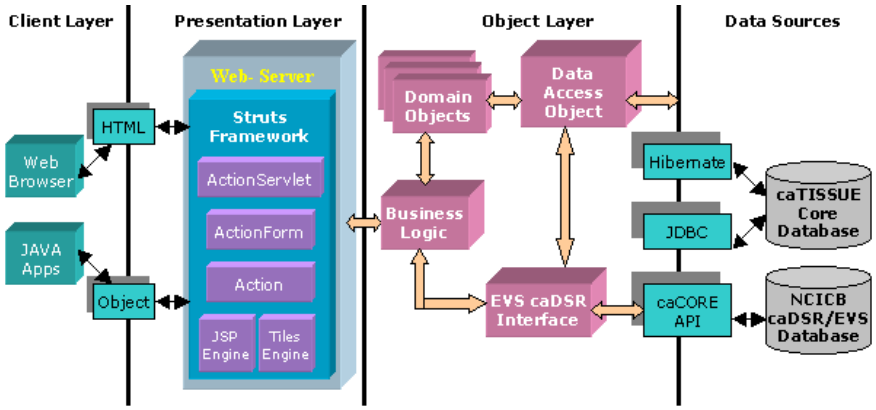


Figure 2: caTissue Architecture [22].

The data entered into the caTissue Suite can be categorized into two groups, Administrative and bio-specimen data. Administrative data is required for setting up the tissue bank. Examples of such data are, Collection Protocols, Distribution Protocols, Biohazards, Sites, Storage containers, Users, Departments, Institutions, Cancer Research Groups, Storage Types, and Array Types. On the other hand, the bio-specimen data holds information relative to the patient's demographics, specimen's quantity, storage locations, and samples distributed by the bank.

The supported specimens can be one of the four different classes 'Molecular', 'Fluid', 'Tissue' and 'Cell'. The three categories of specimens 'New specimen', 'Derivative' and 'Aliquot' are distinguished from each other. Each specimen class is represented by a different UML Class in the model.

[21] caTissue Development Team, Dave Mulvihill , Amy Brink , Laura Jackel, *"caTissue Suite v1.2 User's Guide", v2.3 March 31, 2011.*

[22] "caTissueSuiteV1.1 Technical Guide" document accessible under https://catissue.mskcc.org/catissuecore/caTissueSuiteV11_Technical_Guide.doc

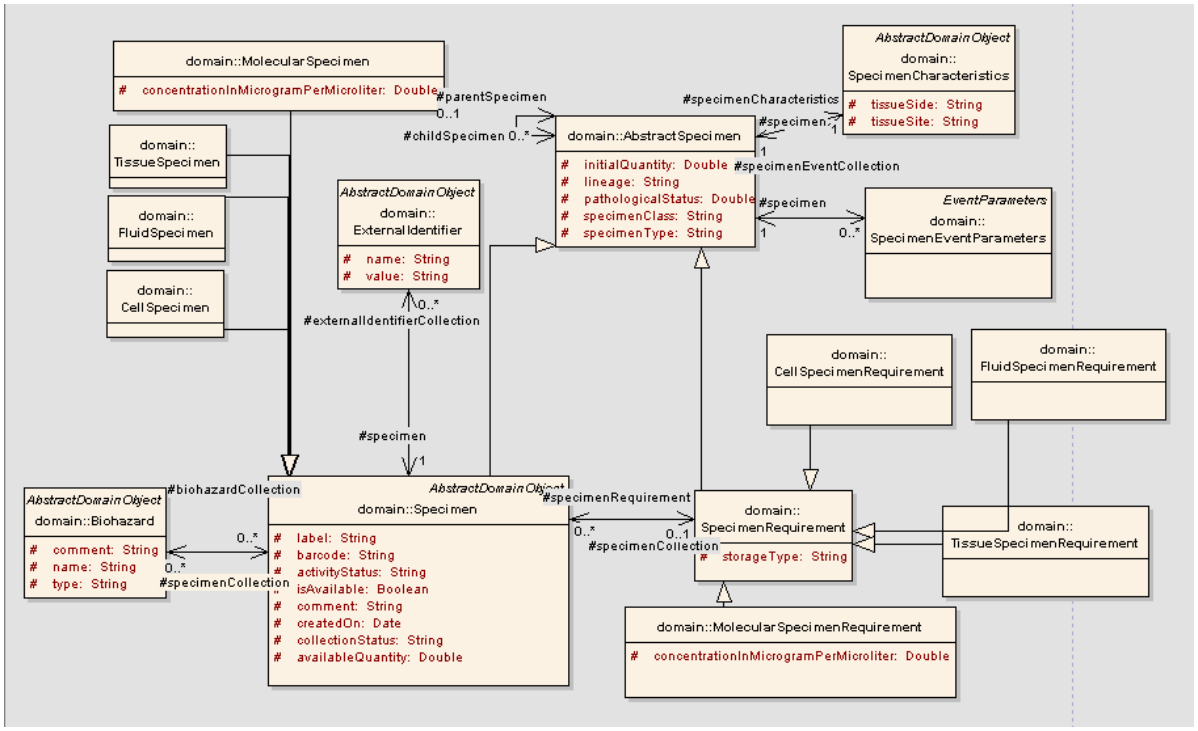


Figure 3: Specimen UML Diagram [23].

Querying for biological specimens within the suite can be achieved in two ways. caTissue provides a Simple and an Advanced Query feature. Simple Query is ideal for queries based on a few parameters and the steps to perform such a query are to perform a simple search and then to view the results.

Advanced Query allows creating and executing more complex queries that contain further parameters, in order to retrieve biological specimens, Specimen Collection Groups, Collection Protocols and participants that the system holds. The steps for performing an Advanced Query are the following:

- Query *any* data
- Security safeguard for regulatory (HIPPA) and proprietary data
- View results in *any* form – drill down mechanism
- Save query with parameterization
- Export data into CSV
- Add specimens to shopping cart to order

[23] "caTissueSuiteV1.1 Technical Guide" document accessible under https://catissue.mskcc.org/catissuecore/caTissueSuiteV11_Technical_Guide.doc

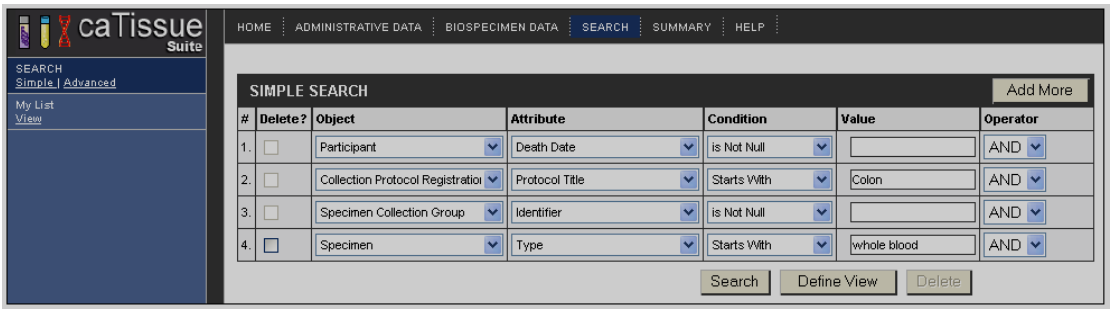


Figure 4: Performing simple search [24].

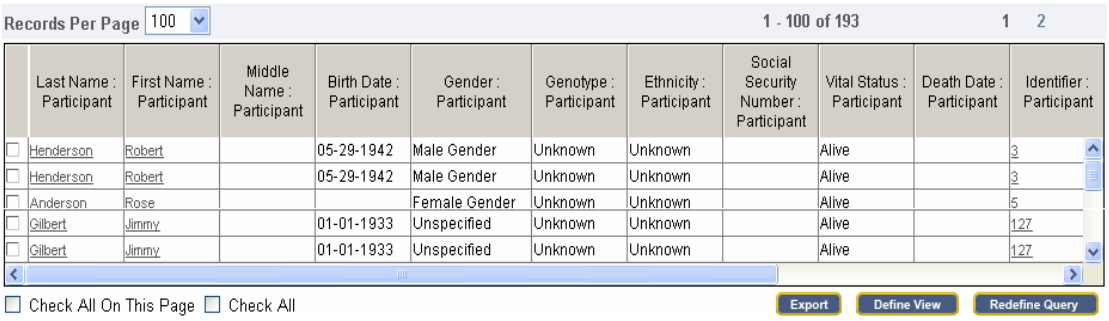


Figure 5: Viewing the simple search results [25].

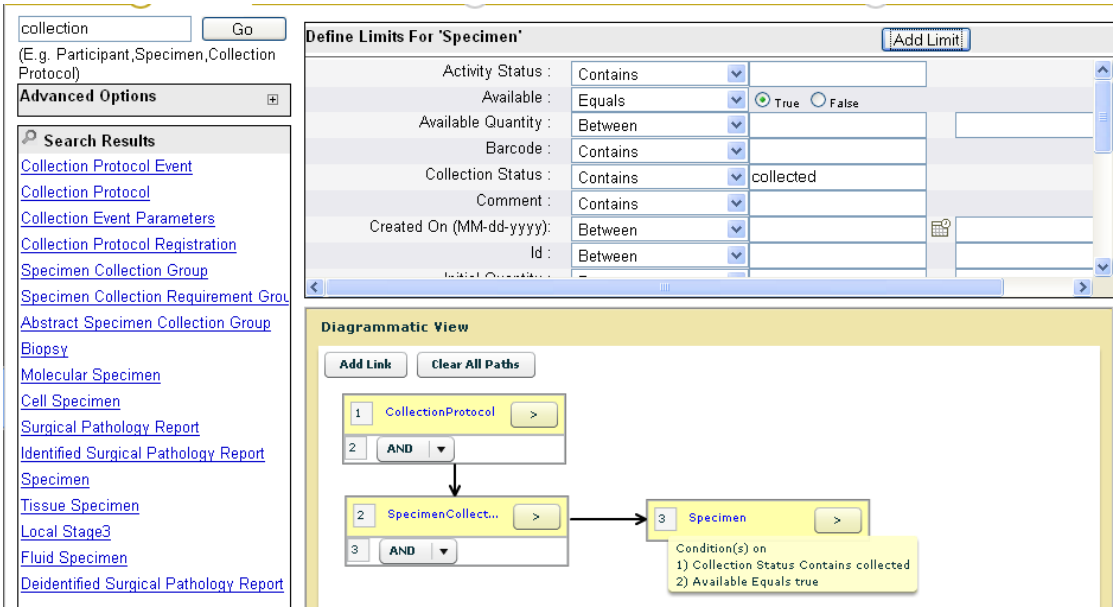


Figure 6: Performing an Advanced Query search [26].

[24] "caTissue Core v1.2.2 (LSD) Overview" presentation accessible under cabigtrainingdocs.nci.nih.gov/caTissue/pps/catissuecore_lsd_overview.ppt

[25] "caTissue Core v1.2.2 (LSD) Overview" presentation accessible under cabigtrainingdocs.nci.nih.gov/caTissue/pps/catissuecore_lsd_overview.ppt

[26] Screenshot taken from caTissue Suite demo web client available under https://catissue.mskcc.org/catissuecore/RedirectHome.do.

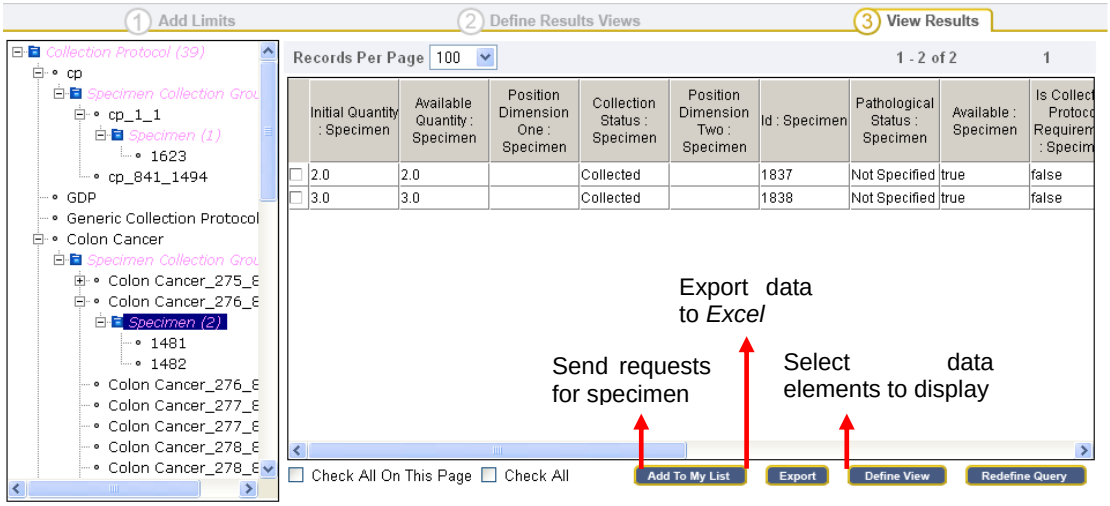


Figure 7: Viewing Advanced Query results, exporting them to excel file or requesting for them [27].

The caTissue Suite provides an application program interface (API) for the potential of interacting with external systems. This API is generated using the caBIG caCORE SDK toolkit and can be used for adding, editing and querying data.

caCORE is an open-source enterprise architecture for the NCI-supported research information systems and utilizes formal techniques from the software engineering and computer science communities. CaCORE has a Model Driven Architecture (MDA) that consists of n-tiers with open APIs and is characterized by the use of controlled vocabularies, wherever possible and registered metadata. Using the MDA and n-tier architecture, ensures the easy access to data by other application systems, however utilizing controlled vocabularies and registered metadata comprises less common conventional software practice and requires specialized tools which are in general unavailable.[28]

The following steps describe how to use these APIs for data entry and query:

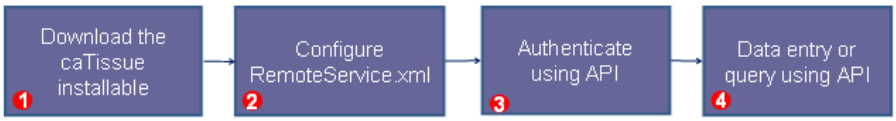


Figure 8: Steps for using APIs for data entry and query [29].

It is possible to define the necessary classes and attributes in a UML model by using a UML modeling tool such as Enterprise Architect (EA). That UML model can furthermore include attribute data types, attribute/class concept code(s), attribute validation(s) and attribute PHI features. As soon as the above components have been defined within the model, it can be exported in an XML file with the appropriate version and export parameters from the EA tool and then imported into the caTissue application by running ANT commands.

After the UML model is being imported, permissible values can be uploaded into the system for all the enumerated attributes of the model by creating a CSV file. Finally, the suite allows the definition of a custom view for the form's format and look and feel on the UI. This is done

[27] Screenshot taken from caTissue Suite demo web client available under https://catissue.mskcc.org/catissuecore/RedirectHome.do.

[28] *"caCORE SDK"*, https://cabig.nci.nih.gov/tools/caCORE_SDK, caBIG® Cancer Biomedical Informatics Grid®, accessed on 22-06-2011.

[29] "caTissueSuiteV1.1 Technical Guide" document accessible under https://catissue.mskcc.org/catissuecore/caTissueSuiteV11_Technical_Guide.doc

by creating a form definition file, holding the description regarding the display of data fields required on the form at the time of data entry. Default values and display properties can also be defined restricting the initially uploaded permissible values to a subset of those values.

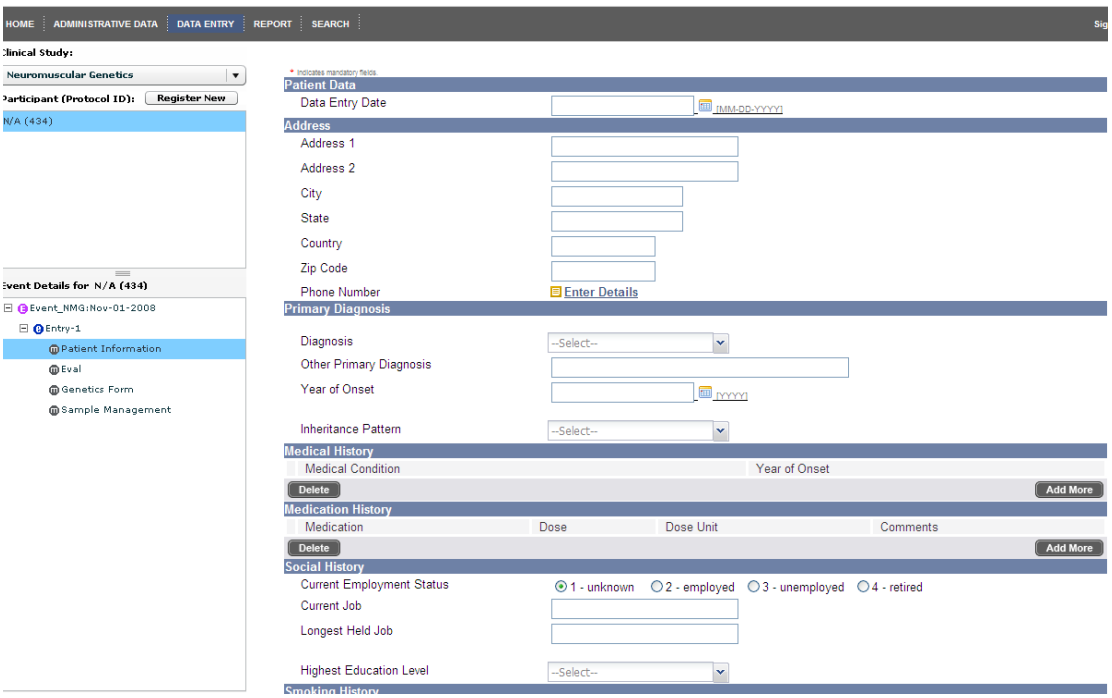


Figure 9: Importing XMI - Data Entry Form [30].

caTissue is part of the caBIG® initiative and was developed with NCI funding.[31] The software is distributed under a free open source license without any restrictions in usage including commercial exploitation. This is a BSD-like license, very similar (almost identical) to the Apache license. More concrete, "a non-exclusive, worldwide, perpetual, fully-paid-up, no-charge, irrevocable, transferable and royalty-free right and license in its rights in the caBIG™ Software, including any copyright or patent rights therein, to (i) use, install, disclose, access, operate, execute, reproduce, copy, modify, translate, market, publicly display, publicly perform, and prepare derivative works of the caBIG™ Software in any manner and for any purpose, and to have or permit others to do so; (ii) make, have made, use, practice, sell, and offer for sale, import, and/or otherwise dispose of caBIG™ Software (or portions thereof); (iii) distribute and have distributed to and by third parties the caBIG™ Software and any modifications and derivative works thereof; and (iv) sublicense the foregoing rights set out in (i), (ii) and (iii) to third parties, including the right to license such rights to further third parties".

CaTissue is a centralized biospecimen management system. It supports the four main specimen types (molecular, fluid, tissue and cell) and it furthermore takes into account derivatives and aliquots. For these kinds of specimen types, the system stores all sample data such as IDs, barcodes, sample history and sample classification data. In addition the system allows data annotation with clinical and pathology information, using appropriate standards. In particular caTissue provides 170+ predefined annotation sets that can capture sample annotation, collection, events and participants. It is possible to define custom

[30] Screenshot taken from caTissue Suite demo web client available under https://catissue.mskcc.org/catissuecore/RedirectHome.do.

[31] *"CaTissue Suite v1.1 Technical Guide Copyright and License page"*, https://cabig-kc.nci.nih.gov/Biospecimen/KC/index.php/CaTissue_Suite_v1.1_Technical_Guide_Copyright_and_License_page, *last modified on 28 January 2010.*

annotation sets and extensions to the caTissue data model by utilizing the annotation designer.

In terms of security, caTissue suite supports the usage of user roles and is capable of supporting virtual repositories within one single installation through the proper configuration of roles and privileges. Likewise, the user can define multiple levels of consent and study-specific consent rules for the stored samples.

The system provides two different ways for searching, either by entering keywords or by generating custom queries, through corresponding graphical user interface that can cope with p-medicines needs. The searched results can be exported to CSV (Comma Separated Values) files, which are a desired feature. However, importing data from similar kind of data sheets is not clear whether it is supported. All participant users are capable of ordering the desired specimens through a shopping-cart like ordering workflow and any physical movements of the specimens in between sites is totally supported by using the Shipping & Tracking Dashboard. From a European perspective and experience, biobanks usually require, beyond simple ordering of specimens, differentiated research project agreements prior to sample allocation. This is not supported by caTissue suite. However, caTissue suite is open source and can be expanded by using the well-defined caBig compatible API.

Certain questioning and speculation is raised by the evaluation assessment reported that was applied by the Board of Scientific Advisors Ad Hoc Working Group on the NCI caBIG® program[32], regarding the attempt and extra cost required for the adoption of this tool. According to that assessment report, the caTissue suite is the most widely adopted caBIG® Life Science tool and has been evaluated and adopted by many cancer centers and other NCI-supported programs. However, the provided prototype required significant modifications and adjustments for many of these adoptions. For that reason, the working group has interviewed several of those adoptions and has revealed several significant barriers to the adoption of this tool by other cancer centers and research programs. In accordance to the evaluation applied by several institutions, caTissue was found not to be user friendly and was lacking in necessary functionality. As a result those institutions developed and implemented custom web-based applications to support their bio-repositories or had to invest considerable resources to bring caTissue on board as NCI required but also spent some more for the capabilities they needed. In addition to those barriers, the caTissue tool was found to be trusted only with rigorously de-identified data and does not support the CHTN standard for non-cancer repositories.

5.2 BBMRI - WP5 Database harmonization and IT-infrastructure

To underpin BBMRI's overall mission (cf. chapter 4.1) of building a pan-European biobanking research infrastructure, BBMRI's WP 5 on Database harmonization and IT-infrastructure[33] ("WP 5") has been striving for "a universal information infrastructure for biobanking in Europe". During the BBMRI Preparatory Phase from February 2007 until January 2011, WP 5 participants have elaborated recommendations[34] on the following issues:

- Requirements for a general information management system for biobanks in Europe
- Systems for maintaining unique and secure identities for specimens, subjects and biobanks

[32] "An assessment of the impact of the NCI Cancer Biomedical Informatics Grid (caBIG®) - Report of the Board of Scientific Advisors Ad Hoc Working Group", March 2011.

[33] http://www.bbmri.eu/index.php/workpackages/wp-5

[34] Litton JE et al. (2010) BBMRI D5.6 Final Report

- Strategy for communication between biobanks, including a common nomenclature, compatible software techniques and appropriate information transmission policies

In general, WP 5 has proposed a federated infrastructure with national or regional hubs and the local biobank databases as spokes, federating data at the meta-data, the aggregated-data and the object-data level (subject and/or sample data). Aspects of this work relevant for p-medicine WP 10 will be outlined and discussed below.

5.2.1 Use Cases and Architecture

As general use cases, WP 5 has implemented the search for biobanks in the BBMRI catalogue and studied the search for cases/specimens by the two Prototypes A and B, the latter being based on the Set Definition Language (SDL) of deCODE genetics. SDL is a language on top of a relational database system that allows defining and reinforcing constraints across database objects. deCODE genetics holds a patent on SDL and a software system that efficiently implements SDL.[35] Prototype A represented an early version of the metadata model and will be discussed in section 5.3.1.2.

5.2.1.1 Search for biobanks

In collaboration with BBMRI WPs 2 and 3, a survey on biobanks has been performed as a starting point. Data from a set of biobank questionnaires (hard copies, > 1000 items) have been collated in the BBMRI catalogue[36] enabling users to retrieve a list with contact data selected from > 200 participating biobanks that might have desired material for a certain study. Although the query tool of the catalogue[37] is still being refined (distributed metadata queries, distributed sample counts), it will never enable search for specific cases and thus cannot provide a model for p-medicine's access to biobanks.

5.2.1.2 Search for cases

Retrieving the pseudonym identifiers of specific cases/annotated samples stored in biobanks, as pursued by the second WP 5 use case, clearly matches the objective of p-medicine's Access-to-biobanks framework. To this end, WP 5 has developed a metadata model for a regional BBMRI hub (Figure 10**: Structure of metadata model for a regional BBMRI hub; for details see Litton et al (2010) .**

) and, based on that model, the so-called Prototype A (see below).

[35] http://www.freepatentsonline.com/y2005/0050030.html

[36] Wichmann HE, Kuhn KA, et al (2011) Comprehensive catalogue of European biobanks. Nature Biotechnology, in press

[37] https://www.bbmriportal.eu/

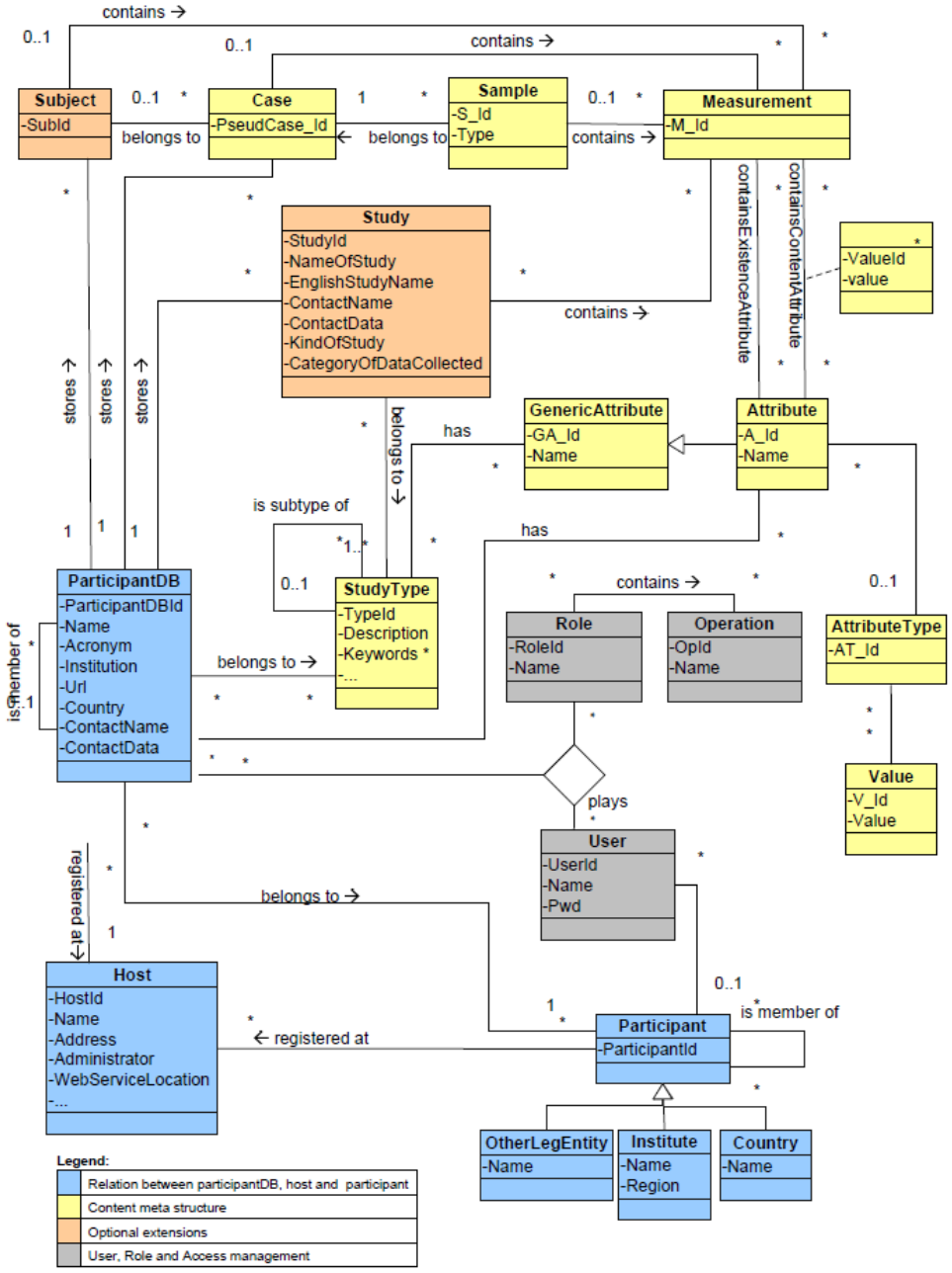


Figure 10: Structure of metadata model for a regional BBMRI hub; for details see Litton et al (2010) [38].

Prototype A (Figure 11**Figure 11**) was implemented by java web services with data exchange via XML structure and SOAP request, with its metadata model (simplified version of Figure 10**: Structure of metadata model for a regional BBMRI hub; for details see Litton et al (2010) .**

) in a MySQL database. It has however been realized only temporarily as an "explorative piece of software" to validate requirements (see 5.3.2) and check feasibility. WP 5 has terminated Prototype A's operability with the BBMRI Preparatory Phase.

[38] Litton JE et al. (2010) BBMRI D5.6 Final Report

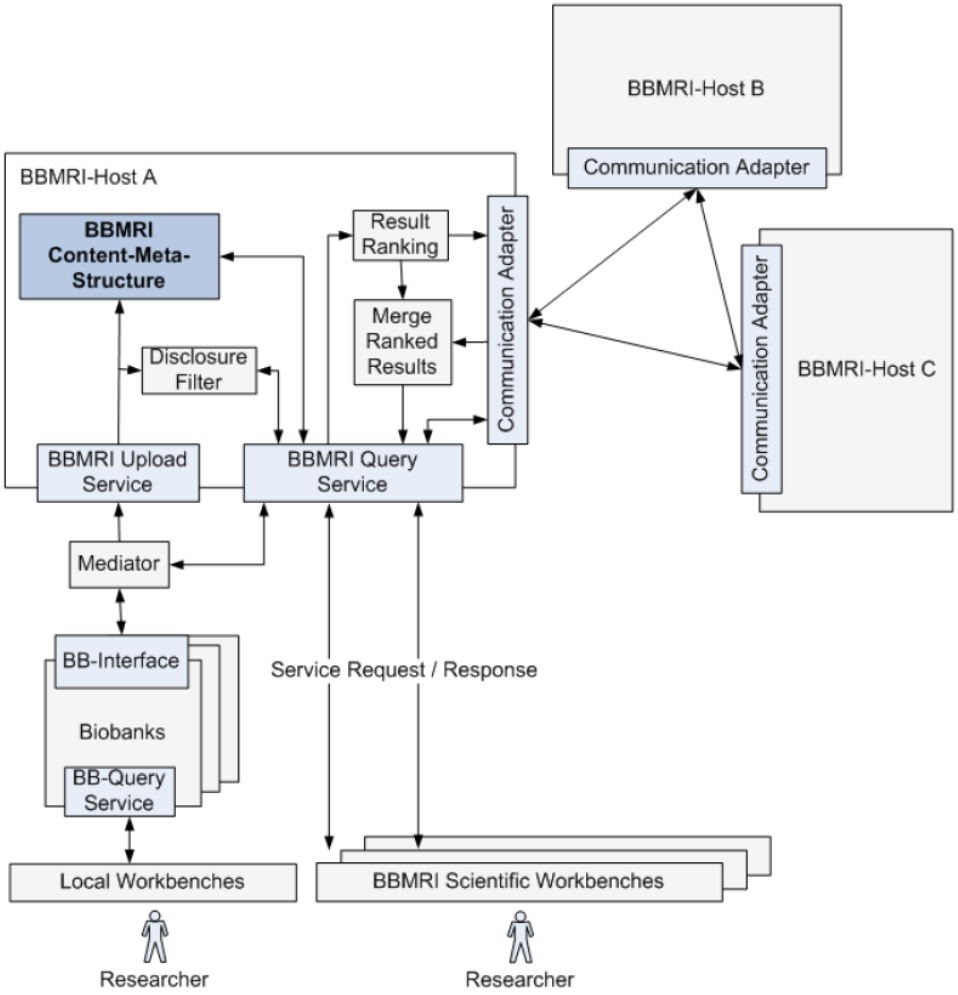


Figure 11: Proposed architecture for Prototype A (BBMRI-Host A); for details see Litton et al (2010) [39].

Several features and modules of the BBMRI Prototype A are matched by the CRIP architecture and toolbox, see chapter 5.6.

5.2.2 Requirements for a federated biobanking infrastructure

WP 5 has filed a list of 32 requirements providing a good checklist for the p-medicine architecting discussion and implementation of the Access-to-biobanks framework.

1. Technical requirements on BBMRI data integration system		
The following general requirements have been identified:		
	R1	Given the complexity, usage of the information system should be divided at least into two complementing parts addressing
	1.1	Needs of resource (here mainly data and samples) discovery and

[39] Litton JE et al. (2010) BBMRI D5.6 Final Report

	1.2	Sharing of original data for research purposes.
	R2	Local database polices, national ELSI regulations and EU data protection act must be followed.
	R3	The identification scheme for samples and subjects can be based on surrogate identifiers maintained by co-operating systems providing context for the identifiers. Global identification scheme is necessary if identifiers are taken outside the context.
	R4	Sample and subject identifiers must be randomized and identifiers should not contain any meaning.
	R5	Each biobank must identify their specimen and related information adequately and persistently.
	R6	The data integration framework must have possibility for data federation without sacrificing benefits of centralized approaches where data is collected into one single database.
	R7	Standard security protocols and measures must be used. There are at least two different security domains related to data discovery (R1-1.1) and data analysis (R1-1.2) having different security requirements due to nature of data.
	R8	All queries and/or access to data services and analysis tools should be logged and data provenance issues taken into account. The auditing information should be stored for determined time.
	R9	Users must be authenticated and authorized, e.g., via federated architecture such as OpenID [2]. Each country must be able to register and manage the credentials of local users.
	R10	Authentication and authorization should be done on a level (like in a local hub) where identification of users is most reliable. A central repository can be used to support access control in cases of possible policy violations.
	R11	Database systems must be kept up-to-date. Access to monthly archives should be accessible at least 10-15 year back in time.
	R12	Core informatics needs related to the hub-and-spokes network are same in all participating network nodes (hubs and biobanks). Developed applications and software can benefit all nodes. Common core needs must be defined in design phase.
	R13	Application programming interfaces (APIs), data formats and vocabularies must be standardized. Existing standards must be used wherever possible.
	R14	Local biobanks should have control on the data they expose (“the mine problem”).
2. Special data federation requirements		
One of the key factors in data integration of distributed systems is the extent of data localization, i.e., how much data is cached or stored outside source databases. Different integration scenarios are presented in Section 1.3.4. The following influencing requirements have been identified:		
	R15	Only *k*-anonymized [1] data and metadata is allowed to leave each biobank node without explicit permission for down-loading detailed data (R2).

	R16	Data processing should be distributed to the source biobank nodes and no identifiable information should leave each node.
	R17	System should have sufficient level of redundancy for minimizing system downtime and increasing data transfer bandwidth.
	R18	Level of independency: National or local networks must be functionally independent from the parent network, i.e., local services should not be hampered by external factors meaning that at least national (or local) metadata must be stored into national (local) hub.
	R19	Data access use cases must be implementable. For analysis purposes it can be essential to collect relevant data into one place. Data sets can be processed faster and kept stable.
	R20	Distribution of data management and curation work. Curation of primary or derived data should be done on sites having knowledge and expertise on the data.
	R21	Metadata should be defined in a way that it can be collected into centralized data marts.
	R22	Metadata for which *k*-anonymity cannot be guaranteed must not be collected outside biobanks (or possibly outside local hubs).
3. **Networking requirements**		
The hub-and-spokes model has been proposed as a basic unit for the data integration architecture because of its simplicity and scalability. In the model, network connections (spokes) are arranged so that all traffic from connected nodes goes through a central hub working as a message broker. A drawback of the approach is that a hub presents a single point of failure. Also, communication can be slow because of the extra step posed by the hub. These problems can be addressed:		
	R23	Increasing redundancy of the system so that hubs can take responsibilities from others. Also services and network connections must be monitored constantly.
	R24	Alternative network strategies, like peer-to-peer connections, should be allowed. This is especially important when transferring actual data since volumes can be huge compared to the metadata.
4. **Data schema and access requirements**		
	R25	The BBMRI network should be shared nothing, meaning that all data that is used to search for any given subject must reside in a single node (biobank and hub). This means that all data derived from samples must "come back home".
	R26	Data access (case R1-1.2) can include manual or semi automated steps where human invention is needed to judge data access and usage rights (called as *disclosure filters* in the architecture model).
	R27	Query and analysis tools should be metadata driven and not custom-written against a fixed data model.
	R28	Query tools and database schemas should support hierarchical and DAG structured vocabularies and ontologies.
	R29	Users must be able to specify dynamically the attributes and the scope of the database that is used in analysis, e.g., aggregation analysis.

	R30	Data schema must support event-based data and query tools should support timebased longitudinal analysis.
	R31	Metadata can be separated based on content, number of cases and existence attributes. Existence attributes can be divided further into *or*-connected quantities and *and*-connected availabilities [3].
	R32	A domain lexicon for biobank informatics must be defined.

Table 2: 5.3.2 BBMRI requirements for a federated biobanking infrastructure[40]

5.2.3 Minimum data set

WP 5 has proposed the minimum data set given in Chapter 7. It should be considered a basis for p-medicine but will certainly be extended according to p-medicine's requirements, as also anticipated by WP 5.

The minimum data set can be seen as an intermediate between the BBMRI questionnaires and the generalized metadata model. The minimum data set is divided in to three levels; the biobank level, the study level and the object level - individual subject/case/sample. The idea is to provide an easy way to present which elements that are considered common in all biobanks. The minimum data set was originally designed at a WP5 meeting in Munich, December 14, 2009, but was heavily revised during discussions at the final WP5 meeting in Klagenfurt, February 4-5, 2010.

All in all, BBMRI has paved the way for a uniform federated pan-European biobanking infrastructure. Making use of this great achievement, p-medicine should strive to align its biobanking access framework with BBMRI and foster seamless interoperability wherever possible.

5.3 i2b2 Hive

5.3.1 General Description

Purpose

Informatics for Integrating Biology and the Bedside (i2b2)[41] is one of the sponsored US initiatives of the NIH Roadmap National Centers for Biomedical Computing with the goal to provide clinical investigators with the software tools necessary to collect and manage project-related clinical research data in the genomics age as a cohesive entity. The i2b2 Hive is the corresponding software suite to construct and manage the modern clinical research chart. The framework enables researcher discovery research based on existing clinical data combined with genomic data. The i2b2 "hive" is a set of software modules called "cells" that have a common messaging protocol that allow them to interact using web services and XML messages. The interoperable scalable framework can be extended for new and unanticipated data types as well as its functionality. The current version supports several diseases like airways disease, hypertension, type 2 diabetes mellitus, Huntington's Disease, rheumatoid arthritis, and major depressive disorder. The platform currently enjoys international adoption by academic health centres, US research networks and industry.

[40] Litton JE et al. (2010) BBMRI D5.6 Final Report

[41] http://www.i2b2.org/

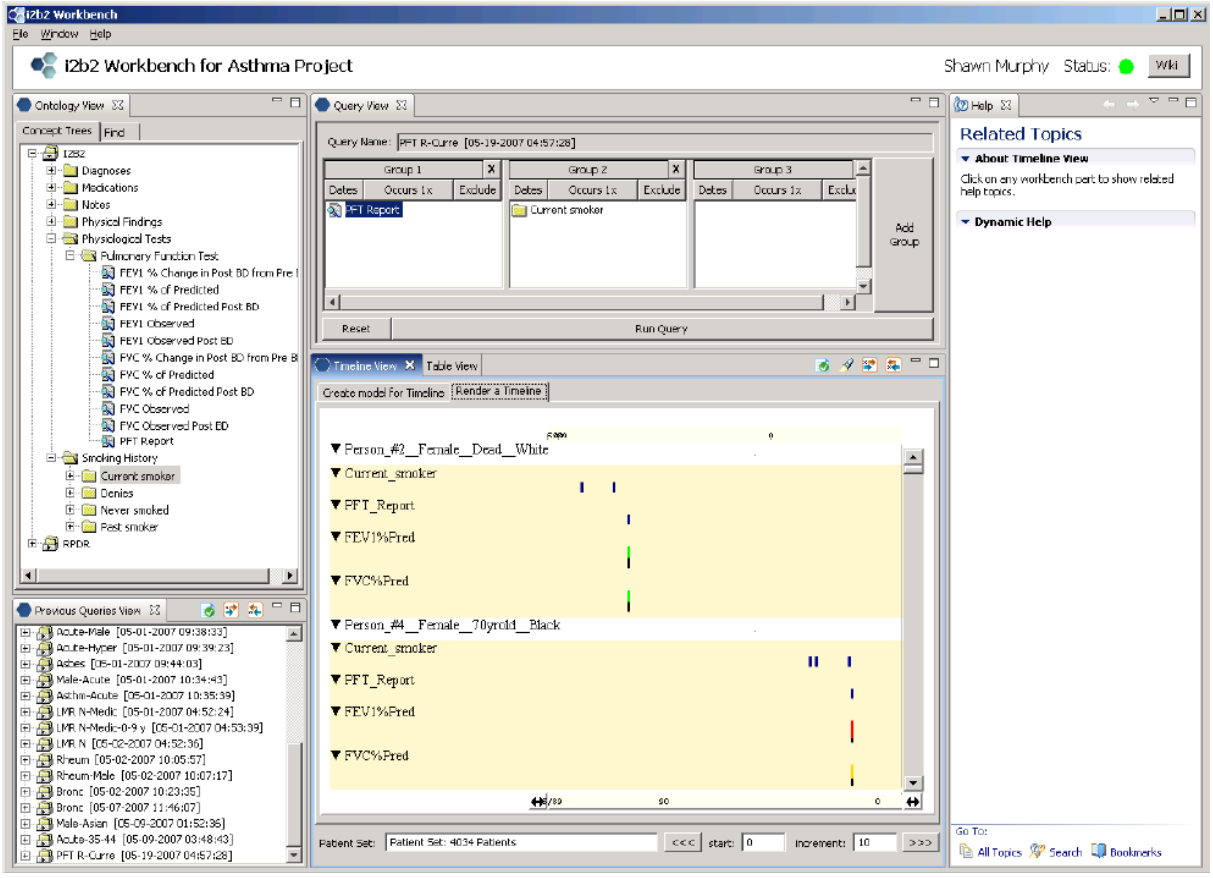


Figure 12: Screenshot of the i2b2 Workbench[42].

Components

The i2b2 Hive is centred around two concepts. The first concept is the existence of services provided by applications that are wrapped into functional units, such that their functionality are exposed as messages that travel to and from the various cells of the hive. The second concept is that of persistent data storage, which is managed by the cell named the "Clinical Research Chart". Each cell can be developed by independent institution to achieve specific analytic goals, which can be integrated into the hive to enhance the functionality.

As a collection, the i2b2 Cells are loosely coupled and generally known to each other only through the use of the Web services.

For human communication, cells usually have a corresponding plug-in that goes onto the i2b2 Navigator. The i2b2 Navigator uses the Eclipse framework. The client applications are plug-ins, and are the most visible part of the Hive. The exposure of these Eclipse plug-ins to the user makes the hive appear as a workbench for managing patient-oriented phenotypic and genotypic data.[43]

[42] Screenshot taken from i2b2 demo web client available under https://www.i2b2.org/software/index.html

[43] Murphy S. N. et. al.: Architecture of the Open-source Clinical Research Chart from Informatics for Integrating Biology and the Bedside. AMIA Annu Symp Proc. 2007; 2007: 548–552. Published online 2007

Figure 13: i2b2 Hive with its components[44].

The basic cells of the i2b2 Hive are shown in Figure 13. Some cells are further described in the following:[45]

- *Data Repository (CRC)* – is a data warehouse that holds the phenotypic and genotypic data of the hive in a structured format. Data queries and visualizations are available through this cell.
- *File Repository* - This cell holds large files of data including radiological images and genetic sequences. The files are generally referenced from the Data Repository Cell.
- *Project Management* - This cell is used to provide user authentication and manage group and role information. It also keeps track of what cells are parts of the hive.
- *Identity Management* - This cell is used to manage a patient's protected health information in a manner consistent with the HIPAA privacy rule[46]. Patient data is available only as a HIPAA defined "Limited Data Set" to most of the hive. This cell contains the “code book” that maps real patient identifiers to arbitrary patient numbers in the CRC. The cell needs the identified data to determine if new patients in the Hive map into old patients in the Hive, and associates their data with the correct coded numbers.
- *Workflow Framework* - This cell is used to process information in steps through various parts of the hive. Most processed information will come to reside in the Data Repository Cell or as a display to the user.

[44] Picture and cell description taken from https://www.i2b2.org/software/index.html

[45] https://www.i2b2.org/. Last visit 2011 Sept 2.

[46] The Health Insurance Portability and Accountability Act (HIPAA) was enacted by the U.S. Congress in 1996. Title II of HIPAA requires the establishment of national standards for electronic health care transactions and national identifiers for providers, health insurance plans, and employers. The HIPAA Privacy Rule regulates the use and disclosure of ‘Protected Health Information’.
(Source: Wikipedia: http://en.wikipedia.org/wiki/Health_Insurance_Portability_and_Accountability_Act.)

- *i2b2 Workbench Application* - The i2b2 Workbench is a collection of client-side components designed as Eclipse-based java plug-ins that communicate with i2b2 Cells and allow the investigator to query, analyze, and display the data of the hive, generally in greater depth than the web client.
- *i2b2 Web Client Application* - The i2b2 Web Client is a collection of client-side components designed as an YUI AJAX-based plug-ins that communicate with i2b2 Cells and allow the investigator to query and display the data of the hive.
- *Ontology Management* - This cell manages the terminology and knowledge information typically used in the hive. It is contacted for, or distributes knowledge to, cells during most of the hives transactions.
- *Natural Language Processing* - This cell manipulates text reports to extract specific terms and knowledge from them. These concepts are then used to achieve various representations of the data.
- Annotator Plug-in - The annotator view will allow non-expert NLP users viewing tools for NLP output
- Correlation Analysis Plug-in - This specialized analysis cell uses mutual information theory to calculate observed correlations within the data of the hive.
- Import Data Plug-in - The import data view will enable users to import data into the dimension and mapping tables.

The i2b2 Hive consists of a number of core cells that establish basic services to support the activities of the Clinical Research Chart (CRC), as well as any number of additional cells to provide enhanced services. It is intended to be a scalable approach for managing an increasing number of independently developed software services to be contributed to the CRC. Fundamentally, the CRC is built to hold medical and medically-oriented genomic data. Any of the various cells of the i2b2 Hive may contribute to placing the data into the CRC, which ultimately occurs by sending XML messages to the CRC.

The XML is defined through a specific schema that contains a set of the patient's related phenotypic and genotypic data as well as possible references to other data objects in the Hive. The document-oriented message allows data to be passed to a cell, have new data added, and then moved to the next cell for additional analysis. In this way fairly complicated workflow patterns can be accommodated when working on particular sets of medical records.

5.3.2 i2b2 ontology framework

The Ontology Management (ONT) is realized through an i2b2 Hive core Cell. This cell manages i2b2 vocabulary definitions and contains concepts and information about relationships between concepts for the entire hive. It is accessed by other cells to give semantic meaning to data.

Vocabularies in the ONT cell are organized in hierarchical structures that represent the (only) relationships between terms. A level in a hierarchy is sometimes referred to as a "node", and a group of related data is called a "category". A category is defined as a set of data for which there is a common rule or rules for querying against the Clinical Research Chart (CRC). The resulting ontology represents a hierarchical taxonomy that is very much driven by the clinical viewpoint and in consequence it is easy for the clinical researcher to browse through it. The top level concepts are 'demographics', 'diagnoses', 'laboratory tests', 'medications', 'procedures', 'providers' and 'reports'.

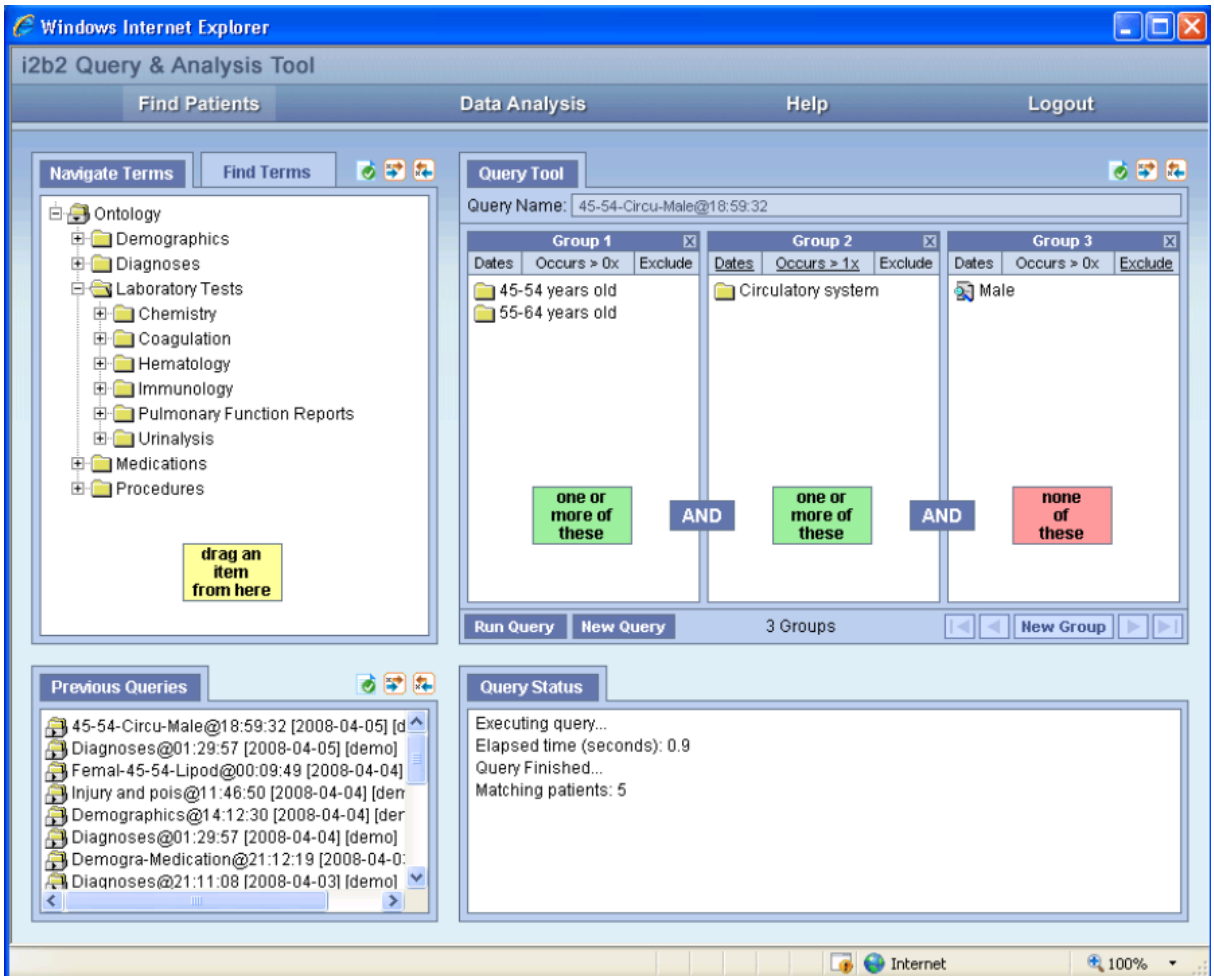


Figure 14: Query interface of the i2b2 Web Client. [47]

The i2b2 data is stored in a relational database and always in a star schema format. A star schema contains one fact and many dimension tables. The fact table contains the quantitative or factual data, while the dimension tables contain descriptors that further characterize the facts. Facts are defined by concept codes and the hierarchical structure of these codes together with their descriptive terms and some other information forms the i2b2 ontology. Vocabularies in the ONT cell may originate as code from different sources. The ONT cell distinguishes these codes from one another by pre-pending a unique prefix to each code.

The US National Center of Biological Ontologies (NCBO) contributes to the i2b2 Hive development. Among others, it aims to integrate NCBO Web services to support the ontology needs of the i2b2 platform: For this aim, NCBO is prototyping mechanisms to "pull" any ontology stored in its BioPortal[48] into the format used by i2b2's ontology cell. The goal is to ensure that users have access to the latest versions of ontologies as well as site specific view of ontologies for use within the i2b2 platform.[49]

[47] Screenshot taken from i2b2 demo web client available under https://www.i2b2.org/software/index.html

[48] http://bioportal.bioontology.org/

[49] http://www.bioontology.org/Biology%20and%20the%20Bedside%20%28i2b2%29. Last visit 2011 Sept 2

5.3.3 Crimson satellite project for sample management within i2b2

i2b2-Crimson is one of currently nine i2b2 satellite projects aiming at providing Hive Cell functionality to allow users to create, import and manage sample registries, to manage subject/sample cohorts and to support the integration of regulatory systems within the hive. i2b2-Crimson is an initiative of the Brigham & Women's Hospital in Boston that complements the hospital's prospective specimen collection services 'Crimson Biospecimen Core'[50]. In consequence, i2b2-Crimson links i2b2 with the web information system of Crimson Biospecimen Core' and potentially other LIMS as shown in Figure 15.

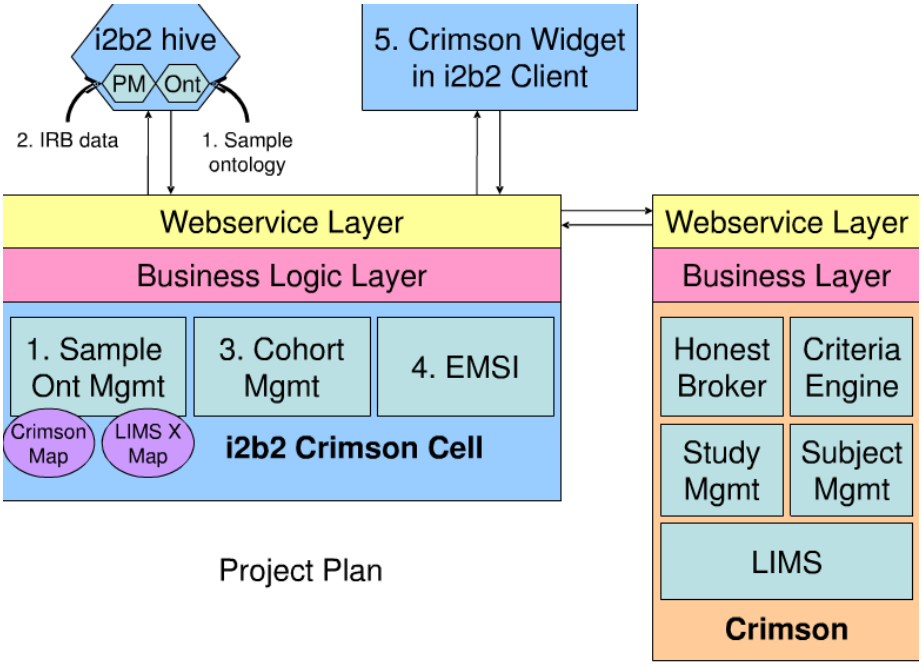


Figure 15: Crimson i2b2 cell. [51]

The development roadmap announces a first version for sample search and request in May 2011. However, only a demonstration is presented on the Crimson Webpage so far. Version 2 shall support multiple sample ontologies while a third version shall allow a federated i2bs Hive sample search[52]. Sample ontology is presented on the Crimson homepage in form of an Excel table. The sample ontology represents a hierarchy that describes different types of specimen. A demonstration of the Crimson functionality in the i2b2 Hive shows enrichment of the i2b2 querying and analysis tool that allows querying and managing samples from different cohorts. It also shows the integration of multiple ontologies such as 'Anatomical Source of a Biosample'.

[50] http://www.partners.org/researchcores/clinical/specimen_BWH.html

[51] Figure taken from i2b2 crimson presentation accessible under http://i2b2crimson.partners.org/i2b2crimson/index.php

[52] http://i2b2crimson.partners.org/

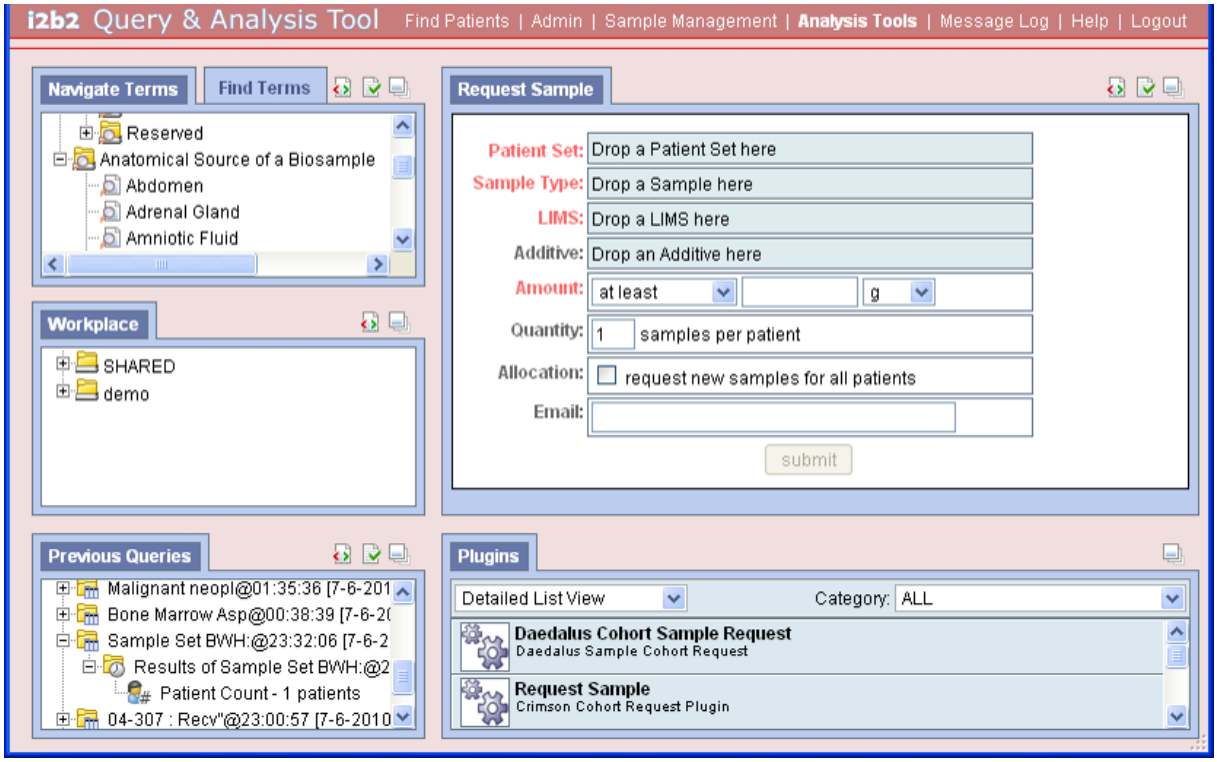


Figure 16: i2b2 Query and Analysis Tool with integrated sample management. [53]

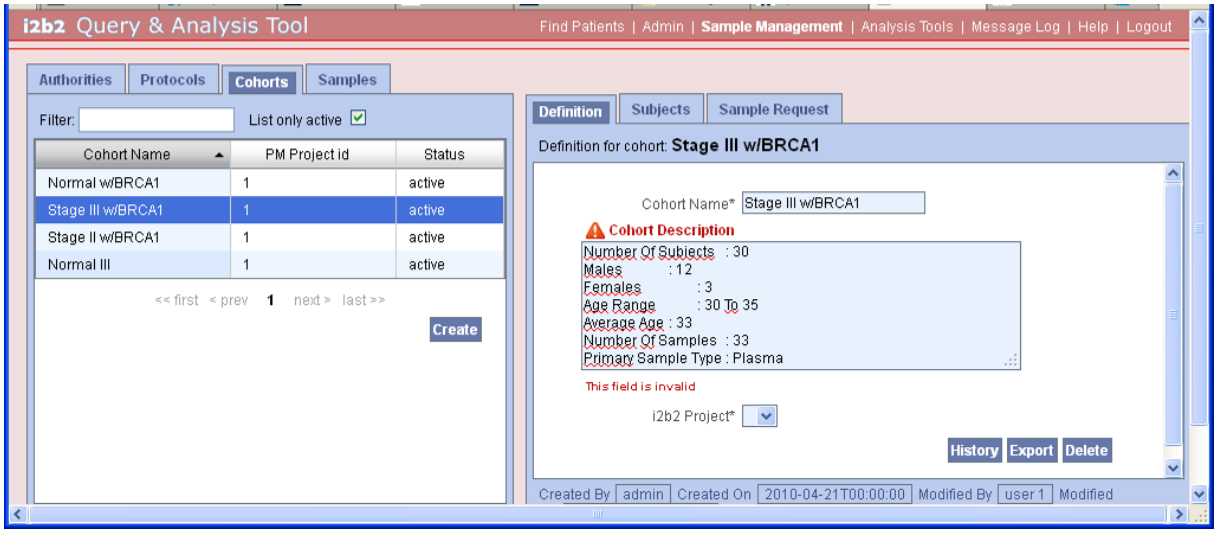


Figure 17: Cohorts definition and sample request in i2b2 Query and Analysis Tool. [54]

5.3.4 Data import, export and query functionality

A special Hive Cell named Import Data Plug-in is available to users to upload specific data into the i2b2 database that comprises observations, patients and encounters. An import wizard guides the user through the import process. Data sources are CSV files that have to follow a specific format. The user guide of the wizard indicates also that XML-Based Clinical

[53] Screenshot taken from i2b2 crimson demo server accessible under http://i2b2crimson.partners.org/i2b2crimson/index.php

[54] Screenshot taken from i2b2 crimson demo server accessible under http://i2b2crimson.partners.org/i2b2crimson/index.php

Experiment Data Exchange Schema (XCEDE) files and a particular database file format may be supported in the future. No information was available about the Export Cell.

The main functionality of the i2b2 Web Client and the i2b2 workbench is dedicated to the querying of the data in the DCRC Data Repository. Queries can be created by selecting and combining terms from the i2b2 ontologies and constraining occurrence or the dates on an item. Queries can be stored together with the query result which can include patient data.

5.3.5 Origin, license model and maturity

The i2b2 Hive application was developed by the i2b2 National Center for Biomedical Computing (NCBC), Boston, US and its scientific and academic partners. The i2b2 NCBC is one of currently seven NIH-funded NCBCs and is co-sponsored by Partners HealthCare System Inc. in Boston, US. The i2b2 project is further maintained by the Brigham and Women's Hospital, Inc. The current version is 1.5.4 and was released during June 2011.

The open source software of i2b2 is distributed as a non-exclusive, royalty-free, irrevocable, non-terminable license. All changes and exchanges have to deliver to the Brigham and Women's Hospital. Any commercial use is prohibited without a written permission.

The NCBC was established in 2004. The first i2b2 version contained a prototype of the CRC and associated stand-alone analytical tools. Several scientific collaborations (e.g. hospitals of Harvard Catalyst CTSA, Morehouse School of Medicine, and University of Utah) turn out a scalable application that enables a close interaction between several cells and the CRC. The current i2b2 version 1.5.2 is available since 22nd June 2011. i2b2 Hive is further developed as a community based open source initiative under the leadership of i2b2 NCBC and Partners HealthCare System Inc. The "i2b2 Academic Users' Group" offers participating institutions a forum in order to support implementing and collaboration work of the i2b2 software. All members participate in a discussion site and are encouraged to attend yearly meetings. Current membership includes around 100 individuals from 43 institutions, including six international institutions. The last meeting was at June 28th – 29th 2011. While eight disease-based driving biology projects served as a testbed of i2b2 Hive further information about the real usage in clinical research could not be found on the i2b2 website.

Overall, i2b2 Hive is a data warehouse based approach of a scientific software platform for the integration and analysis of heterogeneous clinical and related biomolecular data that only partially overlaps with the goals of the p-medicine platform. With the i2b2 Crimson Cell a sample management and access interface has been developed that obviously supports in particular biomaterial provision services of the Brigham & Women's Hospital who grants i2b2 Hive licenses. In principle, similar use cases as in p-medicine's biobank access framework seams to be supported by i2b2 Crimson, however it is unclear whether this software is mature and available as open source project, how much it depends on the Crimson Biospecimen Core and how external biorepositories can be linked to it. In addition, the license model foresees that any modifications of the i2b2 Hive have to be made available as source code to the Brigham & Women's Hospital while commercial exploitation is excluded. It is unclear whether this applies also to additional Hive cells developed by other partners of the community. Furthermore, efforts to adapt the software to the needs of p-medicine and in particular for biobank access and to integrate them with the p-medicine infrastructure could be considerable and is a risk.

5.4 SIMBioMS - System for Information Management in BioMedical Studies

5.4.1 General Description

Purpose

SIMBioMS[55] is a multi-module solution, web-based open source software system for managing data and information in biomedical studies. It provides a solution for the collection, storage, management and retrieval of information about research subjects and biomedical samples as well as experimental data obtained using a range of high-throughput technologies, including gene expression, genotyping, proteomics and metabonomics. The system can be customized and has proven to be successful in several large-scale multi-site collaborative projects. It is compatible with emerging functional genomics data standards and provides data import and export in accepted standard formats. In addition, protocols have been implemented for assay data transfer to the permanent data archives ArrayExpress[56] and European Genotype Archive[57] of the European Bioinformatics Institute.

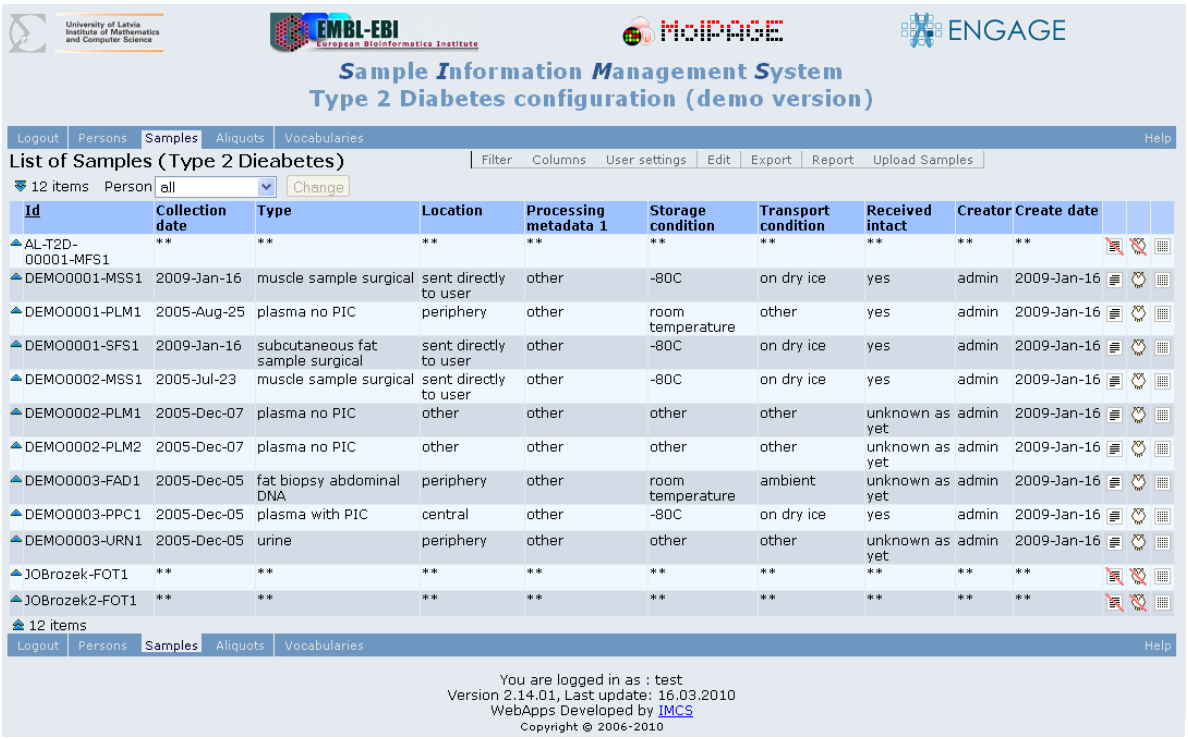


Figure 18: SIMS screenshot. [58]

Components

The system consists of following components:

- *Sample Information Management System (SIMS)*: System to collect phenotypical, environmental and technical information about samples. SIMS also provides a

[55] http://simbioms.org/

[56] http://www.ebi.ac.uk/arrayexpress/

[57] http://www.ebi.ac.uk/ega/

[58] Screenshot taken from SIMS demo server available under http://simbioms-pub.mii.lu.lv/SIMS_T2D/

solution for data anonymization by creating identifiers linked to person's information in a separate module.

- *Assay data and Information Management System (AIMS):* Management of experimental data form high-throughput assays
- *Sample avAILability system (SAIL)*: Indexing of phenotypes availability in different cohorts and collections

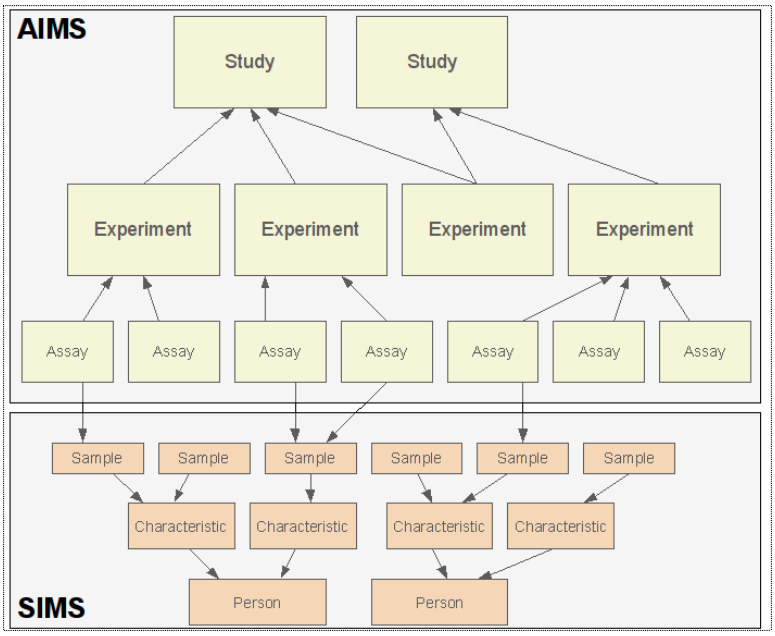


Figure 19: Data logic of SIMS and AIMS and interaction.

SIMBioMS run in Apache Tomcat servlet containers, or other application servers. The data are stored in PostgreSQL databases. The systems are platform independent. Several preconfigured versions, including ones for type 2 diabetes, metabolic syndrome and autoimmune diseases are packed into .war web-application archives. AIMS/SIMS can be installed either as local (e.g. on a laptop) or as centralized databases.[59] The web interface of SAIL was developed in Java using the Google Web Toolkit and Ext-JS widget libraries.

AIMS is linked to SIMS through a three-level hierarchy (Figure 19). A person can be linked to one or more samples, a sample can have one or more aliquots. Each aliquot can have one or more assays performed on it, and each assay can be linked to one or more data files. Assays are grouped in experiments and studies (Type 2 Diabetes, Metabolic Syndrome and Autoimmune Disease), each of which can have one or more data files attached. For instance, raw microarray data files would be normally linked to individual assays, while normalized gene expression matrices to experiments. Assays are technology-specific; the current AIMS configurations include genotyping, sequencing, proteomics and metabonomics.

The system can be configured to support the minimum reporting guideline for biological and biomedical investigations (MIBBI)[60].

[59] Krestyaninova M, Zarins A, Viksna J, Kurbatova N, Rucevskis P, Neogi SG, Gostev M, Perheentupa T, Knuuttila J, Barrett A, Lappalainen I, Rung J, Podnieks K, Sarkans U, McCarthy MI, Brazma A: "A System for Information Management in BioMedical Studies—SIMBioMS", Bioinformatics 2009 Oct 15; 25(20):2768-9. Epub, 2009 Jul 24.

[60] Taylor et al.: "Promoting coherent minimum reporting guidelines for biological and biomedical investigations: the MIBBI project", Nature Biotechnology Volume 26, 8, 2008.

5.4.2 Data harmonisation, data import, export and search functionality

The vocabulary used by SIMS is available in a special tabulator of the system and can be seen in the demo system. It mainly covers the terminology used by biobank software with a focus on human samples and SIMS use cases such as diabetes.

SAIL ("Sample Availability System") is designed to hold phenotype availability information and metadata about samples, experiments and phenotypes, submitted by data owners or databases that contain actual measurement data. SAIL provides a way to browse and summarise complex content across diverse resources and represents as such the search engine of SIMBioMS. This can be used in many different scenarios from designing new studies, understanding previous studies, and finding data and biomaterial from different cohorts that can be combined in meta-analysis studies. SAIL also contains functionality to annotate collections, with tools to create new vocabularies or use terms from standard sample ontologies, and to combine and harmonise vocabularies. The system can handle real valued data as well as availability data, which is useful to match study samples from different cohorts and to find the availability of samples in certain ranges of study parameters or measurements.

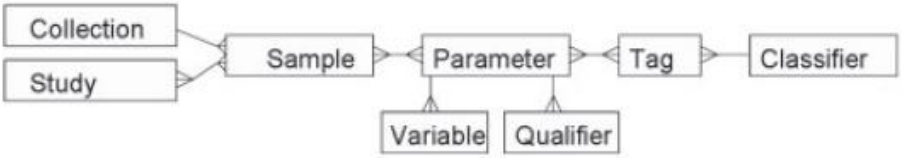


Figure 20: High level data structure in SAIL according to Gostev et al.

As illustrated in Figure 20, Gostev et al explained the concept behind SAIL as follows: Samples are grouped in collections (e.g. cohorts); collections may be annotated with information common to the collection. Parameters (i.e. glucose) are described by variables (i.e. concentration) and qualifiers (i.e. timing: fasting/non-fasting). Parameters can be grouped using tags. Tags are assigned values to identify the parameter group. Tags can be used to group parameters that come from a specific vocabulary, parameters that are used in the definition of a specific disease or parameters that are synonymous with each other. Tags can themselves be grouped into classifiers that can be used for more general parameter classification. For example, tags for different vocabularies can be grouped into the classifier 'Vocabulary'. Similarly, tags for specific types of parameter relationships are grouped into the classifier 'Relation'. When a tag is added to a parameter, the user indicates the value of the tag (such as the specific vocabulary) and the classifier to which the tag belongs ('Vocabulary'). Tags can be used to map external taxonomic structures to data structures in SAIL. Using such a flexible semantic structure SAIL can accommodate parameters from several vocabularies and store relations between them enabling searches that span across more than one data collection as well as for samples only partially matching the search term. Relations can be defined by the users or data providers and can vary from generic forms (i.e. synonym or partial match) to more detailed forms to express associations between specific vocabularies (or even specific for a group of terms)[61].

Browsing and customizable data filtering options allow content exploration and report construction on metadata level. Selected data can be imported into analysis tools, such as Bioconductor.

SIMBioMS provides customizability and compatibility with the microarray metadata format MAGE-TAB and the ISA-Tab format[62], a biological file format for experimental metadata. The

[61] Gostev, M. et al.: "SAIL – a software system for sample and phenotype availability across biobanks and cohorts". Bioinformatics (2011) 27 (4): 589-591.

[62] http://isa-tools.org/

SAIL tool supports importing comma or tab delimited text files in a specific format where column headers references variables.

The execution of queries is not dependant on the relational database, and can be formulated using linked data solutions such as RDF, OWL and SPARQL. However, it was recommended to use a simple customized semantic data structure instead, in order to maximize performance and to be able to flexibly contain annotation structures from different sources.

5.4.3 Origin, license model and maturity

The three software modules of SIMBioMS were developed as a part of the FP6 Integrated Project MolPAGE (Molecular Phenotyping to Accelerate Genomic Epidemiology) and the FP7 collaborative health research project ENGAGE (European Network of Genomic and Genetic Epidemiology)[63]. The copyright of SIMS and AIMS have the Institute of Mathematics and Computer Science (IMCS), University of Latvia and European Bioinformatics Institute (EMBL-EBI), European Molecular Biology Laboratory.

The SIMS, SAIL and AIMS are available under the Affero GNU General Public License.

As of 2009 there were three SIMBioMS instances running to support collaborative projects, including an instance containing data from over 25 000 assays and an instance for population-wide epidemiology studies. Data archives from over 6500 assays have been transferred from Array Express and European Genotype Archive (EGA) [64]. The SIMBioMS platform was discussed during the workshop “Implementation and usage of biobank online inventories, collaborative data availability portals and national BBMRI portals” at the Institute for Molecular Medicine Finland, Helsinki, 24 Nov 2010.

SAIL as the system to browse and search samples and phenotypes across diverse data collections, provides an instant view of biomaterial and phenotype availability in major European biobanks, currently with focus on metabolic syndrome phenotypes in ENGAGE.[65]

In December 2010, the largest of in total four running SAIL instances provided access to data of approximately 189000 samples from 14 collections.

Regular upgrades, 1-2 times a year, of SIMS, AIMS and SAIL are announced on the homepage. The latest SIMS version was released on 14th Apr 2010, the latest AIMS version on 15th Nov 2010, and the latest SAIL version on 19th Nov 2010.[66]

To conclude SIMBioMS is an open source web based platform that integrates capturing phenotypic data with the management of high-throughput data from multiple platforms. At a first glance the SAIL subsystem fulfils requirements for biobank access in p-medicine. Although it has been developed with a focus on biobanks and biological sample collections, its design allows for the integration of data from other biodatabank sources where information can be arranged in annotated records. However, available technical information is rare and it is unclear what efforts would be required to wrap a biobank information system to SAIL. Furthermore, software sources need to be further investigated whether adoption for p-medicine would be an option. As SAIL is provided under the GNU GPL licence its use in p-medicine would have tremendous implications for the exploitation of the p-medicine biobank access framework and the related components.

---

[63] http://www.euengage.org

[64] Krestyaninova M, Zarins A, Viksna J, Kurbatova N, Rucevskis P, Neogi SG, Gostev M, Perheentupa T, Knuuttila J, Barrett A, Lappalainen I, Rung J, Podnieks K, Sarkans U, McCarthy MI, Brazma A: “A System for Information Management in BioMedical Studies—SIMBioMS”, Bioinformatics 2009 Oct 15; 25(20):2768-9. Epub, 2009 Jul 24.

[65] http://www.euengage.org/resources.html. Last visit 2011 Sept 15

[66] http://bioinf.mii.lu.lv/simbioms/simbioms_downloads.html. Last visit 2011 July 11

5.5 P3G DataSHaPER

The Data Schema and Harmonization Platform Epidemiological Research (DataSHaPER)[67] [68] is a suite of practical tools that have been developed by a multidisciplinary consortium of experts, coordinated by three international organizations: Public Population Project in Genomics (P^3G), Promoting Harmonization of Epidemiological Biobanks in Europe (PHOEBE) and Canadian Partnership for Tomorrow Project (CPT). The DataSHaPER provides a flexible and structured approach to the harmonization and pooling of information data between studies and aims to facilitate the prospective harmonization of emerging biobanks, a template for retrospective synthesis and supports the development of questionnaires and information-collection devices, even in cases where data pooling with biobanks is not foreseen.

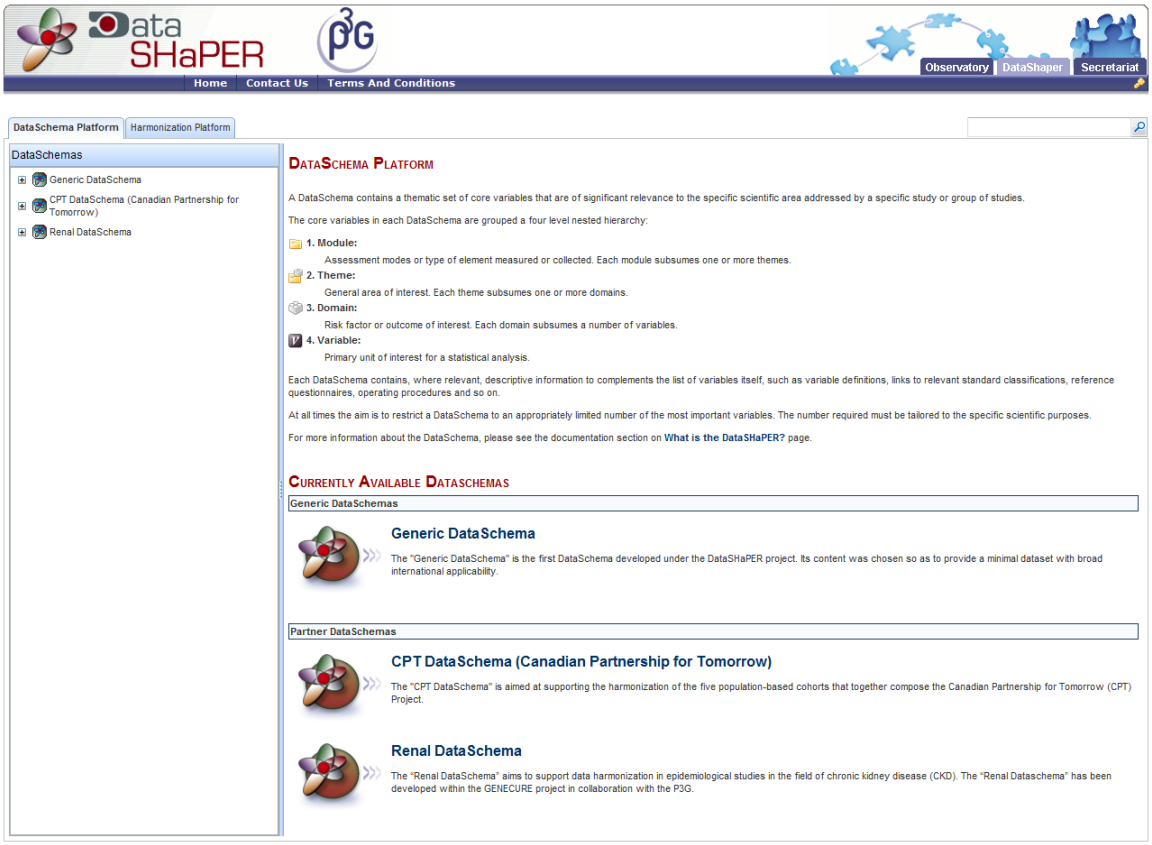


Figure 21: A DataSchema identifies and describes a thematic set of core variables that are of particular value in a specified scientific setting [69].

The main structure of DataSHaPER reflects the following four steps:

- Identifying and documenting the set of variables to be shared;
- Assessing the potential to share each variable between participating studies;
- Defining appropriate algorithms for data processing;

[67] http://www.datashaper.org/

[68] "DataSHapER", http://www.datashaper.org/

[69] Screenshot taken from DataSHaPER demo web client available under http://www.datashaper.org

- Processing and synthesizing real data;

The two primary components of DataSHaPER are the DataSchema and the Harmonization Platforms, which together offer a way of effective data-collection protocols and provide a central reference to facilitate harmonization. The suite supports "prospective as well as "retrospective" harmonization. The DataSchema Platform provides a description of a set of basic variables that have significant value is specific scientific setting.

For each DataSchema, these values are grouped under a four level hierarchy:

- Module, which is the type of elements that have been collected or measured;
- Theme, which constitutes the general domain of interest;
- Domain, which is the risk factor;
- Variable representing the primary unit of interest for a statistical analysis;

Each module subsumes one or more themes, each theme is associated to one or more domains and each domain contains a number of variables. The DataSchema platform consists of a number of DataSchemas and each such schema has its own scientific purpose. Furthermore, the platform has definitions of variables, links to relevant standard classifications and access to reference questionnaires and functional procedures.

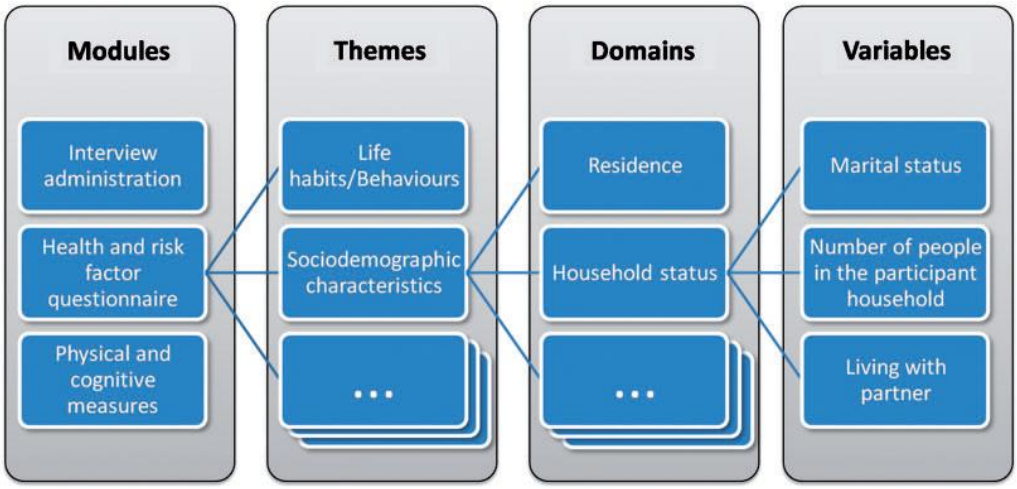


Figure 22: Hierarchical structure of the module, theme and variables related to the 'household status' domain in the Generic DataSchema [70]

DataSchemas in the platform are partnered by the corresponding Harmonisation Units that constitute the foundation for harmonizing studies related to the schemas. This is the Harmonization Platform component and access to it is limited to collaborative context.

The whole suite is written in java, utilizing open-source libraries. The technologies of Spring Framework, Hibernate, Google Web Toolkit and Sesame/Elmo have been used. The user interface is based on AJAX technologies (Asynchronous Javascript and XML), providing enhanced usability. Finally standard formats and ontologies have been used wherever possible. The key information bearing objects (DataSchemas and component elements of the Harmonization platform) are stored using a recognized ontological format, facilitating

[70] "What is a DataSchema?" document accessible under http://www.datashaper.org/download/Protocol_DataSchemaStructure.pdf

interactions with other applications. The DataSchema has adopted the terminology defined by the National Cancer Institute Thesaurus ontology in many relevant cases. The application is accessed freeware and open, however access to under development DataSchemas and relative generated results require authentication and validation of the users by the DataSHaper Team.

According to the terms and conditions section of the datashaper web site[71], the P3G Consortium is committed to placing the tools it develops in the public domain so as to support the international scientific community in improving the health of populations. Therefore, third parties are encouraged to access, read, use and download the Content placed on the Sites for personal, educational, research, public non-commercial use and for not-for-profit use only.

In summary, the datashaper will successfully harmonize data and foster interoperability of epidemiological cohorts and population-based biobanks. It has been neither designed for nor widely applied to disease-specific biobanks like those active with p-medicine. Although validation and adaptation of the datashaper to all types of biobanks would certainly be an interesting project, at present the datashaper tool as such is not deemed an appropriate basis for the p-medicine access-to-biobanks framework.

5.6 CRIP Toolbox

Since 2006, the Central Research Infrastructure for molecular Pathology (CRIP) [72] is regularly integrating data from the Institutes of Pathology / tissue banks of its partners Charité Universitätsmedizin Berlin, Klinikum rechts der Isar / TU Munich, Medical University of Graz and University Hospital Erlangen. Anonymized data on cases and samples stored in these tissue repositories is made accessible over a web-based interactive query tool, allowing users to easily create and launch project requests on-line. Instead of bringing bulk biobank data to users, it is CRIP's mission and policy to bring project requests to biobanks.

With CRIP, a model[73] has been established enabling biobank networks to form operational "meta biobanks" whilst respecting the donors' privacy, biobank autonomy and confidentiality, and the researchers' needs for appropriate biospecimens and information, as well as confidentiality. In 2010, the CRIP model has been adopted for the "Projektportal im Deutschen Biobanken-Register"[74] by the German Ministry for Education and Research (BMBF) laying the ground for a central national German biobanking infrastructure. In 2011, the interactive query tools of both CRIP and biobank suisse[75] have been coupled by web service protocols[76] and synchronized log-in procedures allowing for simultaneous search of both meta biobanks.

[71] "DataSHapER", http://www.datashaper.org/legal/termsAndConditions.htm

[72] http://www.crip.fraunhofer.de/

[73] 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.

[74] www.tmf-ev.de/p2b2

[75] www.biobank-suisse.ch

[76] http://www.crip.fraunhofer.de/en/info/crip_news/crip_bbs

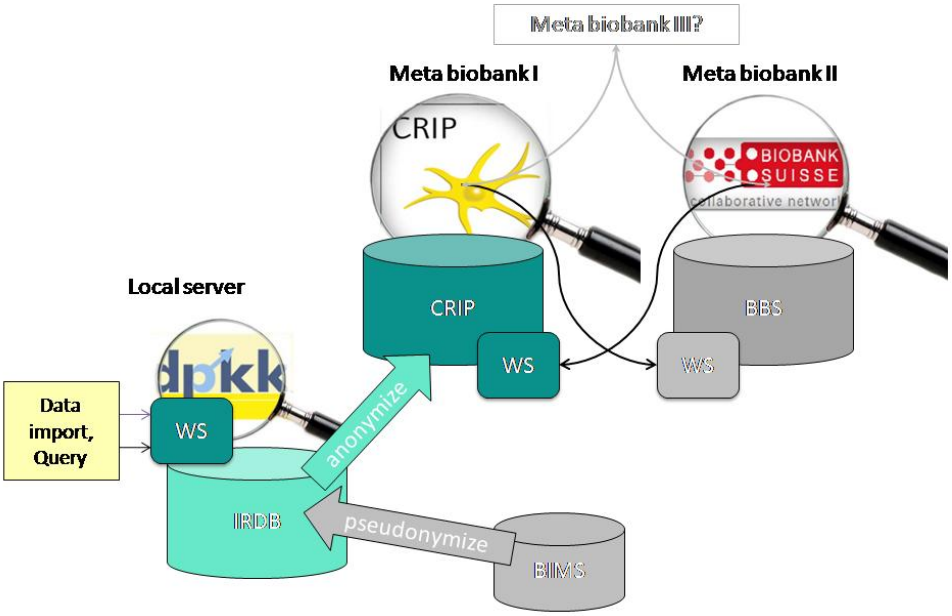


Figure 23: The CRIP Toolbox. All modules operated by CRIP software are shown in green: IRDB = Integrative Research Database; WS = Web service protocol; CRIP = Central database with interactive query and project request tool.

CRIP has been approved as a benchmark for data integration by data protection commissioners[77] and provides a comprehensive toolbox (Figure 23) for case-specific web-based queries including

- The Integrative Research Database (IRDB): A local server with hospital- or consortium-wide clients, harmonizing and anonymizing data for export and minimizing the workload for biobanks to connect.
- Graphical User Interface of the Central Database for Web-based queries, containing anonymous data and displaying them as statistical groups only (CRIP Privacy Regime[78]).
- Search tools for local and central databases
- Project tool including definition and entry of specific project requests.
- Web services for queries and for federating databases.
- Annotation tools for pathology reports (not shown in Figure 23).

The exchange of data using the CRIP tools and web services is mainly based on XML. Data integrity can be checked with the appropriate document type definition, DTD, file.

5.6.1 Integrative Research Database (IRDB)

The IRDB provides means to process pseudonymized data of different origin at the hospital's or biobank's site. The data structure is defined in a metadata repository, describing the format and nature of the data. Data to be imported, in CSV or XML format and selected by "drag and drop", is checked for consistency, reformatted and stored in a local database

[77] http://www.crip.fraunhofer.de/en/ethics_policy/data_protection
[78] http://www.crip.fraunhofer.de/en/ethics_policy/privacy_regime

server, according to the definitions in the metadata repository. Each data set is linked to a site ID, i.e. data from different divisions can be stored and administered in the IRDB. Anonymized data can be exported and uploaded to the central database for each site in CSV or XML format. The software provides basic functionalities, e g. search for data records, in order to exclude them from exports. This is necessary, if a patient has withdrawn his/her consent to use his/her data for research purposes. Other functionality is the data filtering based on a project search profile. For more complex database operations and/or integration into other hospital/consortium applications, the IRDB provides a set of web services.

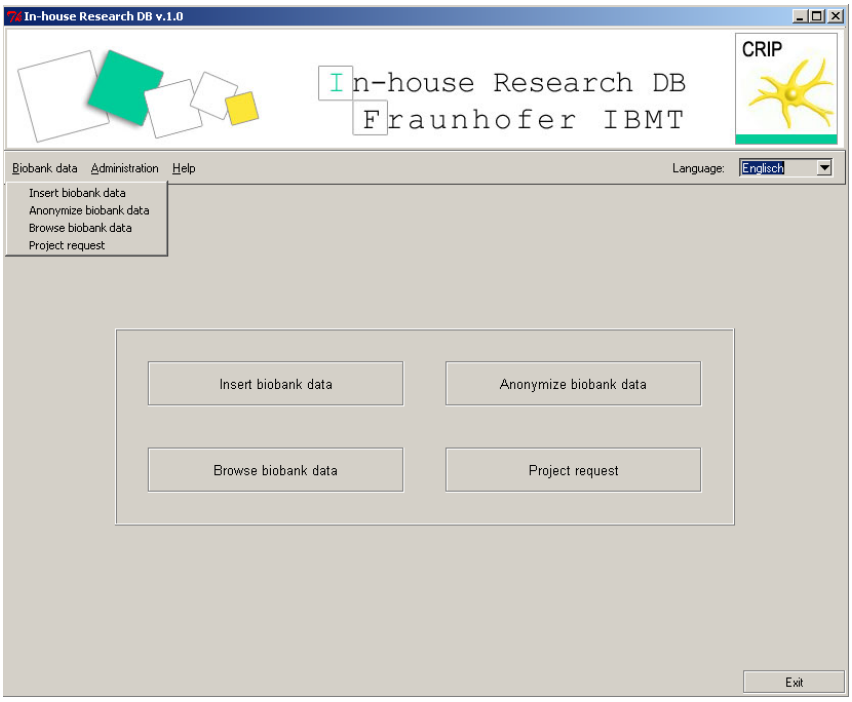
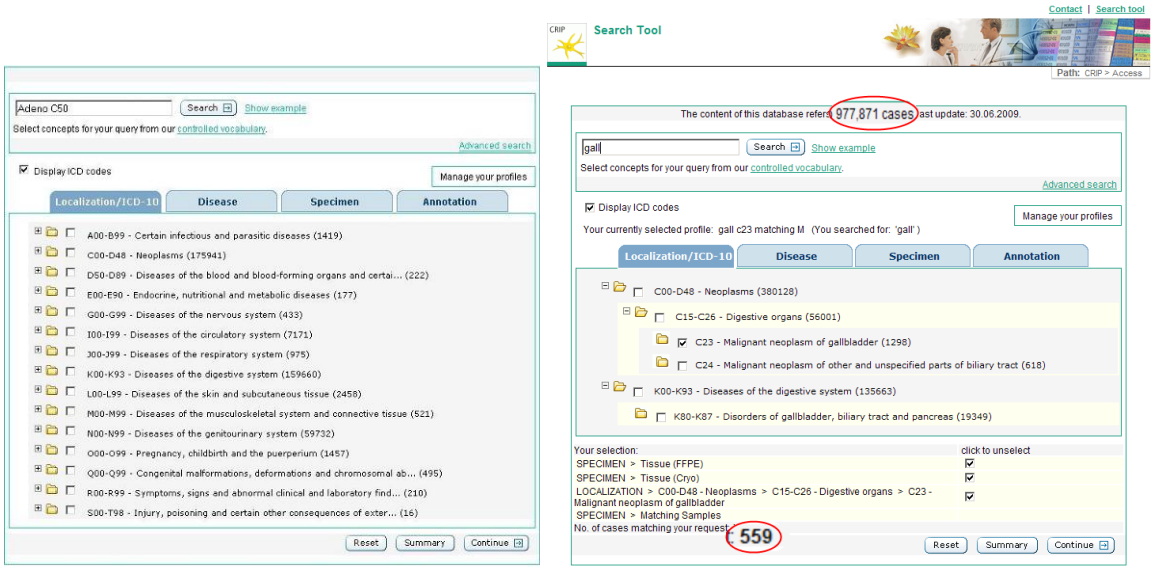


Figure 24: Graphical User Interface of IRDB.

5.6.2 Search tools for local and central databases

Based on the same software modules, the tools for local and central databases provide functionalities to import, search and administer data stored in the database. The import of anonymized data, in CSV or XML, will entirely replace existing data from the same database partner in the database. Subsequently, the search index for case based searches is rebuilt and optimized. The web-based frontend of the database allows the definition of a search profile according to the classification and annotations stored in the database. Search profiles may be saved, changed and refined later.

The frontend provides cards with, by default, four categories: *Disease*, *Localization*, *Annotation*, and *Specimen*. *Disease* and *Localization* are based on ICD10 and ICD-O, but the software allows the definition of other classifications to map the data onto any classification scheme, thus it is possible to offer different “views” on the data. The card *Annotation* is depending on the annotation of the data, provided by the database partner’s IRDB. An upload form accepts uploads from the IRDB and the data is interactively imported into the central database by a data manager, who assures data integrity and quality. The result of searches is presented as statistical groups, guaranteeing a k-anonymized view on the data. With the statistical group the providing hospital or consortium can be shown or the CRIP project tool can be launched to send requests for projects.



a) **b)**

Figure 25: CRIP Search Tool: a) Overview localization/disease; b) selection of specific cases and samples: FFPE and cryo-preserved tissue of malignant neoplasm of gallbladder

5.6.3 Project tool including definition and entry of specific project requests.

The project tool is an add-on to the search tool. A web-based frontend allows the selection of statistical groups from the search result. Afterwards the user can describe the scope of the project and request additional services. The result of the project tool is a *search profile* that can be applied on the data stored in the IRDBs, delivering the list of locally available cases matching the user's project request/search criteria.

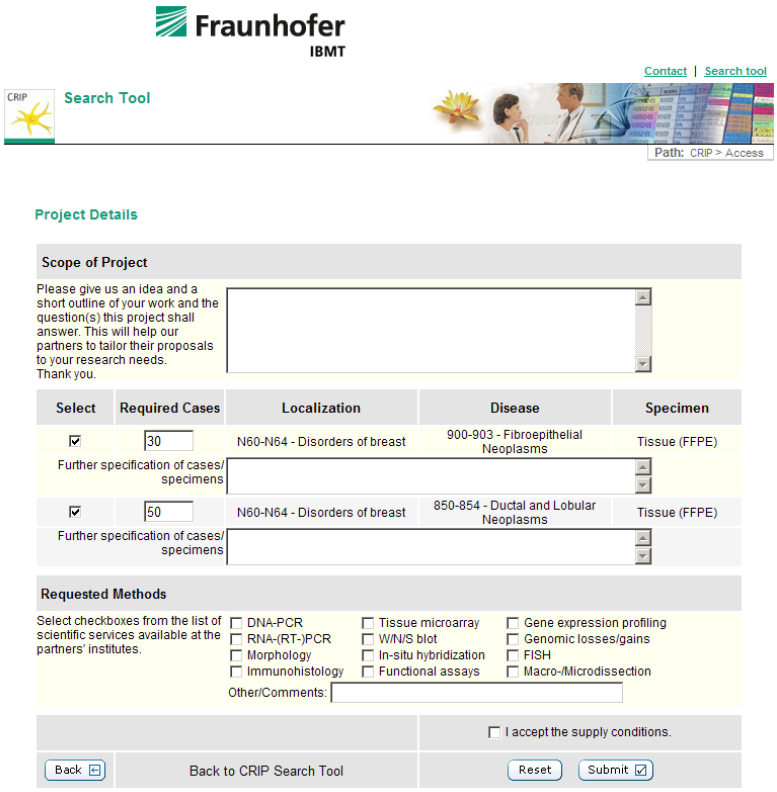


Figure 26: Project Search Tool.

5.6.4 Web services for queries and for federating databases

The CRIP toolbox includes a number of web services. The web services exchange encrypted and secured XML data delivering results by an AJAX protocol. This enables communication not only with other CRIP databases, but also allows integration of local and central databases into a federated database environment as proposed by BBMRI ("hubs and spokes") and recently demonstrated with the biobank suisse. In the latter case a search, executed in the interface of the biobank suisse by an registered user, is propagated to the CRIP central database, and the result is presented in the interface of the biobank suisse, see Figure 23, and vice versa.

5.6.5 Annotation tools for pathology reports

Amongst the many tools provided by the CRIP toolbox, the semi-automated annotation tool provides means to extract classifications and annotations from free text in pathology reports and is needed as a preparatory tool for data, that is not or only insufficiently structured and categorized. Computer understandable classifications, dictionaries, and categories are a prerequisite for organizing such data in meta biobanks. The input into the tool is an XML-based file, with a set of pathology reports. The free text in the report describing the case is filtered based on an expert rule set. This expert rule set has been developed by pathologists from the Medical University of Graz and the Klinikum rechts der Isar / TU Munich and is stored in a knowledge base at Fraunhofer IBMT. The tool retrieves the rules needed for text processing on-line and extends the report with the extracted classifications, annotations (e.g. ICD and TNM). The pathology reports are then saved as an XML-based file.

5.6.6 Licensing and management

Being convinced that, as a prerequisite for successful federated biobanking, proper management is at least as critical and important as state-of-the-art ICT, IBMT does not provide the components of the CRIP toolbox "open source" but, as a rule, regulates cooperation within a biobank network in a so-called "Database Contract". Under this contract the CRIP software tools are licensed free of charge to IBMT's partners and would, of course, be also available to the p-medicine consortium.

Figure 27: Exemplary license for CRIP toolbox (for a partner of "Projektportal im Deutschen Biobanken-Register").

5.6.7 Conclusions

The backbone of the CRIP architecture is built by both the IRDBs located in-house, i.e. behind the firewall, of the biobanks/hospitals, and by the Central Database and Search tool, which are accessible to researchers over the www. CRIP is in line with BBMRI's recommendations on federated biobanking infrastructures (chapter 5.3). In comparison with

the architecture proposed by BBMRI (Figure 17), particularly the following services are already realized by CRIP tools (Figure 28):

- IRDB: “BB-Interface” / „Mediator“ / “Upload Service” / “Disclosure Filter”.
- CRIP-DB: “Content-Meta-Structure”.
- CRIP Search tool: “Result ranking”, “Query service” / “Service request/response”.

Therefore the CRIP tools are also deemed useful components of the upcoming p-medicine Access-to-Biobanks Framework.

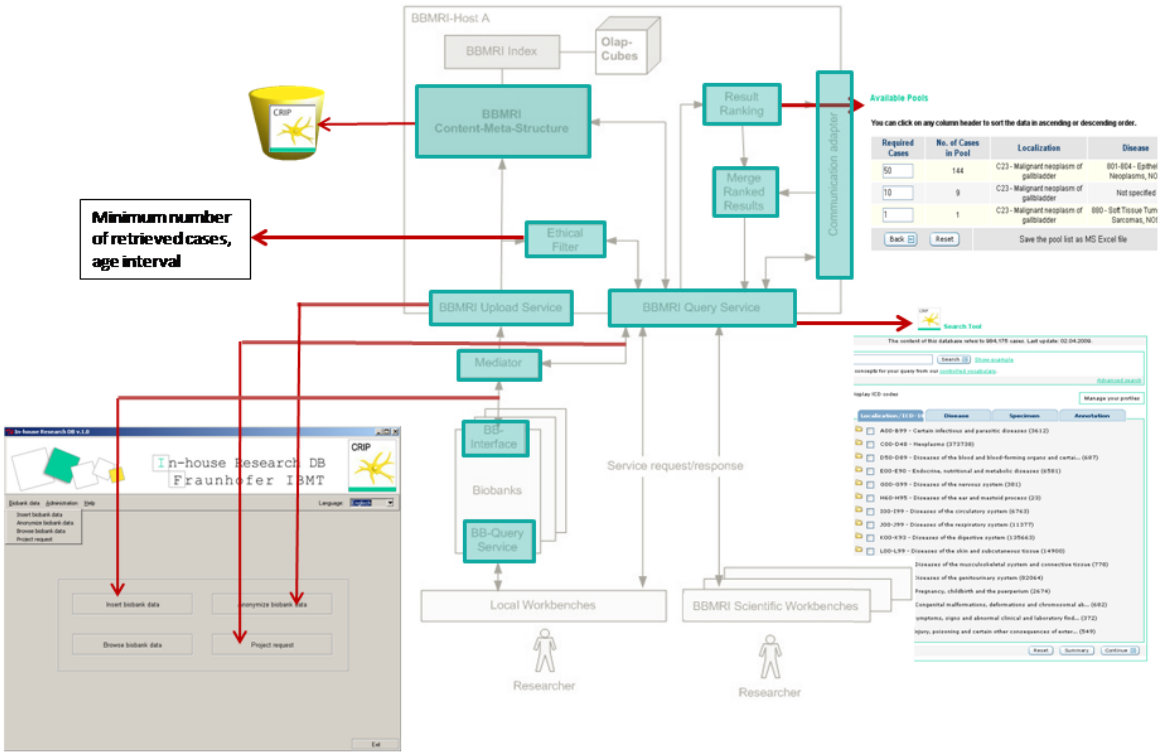


Figure 28: Services proposed for BBMRI Prototype A (in grey) which can be covered by CRIP tools (green).

5.7 eurocryoDB and eurocryoPortals

eurocryoDB and eurocryoPortals represents the non-commercial information system platform of the Fraunhofer Institute for Biomedical Engineering (IBMT) for operators of cryo-biobanks and their customers that manages the acquisition, preparation, storage, retrieval and distribution of biomaterial under quality controlled conditions.

eurocryoDB and eurocryoPortals have been designed as part of IBMT’s patented Fraunhofer Cryostorage Technology (FCT)[79] for high quality biobanking. The fundamental components of FCT represent electronic cryo-vials containing low temperature memory chips that allow attaching data to samples, for instance before shipment or during preservation. Based on these cryo-vials FCT further comprises completely new, self organizing, semi-automated

[79] http://www.sysmex-bioscience.com/Icebreaker-208-2.html, last visited on 15-09-11.

cryo-storage containers and cryo-work benches that ensure maintaining the cool chain in sample logistics. These smart cryo-vials interface with docking devices which offer secure data interaction between laboratory software and interactive samples. Upon receipt of samples the docking device software supports importing sample data from smart vials into the local information system and offers mechanisms to feed back information to the sample provider. In that way aliquot (and data) tracking through a sample-data-sample loop and sample lifecycle management is greatly facilitated.[80]

Figure 29: Smart cryo vial with integrated memory chip and RFID tag fully operational under –196°C.

5.7.1 Architecture

Storage and provision of specimens are centrally managed by the central biobank information and administration system ***eurocryoDB***, which can handle conventional storage inventory as well as seamlessly interact with Fraunhofer's intelligent Cryostorage Technology through the cryo-LIMS <LabOS>[81]. eurocryoDB is a web based solution that administrates and manages all business processes and workflows in the biomaterial logistics of the biobank and their external partners. It administrates samples/aliquots, donors/subjects, customers and the inventory.

The data model of eurocryoDB allows managing different sample collections with different characteristics independently from each other in so-called virtual sample repositories. So far three different types of virtual repositories are supported with their own specific data sets. For each repository an own user group can be specified according to the sophisticated roles and right concept of the system. The data sets of each repository include a shared minimum data set in addition to a fixed repository specific data set and are described through an XML schema with specific nomenclature. The schema simplifies import and export filters to other related LIMS or biobank information systems. Attributes could be set that indicate the legal status of a sample with respect to patient consent and ownership in order to indicate availability of the sample for research.

[80] http://www.ibmt.fraunhofer.de/content/dam/ibmt/de/Dokumente/PDFs/ibmt-produktblaetter/ibmt-telematik-intelligente-gesundheitssysteme/MT_Eurocryo-DB_de.pdf

[81] http://www.sysmex-bioscience.com/Lab-OS-217-2.html, last visited on 15-09-11.

The next release is planned with an integrated editor that allow extending data sets through individual data items, which can be annotated by ontologies. Each specimen can have multiple aliquots at different storage positions, can be linked to a patient and can have one or more data files attached.

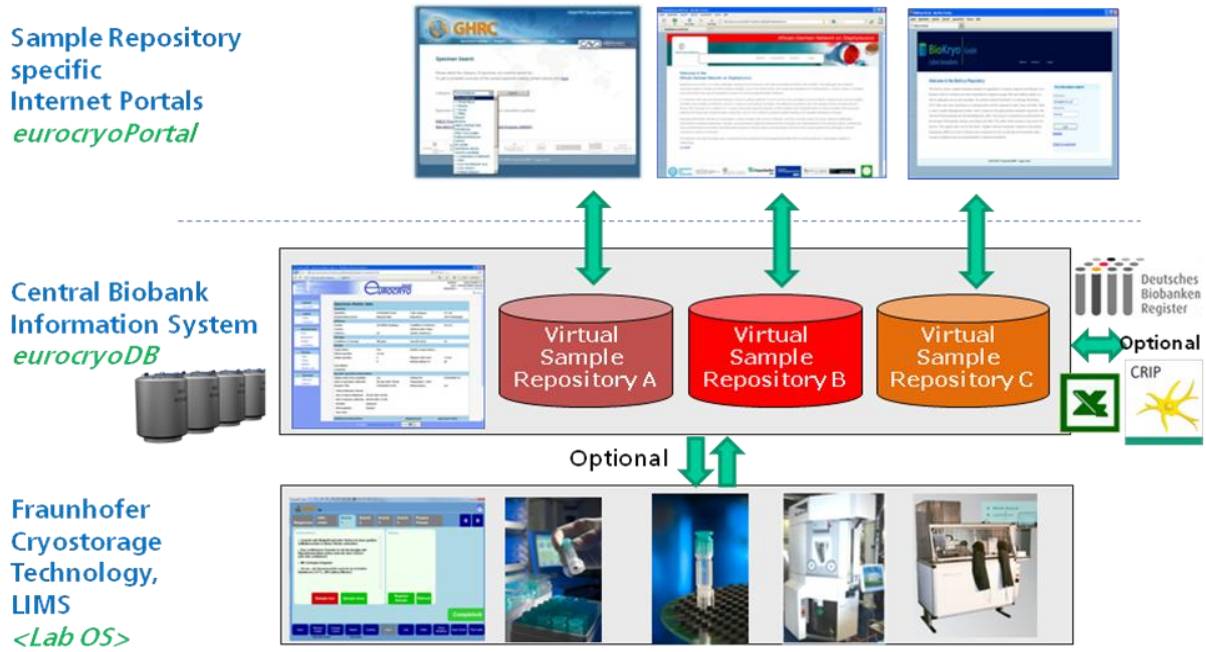


Figure 30: Overview of the eurocryoDB biomaterial management platform.

Each virtual repository of eurocryoDB is optionally associated with a web portal that is the "window" of the repository for external partners and the community. The *web portals* offer material provision services to third parties. The web application includes respective search and order mechanisms for biomaterial through a specific specimen catalog. It is also possible to register samples that are collected by external partners with their data prior to shipping to the biobank and to provide these partners an overview about their individual biomaterial stock in the repository. Orders and registrations are forwarded to the core system eurocryoDB, where they are processed by the biobank staff. Several portals can operate in parallel on one eurocryoDB instance and synchronize their data with it. Each web portal has its typical functional and editorial characteristics and corresponds to a specific repository in the eurocryoDB.

5.7.2 Implementation, interfaces and availability

The eurocryoDB web application runs in an Apache Tomcat servlet container. The data are stored in an Oracle database. Data exchange with the other components is realised through web services technologies. Secure web service is available to download the biomaterial data set of a sample after proper authentication. This service is used by <LabOS> to write the data on a 'smart' sample's memory chip (FCT). In the opposite direction eurocryoDB can receive the storage location of a smart sample in a self-organisating storage system from <LabOS>. In addition to that, eurocryoDB provides an import interface that allows the automatic integration of specimen data and its storage position on the basis of tabular files. Reports can be generated and exported as CSV files. eurocryoDB also provides an interface to Fraunhofer CRIP tools (See chapter 5.7).

The eurocryo Web Portals run also in an Apache Tomcat servlet container, but with a PostgreSQL database.

Originally eurocryoDB has been designed as in-house solution of Fraunhofer IBMT for its cryo biobank research facilities. As such it hosts important biological research resources of Fraunhofer. Among them is the Global HIV Vaccine Research Cryo Repository that has been implemented on behalf of the Bill& Melinda Gates Foundation in order to provide high quality standardized HIV biomaterial provision services AIDS vaccine researchers all over the world.

Sources of eurocryoDB and eurocryoPortal are available to the p-medicine consortium through Fraunhofer IBMT and the application could be adapted to meet the needs of p-medicine's biomaterial access scenario.

5.8 OBiBA – Open Source Software for Biobanks

OBiBa is a collaborative international project that was founded in 2007 as a research project of Génome Québec and its purpose is to develop high-quality open source software for biobanks. It shortly became a core project of Population Project in Genomics Consortium (P^3G), which is an international organization that provides tools and resources to foster collaboration and knowledge sharing for genomic studies.

Opal is a database application that has been developed, that integrates biobank data from various sources into a secure centralized data repository under a uniform model. Its main objective is to achieve seamless data-sharing among biobanks. The tool provides interfaces for managing and querying a data element dictionary and for producing administrative and scientific reports, and supports a variety of other activities required for the analysis of collected data. The main operations of the current version of Opal (version 1.5) are listed below:[82]

- Store data on cohort participants and any other kinds of entity (instruments, samples, etc.);
- Import data from CSV files and Onyx data (Onyx is a web-based application that manages participant baseline interviews at assessment centres or clinics) export files;
- Manage cryptographic keys and data encryption/decryption for secure data transfers between Opal and its data sources;
- Manage participant identifiers from multiple sources;
- Manage, browse, query and export the variable dictionary (metadata) using web interface;
- Derive new variables from other data elements using custom algorithms;
- Create "Views" for organizing data into more manageable components;
- Compute summary statistics (average, standard deviation, frequencies, etc) for any variables;
- Connect to existing SQL databases (where other Participant data may exist);

Using Opal can simplify several daily things, like data importation from Onyx with different or configurable strategies, variable browsing (list, depth, transversal, keywords, filtering), creating basic study activity reports and multiuser support through a web based application. In addition more complex tasks such as annotations with ontologies or keywords, variable refactoring or aggregation and data security are provided. The system supports several interfaces. Those are a command line tool, a web-based graphical user Interface, and interfaces to external systems (Onyx and import from unknown formats in excel or xml) or devices.

[82] OBiBa, http://www.obiba.org/, accessed on 19-07-2011.

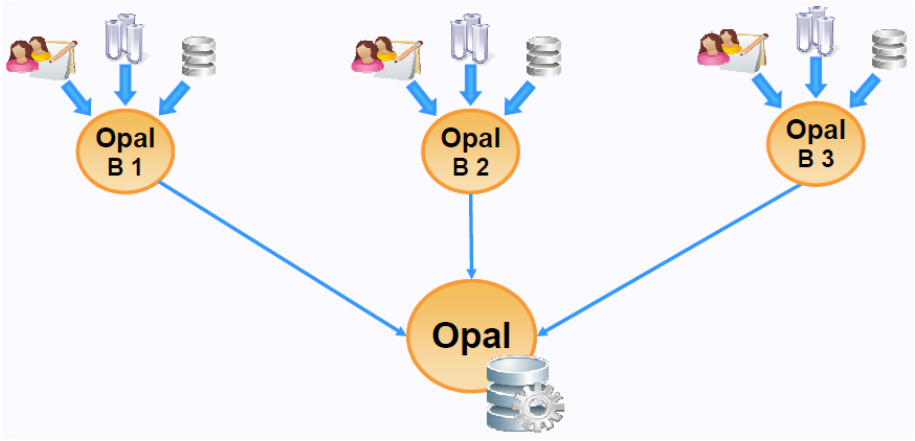


Figure 31: Opal - integrating, exchanging and analysing harmonized data among biobanks [83].

The website[84] announces a next version that will further include:

- An extensive reporting tool for administrative, quality control, and scientific purposes. A set of reports that have been predefined, will be nightly generated and will be available on-line (under restricted access) to help internal staff monitor the recruitment process and control the quality of data collected;
- An Opal R Application Programming Interface (API) that provides data access directly from R scripts;
- Researchers will request for datasets to be generated;
- A comprehensive set of web services to connect and query remotely other Opal instances in order to build federated database networks;
- An implementation of the DataShield method for performing pooled analyses of individual-level data without needing to share them;

In the context of participant privacy, data pooling without informed consent is prevented. In a 1985 pioneer paper, David Chaum[85] stated three general principles to avoid collusions:

- Each organization must use its own participant identifiers (IDs)
- New specific IDs must be used when exchanging data on participants
- IDs are accessible to a very restricted number of people within an organization

Opal has followed these three principles during the implementation process and Figure 33 shows how this is achieved.

The system separates the participant identifiers from the participant's data in two databases[86]. The Opal identifier database holds the participant identifiers and personal data and the opal data database holds anonymous participant's data.

[83] "Opal: An Open Source Platform for Data Integration" presentation accessible under http://www.p3g.org/secretariat/events/2010Montreal/pdf/Vincent%20Ferretti.pdf

[84] http://www.obiba.org/note/122

[85] David Chaum, "Security without Identification Card Computers to make Big Brother Obsolete", Communications of the ACM, 1985.

[86] Opal 1.4.x Documentation, http://wiki.obiba.org/display/OPALDOC/Home, accessed on 19-07-2011.

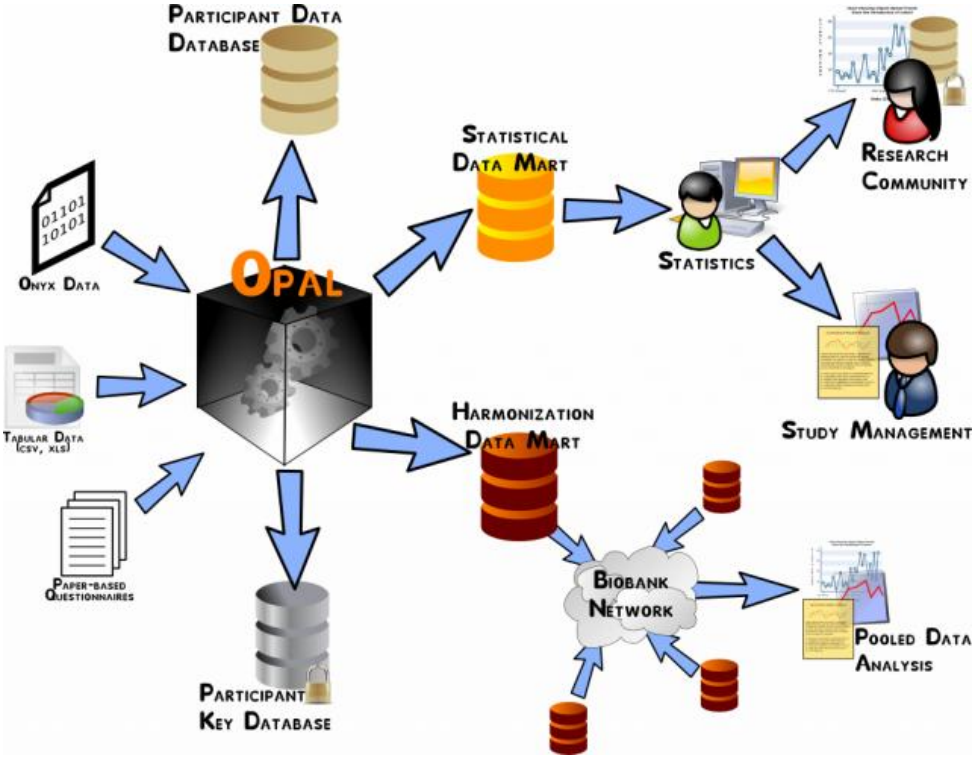


Figure 32: Opal's central role in data storage, analysis, harmonization, and sharing [87].

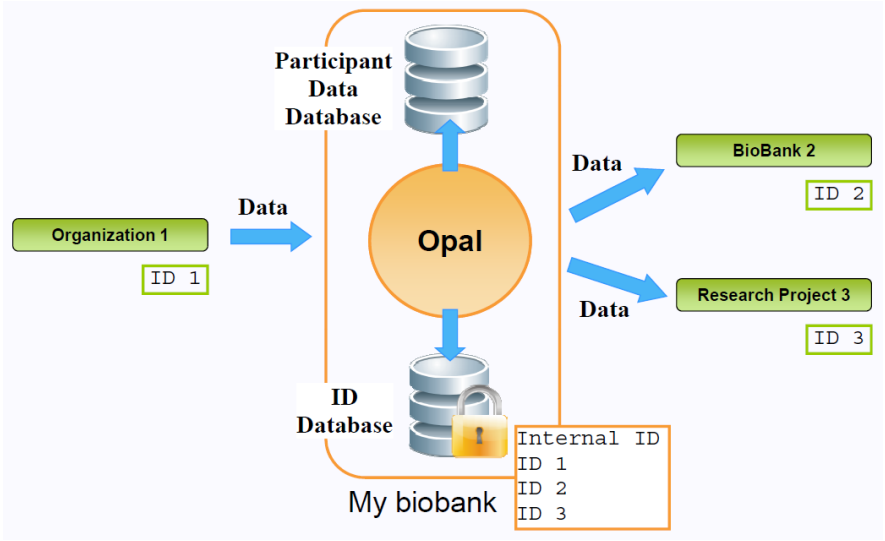


Figure 33: Opal complies with the three stated principles that can ensure the participant's privacy [88].

Furthermore, confidentiality is also ensured within the system by utilizing a comprehensive public key infrastructure (PKI) for the authentication of the organizations and data encryption. More precisely, Opal uses specific private-public key pairs for each organization with which data are to be exchanged and private keys to decrypt automatically received data from organizations.

[87] Picture taken from http://www.obiba.org/

[88] OBiBa Websie – Opal – accessible under http://www.obiba.org/node/131

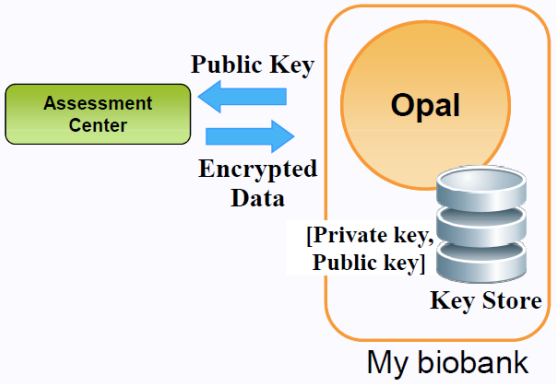


Figure 34: Data encryption within Opal [89].

To summarise, Opal is a database application for biobanks that has been mainly developed in order to capture OBiBa's requirements. Even though Opal seems to be fully compatible with other software, developed within the OBiBa project (like Onyx), it further supports features that suite with p-medicine's scope and task. Hence, importing data from general CSV files, utilizing encryption and decryption operations for securing data, managing participant identifiers from multiple sources, managing, browsing, querying and exporting metadata through a web based interface are some of the features that should fit within the biobank access framework in p-medicine. However, the current version of Opal seems to be still under development and has been noticed lack of sufficient guidelines and user manual for the right usage of the system. Finally, some important features, like a complete reporting tool for administrative purpose and quality control solutions, are still missing and are reported to be implemented in the future.

5.9 Biotracker Biobanking SaaS

Biotracker Biobanking SaaS[90] is a product of Ocimum Biosolutions Inc., a global integrated genomic services company situated in Hyderabad, India. The software is offered as "a fully scalable, configurable, customizable, complete biobanking solution".

Biotracker's Biobanking solution manages the collection, storage, processing and distribution of biological materials as well as the association of critical demographic data, clinical data, and biomarkers to these biospecimens. Biotracker for Biobanking maintains biospecimen genealogy with the ability to perform DNA/RNA extractions and also includes management of cell and tissue micro arrays and laser capture micro dissection.

Biotracker Biobanking's specialized features include a web-based requester module, donor/patient and biospecimen management with full genealogy, comprehensive metadata annotation, controlled vocabulary and synonyms, data and location management and more. The biobanking solution can also support a "virtual" biobanking network where there may be many sites or locations as well as collaborators. Biotracker has also configurable clinical study management attributes.

Biotracker Biobanking is caBIG® Bronze Certified, providing the highest degree of integratability and interoperability, and fulfils regulatory compliancy with HIPAA, 21CRF11 Electronic Signature, CLIA, and GLP/GxP.

[89] OBiBa Websie – Opal – accessible under http://www.obiba.org/node/131
[90] http://www.ocimumbio.com/lims2/products/biotracker-biobanking-saas/

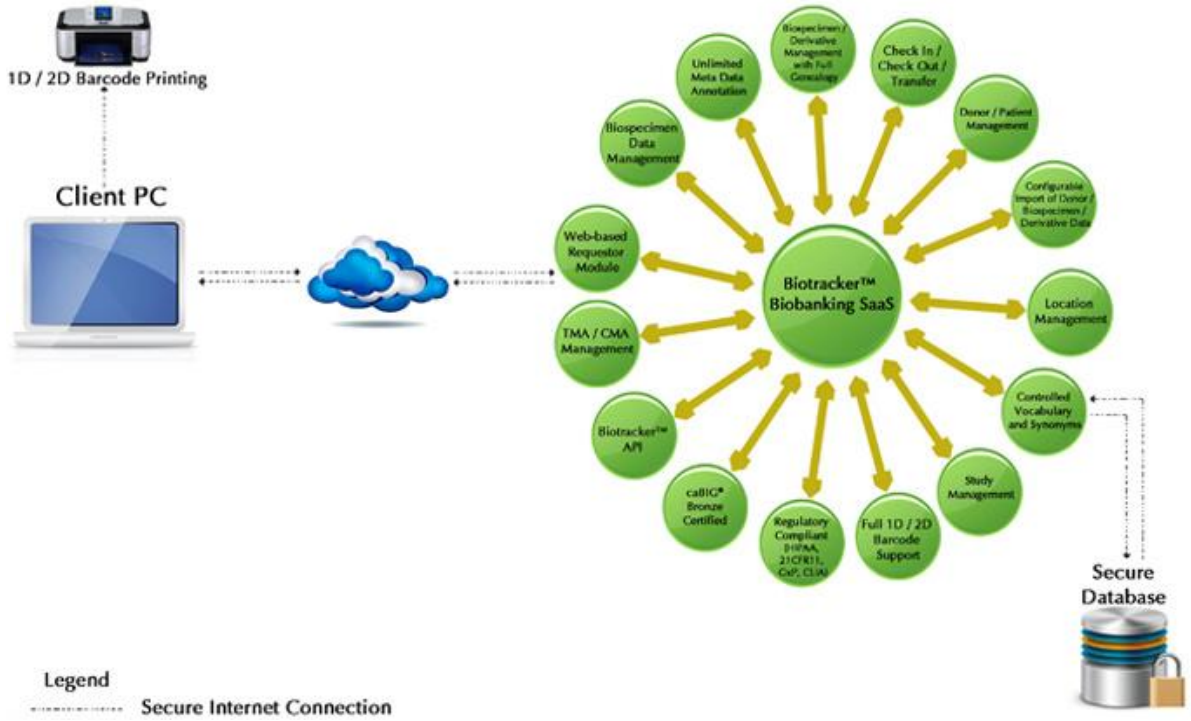


Figure 35: Features of Biotracker Biobanking SaaS.

5.9.1 Data harmonisation, data import, export and search functionality

Meta Data can be defined manually or it is possible to automate the annotation via interfaces with electronic medical records.

The software uses flexible controlled vocabularies from sources such as SNOMED or the caBIG Thesaurus. As a caBIG Bronze certified product the software fulfils the following three criterias:

1. Programmatic access to data through an open API is possible to integrate system with pathology information systems.
2. The terminologies used in the system include definitions of terms that meet caBIG® VCDE workspace guidelines.
3. Data element descriptions are maintained with sufficient definitional depth to enable a subject matter expert to unambiguously interpret the contents of the resource without contacting the original investigator.

Importation of tabular biospecimen data is supported by checking column headings against controlled vocabulary, flagging exceptions and allowing mapping through the user. Querying of biospecimens is facilitated by the use of the controlled vocabulary. A query engine guides users through process of retrieving data. A report wizard is built in and supports CSV, Excel, pdf files or HTML formats.

5.9.2 Origin, license model and maturity

Biotracker Biobanking SaaS is a product of Ocimum Biosolutions. The Indian company is a provider of Life Sciences Laboratory Information Management Systems (LIMS) and genomic solutions and was founded in the year 2000.

Since March 2009, Ocimum Biosolutions offered its complete Biotracker biobanking solution for $ 75,000 available in US and Canada. A similar “Stimulus Package” is also available for European countries for € 58,000. The offer includes Biotracker LIMS for BioRepository

software, up to 3 person months of implementation and customization effort, and the first year support & maintenance, all at a fixed price with on-time delivery[91] [92].

In June 2010, Ocimum Biosolutions launched Biotracker for Biobanking SaaS, the Software as a Service version of its LIMS for biorepositories. As a "Pay-as-you-go"-solution for all current and long-term laboratory automation and data needs, Biotracker SaaS has been particularly designed for small and medium size labs[93]. Further information about the use of the software is not available.

To summarize, Biotracker is a high-level commercial solution for biomaterial management and web based distribution that leverages controlled vocabularies and is compliant with standards. Through its API and import mechanisms for tabular data it could be used also for integrated biobanking although this was not the initial intension with this product. The Biotracker Software-as-a-Service version seams to match at least partially with p-medicine biobank use cases. However, as access to source codes is not foreseen it would be difficult to integrate Biotracker in the p-medicine architecture.

[91] http://www3.ocimumbio.com/news/ocimum-biosolutions-announces-its-biobanking-stimulus-package/; accessed on 16-09-2011

[92] http://www.ocimumbio.com/lims2/?page_id=317; accessed on 16-09-2011

[93] http://www.labbulletin.com/articles/Ocimum-Biosolutions-launches-the-Software-as-a-Service-version-of-its-Biotracker-Biobanking-LIMS, accessed on 16-09-2011

6 Overview on the current legal and ethical rules and guidelines for Biobanks

6.1 Introduction

Human tissues are an important resource for a better understanding of the causes and mechanisms of a large number of diseases. The more thoroughly the human genome is understood, the more it is possible to identify the role not only of external factors such as environmental influences or lifestyle but also hereditary predispositions as a causal factor for diseases. The scientific research in this area, therefore, is not only important for the individual patient but also for large groups of the population. For this reason, so called biobanks have become important in modern medicine.

Defined in most general terms, a biobank is a structured collection of biological specimens and corresponding donor data. [94] The data and material collected in biobanks can differ according to the goals and objectives of biobanks. Biobanks are often used for diagnostic and therapeutical purposes. They also play a central role in the research on causes and mechanisms of diseases. The data and material in biobanks can differ according to the goals and objectives of biobanks.

Experience shows that some of the purposes can only be reached if as many data as possible are gathered. This is most often true for diseases that are caused by complex interactions between genes, environment and lifestyles, such as diabetes, cancer, cardiovascular diseases and dementia. In these cases well-annotated human tissues are an important resource for studying the underlying mechanisms of the disease.[95]

In order to provide the (in certain cases necessary) diversity of samples it is required to form larger biobanks by a cooperation and integration of the already existing (smaller) biobanks. Therefore, the sharing of resources comprised of data, human biological samples and information derived from their analysis is necessary to foster scientific research in the field. Accordingly the collection of genetic data to enhance health care is deemed the main legitimate purpose for their further processing. [96]

The right to freedom of research in this area conflicts with other fundamental rights, especially the rights of informational self-determination and privacy for the donor and possibly also for his blood relatives. Property rights aspects regarding the human tissue might also play a role in biobanks. However, none of the (in many cases constitutional) guarantees of protection which apply to the donor are absolute. The rights rather have to be weighed against each other. It can generally be said, that the law allows for the use of data for scientific purposes given that the scientific interest overweighs the interests of the person

[94] See e.g. OECD, Guidelines on Human Biobanks and Genetic Research Databases (2009), p. 2, where biobanks are defined as „*structured resources that can be used for the purpose of genetic research and which include: (a) human biological materials and/or information generated from the analysis of the same; and (b) extensive associated information.*“ Available at: http://www.oecd.org/dataoecd/41/47/44054609.pdf. See also German Ethics Council, Human Biobanks for research – opinion (2010) p. 24, that sets up three criteria for human biobanks: „ *a) It contains genetic material originating from humans with related data. b) Its samples are electronically linked to personal information (possibly pseudonymized) and further information, in particular relating to health. c) Its samples and data are collected, preserved or used for purposes of scientific research*.“ Available at: http://www.ethikrat.org/files/der_opinion_human-biobanks.pdf.

[95] Cf. Watson/Kay/Smith, Integrating biobanks: addressing the practical and ethical issues to deliver a valuable tool for cancer research, Nature Reviews Cancer 10, 646-651 (September 2010).

[96] Cf. Art 29 Working Paper, p. 7.

affected and that the research purpose can be achieved only in this way or alternatively only with disproportional expense and effort.[97]

Therefore, we need to provide a way for biomedical research on the basis of biobanks guaranteeing the necessary level of protection of the donor´s fundamental rights and, thus, minimize the risks of the participants and their families. The setting of high ethical and legal standards shall also build up trust within the public in biospecimen resource activities and thus ensure the support and the participation of donors in the future. To this end clear guidelines and policies are required to address the challenges of integrating individual institutional or national biobanks and build public trust.

This chapter shall give an overview on the characteristics and risks of biobanks as well as the ethical and legal questions at stake and provide a survey of the status quo of the main legal and good practice guidelines concerning biobanks.

As the guidelines do not use a common terminology some terms shall be defined here in order to facilitate the reading. The term “donor” refers to the person the human body tissues and data derive from. The term “operator” means the entity responsible for the biobank.

6.2 Characteristics and risks of biobanks

6.2.1 Characteristics

6.2.1.1 Dual nature

Biobanks constitute a collection of samples of human body substances (e.g. tissue, blood, DNA), which are linked to personal data and soci-demographic information about the donors of the material as well as data relating to the diseases, the medical treatment conducted, mental dispositions, familial and social situation, environment-related data and data concerning the lifestyle of the donor.[98] Biobanks, thus, have a dual nature. They constitute both a collection of human tissue samples (genetic material) and a collection of data.

6.2.1.2 Anonymisation and (re-)identifiability of genetic data and material

In general, the data protection laws only cover the processing of data concerning identified or identifiable persons.[99] Therefore data that has never been linked to a specific person as well as data for which the link to a specific person has been erased, are not subject to data protection legislation. The notion of “anonymous data” does not only cover data for which the attribution to a specific person is completely impossible, but it also covers data that can only be attributed to a specific person with a disproportionate amount of time, expense and labour.[100]

One particularity of genetic data is that the genetic data show characteristics, which make them singular, in particular compared to other health data, as e.g. data regarding medical treatment conducted. Genetic data provide – or are likely to provide in the future – scientific, medical and personal information throughout the life of a person. Therefore information

[97] See Art 1 and Art. 7 of Directive 95/46 EC.

[98] German Ethics Council, Human Biobanks for research – opinion (2010) p. 7. Also available online: http://www.ethikrat.org/files/der_opinion_human-biobanks.pdf. See also Auray-Blais/Patenaude, A biobank management model applicabel to biomedical research, BMC Medical Ethics, http://www.biomedcentral.com/1472-6939/7/4.

[99] For furhter information see 6.3.1.4.1 infra.

[100] Data is qualified as anonymous if the information concerning personal or material circumstances cannot be attributed to an identified or identifiable individual or if this attribution can only be effected with a disproportionate amount of time, expense and labour. For further information see 6.3.1.4.1.2 infra.

containing genetic data (containing personal information) can precisely be linked to a specific person, if there is a reference data set available containing his/her genetic data.

The human tissue stored in biobanks per se cannot be qualified as personal data, as long as it is not linked to an identified or identifiable person. However, it is possible to extract genetic data from the human tissue samples (genetic material), which than – due to the uniqueness of genetic data – can be linked to a specific person. As sequencing of entire genomes is likely to become a routine procedure[101] more and more genetic data will be generated, which – if they are linked or linkable to an identified person – by matching with other data sets might serve as a basis to re-identify the donor.

6.2.1.3 Long term storage of human tissue and change of purpose

As the setting up of biobanks is expensive biobanks shall not be limited to one research project but rather form the basis for a variety of research projects by securing long-term storage of human body material. Therefore, the purposes of use of materials and data stored in biobanks may evolve over the time as scientific findings may, for instance, show that the human material is no longer suitable for the initial scientific purpose which it was collected for. As a consequence, to provide a clear definition of the scope of a biobank is a problem when biobanks that have collected material across different institutions or countries are merged.

6.2.2 Specific risks of biobanks

6.2.2.1 Risk of re-identification

As already stated above,[102] there is a real risk that genetic data and genetic material that is stored in biobanks might be linked or linkable in the future to a specific person, even if the data set is coded or anonymised.

6.2.2.2 Information gained from genetic data

6.2.2.2.1 Right to know or not to know of potential diseases

One potential risk for the donor is the knowledge about the outcome of the scientific research. The genetic analysis can reveal a disposition of the donor for a certain disease long before the disease breaks out. If this is the case, the doctors could decide to closely monitor the further development of the health of the donor in order to react as soon as possible whenever the disease really breaks out. However such knowledge can be a strain to the donor. On the one hand in many cases it cannot be predicted with certainty, if the disease breaks out at all. On the other hand there might be no or just poor therapy options.

6.2.2.2.2 Possible impact on blood relatives

Genetic data provide or are likely to provide, in the future, scientific, medical and personal information relevant throughout the life of an individual. This information can also have a significant impact on the family of the data subject.

6.2.2.3 Risk of misuse/Breach of secrecy

Considering the complexity and sensitivity of the genetic information, there is a great risk of misuse and reuse for various purposes by the biobank-operator or third parties. Risks of re-use might occur e.g. using the genetic information already extracted, or through additional

[101] Cf. Robertson, The $1000 Genome: Ethical and Legal Issues in Whole Genome Sequencing of Individuals. The American Journal of Bioethics 3(3): InFocus 2003. Available at: http://mitpress.mit.edu/journals/ajob/3/3/robertson.pdf.

[102] See 6.2.1.2 supra.

analysis of the underlying material (e.g. blood sample). The risk of a potential misuse shall not be neglected given that genetic data as well as the information about current and potential future diseases of people are of high interest for public health institutions, pharmaceutical companies, as well as employers and insurance companies.

6.2.2.4 Access for the purposes of criminal prosecution

Finally the crimininal prosecution agencies of the countries, where respective data is stored might have an interest in accessing these data for their purposes. This is in particular problematic regarding the transfer of human tissue to countries, where the recipient of the human bodily material can be obliged by the law to procure data and material of the donor to criminal prosecution agencies, even though there is no such duty in the country where the tissue has been donated.

6.2.2.5 Risk of de facto non-enforceable donor´s rights

Given the advancing integration of biobanks and the fact that the data and material may be used for various scientific research projects it will over the years get more difficult for the donor to identify the operator responsible in order to exercise his/her right for informational self determination.

6.3 Legal guidelines for biobanks

The following overview on the status quo of the existing rules and guidelines for biobanks containing human tissues and related data is divided into three parts. The first part focuses on the documents issued by the European legislator, in particular the respective Directives, which set minimum standards to be implemented by the Member States. The second part examines guidelines and best practice studies for human biobanks, which have been developed by international organisations. In a third step the existing legal frameworks on a national level shall be examined. This third part focuses on the laws of the states where the respective clinical data providing partners within the framework of p-medicine are based, therefore Germany, UK and Italy.

6.3.1 European law

The competence of the EU in the field of health is mainly defined in Art 168 of the Treaty on the Functioning of the European Union[103] (TFEU), which corresponds to Art 152 of the Treaty establishing the European Community[104] (TEC). Accordingly the EU action shall be directed towards improving public health, preventing physical and mental illness and diseases, and obviating sources of danger to physical and mental health in order to establish a high level of human health protection. This also covers inter alia the promotion of research into the causes, the transmission and the prevention of major health scourges.[105]

The European Parliament and the Council shall contribute to the achievement of the objectives by adopting inter alia „*measures setting high standards of quality and safety of organs and substances of human origin, blood and blood derivatives; these measures shall*

[103] Consolidated versions of the Treaty on European Union and the Treaty on the Functioning of the European Union, published in OJ C 83, of 30. March 2010, pp. 47 ss..

[104] Treaty establishing the European Community (consolidated text), published in OJ C 325, of 24 December 2002, pp. 33 ss., on p. 100.

[105] See Art 168 para. 1 TFEU.

not prevent any Member State from maintaining or introducing more stringent protective measures."[106]

On the basis of Art 152 TEC (now Art 168 TFEU) the European legislator has adopted two Directives which are repeatedly mentioned in connection with human biobanks for research. These are **Directive 2002/98/EC** on setting standards of quality and safety for the collection, testing, processing, storage and distribution of **human blood** and blood components and ammending Directive 2001/83/EC and **Directive 2004/23/EC** on setting standards of quality and safety for the donation, procurement, testing, processing, preservation, storage and distribution of **human tissues and cells**.

Another two Directive which also need to be mentioned in this context are the **Directive 2001/20/EC** on the approximation of the laws, regulations and administrative provisions of the Member States relating to the implementation of **good clinical practice** in the conduct of clinical trials on medicinal products for human use and the **Directive 2001/83/EC** on the Community code relating to **medicinal products for human use**. Unlike the Directives mentioned above, these Directives have not been adopted on the basis of ex-Art 152 TEC (now Art. 168 TFEU) but on the basis of ex-Art 95 TEC (now Art. 114 TFEU), which aims on the strengthening of the internal market of the EU.

In the following the applicability of these Directives on human biobanks for scientific research shall be examined. If there are no specific rules on the collection and use of human tissue, the general rules of the **Directive 95/46 EC** (Data protection Directive) may apply, which therefore shall be examined subsequently.

6.3.1.1 Directive 2002/98/EC

This Directive has to be read in combination with Directive 2001/83/EC[107] of the European Parliament and of the Council of 6 November 2001 on the Community code relating to medicinal products for human use. The latter ensures the quality, safety, and efficacy requirements of proprietary industrially-prepared medicinal products derived from human blood or plasma. The use of whole blood, plasma and blood cells of human origin have explicitly been excluded from the regulatory framework of that Directive.[108] Consequently their quality and safety, as far as they are only used for transfusion and not processed as such (e.g. used for the development of medical products), was not subject to any binding Community legislation.[109]

In order to fill this gap and ensure that there is an equivalent level of safety and quality of blood components, whatever their intended purpose, the Directive 2002/98/EC lays down standards of quality and safety of human blood and of blood components, in order to ensure a high level of human health protection.[110] It thus establishes an equivalent level of safety and quality of blood components, technical requirements for the collection and testing of all blood and blood components including starting materials for medicinal products within the Member States.[111] The scope of the Directive is defined in Art. 2:

"1. This Directive shall apply to the collection and testing of human blood and blood components, whatever their intended purpose, and to their processing, storage, and distribution when intended for transfusion.

[106] Art 168 para. 4 lit. a) TFEU.
[107] Published in OJ L 311 , 28.11.2001, p. 67.
[108] See Art 2 para. 4 Directive 2002/98/EC.
[109] See Recital 3 of Directive 2002/98/EC.
[110] See Art.1 Directive 2002/98/EC.
[111] See also Recital 5 of Directive 2002/98/EC.

2. Where blood and blood components are collected and tested for the sole purpose and exclusive use in autologous transfusion and are clearly identified as such, the requirements to be complied with in respect thereof shall be in accordance with those referred to in Article 29(g).”

This shows that although this Directive mainly aims on the prevention of diseases via transfusions, it also covers the collection and testing of human blood and blood components irrespective of their final destination.[112]

According to Art. 5 the Member States shall ensure that activities relating to the collection and testing of human blood and blood components, whatever their intended purpose, and to their preparation, storage, and distribution when intended for transfusion, are undertaken only by the blood establishments which have been designated, authorised, accredited or licensed by the competent authority for that purpose. Each donation of blood or blood components must undergo certain tests listed in Annex IV.[113] The same applies to blood or blood components imported into the Community.

As to data protection Art. 24 of Directive 2002/98/EC states that Member States have to take measures that all data, including genetic information, collected within the scope of the Directive to which third parties have access have been render anonymous so that the donor is no longer identifiable. For that purpose, they shall ensure:

(a) that data security measures are in place as well as safeguards against unauthorised data additions, deletions or modifications to donor files or deferral records, and transfer of information;

(b) that procedures are in place to resolve data discrepancies;

(c) that no unauthorised disclosure of such information occurs, whilst guaranteeing the traceability of donations.

Unlike the general safeguard measures provided by Directive 95/46 EC, the safeguards provided by the Directive 2002/98/EC shall not only be applicable to personal data, but prevent any unauthorised changes to donation registries, or processing records, or the unauthorised disclosure of information.[114] The technical requirements and standards laid down by this Directive have been specified by three Directives issued by the Commission:

[112] Art. 2 Directive 2002/98/EC:

„1. This Directive shall apply to the collection and testing of human blood and blood components, whatever their intended purpose, and to their processing, storage, and distribution when intended for transfusion.

2. Where blood and blood components are collected and tested for the sole purpose and exclusive use in autologous transfusion and are clearly identified as such, the requirements to be complied with in respect thereof shall be in accordance with those referred to in Article 29(g).“

[113] Annex IV: „The following tests must be performed for whole blood and apheresis donations, including autologous predeposit donations:

- ABO Group (not required for plasma intended only for fractionation)
- Rh D Group (not required for plasma intended only for fractionation)
- testing for the following infections in the donors:
- Hepatitis B (HBs-Ag)
- Hepatitis C (Anti-HCV)
- HIV 1/2 (Anti-HIV 1/2)

Additional tests may be required for specific components or donors or epidemiological situations.“

[114] See Recital 25 Directive 2002/98/EC.

The **Commission Directive 2004/33/EC**[115] lays down technical requirements on the suitability of blood and plasma donors and the screening of donated blood in the EU. It defines what information has to be provided to prospective donors, the information that is required from donors as well as the conditions of storage, transport and distribution conditions of blood and blood components. The prospective donors inter alia have to be provided information on the protection of personal data, in particular that no information concerning the identity of the donor or the donor's health and the results of the tests performed will be disclosed without authorisation as well as information on the responsibility of the blood establishment to inform the donor, through an appropriate mechanism, if test results show any abnormality of significance to the donor's health.[116] Furthermore blood establishments have to obtain information as to the identity as well as the health and medical history of the donor.[117]

Commission Directive 2005/61/EC[118] lays down the technical requirements to ensure traceability of blood and blood components from the donor to the recipient, and the procedures for traceability and notification of serious adverse reactions. The term "**traceability**", as defined by Art. 1 of the Commission Directive, means "the ability to trace each individual unit of blood or blood component derived thereof from the donor to its final destination, whether this is a recipient, a manufacturer of medicinal products or disposal, and vice versa". In order to ensure the traceability of blood and blood components Member States have to provide for accurate identification procedures, record maintenance and an appropriate labelling system (Art. 2 para. 1). The **traceability system** in place shall enable the blood establishment to trace blood components to their location and processing stage (Art. 2 para. 2). Therefore, the blood establishment shall be able to uniquely identify each donor, each blood unit collected and each blood component prepared, whatever its intended purpose, and the facilities to which a given blood component has been delivered (Art. 2 para. 3). Furthermore the blood establishments shall record each blood unit or blood component and the final destination (Art. 2 para. 4). Finally every blood establishment has to have a unique identifier that enables it to be precisely linked to each unit of blood that it has collected and to each blood component that it has prepared (Art. 2 para. 5).

The **Commission Directive 2005/62/EC**[119] lays down the specific technical requirements for a **quality system** for blood establishments. Accordingly Member States have to ensure that the quality system in place in all blood establishments and hospital blood banks complies with the Community standards and specifications set out in the Annex to the Directive.

6.3.1.2 Directive 2004/23/EC

The Directive 2004/23/EC[120] applies "to the donation, procurement, testing, processing, preservation, storage and distribution of human tissues and cells intended for human

[115] Commission Directive 2004/33/EC of 22 March 2004 implementing Directive 2002/98/EC of the European Parliament and of the Council as regards certain technical requirements for blood and blood components, OJ L 91 of 30.3.2004.

[116] What information has to be provided to the prospective donor is defined in Annex II, Part A of the Commission Directive 2004/33/EC.

[117] See Annex II, Part B of the Commission Directive 2004/33/EC.

[118] Commission Directive 2005/61/EC of 30 September 2005 implementing Directive 2002/98/EC of the European Parliament and of the Council as regards traceability requirements and notification of serious adverse reactions and events, OJ L 256 of 1.10.2005.

[119] Commission Directive 2005/62/EC of 30 September 2005 implementing Directive 2002/98/EC of the European Parliament and of the Council as regards Community standards and specifications relating to a quality system for blood establishments, OJ L 256 of 1.10.2005.

[120] Published in OJ L 102 of 07.04.2004, pp. 48 ss., changed by the Regulation (EC) No 596/2009 of the European Parliament and of the Council of 18 June 2009 adapting a number of instruments subject to the

applications and of manufactured products derived from human tissues and cells intended for human applications".[121] It aims to facilitate exchange of human material within the EU, whilst ensuring a high level of health protection and reassuring patients that tissues and cells derived from donation in another Member State, or originating outside of the EU, carry the same guarantees as those in their own country.

Therefore human biobanks containing human body material that is intended for research purposes only are not covered by the scope of this Directive. If the human tissue held in a biobank is intended for human use, the rules set by Directive 2004/23/EC establishing specific standards and technical requirements for each one of the steps in the human tissue and cell application process have to be taken into account.

According to Art. 5 the Member States shall ensure that tissue and cell procurement and testing are carried out by persons with appropriate training and experience and that they take place in conditions accredited, designated, authorised or licensed for that purpose by the competent authority or authorities. The accreditation, designation, authorisation or licensing of tissue establishments and tissue and cell preparation processes are specified in Art. 6. Art. 7 provides for inspections and control measures to be taken by the competent authorities. According to Art. 8 the Member Stats shall provide that tissue establishments shall implement a **donor identification system** in order to ensure the traceability of the human body material. Furthermore tissue establishments shall be registered and meet reporting obligations. Special notification obligations exist in case of serious adverse events and reactions.[122]

The Directive stipulates in Art. 13 that the procurement of human tissues or cells shall only be licit after **all mandatory consent or authorisation requirements in force** in the respective Member State has been met. In addition it stipulates that Member States shall take all necessary measures to ensure that donors, their relatives or any persons granting authorisation on behalf of the donors are provided with all **appropriate information**. The information to be provided on the donation of cells and tissues is specified in the Annex of the Directive. Accordingly the information has to be given by a trained person able to transmit it in an appropriate and clear manner, using terms that are easily understood by the donor.[123] The information hast to cover inter alia the purpose and nature of the procurement, its consequences and risks; analytical tests, if they are performed; recording and protection of donor data, medical confidentiality; therapeutic purpose and potential benefits and information on the applicable safeguards intended to protect the donor.[124]

Data protection and confidentiality issues are regulated by Art. 14 of the Directive. Accordingly Member States shall take all necessary measures to ensure that all data, including genetic information, collated within the scope of this Directive and to which third parties have access, have been rendered anonymous so that neither donors nor recipients remain identifiable for the third party. Therefore the Member States shall ensure inter alia

---

procedure referred to in Article 251 of the Treaty to Council Decision 1999/468/EC with regard to the regulatory procedure with scrutiny – Adaptation to the regulatory procedure with scrutiny, OJ L 188, 18.7.2009, pp. 75 ss.

[121] Art. 2 (1) of the Directive 2004/23/EC. The term "human application" is defined as "the use of tissues or cells on or in a human recipient and extracorporal applications". Art 3 lit. l of the Directive 2004/23/EC.

[122] The term „serious adverse event" is defined in Art. 3 lit. m) and means „any untoward occurrence associated with the procurement, testing, processing, storage and distribution of tissues and cells that might lead to the transmission of a communicable disease, to death or life-threatening, disabling or incapacitating conditions for patients or which might result in, or prolong, hospitalisation or morbidity". The term "serious adverse reaction" is defined in Art. 3 lit. n) as „an unintended response, including a communicable disease, in the donor or in the recipient associated with the procurement or human application of tissues and cells that is fatal, life-threatening, disabling, incapacitating or which results in, or prolongs, hospitalisation or morbidity".

[123] Part A (2) Annex, Directive 2004/23/EC.

[124] Part A (3) Annex, Directive 2004/23/EC.

that data security measures are in place, as well as safeguards against any unauthorised data additions, deletions or modifications to donor files or deferral records, and transfer of information and that no unauthorised disclosure of information occurs whilst guaranteeing the traceability of donations.

According to Art. 17 of the Directive every tissue establishment shall designate a responsible person who shall at least possess of a diploma, certificate or other evidence of formal qualifications in the field of medical or biological sciences awarded on completion of a university course (or equivalent) and have at least two years practical experience in the relevant fields.

These **technical requirements and standards** are specified in the Directives 2006/17/EC and 2006/ 86/EC and 2010/453/EU issued by the Commission in order to implement the Directive.

The **Commission Directive 2006/17/EC**[125] establishes specific technical requirements for each step in the human tissue and cell preparation process, in particular regarding the requirements for the procurement of human tissues and cells, the selection criteria for donors of tissues and cells, and laboratory tests required for donors, tissue and/or cell donation and procurement procedures and reception at the tissue establishment as well as requirements for direct distribution to the recipient of specific tissues and cells.

Commission Directive 2006/86/EC[126] defines the technical standards and requirement as regards traceability requirements, notification of serious adverse reactions and events and certain technical requirements for the coding, processing, preservation, storage and distribution of human tissues and cells, and products derived from human tissues and cells. The requirements for accreditation, designation, authorisation or licensing of tissue establishments are regulated in Annex I of the Commission Directive, which provides detailed prerequisites as to the organsation and management, the presonell, the equipment and materials, facilities/permisses as well as the documentation and records.

Art. 10 stipulates that a single European identifying code (European coding system) shall be allocated to all donated material at the tissue establishment, to ensure proper identification of the donor and the traceability of all donated material and to provide information on the main characteristics and properties of tissues and cells. The code shall incorporate at least the information set out in Annex VII.[127]

Commission Decision 2010/453/EU[128] establishes guidelines concerning the implementation of inspection and control measures in the field of human tissues and cells. It defines the responsibilities and the qualification and training requirements of inspectors. These guidelines, however, are not legally binding.

---

[125] Commission Directive 2006/17/EC of 8 February 2006 implementing Directive 2004/23/EC of the European Parliament and of the Council as regards certain technical requirements for the donation, procurement and testing of human tissues and cells, published in OJ L 38, 9.2.2006, p. 40.

[126] Commission Directive 2006/86/EC of 24 October 2006 implementing Directive 2004/23/EC of the European Parliament and of the Council as regards traceability requirements, notification of serious adverse reactions and events and certain technical requirements for the coding, processing, preservation, storage and distribution of human tissues and cells, published in OJ L 294, 25.10.2006, p. 32.

[127] For further information see Reynolds/Warwick/Poniatowski/Trias, European coding system for tissues and cells: a challenge unmet? Cell Tissue Bank (2010) 11:353–364, also available at: http://www.springerlink.com/content/917177124nll1215/fulltext.pdf.

[128] Commission Decision 2010/453/EU of 3 August 2010 establishing guidelines concerning the conditions of inspections and control measures, and on the training and qualification of officials, in the field of human tissues and cells provided for in Directive 2004/23/EC of the European Parliament and of the Council [notified under document C(2010) 5278], published in OJ L 213, 13.8.2010, p. 48.

6.3.1.3 Directive 2001/20/EC

The Directive 2001/20/EC „establishes specific provisions regarding the conduct of clinical trials, including multi-centre trials, on human subjects involving medicinal products ..., in particular relating to the implementation of good clinical practice. This Directive does not apply to non-interventional trials“.[129] The term clincal trial is defined by Art. 2 para. 1 of the Directive. It describes „any investigation in human subjects intended to discover or verify the clinical, pharmacological and/or other pharmacodynamic effects of one or more investigational medicinal product(s), and/or to identify any adverse reactions to one or more investigational medicinal product(s) and/or to study absorption, distribution, metabolism and excretion of one or more investigational medicinal product(s) with the object of ascertaining its (their) safety and/or efficacy“.

As clinical trials are investigations in humans intended to discover or verify the effects of one or more investigational medicinal products, the Directive does not specificly refer to research based on biobanks. The same is true for the **Commission Directive 2005/28/EC** (Good Clinical Practice Directive) which further concretises the Directive 2001/20/EC by laying down principles and detailed guidelines for good clinical practice as regards investigational medicinal products for human use, as well as the requirements for authorisation of the manufacturing or importation of such products.

6.3.1.4 Directive 95/46 EC (Data Protection Directive)

The EC Directive 95/46 on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data, adopted by the European Parliament and the Council on 24.10.1995, is by far the most comprehensive and complex international policy instrument as to the protection of personal data.

6.3.1.4.1 Scope

6.3.1.4.1.1 Personal data

The Data Protection Directive is only applicable to personal data. The term “personal data” is defined in Article 2 lit. a). It covers “any information relating to an identified or identifiable natural person”. This natural person is called “data subject”. Furthermore an “identifiable person” is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity. Data however that does not refer to a natural person is not subject to the processing restrictions of the Directive. Accordingly data not referring to a natural person anymore (anonymous data) as well as data concerning objects are not covered by Directive 95/46 EC.

6.3.1.4.1.2 Anonymous data

The Directive 95/46 EC considers data as anonymous only, if the data subject is no longer identifiable, i.e. the link that refers to the data subject is **irrecoverably erased or has never existed**. Anyhow the German legislation for instance seized the suggestion of the proposal and, unlike the European legislation, implemented the “excessive effort” in its definition of anonymous data.[130]

However the European perception regarding “anonymous data” seems to change. In 2003 the European Commission published its “First report on the implementation of the Data Protection Directive”. Referring to a document of the European Privacy Officers Forum (EPOF) from 2002 the Commission pointed out, the interpretation of certain provisions of Directive 95/46 EC had to be reasonable and flexible. Whereas the EPOF stated that the

[129] Art. 1 para. 1 Directive 2001/20/EC.

[130] See also sect. 3 para. 6 of the Federal Data Protection Act (BDSG).

definition of anonymization should be pragmatic and it should emphasise that the capability of identification must be subject to the reasonableness standard. EPOF pointed out that the German definition indeed would satisfy both requirements. These statements gave reason to assume that the European Commission approves a definition of "anonymous data" that includes an "excessive effort".

Meanwhile the "**excessive effort**" has been introduced to EU law.[131] Therefore information concerning personal or material circumstances that can only with a disproportionate amount of time, expense and labour be attributed to an identified or identifiable individual is anonymous data.

6.3.1.4.1.3 Pseudonymous data

In contrast to some national data protection regulations, the Directive 95/46 EC does not know the term "pseudonymous data". Pseudonymous data is therefore to be seen as personal data.

The German Federal Data Protection Act for example defines in section 3 para. (6a) pseudonymising as "replacing a person's name and other identifying characteristics with a label, in order to preclude identification of the data subject or to render such identification substantially difficult". Especially in a medical research project, the use of pseudonymous data often is mandatory and beneficial for the patient, because it is possible to re-identify the patient in case of newly developed treatments.

6.3.1.4.1.4 Applicability of the Directive 95/46 EC to biobanks

In order to assess the applicability of the Data Protection Directive to biobanks, it has to be examined whether the biobank contains personal data, i.e. data directly or indirectly (by a reference or by additional information) related to an individual.

The wording calls for a wide interpretation. According to the Explanatory Memorandum, the definition should be as broad as possible "so as to include all information"25. With regard to content, it includes any sort of information26. Traces, however, that have been left behind, such as e.g. wheel traces or finger prints, blood traces, hair or other human body material are not regarded as information related to a person and therefore are not per se covered by the scope of the Directive 95/46 EC. These traces however can serve as a starting point to generate information linked or linkable to a person, by e.g. genetic analysis. As already stated above, the DNA is a unique information about a person, this data is deemed to be linkable to a specific person. As the functioning of a biobank generally implies the extracting of genetic data the rules of the Directive 95/46 EC have to be taken into account.[132]

6.3.1.4.2 Principles set up for the processing of personal data

For the processing of personal data the Directive 95/46 EC establishes several principles regarding the fair and lawfull processing of data and provides for technical and organisational measures to ensure the security and safety of personal data.

6.3.1.4.2.1 Fair and lawful data processing

According to Art. 6 of the Directive, personal data must be **collected for specified, explicit and legitimate purposes** and not further processed in a way incompatible with those purposes. In addition, personal data must be adequate, relevant and not excessive in relation to the purposes for which they are collected and further processed (**proportionality principle**). This requires the evaluation of proportionality and legitimacy, taking into account

[131] See Art. 2 k of the Council Framework Decision 2008/977/JHA of 27 November 2008 on the protection of personal data processed in the framework of police and judicial cooperation in criminal matters.

[132] See Simitis, Bundesdatenschutzgesetz (7th edition, 2011), § 3 Rz 5.

the risks for the protection of fundamental rights and freedoms of individuals and notably whether or not the intended purpose could be achieved in a less intrusive way.

The Directive prohibits further processing that would be incompatible with the purpose the data was collected for. Accordingly the compliance with these principles imply a clear determination of the purpose for which genetic data and human tissue samples are collected and further processed. Nevertheless, the Directive provides exemptions to such prohibition for historical, statistical or scientific purposes provided that Member States put in place appropriate safeguards.

6.3.1.4.2.2 Criteria for legitimate data processing

The criteria for making data processing legitimate are established in Art. 7, according to which personal data may be processed only if the data subject has unambiguously given his/her consent or processing is necessary inter alia for the performance of a contract to which the data subject is party or for compliance with a legal obligation to which the controller is subject on, in order to protect the vital interests of the data subject or in cases where processing is necessary for the purposes of the legitimate interests pursued by the controller or by the third party or parties to whom the data are disclosed, except where such interests are overridden by the interests for fundamental rights and freedoms of the data subject which require protection under Art.1 para. 1 of the Directive 95/46 EC.

Furthermore, as biobanks contain patient data Art. 8 of the Directive is of relevance. Para. 1 states, Member States shall prohibit the processing of special categories of data, i.e. data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, trade-union membership, and the processing of data concerning health or sex life. Genetic data and any medical patient data is health data in the sence of the Directive.[133] Such sensitive data shall only be processed, if according to para. 2

- the data subject has given his explicit consent to the processing of those data or
- the processing is necessary for the purposes in the field of employment law or
- it is necessary to protect the vital interests of the data subject or
- processing is carried out in the course of its legitimate activities with appropriate guarantees by a foundation, association or any other non-profit-seeking body with a political, philosophical, religious or trade-union aim or
- the processing relates to data which are manifestly made public by the data subject or is necessary for the establishment, exercise or defence of legal claims.

With respect to the scientific research that is planed to be carried out in the course of p-medicine only the first exemption is applicable. According to para. 2 the processing of health data would therefore only be legitimate, if the data subject has given his consent.

Additionally para 3 to 5 provide further exemptions (in case of para. 4 only if Member States choose to lay down such exemptions).

Para. 3 provides exemptions in case of preventive medicine, medical diagnosis, the provision of care or treatment or the management of health-care services and where those data are processed by a health professional subject under national law. At first sight this exemption could be applicable in case of p-medicine. Nevertheless the data processing during and after p-medicine will focus on medical research whereas para. 3 covers data processing in medical care, meaning it is required for a concrete treatment of a concrete patient (data subject). Hence para. 3 does not provide an exemption to the general prohibition of the processing of sensitive data.

[133] See Working Document of the Art. 29 Data Protection Working Party: Working Document on Genetic Data, p. 5 (available at: http://ec.europa.eu/justice_home/fsj/privacy/docs/wpdocs/2004/wp91_en.pdf).

However according to Art. 8 para. 4 Member States may lay down additional exemptions by national law or by decision of the supervisory authority for reasons of substantial public interest, subject to the provision of suitable safeguards.

Art. 8 para. 4 therefore does not provide an exemption per se, but empowers Member States to introduce national exemptions for reasons of substantial public interest and subject to the provision of suitable safeguards. Examples for such a substantial public interest are introduced by Recital (34) of the Directive:

"(34) Whereas Member States must also be authorized, when justified by grounds of important public interest, to derogate from the prohibition on processing sensitive categories of data where important reasons of public interest so justify in areas such as public health and social protection - especially in order to ensure the quality and cost-effectiveness of the procedures used for settling claims for benefits and services in the health insurance system - ***scientific research*** *and government statistics; whereas it is incumbent on them, however, to provide specific and suitable safeguards so as to protect the fundamental rights and the privacy of individuals;"*

According to Recital (34) scientific research as it will be undertaken in the course of p-medicine is a possible example for an important public interest in the meaning of Directive 95/46 EC. Member States can therefore introduce regulation permitting the processing of sensitive personal data for scientific research purposes under the condition to provide specific and suitable safeguards so as to protect the fundamental rights and the privacy of individuals. The exemptions can be introduced either by national law or by decision of the supervisory authority. However the disadvantage of para. 4 for European scientific research projects is that it is the free choice of each Member State whether he decides to introduce such exemptions in his national law at all and if so, under which preconditions they are introduced.

Furthermore these exceptions are only hypotheses where the legitimacy of the data processing is formally assumed. The legitimacy of the processing of sensitive data is not given when only formally fitting into one of these exemptions. In fact the balance of the interests deriving from the processing of sensitive data has to be assessed in every concrete case.

6.3.1.4.2.3 Security policy

The Directive 95/46 EC states in its Art. 17 that Member States shall provide that the controller/processor must implement appropriate technical and organizational measures to protect personal data against accidental or unlawful destruction or accidental loss, alteration, unauthorized disclosure or access, in particular where the processing involves the transmission of data over a network, and against all other unlawful forms of processing. It is not easy to distinguish technical from organisational security measures as both elements are interrelated.While technical measures deal with the practical use of the methods implemented to secure the data being processed (including prevention of physical access to the hardware, such as secure premises and access control), organisational measures refer to a set of rules to enable data security by regulating authorisation and authentication procedures.[134] Technical measures include the use of encryption, secure connections, firewalls or access by biometric identification or similar methods. Organisational measures are access policies and identity management for the IT system processing the data.[135]

[134] Terstegge, in: Büllesbach/Poullet/Prins, Concise European IT Law, Alphen aan den Rijn 2006, Directive 95/46/EC, Art. 17 no. 1.

[135] The institute for Legal Informatics of the Leibniz University Hanover provides a sound examination of data security measures in EU FP7 framework project OPTIMIS dealing with appropriate technical and organisational measures to protect personal data to be taken in the field of cloud computing. See OPTIMIS, D7.2.1.2 – Cloud

Which security measures have to be provided in a specific case depends on the "state of the art" and the cost of their implementation, as such measures shall ensure a level of security appropriate to the risks represented by the processing and the nature of the data to be protected. The Directive 95/46 EC does not regulate what precisely „**state of the art**" technology is. As data security is to a large extent a technical matter, the Directive contains quite general requirements in order to be technology neutral.[136] For guidance on specific implementation of data security concepts, the Directive 95/46 EC relies on the domains of computer science and information security. In principal, it is therefore up to the developers to decide which security measures can be considered state of the art.[137] However, in the majority of cases, good technical or management practices are related to current standards.[138] Therefore, a recognised standard should be guidance for the security concept of p-medicine. Sensitive data as mentioned in Art. 8 para. 1 Directive 95/46 EC may require even more sophisticated security measures,[139] while other data may require less strict measures. Therefore, one possible security measure could be to separate these data in order to apply a higher standard on more sensitive data. [140]

The data controller however is not obliged to provide the highest security standards according to the state of the art. Security measures rather have to be appropriate with regard to the anticipated risks inherent in the data processing, as well as with regard to the nature of data and the costs of their implementation.[141] When assessing appropriate security measures for biobanks, the potential risks for the donors and their blood relatives related to the collection and processing of genetic data and the corresponding samples as outlined in 6.2.2 above as well as the technical and organisational measures to prevent these risks have to be thoroughly evaluated. Due to the technical development, the state of the art can change over the time, so that the security measures undertaken shall be reviewed periodically.

6.3.1.5 Documents by the Art. 29 Data Protection Working Party

The Working Party set up in accordance with Art. 29 of the Directive issued two documents concerning genetic data so far. In its **Opinion 6/2000 on the Human Genome and**

Legal Guidelines: Data Security, Ownership Rights and Domestic Green Legislation, pp. 17 ss., available at: http://www.optimis-project.eu/sites/default/files/content-files/document/d7212-optimis-cloud-legal-guidelines.pdf.

[136] See Meints, The Relationship between Data Protection Legislation and Information Security Related Standards, in: The Future of Identity in the Information Society: 4th IFIP WG 9. 2, 9. 6, 11. 6, 11. 7/FIDIS International Summer School, Brno, Czech Republic, September 1-7, 2008, Revised Selected Papers, Vol. 298, p. 254, 256.

[137] See OPTIMIS, D7.2.1.2 – Cloud Legal Guidelines: Data Security, Ownership Rights and Domestic Green Legislation, pp. 27 ss., available at: http://www.optimis-project.eu/sites/default/files/content-files/document/d7212-optimis-cloud-legal-guidelines.pdf.

[138] Cf. Meints, The Relationship between Data Protection Legislation and Information Security Related Standards, in: The Future of Identity in the Information Society: 4th IFIP WG 9. 2, 9. 6, 11. 6, 11. 7/FIDIS International Summer School, Brno, Czech Republic, September 1-7, 2008, Revised Selected Papers, Vol. 298, p. 254, 265.

[139] Terstegge, in: Bullesbach/Poullet/Prins, Concise European IT Law, Alphen aan den Rijn 2006, Directive 95/46/EC, Art. 17 no 2.

[140] See also Recommendation R(97)5 of the Committee of Ministers to Member States "on the Protection of Medical Data", which states in 9.1. that "appropriate technical and organizational measures shall be taken to protect personal data … against accidental or illegal destruction, accidental loss, as well as against unauthorised access, alteration, communication or any other form of processing. Such measures shall ensure an appropriate level of security taking account, on the one hand, of the technical state of the art and, on the other hand, of the sensitive nature of medical data and the evaluation of potential risks."

[141] Terstegge, in: Bullesbach/Poullet/Prins, Concise European IT Law, Alphen aan den Rijn 2006, Directive 95/46/EC, Art. 17 no 2.

Privacy[142] the Working Party emphasised the necessity of coupling new genetic technologies with adequate safeguards to protect the right for privacy. However it did not specify them.

In its **Working Document on Genetic Data**, issued in 2004,[143] the Working Party identifies areas of concern related to the processing of genetic data from a data protection perspective and contribute to a more uniform approach in the light of the national measures adopted in this field under Directive 95/46 EC. Given the special nature and characteristics of genetic data and the impact their use may have on the individual's life and on the members of his family, it is deemed very important to determine the purposes for which genetic data may be processed. The Working Party distinguishes five puporses for the collection and processing of genetic data, which are inter alia medical health care and medical and scientific research.

Within the field of **medical health care** genetic data may be used for the purpose of "diagnostic genetic tests" serving to clarify the causes of an already clinically manifest illness, and "predictive genetic tests" that are designed to identify genetic changes which are highly likely to lead to an illness at a later point in the life of the person tested. In both cases, the data subject should be duly informed about the necessity of carrying out such tests and give its explicit consent for that purpose and for the processing of its genetic data (see Art. 8 para. 2 lit. a). **Informed consent is particularly crucial** in the field of genetic testing as the information that individuals will receive about themselves could have serious implications. The Working Party does not only affirm a right to know, but also a right not to know, in case the data subject chooses not to be informed about the results of the genetic testing, in particular when the disease is highly serious and at the time there are no scientific means to prevent or to treat it.[144]

Regarding the use genetic data in the field of **medical and scientific research** the Working Party points out the establishment of large genetic research databases is a **potential cause of concern from a data protection perspective**. As biobanks are an ongoing study, they are potentially subject to a number of different uses or appetites, given that many of the research uses are secondary to some of the original purposes. Therefore the Working Party outlines the importance of a thorough examination of issues like "a) the further processing of the data for purposes that may have not even been conceived at the time of their collection, b) the duration of storage of genetic data, and c) the appropriate security measures".[145]

In this constext the Working Party states that practices of anonymisation could be implemented in order to address data protection concerns. However, it is also pointed out, that anonymisation may not be feasible as regards genetic data.[146] Therefore the Working Party stipulates that a **high level security measures, both organisational and technical, should be taken** to protect the data contained in biobanks, in compliance with Article 17 of the Directive. For instance, operators should be encouraged to carry out surveys of potential risks, establish policies for security, inform and train staff on an on-going basis, set up restricted access controls in order to avoid undue access by administrative staff and other persons, etc[147]

---

[142] Available at: http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2000/wp34en.pdf.

[143] Available at: http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2004/wp91_en.pdf.

[144] Art. 29 Working Party, Working Document on Genetic Data, 12178/03/EN, WP 91, p. 9, available at: http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2004/wp91_en.pdf.

[145] Art. 29 Working Party, Working Document on Genetic Data, 12178/03/EN, WP 91, p. 11, available at: http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2004/wp91_en.pdf.

[146] See 6.2.1.2 supra.

[147] Art. 29 Working Party, Working Document on Genetic Data, 12178/03/EN, WP 91, pp. 11 s., available at: http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2004/wp91_en.pdf.

Finally, procedures should be put in place in order to ensure that genetic data are only processed under the supervision of qualified professionals who are entitled to such processing on the basis of specific authorisations and rules.[148]

In a very recent **opinion, adopted on 13 July 2011**, the Working Party seeks to clarify the definition of the term **"consent"** and how it should be interpreted and used under European data protection laws.[149] The guidelines are in large part a compilation of recommendations previously made by the Article 29 Working Party for particular forms of processing, such as collection of patient data for electronic health records, transfer of data to third parties, processing of passenger name records, etc.[150]

The Article 29 Working Party heavily relies on the notion of "indication" of the data subject's wishes, which is laid down in the definition of consent given by the 1995 Directive and concludes that **positive action** would be required to demonstrate consent. Therefore only statements or actions, **not mere silence or inaction**, can constitute valid consent.[151] Consequently the sending of an e-mail to a data subject informing him/her of changes to the privacy policy or stating that the processing of his/her data will be undertaken unless he/she objects within a defined period of time would not be sufficient to constitute the data subject´s consent to the new policy or the contemplated processing. The consent would have to be evidenced by an affirmative clicking of a box or any other relevant positive act.

In addition, to be able to make informed choices, data subjects need to be informed about the data processing and **good quality and accessibility of the information** is paramount to this. According to the Article 29 Working Party, a general consent to any and all transfers to unspecified third parties would not be sufficiently specific to constitute valid consent. The Article 29 Working Party pointed to the 2010 opinion of the Advocate General in a case involving agricultural funds in Europe, in which the Advocate General held that a broad consent in the fund's terms and conditions was not sufficiently precise to conclude that the beneficiary of the fund had given unambiguous consent to the publication of his or her name. Finally, data controllers should be able to demonstrate that they have obtained valid consent.

6.3.2 International Good Governance Guidelines

Over the last 15 years, an increasing number of international bodies have developed relevant guidelines or statements of principle as to the activities concerning human bodily tissues and genetic data. Those bodies can be grouped into four different types.[152] There are (i) bodies that are representing the vast majority of the countries, such as the United Nations and its specialised agency UNESCO; (ii) the Council of Europe, a regional body that represents countries within Europe as well as other countries that are prepared to sign up to its conventions; (iii) international scientific organisations, in particular the Human Genome Organisation (HUGO); and (iv) bodies that represent the industrialised nations, such as the OECD.

Many of the documents and guidelines adopted by the bodies mentioned do not specificly refer to (disease based) biobanks for scientific research, but set up general principles for the

[148] Art. 29 Working Party, Working Document on Genetic Data, 12178/03/EN, WP 91, pp. 13, available at: http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2004/wp91_en.pdf.

[149] Art. 29 Working Party, Opinion 15/2011 on the definition of consent, 01197/11/EN, WP187.

[150] Goralczyk, MMR-News

[151] For example, when a data subject registers with a social network and the default settings of his or her profile make all personal information viewable to all "friends of friends", it cannot be inferred that this user has given his or her consent.

[152] See Knoppers/Abdul-Rahman/Bédard, Genomic Databases and International Collaboration, King´s Law Journal 2007, 18 KLI p. 291 (p. 294).

protection of human rights in the field of medical research based on genetic data. Furthermore, the majority of these documents are not legally binding.[153] Nevertheless, in the absence of legally binding rules, these international good governance guidelines indicate ways how to deal with above mentioned the risks related with the research in the filed of genetic data based on human biobanks.[154]

6.3.2.1 UNESCO

The UNESCO has issued several declarations regarding the field of genomic research within the last decades.[155] The most important of them regarding human biobanks are the Universal Declaration on Human Genome and Human Rights (1997), the International Declaration on human genetic data (2003) and the Universal Declaration on Bioethics and Human Rights (2005).

6.3.2.1.1 Universal Declaration on Human Genome and Human Rights (1997)[156]

Already in 1997 the UNESCO issued its Universal Declaration on the protection of the human genome and human rights („Human Genome Declaration“) outlining the basic ethical principles for the proper conduct of human genome research in general. It was the first universal instrument to establish an ethical framework for activities in this area, whether for research itself or for the application of findings. It is thus a vital part of the intense standard-setting activity that characterizes bioethics today.

According to Art. 5 of the Declaration research, treatment or diagnosis affecting an individual´s genome shall be undertaken only after rigorous and prior **assessment of the potential risks and benefits**. Furthermore, the prior, free and **informed consent** of the person concerned shall be obtained. If according to the law a person does not have the capacity to consent, research affecting his/her genome may only be carried out for his or her direct health benefit, subject to the authorization and the protective conditions prescribed by law. Accordingly **research without expected direct health benefit may only be undertaken by way of exception**, with the utmost restraint, exposing the person only to a minimal risk and minimal burden and if the research is intended to contribute to the health benefit of other persons in the same age category or with the same genetic condition.

Art. 7 of the Declaration provides that genetic data associated with an identifiable person and stored or processed for the purposes of research or any other purpose must be held confidential in the conditions set by law.

6.3.2.1.2 International Declaration on human genetic data (2003)

The International Declaration on human genetic data from 16.10.2003[157] aims to ensure the “respect of human dignity and protection of human rights and fundamental freedoms in the collection, processing, use and storage of human genetic data, human proteomic data and of the biological samples from which they are derived” (Art. 1). The Declaration, thus, explicitly covers not only the storage of genetic data, but also of **biological samples**, which is the case in biobanks. The provisions of this Declaration however, do not apply to the use of data

[153] E.g. in the case of both UNESCO and the Council of Europe, the declarations, conventions and treaties that they produce can only take effect once they have been signed and then implemented into national law by each country.

[154] See 6.2.2 supra.

[155] For further information see http://www.unesco.org/new/en/social-and-human-sciences/themes/bioethics.

[156] Available at: http://unesdoc.unesco.org/images/0011/001102/110220e.pdf#page=47.

[157] Available at: http://unesdoc.unesco.org/images/0013/001331/133171e.pdf#page=45.

stored in biobanks for investigation, detection and prosecution of criminal offences and in parentage testing that are subject to domestic law.[158]

Art. 5 of the Declaration sets up **four purposes**, for which human genetic data and samples may be collected, processed, used and stored. These are the purposes of (i) diagnosis and health care, including screening and predictive testing; (ii) medical and other **scientific research**, including epidemiological, especially population-based genetic studies, as well as anthropological or archaeological studies, collectively referred to hereinafter as "medical and scientific research"; (iii) forensic medicine and civil, criminal and other legal proceedings, taking into account the provisions of Article 1(c); (iv) or any other purpose consistent with the Universal Declaration on the Human Genome and Human Rights and the international law of human rights.

According to Art 6. genetic data and samples should be dealt with only on the basis of **transparent and ethically acceptable procedures**. In addition, clear, balanced adequate and appropriate information shall be provided to the donor, whose consent is sought. Such information shall, in particular, specify the purpose for which human genetic data and human samples are used and stored.

The collection of human genetic data and samples is governed by the **principle of informed consent**.[159] There are, however exceptions for minors and adults incapable of giving informed consent. In these cases consent shall be given by the legal representative in the best interest of the person concerned.[160] The consent given may generally be withdrawn by the data subject, unless such data are anonymised, i.e. irretrievably unlinked to an identifiable person.[161]

Art. 13 states that no one should be denied access to his/her own genetic data unless such data are irretrievably unlinked to that person as the identifiable source or unless domestic law limits such access in the interest of public health, public order or national security.

In Art. 14 the Declaration sets up rules on the privacy and confidentiality of data and samples. Human genetic data and biological samples linked to an identifiable person should be protected and in general not be disclosed to third parties (in particular, employers, insurance companies, educational institutions and the family). As a general rule **human genetic data and biological samples collected for the purposes of scientific research should not be linked to an identifiable person**. These data and samples, however, shall be linked to an identifiable person, only if necessary to carry out the research and provided that the privacy of the individual and the confidentiality of the data or biological samples concerned are protected in accordance with domestic law. In any case human genetic data and samples should not be kept in a form which allows the data subject to be identified for any longer than is necessary for achieving the purposes for which they were collected or subsequently processed. However, even when such data or biological samples are unlinked to an identifiable person, the necessary precautions should be taken to ensure the security of the data or biological samples.

The rules governing cases where the initial purpose for which the data was collected shall be changed are set out by Art. 16 of the Declaration. Accordingly these data must not be used for a different purpose that is incompatible with the original consent, unless the prior, free, informed and express consent of the person concerned is obtained or unless the proposed use corresponds to an important public interest reason and is consistent with the international law of human rights. However, if the consent of the person concernced cannot

[158] Art. 1 sect. 3 International Declaration on human genetic data.

[159] Art 8. lit. a) International Declaration on human genetic data.

[160] Art. 8 lit. b) to d) International Declaration on human genetic data.

[161] Art. 9 International Declaration on human genetic data.

be obtained or in the case of data irretrievably unlinked to an identifiable person, human genetic data and samples may be used in accordance with domestic law or following the consultation of the competent ethics committee.[162] Thus, the **consultation of the ethics committee** shall be **required in case of a change of purpose even if only anonymised data is processed**.

According to Art. 17 archived samples may be used to produce human genetic data with the free, informed and express consent of the data subject. However, if such data have significance for medical and other scientific research or public health purposes, it may - upon consultation with the competent ethics committee - be used for those purposes even in the absence of consent of a person. Such data shall then be made irretrievably unlinked to an identifiable person.

Finally the Declaration proposes to the Member States to establish a monitoring mangagement framework (Art. 20). Such a framework should, according to the Declaration, be based on the principles of independence, multidisciplinarity, pluralism and transparency.

6.3.2.1.3 Universal Declaration on Bioethics and Human Rights (2005)

The Universal Declaration on Bioethics and Human Rights is the third instrument related to bioethics adopted by UNESCO. It is also the most encompassing until now. While the focus of the Universal Declarations of 1997 and 2003 has been on genetics and life sciences, the Universal Declaration covers many areas of bioethics. Its core principle is that „human dignity, human rights and fundamental freedoms are to be fully respected“ (Art. 3.1). Furthermore, it contains provisions on inter alia, prior informed consent (Art. 6 ss.), confidentiality (Art. 9), non-discrimination (Art. 10 ss.), social responsibility (Art. 14) and benefit-sharing (Art. 15). It also provides recommendations to states on how to apply the principles of the Declaration (Art. 18 ss.)

6.3.2.2 Council of Europe

6.3.2.2.1 Convention on Human Rights and Biomedicine (1997)

The Convention for the protection of Human Rights and dignity of the human being with regard to the application of biology and medicine (Convention on Human Rights and Biomedicine) has been opened for signature in Oviedo in 1997. As this Convention is an international treaty by nature it is not directly applicable, until it has been signed and ratified by the signatory states. However, several Member States, including Germany[163] and UK, have not signed up to the Convention yet. Italy has signed the Convention on the 4th of April 1997, but ratification is still pending.[164]

As the above mentioned Declarations issued by the UNESCO, the Convention establishes the principle of informed consent as a general rule for interventions in the health field (Art. 5) and provides special rules for persons incapable to consent (Art. 6) as well as on the protection of persons who have a mental disorder (Art. 7).

---

[162] See Art. 16 lit. b) of the International Declaration on human genetic data.

[163] In Germany it was primarily criticed that the Convention remains in certain points vague and unclear. It was therefore feared by the critics that it would allow, under certain circumstances, research on persons unable to consent as babies, comatose or mentally handicapped patients, even without their consent or a specific benefit for them. For further information see for e.g. Deutscher Bundestag, Enquete-Kommission Ethik und Recht der modernen Medizin, Protocol 15/41, 15. Wahlperiode, of 4. july 2005. In Germany the ratifcation process is still pending. See Deutscher Bundestag, Unterrichtung durch die Bundesregierung - Bericht der Bundesregierung über den Stand der Unterzeichnung und Ratifizierung europäischer Abkommen und Konventionen durch die Bundesrepublik Deutschland für den Zeitraum März 2009 bis Februar 2011, 28. 03. 2011, Drucksache 17/5315, pp. 4 s., available at: http://dip.bundestag.de/btd/17/053/1705315.pdf.

[164] http://conventions.coe.int/Treaty/Commun/ChercheSig.asp?NT=164&CM=&DF=&CL=ENG.

In it´s Chapter V the Convention contains provisions as to scientific research in the field of biology and medicine: As a general rule scientific research shall be carried out freely, subject to the provisions of this Convention and the other legal provisions ensuring the protection of the human being. Art. 16 determines the protection of persons undergoing research. Accordingly research on a person may only be undertaken if all the following conditions are met:

- *there is no alternative of comparable effectiveness to research on humans;*
- *the risks which may be incurred by that person are not disproportionate to the potential benefits of the research;*
- *the research project has been approved by the competent body after independent examination of its scientific merit, including assessment of the importance of the aim of the research, and multidisciplinary review of its ethical acceptability;*
- *the persons undergoing research have been informed of their rights and the safeguards prescribed by law for their protection;*
- *the necessary consent as provided for under Art. 5 has been given expressly, specifically and is documented. Such consent may be freely withdrawn at any time.*

Research on a person without the capacity to consent as stipulated in Art. 5 may be undertaken only if all the following conditions are met:

- *the conditions laid down in Art. 16, sub-paragraphs i to iv, are fulfilled;*
- *the results of the research have the potential to produce real and direct benefit to his or her health;*
- *research of comparable effectiveness cannot be carried out on individuals capable of giving consent;*
- *the necessary authorisation provided for under Article 6 has been given specifically and in writing; and*
- *the person concerned does not object.*

Thus, the expectation of a real and direct benefit is a precondition for carrying out research on a person without the capacity to consent. Art 17. sect. 2, however, provides an exception from this rule, if the research aims to contribute, through significant improvement in the scientific understanding of the individual's condition, disease or disorder, to the ultimate attainment of results capable of conferring benefit to the person concerned or to other persons in the same age category or afflicted with the same disease or disorder or having the same condition, given that the research entails only minimal risk and minimal burden for the individual concerned and the compliance to the other conditions mentioned by Art. 17. sect. 1.

Finally Art. 22 provides that if a part of the human is removed in the course of an intervention, it may be stored and used for a purpose other than that for which it was removed, only if this is done in conformity with appropriate information and consent procedures.

6.3.2.2.2 Recommendation on research on biological materials of human origin (2006)

The Recommendation Rec(2006)4[165] of the Council of Europe on research on biological materials of human origin applies to the full range of research activities in the health field involving the removal of biological materials of human origin to be stored for research use or

[165] Recommendation Rec(2006)4 of the Committee of Ministers to member states on research on biological materials of human origin, adopted by the Committee of Ministers on 15 March 2006 at the 958th meeting of the Ministers' Deputies, available at: https://wcd.coe.int/ViewDoc.jsp?id=977859.

other purposes, including material removed for a previous research project (Art. 2). According to the Recommendation the use of biological material of human origin may be accompanied by the use of associated personal data.

Unlike the Directive, the scope of this recommendation is not limited to identifiable data[166], but **also covers the use of non-identifiable**[167] human materials. Within the group of identifiable biological materials the Recommendation further distinguishes between "coded materials", where the user has direct access to the code, and "linked anonymised materials", where the code is under the control of a third party. The notion of linked anonymised materials corresponds to the notion of pseudonymised data mentioned above.[168] The notion "unlinked anonymised materials" refers to biological materials which, alone or in combination with associated data, do not allow, with reasonable efforts, the identification of the persons concerned.

The Recommendation establishes the **primacy of anonymisation** of data as far as appropriate to the research activities concerned. As a consequence, any use of biological materials and associated data in an identified, coded, or linked anonymised form should be justified by the researcher (Art 8).

The principles for the obtaining human biological materials for research are set out in Chapter III of the Recommendation. Accordingly information and consent or authorisation to obtain such materials should be as specific as possible. Furthermore an intervention in order to obtain human biological materials for storage for research purposes shall only be carried out in compliance with the Additional Protocol to the Convention on Human Rights and Biomedicine, concerning biomedical research.[169] According to this additional protocol research on human beings may only be undertaken if there is no alternative of comparable effectiveness (Art 5). Research shall not involve risks and burdens to the human being disproportionate to its potential benefits (Art 6). In addition the research project has to be approved by the competent body after independent examination of its scientific merit, including assessment of the importance of the aim of research, and multidisciplinary review of its ethical acceptability (Art. 7).

In Art. 14 the Recommendation sets out general principles applicable to all collections of biological materials, which are as follows:

- *Designation of a person and/or institution responsible for the collection;*
- *Specification of the purpose(s) of the collection.*
- *The principles of transparency and accountability should govern its management, including access to and use and transfer of its biological materials and disclosure of information;*
- *Documentation of each sample of biological material in the collection (including information on any relevant consent or authorisation);*
- *Establishment of clear conditions governing access to, and use of, the samples.*
- *Setting up of quality assurance measures (including conditions to ensure security and confidentiality during storage and handling of the biological materials).*

[166] Art 3 i) Rec(2006)4: „Identifiable biological materials are those biological materials which, alone or in combination with associated data, allow the identification of the persons concerned either directly or through the use of a code."

[167] Art 3 ii) Rec(2006)4: "Non-identifiable biological materials, hereafter referred to as "unlinked anonymised materials", are those biological materials which, alone or in combination with associated data, do not allow, with reasonable efforts, the identification of the persons concerned."

[168] See above 6.3.1.4.1.3.

[169] Council of Europe Treaty Series, No. 195.

Art. 16 contains a provision regarding the **transborder flows** (even within the EU) of human bodily material and respective data. Accordingly biological materials and associated personal data should only be transferred to another state if that state ensures an **adequate level of protection**.

The use of biological materials in research projects is subject to Art. 21 ss.. Art. 21 provides – as a general rule – that research on biological materials should only be undertaken if it is within the scope of the **consent** given by the donor. The donor may place restrictions on the use of his or her biological materials and freely refuse consent for the use in a research project of his/her identifiable biological materials, or withdraw consent, at any time.

If the intended use of identifiable biological materials in a research project is not within the scope of prior consent, if any, given by the donor, reasonable efforts should be made to contact the person in order to obtain consent to the proposed use. If this is not possible, the biological material must only be used, if a) the research addresses an **important scientific interest**; b) the aims of the research **could not reasonably be achieved using biological materials for which consent can be obtained**; and c) there is no evidence that the **donor has expressly opposed** such research use.

In contrast to materials linkable to a certain person, unlinked anonymised biological materials may be used in research provided that such use does not violate any restrictions placed by the person concerned prior to the anonymisation of the materials. Anonymisation should be verified by an appropriate review procedure. (Art. 23). According to Art. 24 of the Recommendation research should be subject to independent review, meaning that it schould only be undertaken if the research project has been subject to an independent examination of its scientific merit, including assessment of the importance of the aim of the research, and verification of its ethical acceptability. National law may additionally require approval by a competent body.

Finally Art. 25 the Recommendation states that the principles of chapter VIII (confidentiality and right to information) of the Additional Protocol concerning biomedical research[170] should be applied to any research project using biological materials and associated personal data. These principles not only provide that any information of a personal nature collected during biomedical research shall be considered as confidential and treated according to the rules relating to the protection of private life and protected against inappropriate disclosure (Art. 26). They also provide that research participants shall be entitled to know any information collected on their health and other personal information collected for a research project. In addition Art. 27 establishes a duty of care, meaning that, *"if research gives rise to information of relevance to the current or future health or quality of life of research participants, this information must be offered to them. That shall be done within a framework of health care or counselling. In communication of such information, due care must be taken in order to protect confidentiality and to respect any wish of a participant not to receive such information*." Finally, Art. 28 of the Additional Protocol provides for the availability of the research results.

6.3.2.3 Human Genome Organisation

Already in 1995 the HUGO Ethics Committee issued a first **Statement on the Principled Conduct of Genetic Research**[171]. In this Statement the Committee formulates principles regarding inter alia information, consent, respect of the choices made by the donor with

[170] Additional Protocol to the Convention on Human Rights and Biomedicine, concerning Biomedical Research Strasbourg, 25.I.2005, available at: http://conventions.coe.int/Treaty/en/Treaties/html/195.htm.

[171] Available at: http://www.hugo-international.org/img/statment%20on%20the%20principled%20conduct%20of%20genetics%20research.pdf.

regard to the use of their data and samples as well as collaboration in the field of human genetic research.

In the subsequent **Statement on DNA Sampling: Control and Access 1998**[172] the Ethics Committee of HUGO addresses several ethical issues pertinent to sample collection and sharing in genetic research. The participants shall be offered choices in the consent process reflecting the potential uses of the DNA sample and its information. In addition the Committee states that **samples obtained during medical care and stored may be used for research even without explicit prior consent of the patient**, if there is **general notification** of such a policy, the patient has not objected, and the sample to be used by the researcher has been coded or anonymized. However, samples obtained before notification of a policy may be used for other research only if the sample has been coded or anonymized prior to use.

In its **Statement on Human Genomic Databases 2002**[173] the HUGO Ethics Committee underlines once again the importance of genomic databases as global public goods and a public resource bringing benefits accessible to all humans. Therefore, the free flow of data and the fair and equitable distribution of benefits from research using genomic databases should be encouraged.

Furthermore, the HUGO Ethics Committee points out, that the choices of individuals, families and communities with regard to donation, storage and uses of samples and the information derived therefrom and privacy should be respected. As regards the instrument of "**informed consent**" the Committee, however, proposes a **broad notion**, including notification of uses (actual or future), or opting out, or, and in some cases, even blanket consent. Furthermore the Committee states that the participants should be informed of the **degree of identifiability** of their data (e.g. coded, anonymized, aggregate, etc.) and of the **security mechanisms in place to ensure confidentiality**. Finally participants should be told that samples or the information derived therefrom may be shared with other researchers including those from other countries, with commercial entities and through publication and availability on the internet.

6.3.2.4 OECD

The OECD has issued numerous publications, guidelines, statistics and working papers in the areas of biotechnology, health, science, innovation, transborder personal data protection, and access to digital research data from publicly funded projects. In 2006 the OECD published a report detailing the findings of a working party convened specifically to examine the management and governance of human genetic research databases.[174]

One of the most recent guidelines concerning human biobanks, the **Guidelines on Human Biobanks and Genetic Research in Databases**, has been adopted by by the OECD Council on 22 October 2009.[175] On over 50 pages these guidelines provide detailed principles, illustrated with "best practices" examples for biobanks and respective annotations giving additional explanation. Thus, the guidelines provide comprehensive guidance for human biobanks, reaching from general principles, over the establishment of human biobanks, the governance, management and oversight, the terms for participation, the contents of biobanks and the protection of human biological material and data, the access to human biobanks, the qualification, education and training of persons working in the field of human biobanks, to the discontinuation of human biobanks as well as the disposal of

[172] Available at: http://www.hugo-international.org/img/dna_1998.pdf.

[173] HUGO Ethics Committee, Statement on Human Genomic Databases (adopted in December 2002). Available at http://www.hugo-international.org/img/genomic_2002.pdf.

[174] See http://www.oecd.org/document/50/0,3746,en_2649_34537_37646258_1_1_1_1,00.html.

[175] Available at: http://www.oecd.org/dataoecd/41/47/44054609.pdf.

materials and data. It also contains provisions concerning the custodianship, benefit sharing and intellectuall property issues. As it would not be usefull to examine these principles in detail at this early stage, only the most important of these principles shall be outlined in the following.

The governance of human biobanks, according to principle 3 shall be governed by the **principle of transparency and accountability**. In addition, the operators shall clearly formulate its governance structure, which is publicly available. For the use of human biological materials or data in a manner not anticipated in the original informed consent process, a review process shall be installed including research ethics committees or comparable oversight mechanisms. The same procedure shall be in place for the use of human biological materials or data where consent was obtained using a broader or layered format for uses unspecified at the time of collection.[176]

As to the **terms of participation** the OECD guidelines on human biobanks promote the principle of informed consent, which in general shall be obtained from the participant. The operator may, in certain cases, provide for obtaining consent/authorisation form an appropriate substitute decision-maker, or for obtaining waiver of consent from a research ethics committee or an appropriate authority, in accordance with applicable law and ethical principles pertaining to the protection of human subjects.[177] In addition, the guidelines provide that the operators should disclose to participants, insofar as possible, the exceptional conditions under which researchers may be provided access to human biological materials or data that is not coded or anonymised.[178] Furthermore, information shall be provided as to under what circumstances the operators may be legally obliged to provide their human biological materials and data to third parties (e.g. law enforcement agencies, employers, insurance providers) for non-research purposes.[179]

According to principle 5. of the guidelines the operator of the human biobank shall be responsible to ensure that the collection and use of the stored material and data are scientifically, legally and ethically appropriate. According to principle 6. the operator shall, furthermore, establish and implement specified policies and procedures in order to protect the data and human material from unauthorised access as well as from the (re-)identification of the donor. In addition the guidelines state that the operator shall inform the donors on the duration of storage of their material and data.

Transparency is also a major issue as regards to access of data stored in a biobank. Access to biobanks shall be based on objective and clearly articulated criteria, and consistent with the donors informed consent. A transfer of data and samples stored in a biobank shall only be licit, if the recipient has adequate standards in place regarding privacy and confidentiality.[180] Therefore, the operator shall set out the terms of access to data (access agreement) and specimen and samples collected from donors (material transfer agreement). Users of data should sign confidentiality agreements when access pertains to data that are not publicly available.[181] In order to enable the tracking of data and sample usage, the participant's consent on the type of research for which their human biological materials and data can be used should be incorporated into the biobank.

In respect to the custodianship, benefit-sharing and intellectual property issues the guidelines state that biobanks should have a clear policy on who will retain rights over human biological materials and data, and on the nature of these rights, which may differ from jurisdiction to

[176] See point 3.1. OECD Guidelines on Human Biobanks and Genetic Research Databases.

[177] See principle 4.B. OECD Guidelines on Human Biobanks and Genetic Research Databases.

[178] Principle 4.E. OECD Guidelines on Human Biobanks and Genetic Research Databases.

[179] Principle 4.F. OECD Guidelines on Human Biobanks and Genetic Research Databases.

[180] Principle 7.B. OECD Guidelines on Human Biobanks and Genetic Research Databases.

[181] See points 7.5. and 7.6. OECD Guidelines on Human Biobanks and Genetic Research Databases.

jurisdiction. Furthermore the operator of a biobank shall develop a clear policy on any intellectual property rights that may arise from the use of the biobank.

6.3.3 National law

Although these (international) guidelines set up principles for the running of a biobank that reflect a common understanding of how the donor´s fundamental rights to privacy and informational self determination should be protected, the operator´s have to respect the applicable domestic legislation, regulation and policies in the first place.

At a national level the regulatory conditions for are uneven within the EU. While some Member States have explicitly regulated the field of biobanks with all the safeguards and restrictions associated, in most Member States there is no specific legislation on biobanks. As biobanks contain (sensitive) personal data, the national law´s as to the protection of personal data apply, by which the (mininum) standards set out by the Diretive 95/46 EC had to be transposed into the national laws of the Member States.

6.3.3.1 Germany

6.3.3.1.1 National law

In Germany some aspects of the handling of genetic material have been regulated by the Gesetz über genetische Untersuchungen bei Menschen[182] (Human Genetic Examination Act, also known as Genetic Diagnosis Act) which has entered into force in February 2010. This Act, however, is not applicable to genetic examinations and analyses that are undertaken for the purposes of scientific research.[183]

There is therefore currently no specific legislation related to biobanks in Germany. However whether there is a need for a specific law with respect to biobanks in Germany or not is under political discussion at the moment.[184] Whereas the German Ethics Council[185] and some of the political parties consider a specific regulation as essencial,[186] inter alia the Deutsche Forschungsgemeinschaft (DFG)[187], Germany's largest research funding organisation, states that taking existing international guidelines and the recommendations of the German Ehtics Council into account there is no need for a specific law at present.[188]

However, as there hasn't been a decision towards a specific law for biobanks in Germany yet, more general rules are applicable. Again the general rules on data protection, mainly established by the National Data Protection Act (Bundesdatenschutzgesetz, BDSG), apply to the collection and processing of data and human samples stored in biobanks. In addition due to the federal constitution of Germany the 16 federal states have their own general data protection law applicable to the public sector in their own territory. However as both the National Data Protection Act as well as the federal data protection acts have transposed the European Data Protection Directive (95/46 EC) metioned above, the set principles are almost

---

[182] ber genetische Untersuchungen bei Menschen (Gendiagnostikgesetz - GenDG), 31.07.2009, published in BGBl. I p. 2529, 3672, also available at: http://www.gesetze-im-internet.de/bundesrecht/gendg/gesamt.pdf.

[183] See sect. 2 para. 2 Human Genetic Examination Act 2010.

[184] Also see German Ethics Council, Human biobanks for research – Opinion (2010) p. 47, available at: http://www.ethikrat.org/files/der_opinion_human-biobanks.pdf.

[185] See 6.3.3.1.2

[186] http://www.aerztezeitung.de/politik_gesellschaft/arzneimittelpolitik/article/631724/spd-regierung-soll-biobanken-gesetz-pruefen.html .

[187] http://www.dfg.de/en/index.jsp .

[188] http://www.dfg.de/en/service/press/press_releases/2011/press_release_no_12/index.html .

the same. As foreseen in art 8 para 4 of Directive 96/46 EC Germany laid down exemptions for the processing of personal data with respect to scientific research (§ 40 BDSG).

Beside these general provisions there are in each federal state specific data protection laws and orders applicable to hospitals and the health sector, which provide requirements when it comes to scientific research. These legal requirements can vary substantially. Often a distinction between inhouse research and research outside the relevant medical institution is made.

The Hamburg Hospital Act,[189] for instance, provides specific binding rules on the use of data in research projects as well as the collection of samples and data. According to Sect. 12 of the Hamburg Hospital Act employees (doctors) of the hospital who have legitimately collected patient data may process these data to carry out their own scientific research project, if the legitimate interests of those affected are not jeopardized. Furthermore patient data may be transferred to a third party for a specified research project, if the data is anonymised. If the research purpose requires the processing of personal data, the transfer shall be primarily based on the consent of the patient. In case the consent cannot be obtained at reasonable cost, data may be transferred even without the patients consent if the public interest in carrying out the research project overweights the legitimate interests of the person concerned and the research purpose can not be achieved in other ways. If anonymisation is not possible, data shall only be used in a pseudonymized form. In addition it is provided that data and samples used for research have to be organizationally separated from data used for medical treatment (separation of data).

According to sect. 12a the collection of samples and patient data for general research purposes is permitted upon the donors´ informed consent. The integration of data and samples gained from medical treatment into a biobank, however, does not require the donors ´consent, if the data and samples have been anonymised. The transfer of data and samples to a third person for the purpose of scientific research shall be effected in anonymised (or at least pseudonymised) form. If the collection of data and samples shall be used for the purpose of genetic research, it shall be examined if the risk of a potential re-identification of the donors requires that pseudonymisation is effected by a (trusted) third party.

With respect to p-medicine, where the University of Kiel (CAU) and Berlin Charité are in our understanding primarily foreseen to provide access to their biobanks, there are no specific rules on biobanks governing their behavior. As regards Berlin Charité the Landeskrankenhausgesetz[190] (LKG) Berlin provides a general rule on data protection in sect. 27. Accordingly the hospital management has to ensure that patient data may only be processed as far as necessary. For the purpose of scientific research, however, patient data might be disclosed – anonymisation provided – even without the patients consent. In Schleswig-Holstein there is no specific Hospital Act containing provisions as to the processing of patient data. Consenquently the general rules of the Landesdatenschutzgeszt Schlewig-Holstein (Data Protection Act for Schleswig-Holstein) for the use of data in the public sector apply.[191]

[189] Hamburgisches Krankenhausgesetz (HmbKHG), of 17. April 1991, published in HmbGVBl. 1991, p. 127, available at: http://landesrecht.hamburg.de/jportal/portal/page/bshaprod.psml;jsessionid=E91840A794B9019F64E9050396773DB4.jp75?showdoccase=1&doc.id=jlr-KHGHArahmen&doc.part=X&doc.origin=bs&st=lr

[190] http://www.berlin.de/imperia/md/content/lageso/gesundheit/krankenhausaufsicht/lkg.pdf?start&ts=1253516283&file=lkg.pdf.

[191] See https://www.datenschutzzentrum.de/medizin/krankenh/patdskh.htm.

6.3.3.1.2 German Ethics Council[192]

6.3.3.1.2.1 Opinion: Research and biobanks (2004)[193]

Already in 2004 the German Ethics Council, formerly named National Ethics Council, issued an opinion related to research and biobanks, stressing the need for regulation in the fields of biobanking.[194] In its opinion the German Ethics Council set ups the consent of the donor "as the foundation of biobanks"[195]. Any collection of bodily tissues and personal data shall require the informed consent of the donor. This shall also be true when samples and data have been lawfully obtained for other purposes (e.g. diagnostic or therapeutic purposes) and shall subsequently be used for scientific research. In this case the requirement of consent may be waived if samples and data are completely anonymised, unless the donor has expressed a contrary wish at the time of collection of the samples.

In addition, the German Ethics Council outlines for the necessity consent scenarios facilitating long term access and usage of samples and data stored in biobanks, meaning that, under certain circumstances, the donors may agree to the use of their samples and data for the purpose of medical research in general. This "general consent" may be valid for an indefinite period of time and allow the transfer to third parties for research purposes.[196]

The donor should have the right to revoke its consent at any time. In this case it should, however, be possible to agree with the donor that samples and data do not have to be destroyed, if they are anonymised. The use of anonymised data by researchers would then still be possible.

In order to mitigate the potential risks for the donors the German Ethics Council proposes to introduce a legally binding secrecy for biobanks, meaning that once the data and samples are collected for research purposes, their use for other purposes is prohibited.

6.3.3.1.2.2 Opinion: Human biobanks for research (2010)

In its recent Opinion "Human biobanks for research"[197] published by the German Ethics Council in 2010, the Council underlined the need for a specific regulation of biobanks, as biobanks are not only increasing in number but also contain an increasing amount of socio-demographic data, genetic data and information on lifestyle. As a result the records are becoming more and more individualised, which leads – together with other factors – to a growing re-identifiability of the data subjects.[198] Furthermore, the opinion addresses issues like building networks of biobanks and the internationalisation of scientific research in biobanks as well as privatisation and commercialisation issues and questions of the extension of purposes and third party access.[199]

In this opinion the German Ethics Council proposes a five-pillar concept for biobank legislation:

---

[192] Further documents adopted by the German National Ethics Council are available at: http://www.ethikrat.org/themen/biobanken (in german and english).

[193] German National Ethics Council, Biobanks for research – Opinion (2004), available at: http://www.ethikrat.org/_english/publications/Opinion_Biobanks-for-research.pdf.

[194] For further information see e.g. Taupitz, The Use of Human Bodily Substances and Personal Data for Research: The German National Ethics Council´s Opinion, JIBL Vol. 03 I 2006, pp. 25 ss.

[195] German National Ethics Council, Biobanks for research – Opinion (2004) p. 48.

[196] German National Ethics Council, Biobanks for research – Opinion (2004) pp. 11 ss..

[197] German Ethics Council, Human biobanks for research – Opinion (2010), available at: http://www.ethikrat.org/files/der_opinion_human-biobanks.pdf.

[198] See German Ethics Council, Human biobanks for research – Opinion (2010), pp. 9 ss.

[199] See German Ethics Council, Human biobanks for research – Opinion (2010), pp. 11 ss.

The first pillar concerns the introduction of a **biobank secrecy**, which shall not only protect the personality rights of the data subject against private abuse but also against government encroachments. Therefore the biobank secrecy shall include a duty of professional discretion as well as a right to refuse to give evidence and a prohibition of seizure.[200]

The second pillar is dedicated to the **definition of permissible use of data and samples** stored in biobanks. The German Ethics Council states that the collection and use of data and samples shall – as a principle – be defined by the consent of the donor. This shall also apply in general to samples and data which are to be entered in a biobank only after a planned anonymisation or pseudonymisation.[201] Furthermore the donor should be able to define the specific purposes for which the data and samples may be used (limitation of purposes). Finally the opinion sets out minimum information duties as to the purpose of the biobank and possible transfers to third parties.[202]

Subject to the third pillar is the involvement of **ethics commissions**, which shall evaluate periodically the activities of the biobank on the basis of a report giving detailed information on the past biobank activities, in particular on the projects carried out.[203]

The fourth pillar **"Quality assurance"** outlines the measures to be taken in order to safeguard the donors´personality rights for the complete existence of human samples and respective data. This comprises technical and organisational measures against unlawful access as well as the duty of anonymisation or pseudonymisation of samples and data as soon as possible. Furthermore, operators shall provide clear regulations as to the access to biobanks. For biobanks without restriction in purpose and duration further quality assureance measures shall be established according to the opinion.[204]

The fith pillar concerns the requirement of **transparency for biobanks**, according to which all biobanks shall have a complete documentation of the way the relevant samples and data are processed.[205]

Finally the German Ethics Council addresses legal question in relation to the **protection of donors in international cooperation** dealing with biobank based research. The Council stresses that the provisions for biobanks established in Germany possibly lose their effect if the data and samples are transferred outside the EU, in particular if the country of the recipient does not provide respective provisions with regard to biobank secrecy. In these cases the danger might exist that a foreign authority which having access to biobank samples and data under its own law transferres the information thus obtained to German criminal investigation authorities and these use the data for criminal prosecution.[206] In order to prevent this risk the Council proposes inter alia that "there should be a strict separation of biobank materials and data on the one hand and reference lists which can be used to allocate the pseudonymized samples and data to the relevant donors on the other hand. The reference lists, or the connections to personal data contained in them, should not be permitted to be transferred abroad."[207]

[200] German Ethics Council, Human biobanks for research – Opinion (2010), pp. 28 ss and pp. 32 s..

[201] German Ethics Council, Human biobanks for research – Opinion (2010), pp. 37 ss..

[202] German Ethics Council, Human biobanks for research – Opinion (2010), pp. 39 ss..

[203] For further information see German Ethics Council, Human biobanks for research – Opinion (2010), pp. 40 s..

[204] German Ethics Council, Human biobanks for research – Opinion (2010), pp. 41 ss..

[205] German Ethics Council, Human biobanks for research – Opinion (2010), pp. 43 ss..

[206] German Ethics Council, Human biobanks for research – Opinion (2010), pp. 44 ss..

[207] German Ethics Council, Human biobanks for research – Opinion (2010), pp. 44 s..

6.3.3.2 United Kingdom

In the UK the activities concerning the removal, storage, use and disposal of human tissue are regulated within the Human Tissue Act 2004 (HTA 2004).[208] The scope of this act does not only cover the activities concerning the tissue of persons alive but also of deceased persons. The HTA 2004 is primarily based on the principle of consent. It provides a list of purposes for which consent is required, so called „Scheduled Purposes"[209]. Not all of these purposes concern the use of human tissues from persons alive.

The HTA 2004 establishes the requirement of (appropriate) consent as the fundamental principle and underpins the lawful removal, storage and use of body parts, organs and tissue. Carrying out activities covered by the act without appropriate consent constitutes an offence under sect 5. HTA 2004.

The notion of "appropriate consent" is closer defined by sect. 2 and 3 of Part I HTA 2004, which differentiate according to whether the data subject is a child or an adult. In the case the data subject is a child (that is alive) "approriate consent" means the consent of the child or a person having parental responsibility for the child (sect. 2). If the data subject is an adult (that is alive) "appropriate consent" means his/her consent. Activities involving material from adults lacking capacity to consent are regulated in sect. 6, which enables the Secretary of State to specify the circumstances in which it is to be deemed have consent for activities regulated by the HTA 2004.

However the requirement of consent can be waived under certain circumstances, specified in sect. 7. Subsect. 1 to 3 of this section allow the Human Tissue Authority to give a direction deeming consent to be in place in relation to relevant samples from a living person who is either untraceable, or who has not responded to requests for consent to use of his/her material, but where the material could be used to provide information which may be relevant to another person.[210] Furthermore, subsect. 4 enables the Secretary of State to make regulations which would provide a similar power for a court to deem consent to be in place where relevant material or a body could be used for health-related research.[211]

Moreover, the HTA 2004 prescribed the establishment of a Human Tissue Authority. The HTA was established on 1st April 2005.[212]

6.3.3.3 Italy

In Italy, there is no specific statutory law regarding biobanks for scientific research.[213] As the functioning of a biobank generally implies the processing of personal data the Italian data protection law has to be taken into account. The **Italian Data Protection Act**[214] was enacted

[208] The HTA 2004 replaced the Human Tissue Act 1961, the Anatomy Act 1984 and the Human Organ Transplants Act 1989 as they relate to England and Wales, and the corresponding Orders in Northern Ireland. It also covers the mentioned activities in Wales and Northern Ireland. There is separate legislation in Scotland, where the mentioned activities are governed by the Human Tissue (Scotland) Act 2006. For further information see: http://www.hta.gov.uk/_db/_documents/Information_about_HT_(Scotland)_Act.pdf.

[209] See art. 1 Human Tissue Act 2004.

[210] These exceptions are expected to be rarely-used powers. However, they may be important where valuable information could be obtained about the treatment and diagnosis of the applicant for the direction. See explanatory notes to sect. 7 of the Human Tissue Act 2004.

[211] See explanatory notes to sect. 7 of the Human Tissue Act 2004.

[212] For more information see http://www.hta.gov.uk.

[213] The existing national legislation in this field mainly applies to the collection of tissues, blood and organs for human applications. Cf. E Stefanini, " The Need for Italian Biobank Regulation", (2010) 7:1 SCRIPTed 71, http://www.law.ed.ac.uk/ahrc/script-ed/vol7-1/stefanini.asp.

[214] Law no. 675 of 31 December 1996, available at: http://www.dataprotection.it/codice_privacy_english.htm. The act covers both electronic and manual files, for both government agencies and the private sector.

in 1996 implementing the Directive 95/46 EC. This Act, however, provided only the general regulatory framework that applied to data protection, and was supplemented by a number of later acts. The **Italian Data Protection Code**[215] which entered into force on January 1st 2004, brought all the laws and regulations together that had previously governed data protection. It provides differentiated requirements for the processing of "personal data in the health care sector" covered by sect. 76[216] and the "processing of genetic data, regardless of the entity processing them" covered by sect. 90. According to the latter the processing of genetic data shall be allowed exclusively in the cases provided for in ad-hoc **authorisations granted by the Garante** (the Italian Data Protection Authority), after having consulted with the Minister for Health who shall seek, to that end, the opinion of the Higher Health Care Council.[217]

Based on sect. 90 of the Italian Data Protection Code the Italian **Data Protection Authority** issued a **General Authorisation for the processing of genetic data**, on 22 February 2007.[218] In this context the term "genetic data" comprises "any data that, regardless of its type, concerns an individual's genotypic characteristics, or the pattern of inheritance of such characteristics within a related group of individuals"[219] The General Authorisation also refers to "biological samples" as "any sample of biological material containing information on an individual's genotypic characteristics".[220]

In points 2. and 3. the General Authorisation sets **specific limitations** for the use of genetic data. Accordingly genetic data may exclusively be processed by certain **bodies** (e.g. health sector professionals, public and private health bodies, genetic laboratories, natural and legal persons who have research purposes, genetic consultants, pharmacists, defending counsels, and certain international organisations)[221] for certain **purposes**, inter alia medical and scientific needs.[222] In addition the processing of genetic data must be indispensable to carry out the intended research project.

Point 3. of the General Authorisation states:

---

[215] Law no. 196 of 30 June 2003, available in an English version at: http://www.dataprotection.it/codice_privacy_english.htm.

[216] "1. Health professionals and public health care bodies may process personal data disclosing health, also within the framework of activities in the substantial public interest pursuant to Section 85,

a) with the data subject's consent, also without being authorised by the Garante, if the processing concerns data and operations that are indispensable to safeguard the data subject's bodily integrity and health,

b) also without the data subject's consent, based on the Garante's prior authorisation, if the purposes referred to under a) concern either a third party or the community as a whole.

2. In the cases referred to in paragraph 1, consent may be given in accordance with the simplified arrangements referred to in Chapter II."

[217] This authorisations shall also specify the additional information that has to be provided to the donor with particular regard to the purposes sought and the results to be achieved also in connection with the unexpected information that may be made known on account of the processing as well as with the donors right to object to the processing on legitimate grounds.

[218] Available at: http://www.garanteprivacy.it/garante/doc.jsp?ID=1395420.

[219] Point 1. lit. a) of the General Authorisation for the processing of genetic data (2007), available at: http://www.garanteprivacy.it/garante/doc.jsp?ID=1395420.

[220] Point 1. lit. b) of the General Authorisation for the processing of genetic data (2007), available at: http://www.garanteprivacy.it/garante/doc.jsp?ID=1395420.

[221] See point 2. of the General Authorisation for the processing of genetic data (2007), available at: http://www.garanteprivacy.it/garante/doc.jsp?ID=1395420.

[222] See point 3. of the General Authorisation for the processing of genetic data (2007), available at: http://www.garanteprivacy.it/garante/doc.jsp?ID=1395420.

"Such genetic data may be processed as are closely relevant to the purposes mentioned below, where they may not be achieved, on a case by case basis, by processing either anonymous data or personal data of a different nature: …

c. scientific and statistical research with a view to protecting the community's health in the medical, biomedical and epidemiological sectors, providing that the availability of exclusively anonymous data on population samples does not allow the research purposes to be achieved, whereby the said research shall be carried out with the data subject's consent except for the statistical surveys and/or scientific researches provided for by law."[223]

Furthermore the General Authorisation stipulates processing mechanisms for genetic data and samples. The operators shall configure the collection and use of biological samples as well as the processing of genetic data in such a manner as to prevent infringements of the data subjects' rights, fundamental freedoms, and dignity. The said activities shall be carried out lawfully and fairly for specific purposes to be set out in pursuance hereof and notified to data subjects.

In order to minimise the risks of accidental disclosure and/or unlawful/unauthorised access the donors shall only be identifiable when necessary.

As a general rule, data shall be processed in an anonymous form. Where the purposes for which genetic data are processed may not be achieved without identifying the donor the operator shall take specific measures to keep identification data separate ever since collection. In addition the Italian Data Protection Authority states in its General Authorisation that any scientific and/or statistical research on the basis of genetic data and biological samples shall be carried out pursuant the relevant sector-related standards in order to assure that the data and biological samples are used for suitable scientific purposes.[224]

Furthermore sect. 4.3. sets out security measures to be taken in connection with preservation and security of genetic data and/or biological samples. Accordingly **access** to biobanks shall be controlled by security staff and/or electronic devices envisaging specific identification procedures (including biometrics). Any person admitted after closing time shall have to be identified and his/her data recorded. In addition **the preservation, use, and transportation** of biological samples shall be carried out in a manner that ensures their **quality, integrity, availability and traceability**. The electronic transfer of genetic data shall be organised via certified electronic mail after encrypting and electronically signing the information. Web application-based communication channels may be used if they provide secure communication channels. Genetic data and biological samples contained in lists, registers and/or databases shall be processed with the help of **encryption techniques** and/or by means of **identification codes**. Apart from these specific security measures, all the other (general) obligations laid down in sect. 11, 14, 22, 31 ss. of the Italian Data Protection Code and the technical arrangements concerning minimum security measures as set out in the technical specifications annexed to the said Code are to be taken into account. Finally the Data Protection Authority states that the relevant obligations shall also be fulfilled as regards biological samples.[225]

According to the General Authorisation any processing of genetic data requires the consent of the donor upon prior information about, in particular:

- all specific purposes to be achieved;
- the possible findings, also including any unexpected ones that might come from the processing of genetic data;

[223] Available at: http://www.garanteprivacy.it/garante/doc.jsp?ID=1395420.

[224] See point 4.2 of the General Authorisation for the processing of genetic data (2007), available at: http://www.garanteprivacy.it/garante/doc.jsp?ID=1395420.

[225] See point 4.3.

- whether the data subject is allowed to limit the scope of communication of his or her genetic data and the transfer of biological samples, including their possible use for other purposes;
- the retention period of genetic data and biological samples.

In cases of processing for scientific and statistical research purposes, the information notice shall also specify:

- that the consent must be given freely and may be withdrawn at any time without this being in any manner detrimental and/or prejudicial to the data subject, except where the data and biological samples do not allow the data subject to be identified any longer;
- what arrangements have been made to allow data subjects to be only identifiable for as long as is necessary;
- whether the data and/or biological samples may be retained and used for other scientific and statistical research purposes - as long as they are known – which shall be appropriately specified also with regard to the categories of entity the data may be communicated and/or the samples transferred; and
- how data subjects can access the information contained in the research project.

The "consent" requirement can be derogated in limited cases: 1) with reference to medical purposes, when the treatment is necessary to protect the health of the interested party or of his relatives; 2) with reference to judicial purposes, when the right that is to be defended in court has almost the same grade of the right to privacy; or 3) with reference to scientific purposes or work-safety purposes, when provided by law.

The scope of the consent given by the donor depends largely on the completeness and specificity of the information received. However, the utilisation of the samples and the relevant genetic data for additional research projects, in relation to which the consent has not been collected, is allowed if the scientific purposes sought are directly connected with those expressly consented to by the data subjects.[226]

In addition, the General Authorisation entitles the data subject who undergoes a genetic test the **right not to know** the results, in order to avoid serious psychological consequences in case of unexpected findings. Sect. 90 of the Italian Data Protection Code provides that "The authorisation…shall also specify the additional items of information that should be contained in the information notice… with particular regard to the purposes sought and the results to be achieved also in connection with the unexpected information". In this context the expressed wish of the donor not to be informed about the research results has to be taken into account. However this right not to know is not an absolute one, since it may be restricted when disclosure is necessary to avoid risks of serious harm to the patient or to other persons.[227]

[226] Cf. E Stefanini, " The Need for Italian Biobank Regulation", (2010) 7:1 SCRIPTed 71, http://www.law.ed.ac.uk/ahrc/script-ed/vol7-1/stefanini.asp.

[227] Cf. E Stefanini, " The Need for Italian Biobank Regulation", (2010) 7:1 SCRIPTed 71, http://www.law.ed.ac.uk/ahrc/script-ed/vol7-1/stefanini.asp

6.4 Summary

The existing provisions and guidelines provide a whole variety of measures and principles. Whereas the (very general) provisions have to be met the more specific guidelines give scope to set up a framework that faces the risks related to biobanks on the one hand and facilitates the international cooperation and scientific exchange in research with human tissues and respective data of the donors on the other hand. The overview shows clearly that the measures to be taken can vary according to the content and the purposes envisaged with respect to the biobank. Therefore, the steps and measures to be taken to establish an adequate level of data protection and donor protection will have to be developed in close cooperation with the needs of the clinical and technical partners in p-medicine.

One fundamental ethical and legal requirement for the use of human tissue samples and respective data is that it shall primarily be based on the **informed consent of the donor**.[228] Accordingly the donors have to be provided **sufficiently clear information** as to for what purposes and how their data will be used and who and under what conditions access to their data shall be granted. In order to supervise the compliance with the donors´ consent, it is proposed to connect it with the terms of use and the respective data/samples, so that relevant information is available every time the biobank is used. These data should also be transferred together with the samples and data to other institutions.

Whereas the Directive 95/46 EC states that data may only be processed for explicitly defined purposes, this requirement is not easy to handle in the context of biobanks, which by their nature, shall form the basis not only for a specific research project but also for the upcoming research in the field of human genetic data. In fact, many of the research uses are secondary to some of the original purposes and, therefore, cannot be foreseen at the time the data and samples are collected. Consenquently already a slight change of the use of their samples and data is not covered by the donors´ consent. The donors therefore either have to be re-contacted every time the data are used for a different research project in order to recieve consent to process data for these new purposes, which would be expensive and time consuming, or the donors ´consent may be waived by an ethics commitee considering the risks for the donors and the potential benefits for the public.

Another fundamental principle for biobanks is that data shall primarily be used in **anonymised** form.[229] However, anonymisation is not appropriate in all cases, e.g. where the development of the donor shall be observed. If anonymisation is not possible data and material shall be used in **pseudonymised** form.

The **requirement for informed consent may be waived** in some circumstances by an authorised entity such as a research ethics committee in accordance with the applicable law and ethical principles pertaining to the protection of human subjects and will vary from jurisdiction to jurisdiction.

In general it is stated that the donors should have the **possibility to limit or revoke their consent at any time**. However, the guidelines provide different consequences in cases where donors´ revoke their consent to the use of data. Although the initial basis of collection and use of samples and data falls away, there is generally no obligation according to the guidelines to destroy the samples and delete the data of the donor. The operators rather are obligated to anonymize samples and data, so that the donor cannot prevent the use of his

[228] See Art. 5 ss. of the Council of Europe, Convention on Human Rights and Biomedicine (1997); Art. 5 UNESCO, Universal Declaration on Human Genome and Human rigths (1997); Art. 8 UNESCO, International Declaration on human genetic data (2003); Art 6 ss. UNESCO, Universal Declaration on Bioethics and Human Rights (2005).

[229] See Art. 14 lit. c) International Declaration on Human Genetic Data (2003);

data once the data is anonymised. This, however, can be problematic, considering that firstly the anonymization itself is a processing of data in the sence of the Directive and therefore needs a legal basis and secondly in certain cases an absolutely secure anonymisation is not possible.

Access to the biobank shall only be based on objective and clearly defined criteria and should be in line with the donors´ informed consent. Furthermore it is stated that researchers shall only have access to data and samples that are anonymised or – if anonymisation is not possible – pseudonymised, so that the donor cannot be identified by the researcher. In addition the researchers shall be (by law or contract) obligated to refrain from any attempt to re-identify the donors.

It is commonly agreed that biobanks have to establish highest **data security standards** by providing both technical and organisational measures to protect the data contained in biobanks against **accidential loss and destruction** as well as against **unlawful access** to the data and samples stored. For instance, data controllers should be encouraged to carry out surveys of potential risks, establish policies for security, inform and train staff on an on-going basis, set up restricted access controls in order to avoid undue access by administrative staff and other persons, etc. Furthermore the access to and changes of data and samples shall be logged.

The operators of biobanks shall provide for **transparent operating procedures and policies** for the procurement, collection labelling, registration, processing, storage, tracking, retrieval, transfer, use and destruction of human biological materials, data and information. Such technical and organisational policies have to be re-**evaluated periodically.** Finally, the operators shall clearly formulate its governance structure, which is to hold publicly available.

7 IT Standardization in integrated biobanking

Biobank integration into a larger infrastructure or federating biobanks for facilitated access to biomaterial and data requires standardization of various aspects on several levels. Some of them are:

- Biomaterial:
  - Sample quality: Best practices for sample collection, processing, storage and distribution have to be implemented in order to ensure usability of the biomaterial for research.
  - Unique sample identification and cross-laboratory sample tracking mechanisms
- Data
  - Data quality
  - Data harmonization; terminology and taxonomy for sample classification
  - Data exchange between labs and related data set and communication standards
  - Data linkage with other data sources
  - Linkage of samples and their data to associated research and their results
- Ethical and legal aspects:
  - Privacy protection
  - Access management
  - Informed consent and possibility for later withdrawal

While various institutions have issued best-practices-guidelines for the management of biomaterial repositories, which cover only some of the above aspects on the biomaterial level and on the ethical and legal level, these standardisation aspects have not yet been sufficiently addressed for federated biobanks and respective meta biobank infrastructures.

BBMRI developed a strategy for communication between biobanks, including a common nomenclature, compatible software techniques and appropriate information transmission policies. The strategy proposes a generalized metadata model for regional metadatabases named BBMRI Hubs. The model holds information about the relation between participating biobanks, their content, user roles and rights. A content meta structure holds a sufficient set of needed information structures and schema elements to boost query answering. Schema elements from local biobanks shall be mapped with the BBMRI Content-Meta structure. The resulting implementation model is a federated hubs and spokes network as further described in Chapter 5.2 ‘BBMRI - WP5 Database harmonization and IT-infrastructure’.

7.1 Harmonisation of data sets for federated biobank infrastructures

7.1.1 BBMRI minimum data set and biobank lexicon

A minimum data set was proposed as an intermediate between BBMRI questionnaires and the generalized metadata model. The dataset is composed of a limited number of attributes describing i) biobanks, ii) studies and iii) subjects, cases and samples within biobanks.

Data describing biobanks		
Definition	**Allowed values**	**Explanation**
BiobankAcronym	ASCII	
NameOfBiobank	Free text in English	
Institution	Free text in English	
URL		
Country	ISO-standard (3166 alpha2), two letter code	
ContactName	Free text in English	
ContactData	Free text in English	Address, Phone (E.164, No. 905 – 1.IV.2008), e.g., +46 8 524 877 59, Mail
Data describing studies		
Definition	**Allowed values**	**Explanation**
NameOfStudy	Free text in any language	
EnglishStudyName	Free text in English	Translation of study name in English
ContactName	Free text in English	
ContactData	Free text in English	Address, Phone (E.164, No. 905 – 1.IV.2008), e.g., +46 8 524 877 59, Mail
KindOfStudy	Population-based, specific-disease, broad-spectrum of diseases	If "specific-disease", note ICD10
CategoriesOfDataCollected	[ClinicalDataAvailable, Diagnosis, Health information, Physiological/biochmical measures, Sociodemographic char., Socioeconomic char., Life habits/Behav., Physical environment]	Can be several values
Data describing subjects/cases/samples within biobanks		
Definition	**Allowed values**	**Explanation**
AgeGroup	Interval [a,b], a>0, b<200, b>=a	a and b should be selected so that k-anonymity is guaranteed. Age group of donor at time for sample collection, number of age groups determined by biobank
Gender	Male, Female, Other	Gender of subject
SampleType	DNA, cDNA/RNA, whole blood, blood cells isolates, serum,	Type of sample. From the BBMRI core question.

	plasma, fluids, tissues cryo, tissues paraffin-imbedded, cell-lines	
SampleDate	ISO-standard (8601) time format	Date when sample was harvested
ClinicalDataAvailable	Yes/No	There exists clinical data related to the sample
OrganCategory	From the BBMRI Detailed descr bio samples	
OmicsDataAvailable	Yes/No	Genomics, proteomics etc
RestrictionsOnSampleUse	None, Consent participant, IRB approval, Approval of owner of collection	Can be several values
NOTES: Time stamp and version control are part of the metadata schema and upload services		

Table 3: BBMRI Minimum Data Set for a federated biobanking infrastructure [230]

In addition to the BBMRI Minimum Data Set BBMRI members have released a multi-language biobank lexicon which defines the nomenclature used in the context of biobanking. The lexicon comprises around 90 concepts ranging from aliqot, anonymization, audit to unidentifiable, virtual biobank and wrapper. The main intention is to define the common language used in that domain through descriptive definitions of the concepts. However, the concepts have no meaning for data integration in federated biobanks

7.1.2 i2b2 Crimson Sample Ontology

Sample ontology is presented on the Crimson homepage in form of an Excel table. The sample ontology represents a hierarchy that describes different types of specimen. Around 400 concepts with characteristics of samples are distinguished in the Excel sheet. A demonstration of the Crimson functionality in the i2b2 Hive shows enrichment of the i2b2 querying and analysis tool that allows querying and managing samples from different cohorts. It also shows the integration of multiple ontologies such as 'Anatomical Source of a Biosample'.

7.1.3 Data Harmonization Approach of the CRIP Toolbox

The IRDB component of the CRIP Toolbox (cf. chapter 5.6) automatically runs a drive-by data harmonization including:

- validation of data value types (text, numbers, dates)

- mapping of values to catalog values

- calculation of values (e.g. patient's age at asservation from date of birth and date of asservation)

- rewriting values making use of regular expressions (e.g. date translations)

[230] Litton JE et al. (2010) BBMRI D5.6 Final Report

- preprocessing of import files

- mapping of data keys/field names

- error analysis

These services spare the network's biobank partners adapting or mapping their local data set definition into a mandatory central specification if and when participating in the network. Through this user-friendly tool, information on cases and samples can be made available over the web to external research partners without superimposing the local coding/annotation specifications/guidelines with external rules and requirements. Local systems are not affected by network implementation and stay operative all the time.

7.1.4 ACGT Minimal basic dataset for research related human biomaterial repositories

A database for bio-banking is described in D2.2 of the ACGT project[231]. The tables of this database are given here and can be seen as a minimal basic dataset for bio-banks.

The dataset can be divided into different parts:

- Dataset for managing a bio-bank (logistics)
- Dataset for describing the stored biomaterial
- Dataset for describing the research that is done by whom
- Dataset dealing with the requirements for a bio-bank taking part in ContraCancrum

The description of the format of the data is given in the column 'type'. Reference to the TCGA data portal (http://cancergenome.nih.gov/dataportal/data/about/) is given.

Biobank logistics

Context	Category		Type	Format/ Coding/Options	TCGA Reference
Bio-bank	Name of the Bio-bank		A50		
	ID of the Bio-bank in ContraCancrum		I5		
	Localisation		I5	Institution-ID	
	Responsible person		I5	User-ID	
	commercial Bio-bank		I2	-1: not yet known 1: no 2: yes	
		If yes, costs for storage / vial and year	R8.2	€ for storing material	
	service provided		I2	-1: not yet known 1: no 2: yes	
		If yes, please specify	A254		

[231] ACGT: D2.2. User requirements for an ontology based clinical data management system and for the Trial Builder. pp 54-61, 15th September 2007

Opening Date	Date since running		D8		
Specification, what the Bio-bank is able to do (multiple entries possible)	Normal tissue	will be stored	I2	-1: not yet known 1: no 2: yes	
		If yes, specify	A25	Which tissue	
			I2	-1: not yet known 1: cell culture 2: DNA 3: RNA 4: Protein 5: Paraffin material 6: other	
		If other, specify	A25		
	Tumour material	will be stored	I2	-1: not yet known 1: no 2: yes	
		If yes, specify	A25	Which tissue	
			I2	-1: not yet known 1: cell culture 2: DNA 3: RNA 4: Protein 5: Paraffin material 6: other	
		If other specify	A25		
	other tissue	will be stored	I2	-1: not yet known 1: no 2: yes	
		If yes, specify	A25	Which tissue	
			I2	-1: not yet known 1: cell culture 2: DNA 3: RNA 4: Protein 5: Paraffin material 6: other	
		If other specify	A25		
Size of Bio-bank	Number of samples that can be stored		I7		
	Number of samples that are already stored		I7		
Information about the Bio-bank	provided		I2	-1: not yet known 1: no 2: yes	
		If yes, specify	M	Memo	

		Kind of Information	I2	-1: not yet known 1: only general Information 2: Templates for contracts	
		Shipping conditions	I2	-1: not yet known 1: no 2: yes	
		Homepage	A254	http://	
		Contact email	A100	Include @	
Bio-bank committee	Names	User-ID	I5	Multiple entries	

Table 4: ACGT Minimum Data Set for biobank logistics.

Stored biomaterial from a patient

Context		Category	Type	Format/ Coding/Options	TCGA Reference
Patient Pseudonym		Pseudonym	A25	Pseudonym	
Bio-bank-ID		Bio-bank-ID	I5		
Informed consent		Type	I2	-1: not yet known 1: donated material, every analysis possible 2: analysis restricted 3: new informed consent for each analysis needed	
Bio-material	collected	Date	D8	DDMMYYYY	
	stored	Date	D8	DDMMYYYY	
Material*		Material-ID	I15		
		type	I2	-1: not yet known 1: blood 2: plasma 3: serum 4: tumour tissue 5: normal tissue 6: bone marrow 7. cerebrospinal fluid 8: urine 9: other	
		If other, specify	A50		
storage of material before processing and definite storage		Done	I2	-1: not yet known 1: no 2: yes	
		If yes:	I2	-1: not yet known 1: room temperature 2: refrigerator	

	Storage before processing			3: deep-frozen: - 20°C 4: deep-frozen: - 80°C 5: stabilisation agent added 6: directly processed 7: other	
	If stabilization agent added: please specify		A100		
	If other, specify		A50		
	Shipping/Transport necessary		I2	-1: not yet known 1: no 2: yes	
	If yes: Please specify	Kind of transport	I2	-1: not yet known 1: without cooling 2: with cooling 3: deeply frozen	
		Label for temperature	I2	-1: not yet known 1: no 2: yes	
		Condition of material after transport	I2	-1: not yet known 1: good 2: defrosted 3: deeply frozen during whole transport 4: cold chain interrupted	
		Duration [min]	I5		
Corresponding Histology	Done		I2	-1: not yet known 1: no 2: yes	
	If yes: Please specify	Histology	A50	*including normal tissue*	
		ICD-10	A20		
		Comment	M	Memo	
		Tumour	I2	-1: not yet known 1: inhomogeneous tumour 2: homogeneous tumour 3: only normal tissue	
		Tumour cells	I2	-1: not yet known 1: no 2: vital 3: only necrotic	
Time before processing	Minutes		I5		
Processing	Primary method	general	I2	-1: not yet known 1: nothing 2: cell isolation 3: cell sorting 4: cell isolation and sorting	

				5: other	
		If other, specify	A50		
		Method specification	M	Memo	
	Secondary method	general	I2	-1: not yet known 1: cell culture 2: controlled frozen for cell culture 3: DNA extraction 4: RNA extraction 5: protein extraction 6: membrane extraction 7: mitochondria extraction 9: other	
		If other, specify	A50		
		Method specification	M	Memo	
DNA	Quality		I2	-1: not yet known 1: 100 % pure 2: 90 % pure 3: < 90 % pure	
	Total amount isolated [µg]		R6.2		
	Number of vials		I2		
	Vial-ID**		A29	Multiple entries possible	PK dna_id
	Storage	Date	D8		
		Temperature			
		Place: Institution ID	I15		
RNA	Quality		I2	-1: not yet known 1: 100 % pure 2: 90 % pure 3: < 90 % pure	
	Total amount isolated [µg]		R6.2		
	Number of vials		I2		
	Vial-ID**		A29	Multiple entries possible	PK rna_id
	Storage	Date	D8		
		Temperature			

		Place: Institution ID	I15		
Protein	Quality		I2	-1: not yet known 1: 100 % pure 2: 90 % pure 3: < 90 % pure	
	Total amount isolated [µg]		R6.2		
	Number of vials		I2		
	Vial-ID**		A29	Multiple entries	
	Storage	Date	D8		
		Temperature			
		Place: Institution ID	I15		
other	Specify material		A10		
	Quality		I2	-1: not yet known 1: 100 % pure 2: 90 % pure 3: < 90 % pure	
	Total amount isolated [µg]		R6.2		
	Number of vials		I2		
	Vial-ID**		A29	Multiple entries	
	Storage	Date	D8		
		Temperature			
		Place: Institution ID	I15		
Paraffin	ID of the institution		I15		
	Histological No.		A25		
	Responsible physician User ID		A20		
	Paraffin Block available		I2	-1: not yet known 1: no 2: yes	
	Number of slides		I3		
	Slide-ID**		A29	Multiple entries	PK-slide_id

* For each material of a patient an extra dataset has to be provided

** Vial ID will be automatically generated: Material-ID + Kind of material + No. of the vial (for example: 123456-DNA-5: meaning this is the 5th vial containing DNA from material with the ID of 123456)

Table 5: ACGT Minimum Data Set for biomaterial description.

Research that is done by whom using material from the bio-bank

Context	Category	Type	Format/ Coding/Options	TCGA Reference
Bio-bank ID	Institution-ID	I5	Multiple entries possible	
Material ID	Material-ID	I15	Multiple entries possible	
Number of vials needed	Number	I5		
Vial ID	Vial-ID	A25	Multiple entries possible	
Institution doing the research	Institution-ID	I5		
Main Researcher	User-ID	I5		
Project	Name of project	A254		
	Acronym of project	A100		
	Main question	M	Memo	
	Method	M	Memo	
	Research protocol available	I2	-1: not yet known 1: no 2: yes	
	Research approved by EC/IRB	I2	-1: not yet known 1: no 2: yes	
		I5	ID	
	Research financed by	I5	ID	
Request for material	approved	I2	-1: not yet known 1: no 2: yes	
	Date of request	D8	DDMMYYYY	
	Date of approval	D8	DDMMYYYY	
	Contract signed	I2	-1: not yet known 1: no 2: yes	
	Contract signed – Date	D8	DDMMYYYY	
	Date material is shipped	D8	DDMMYYYY	
Result	published	I2	-1: not yet known 1: no 2: yes	
	Link to PubMed	A254	Multiple entries possible	

Table 6: ACGT Minimum Data Set for research description of biomaterial requester.

Requirements for Participation of a Bio-bank / Molecular Biological Laboratory

Context	Category	Type	Format/ Coding/Options	TCGA Reference
Bio-bank ID	Institution-ID	I5		
Bio-bank fulfils legal requirements to participate in the project		I2	-1 not yet known 1: no 2: yes	
Bio-bank fulfils ethical requirements to participate in the project		I2	-1 not yet known 1: no 2: yes	
Bio-bank has SOPs according to GLP criteria		I2	-1 not yet known 1: no 2: yes	
Bio-bank is a registered member of the project		I2	-1 not yet known 1: no 2: yes	
Bio-bank has already participated in a trial of the project		I2	-1 not yet known 1: no 2: yes	
Responsible Person of the Bio-bank is registered in p-medicine		I2	-1 not yet known 1: no 2: yes	
	If, yes, User-ID	I5		
Date when contract with project was signed	Date	D8		
Person who signed contract for project	User-ID	I5		
Person who signed contract for the Bio-bank	User-ID	I5		

Table 7: ACGT Minimum Data Set for the participation of a biomaterial repository in the project.

Molecular biological and histological data

Context	Category	Type	Format/ Coding/Options	TCGA Reference
Material used	Institution-ID	I5	Multiple entries possible	
	Material-ID	I15	Multiple entries possible	
	Vial-ID / Slide-ID	A25	Multiple entries possible	
Pathology	Section location	R10.2		SECTIONLOCATION
	Number proliferation cells	R10.2		NUMBERPROLIFERATINGCELLS
	Percent tumour cells	R10.2		PERCENTTUMOURCELLS
	Percent normal cells	R10.2		PERCENTNORMALCELLS
	Percent necrosis	R10.2		PERCENTNECROSIS
	Percent stromal cells	R10.2		PERCENTSTROMALCELLS
	Percent lymphocyte infiltration	R10.2		PERCENTLYMPHOCYTEINFILTRATION
	Percent monocyte infiltration	R10.2		PERCENTMONOCYTEINFILTRATION
	Percent granulocyte infiltration	R10.2		PERCENTGRANULOCYTEINFILTRATION
	Percent neutrophile infiltration	R10.2		PERCENTNEUTROPHILINFILTRATION
	Percent eosinophile infiltration	R10.2		PERCENTEOSINOPHILINFILTRATION
	Endothelial proliferation	A50		ENDOTHELIALPROLIFERATION
	Nuclear pleomorphism	A50		NUCLEARPLEOMORPHISM
	Palisading necrosis	A50		PALISADINGNECROSIS
	Cellularity	A50		CELLULARITY
	Percent p53 staining	R10.2		
	Percent ki67 staining	R10.2		

Table 8: ACGT Minimum Data Set for molecular and histological data.

7.2 Data sets proposed by p-medicine's biobanking use case owners

The use case owners for integrated biobanking have suggested data sets according to the nomenclature and variables used in their corresponding biomaterial resources that should be taken for federated biobanking with their research partners. University Hospital Schleswig-Holstein of the Christian-Albrecht-Universität zu Kiel (CAU) proposed two greatly overlapping

data sets for acute and relapse lymphoblastic leukaemia, while the Biocenter of the University Würzburg that collects, preserves and analyses the samples of the German SIOP Wilms tumor trial proposes a different table of relevant data items for the SIOP Wilms tumour samples. They are shown in the following subchapters.

Further analysis in collaboration with WP4 'Standardisation, Semantic Interoperability and Data Integration' is required in order to harmonise the different data sets and to elaborate a more general approach for the integration of biobank resources of different research communities with their trials and projects. A biobank ontology that focus on integrated biobanking for clinical research and that can be understood and maintained by clinicians would be required for this purpose.

7.2.1 P-medicine proposed dataset for lymphoblastic leukaemia biobanking scenario

Acute lymphoblastic leukaemia

Description	Variable Name	Values
Study protocol	protocol	standardized text
Treatment arm	arm	standardized text
follow-up min. 5 years/event	fu_5y	no/yes
Relapse study	rel_stud_ny	no/yes
Immunphenotype	immunphen	precursorB/T-ALL/other
Sex	sex	male/female
WBC at diagnosis	leuko	
Age (years)	age	
Blasts BM	blast_bm	
Blasts PB	blast_pb	
CNS involvement	cns	no/yes
Testicular involvement	testis	no/yes
Other extramedullary site involved	other_extram	no/yes
Main cytogenetic aberration	cytogen	standardized text
TEL/AML1	tel	no/yes
BCR/ABL	bcr	no/yes
MLL/AF4	mllaf4	no/yes
MLL/ENL	mllenl	no/yes
MLL/AF9	mllaf9	no/yes
DNA-Index	icp	
IKZF1	ikzf1	not mutated/mutated
CRLF2	crlf2	not mutated/mutated
NOTCH1	notch1	not mutated/mutated
TP53	tp53	not mutated or deleted/mutated only/deleted only/mutated and deleted
TPMT	tpmt	WT/heterozygote/homo.
Karyotype available	karyotyp_ny	not done/done
MLPA screening	mlpa_ny	no/yes
MLPA kit		standardized text
Gene expression profiling	gep_ny	no/yes
Gene expression platform		standardized text

SNP germline	snp_germ_ny	no/yes
SNP germline platform		standardized text
SNP tumour	snp_tumour_ny	no/yes
SNP tumour platform		standardized text
Epigenetic profiling	epi_prof_ny	no/yes
Epigenetic platform		standardized text
Whole exome seq	exo_seq	no/yes
Sequencing platform		standardized text
Whole genome seq	gen_seq	no/yes
Sequencing platform		standardized text
Micro RNA profiling	mir_prof_ny	no/yes
Micro RNA platform		standardized text
Prednison response	predresp	good/poor
MRD BFM TP1 available	mrd_TP1_ny	no/yes
MRD BFM TP2 available	mrd_TP2_ny	no/yes
MRD other time point(s)	mrd_other	standardized text
Nonresponse after PROT Ia	NR33	no/yes
Available types (for the different time points)		
Bone marrow	bm	no/yes
Blood	pb	no/yes
Cerebrospinal fluid	csf	no/yes
Tissue	tissue	no/yes
Other material type	other_mat	no/yes
Xenoprimograft after first passage	xeno_firstpa	no/yes
Xenoprimograft after second passage	xeno_secpa	no/yes
Avalable material RNA	RNA	no/yes
DNA	DNA	no/yes
cDNA	cDNA	no/yes
Cells	Cells	no/yes
Serum	Serum	no/yes
Plasma	Plasma	no/yes
Protein	Protein	no/yes
Smears	Smears	no/yes
Cytospins	Cytospins	no/yes
Relapse occurred	relapse	no/yes
Death	death	no/yes
Treatment related death frontline	trm	no/yes
Secondary malignancy (before relapse)	smn	no/yes

Table 9: Proposed data set for acute lymphoblastic leukaemia.

Relapse lymphoblastic leukaemia

Description	Variable Name	Values

Study protocol	protocol	standardized text
Treatment arm(s)	arm	standardized text
follow-up min. 5 years/event	fu_5y	no/yes
Immunphenotype	immunphen	precursorB/T-ALL/other
Sex	sex	male/female
WBC at diagnosis	leuko	
Age (years)	age	
Blasts BM	blast_bm	
Blasts PB	blast_pb	
CNS involvement	cns	no/yes
Testicular involvement	testis	no/yes
Other extramedullary site involved	other_extram	no/yes
Main cytogenetic aberration	cytogen	standardized text
TEL/AML1	tel	no/yes
BCR/ABL	bcr	no/yes
MLL/AF4	mllaf4	no/yes
MLL/ENL	mllenl	no/yes
MLL/AF9	mllaf9	no/yes
E2A/PBX1	e2apbx1	no/yes
DNA-Index	icp	
IKZF1	ikzf1	not mutated/mutated
CRLF2	crlf2	not mutated/mutated
TP53	tp53	not mutated or deleted/mutated only/deleted only/mutated and deleted
NOTCH1	notch1	not mutated/mutated
Karyotype available	karyotyp_ny	not done/done
MLPA screening	mlpa_ny	no/yes
MLPA kit		standardized text
Gene expression profiling	gep_ny	no/yes
Gene expression platform		standardized text
SNP germline	snp_germ_ny	no/yes
SNP germline platform		standardized text
SNP tumour	snp_tumour_ny	no/yes
SNP tumour platform		standardized text
Epigenetic profiling	epi_prof_ny	no/yes
Epigenetic platform		standardized text
Whole exome seq	exo_seq	no/yes
Sequencing platform		standardized text
Whole genome seq	gen_seq	no/yes
Sequencing platform		standardized text
Micro RNA profiling	mir_prof_ny	no/yes
Micro RNA platform		standardized text
Cytologic response	cytresp	no/yes

MRD after induction available	mrd_ind_ny	no/yes
MRD before SCT available	mrd_befsct_ny	no/yes
MRD other time point(s)	mrd_other	standardized text
MRD at relapse diagnosis (submicroscopic BM involvement)	subm_bminv	no/yes
Available material		
Time point of sampling		
Relapse diagnosis	reldia	no/yes
after induction	after_ind	no/yes
before SCT	before_sct	no/yes
Available types (for the different time points)		
Bone marrow	bm	no/yes
Blood	pb	no/yes
Cerebrospinal fluid	csf	no/yes
Tissue	tissue	no/yes
Other material type	other_mat	no/yes
Xenoprimograft after first passage	xeno_firstpa	no/yes
Xenoprimograft after second passage	xeno_secpa	no/yes
Available fractions (for the different types)	RNA	no/yes
DNA	DNA	no/yes
cDNA	cDNA	no/yes
Cells	Cells (MNC)	no/yes
Plasma	Plasma	no/yes
Protein	Protein	no/yes
Smears	Smears	no/yes
Cytospins	Cytospins	no/yes
Secondary relapse	relapse	no/yes
Death	death	no/yes
Treatment related death after relapse	trm	no/yes
Secondary malignancy (after relapse)	smn	no/yes
For all variables: alternate response codes: not done, unknown		

Table 10: Proposed data set for relapse lymphoblastic leukaemia.

7.2.2 P-medicine proposed dataset SIOP Wilms Tumor samples

The following data sets for SIOP Wilms Tumor are mainly motivated from the perspective of biomaterial management rather than federated biobanking (biobanking use case BA_2 Managing biomaterial data in ObTiMA).

Material related data items

touch preps	number	number: ~ 1-10, no order	Identifier 10 Ids
blood	number of vials	ca. 1-3 (with date) type (EDTA, heparin,unknown) volume (1-10 ml for each) DNA extracted (yes/no) amount DNA (microgram/ml) free text	3 Ids per ID per ID per ID per ID per ID
normal kidney	number of samples	1-3 (with date) amount per sample (ml) frozen -80, -20, thawed, in culture medium DNA extracted (yes/no) amount DNA (microgram/ml) DNA quality (text) free text	3 Ids per ID per ID per ID per ID per ID per ID
tumor	number of samples	1-5 (with date) amount per sample (ml) frozen -80, -20, thawed, in culture medium) additional identifiers (free text) DNA extracted (yes/no) amount DNA (sampleID) DNA quality (text) RNA extracted (yes/no) amount RNA (sample ID) RNA quality (text or RIN)	5 Ids per ID per ID per ID per ID per ID per ID per ID per ID per ID
material for culture		0-2 (+ free text) cultured, successful? (Y/N + text) frozen in DMSO? (number of vials, text)	2 ID per ID per ID
additional material	free field		
Multiple extractions of DNA and RNA from a single sample can occur. Sub-IDs of should be assigned to these derived materials that maintain the link to the original sample.			

Table 11: Proposed data set for SIOP Wilms tumor sample.

Analyses data

allele loss	chromosome arm 1p chromosome arm 11p chromosome arm 16q Further chromosoms may be added in the future!	marker1 marker2 marker3 extendable, 0-10 markers marker1 marker2 marker3 extendable, 0-10 markers marker1 marker2 marker3 extendable, 0-10 markers	
CTNNB1	exon3 size exon3 sequence	ok/altered/nd wt/het/hom (+text)	
WT1	deletion analysis mutation analysis	ok/altered/nd (+text) ok/altered/nd (+text)	
mRNA expression		yes/no (+text)	
addl. tests		text	

Table 12: Proposed data set for analysis data with SIOP Wilms tumor sample.

8 P-medicine approach for a biobank access framework

8.1 IT Approach for the p-medicine biobank access framework

In this section we propose a technical solution for the p-medicine biobank access framework according to the use cases described in Chapter 3. We will firstly summarize the main functionality based on the use cases and then describe the basic architecture of the biobank access framework.

8.1.1 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. The framework harmonizes the data according to a standard biobank data set as described in Chapter 7. In particular the p-medicine biobank framework will support the following main functionalities, which are derived from the four use cases described in Chapter 3:

1. **Offering Biomaterial for Research**

   A biomaterial owner is supported in offering his biomaterial and related data to open or closed research communities, according to legal aspects. The offered data can be stored in any arbitrary biobank management system. The owner has the possibility to select which of his biomaterial he wants to offer to which research communities. Furthermore, biomaterial owners can push their data into the p-medicine data warehouse in order to link the data to other biomedical data sources.

2. **Searching and Requesting Biomaterial for Research**

   A researcher is enabled to search the biomaterial that is offered within his communities. He can get information about the available quantity and data that is related to the material. It is furthermore possible for him to request biomaterial for a research project. For this purpose the project needs to be described in detail. The biobank access framework forwards the request to the biomaterial owner.

3. **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 pre-defined but adjustable case record forms for patient's biomaterial according to a standard biobank dataset are provided in ObTiMA. 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 that comply with the standard biobank dataset.

8.1.2 Basic Architecture of the p-medicine Biobank Access Framework

We have designed the biobank access framework as a set of loosely coupled components, which are depicted in Figure 36. The main component of the framework is ***p-BioSPRE***, the p-medicine Biomaterial Search and Project Request Engine, which is a metabiobank to share biomaterial for research purposes. Furthermore, the framework comprises the ***p-Biobank Wrappers,*** which are tools to support biobank owners to offer their biomaterial and related data, which can be stored in any biobank information system, in p-BioSPRE and manage associated requests. In order to enable users of the p-medicine trial management system

ObTiMA to integrate biomaterial data in clinical trials and offer it in p-BioSPRE a ***Trial Biomaterial Manager*** is provided.

As described in Chapter 5, the solution for the p-medicine access framework is technically based on the CRIP toolbox. In the following sections we will describe the different components in more detail.

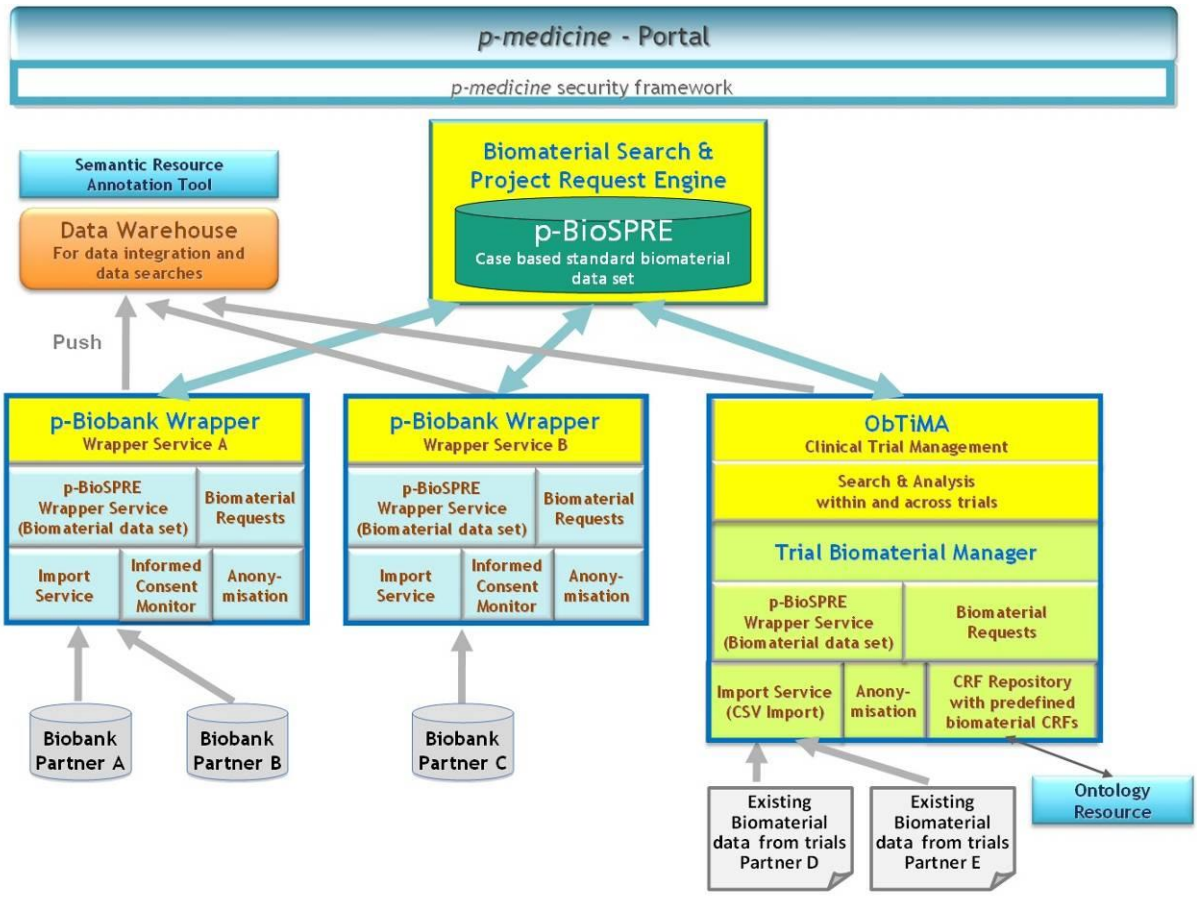


Figure 36: Basic Architecture of the p-medicine Biobank Framework.

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

P-BioSPRE is a metabiobank that provides researchers the possibility to search for and request biomaterial that fits their research purposes. Technically p-BioSPRE bases on the CRIP metabiobank. It is a web application and can be accessed via the p-medicine portal. The application's management of user roles and rights is compliant to the p-medicine security framework.

P-BioSPRE provides a search interface that enables authorized users to search for biomaterial according to the standard biobank data set. A user is authorized to search biomaterial and related data if he has a p-medicine user account and is a member of the research community the biomaterial is offered to. According to the legal requirements described in Chapter 6, the biomaterial data that is provided in p-BioSPRE is anonymized.

When the user has found appropriate biomaterial for his research, he can request to order the material. For this purpose p-BioSPRE provides request forms that enable the user to specify the amount of biomaterial he needs and to define his research project in detail. P-BioSPRE forwards the request to the biomaterial owner, who can then contact the researcher and decide about the request.

8.1.2.2 p-BioBank Wrappers

Biomaterial data is uploaded into p-BioSPRE from so called p-Biobank Wrappers. A p-Biobank Wrapper enables a user to share his biomaterial and related data in p-BioSPRE within an open or closed research community. Technically a p-Biobank Wrapper is based on the Integrative Research Database from the CRIP toolbox. It is a local server that is installed at the site of a biomaterial owner and configured to link one or more of his biobank management systems, in which data is stored that he wants to share. As a precondition to link a biobank management system to a p-Biobank Wrapper the system needs to implement an export interface that allows exporting the biomaterial data according to the standard biobank data set.

The p-Biobank Wrapper provides an interface to support a biomaterial owner to import sample data from the linked biobank management systems. The process is realized by an import service that pseudonymizes the imported data according to the p-medicine concept via a trust center. It is specified for each of the biobank samples which kind of research the patient allows for his samples. An according Informed Consent Monitor will be developed in WP 14.

The biobank owner can select the imported data that he wants to share with certain research communities and upload the data to p-BioSPRE. During the upload process the data is anonymized. It has to be noted, that if biomaterial data for the same patient is imported from different biobank information systems into a p-Biobank Wrapper, the data is still linked, since for a patient always the same pseudonym is used. The information that the data belongs to the same patient is preserved when the data is anonymized and uploaded into p-BioSPRE. This information is lost, when the data is uploaded from different p-Biobank Wrappers.

The p-Biobank Wrapper enables a biobank owner to manage his biomaterial requests. When a user in p-BioSPRE requests to order biomaterial, the according biomaterial owner can view that request in his instance of the p-Biobank Wrapper and contact the researcher.

Furthermore, the p-Biobank Wrapper comprises push services, which enable a biobank owner to push selected biomaterial data into the data warehouse and share for integration with other biomedical data sources.

8.1.2.3 ObTiMA Trial Biomaterial Manager

The trial biomaterial manager is developed as a component of the web based trial management system ObTiMA to enable management of biomaterial data in clinical trials and sharing selected biomaterial data.

The trial biomaterial manager provides users an interface to manage biomaterial data in clinical trials according to the standard biomaterial data set. For this purpose predefined biomaterial CRFs are provided that can be adjusted to the user needs. The data items of the CRFs correspond to the standard biomaterial data set and they are 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 with existing biomaterial data. A precondition for the import is that the excel files comply with the standard biobank data set.

A search interface is provided to get an overview about the available biomaterial. Furthermore, the search interface allows to link clinical data and biomaterial data within clinical trials and across trials for further analysis.

The trial biomaterial manager realizes the functionality of the p-Biomaterial Wrappers for ObTIMA. A user can select the biomaterial data that he wants to share, and upload the data to p-BioSPRE. For this purpose upload services are integrated, which take care that during the upload process the data is anonymized. ObTiMA enables a biobank owner to manage

his biomaterial orders. When a user in p-BioSPRE requests to order biomaterial that was uploaded from an ObTiMA instance, the biomaterial owner can view that request in his instance of ObTiMA and contact the researcher. Similar as the p-Biobank Wrapper ObTiMA integrates push services that are able to push selected data into the data warehouse.

8.2 Envisaged legal support for sharing biomaterial and related data within p-medicine infrastructure

The biobank access framework of p-medicine shall not build a new biobank but facilitate the access to already existing biobanks for researchers. For this reason a metabiobank will be set up, in which data as to existing samples will be made available by biobank operators to researchers. Registered researches will have the possibility to search this metabiobank for suitable human tissue samples and request access to these data and related samples. As for now we presume that the data regarding the existing samples at the participating biobanks are anonymous, meaning that they cannot be linked directly or indirectly to a specific donor with reasonable effort in time and/or money.[232]

LUH will contribute to these efforts mainly in two ways. On the one hand we will establish the legal framework in order to facilitate the cooperation between the parties involved by providing the common contractual setting for the cooperation between the parties. Furthermore LUH serves as a helpdesk for all data protection issues arising within p-medicine.

In the following the main contractual agreements that will have to be concluded between the participants of the p-medicine biobank platform shall be outlined.

The parties involved can be divided into three groups. These are:

1. The biobank operators holding the human tissue and related data.
2. The researchers who are searching for human biological samples and related data that are suitable for the research project planned. The result of every search will show whether suitable samples for the envisaged project could be found. This information shall serve as the basis for the contact with relevant biobank operators.
3. The metabiobank operator who gathers the data on the available human samples stored in the different biobanks and makes these data available to the researchers.

In order to facilitate and regulate the interaction between the three parties mainly three contractual agreements will be required.

The first agreement regards the relation between the biobank operators and the metabiobank operator. In this agreement the conditions of participation of biobank operators wanting to share their samples and related data will be regulated. Furthermore intellectual property issues regarding the software needed to transfer the statistical data, as well as regarding the intellectual property rights concerning the database will have to be taken into account.

The second agreement regards the relation between the researchers and the metabiobank operator. In this agreement the rights and duties of the researchers accessing the metabiobank shall be regulated. Furthermore, within this agreement legal aspects of the use of the data gained from the metabiobank as well as intellectual property rights regarding the software in place will have to be taken into account.

Both agreements shall also address common issues, as e.g. the relationship between the biobank operators and the researchers. In order to facilitate the communication between these two parties it seems necessary to set up rules governing their correspondence in order

[232] For further information tot he legal term „anonymous data“ see subchapter 6.3.1.4.1.2 above.

to avoid unnecessary delays which would hinder the cooperation between the researchers and biobank operators and thus hinder the access to the existing biobanks.

There will be no obligation for the biobank operators to grant access to all researchers interested in their samples. In case the biobank operator decides to grant the researchers access to all or some human biological samples and related data held in the biobank, another agreement will have to be concluded between these two parties. As the potential projects can vary to a large extent the transfer of human biological samples and/or corresponding personal data the conditions of every specific project will have to be negotiated and agreed upon separately for each project.

Appendix 1 – i2bs Crimson Sample Ontology

C_FULLNAME	C_NAME
\Samples\	Samples
\Samples\Primary_Specimen\	Primary Samples
\Samples\Primary_Specimen\Fluids\	Fluids
\Samples\Primary_Specimen\Fluids\BAL\	BAL
\Samples\Primary_Specimen\Fluids\Bronchoalveolar Lavage\	Bronchoalveolar Lavage
\Samples\Primary_Specimen\Fluids\Blood\	Blood
\Samples\Primary_Specimen\Fluids\Whole_Blood\	Whole Blood
\Samples\Primary_Specimen\Fluids\Blood\Additive\	Additive
\Samples\Primary_Specimen\Fluids\Blood\Additive\None\	None
\Samples\Primary_Specimen\Fluids\Blood\Additive\EDTA\	EDTA
\Samples\Primary_Specimen\Fluids\Blood\Additive\Lithium_Heparin\	Lithium Heparin
\Samples\Primary_Specimen\Fluids\Blood\Additive\Sodium_Heparin\	Sodium Heparin
\Samples\Primary_Specimen\Fluids\Blood\Additive\Acid_Citrate_Dextrose\	Acid Citrate Dextrose
\Samples\Primary_Specimen\Fluids\Blood\Additive\Sodium_Citrate\	Sodium Citrate
\Samples\Primary_Specimen\Fluids\Blood\Additive\Thrombin_pro-coagulant\	Thrombin pro-coagulant
\Samples\Primary_Specimen\Fluids\Blood\Additive\Serum_Separator_Gel\	Serum Separator Gel
\Samples\Primary_Specimen\Fluids\Blood\Additive\Plasma_Separator_Gel\	Plasma Separator Gel
\Samples\Primary_Specimen\Fluids\Blood\Source_Container\	Source Container
\Samples\Primary_Specimen\Fluids\Blood\Source_Container\Blood Culture\	Blood Culture
\Samples\Primary_Specimen\Fluids\Blood\Source_Container\Blood Culture\Aerobic Bottle\	Aerobic Bottle
\Samples\Primary_Specimen\Fluids\Blood\Source_Container\Blood Culture\Anaerobic Bottle\	Anaerobic Bottle
\Samples\Primary_Specimen\Fluids\Blood\Source_Container\Light_Green_Top_[Sodium_Heparin]\	Light Green Top [Sodium Heparin]
\Samples\Primary_Specimen\Fluids\Blood\Source_Container\Green_Top_[Lithium_Heparin]\	Green Top [Lithium Heparin]
\Samples\Primary_Specimen\Fluids\Blood\Source_Container\Lavender_Top_[K2-EDTA]\	Lavender Top [K2-EDTA]
\Samples\Primary_Specimen\Fluids\Blood\Source_Container\Lavender_Top_[K3-EDTA]\	Lavender Top [K3-EDTA]
\Samples\Primary_Specimen\Fluids\Blood\Source_Container\Pale_Yellow_Top_[Acid-Citrate-Dextrose]\	Pale Yellow Top [Acid-Citrate-Dextrose]
\Samples\Primary_Specimen\Fluids\Blood\Source_Container\Blue_Top_[Sodium_Citrate]\	Blue Top [Sodium Citrate]
\Samples\Primary_Specimen\Fluids\Blood\Source_Container\Yellow_Top_[Sodium_Polyanethol_Sulfonate]\	Yellow Top [Sodium Polyanethol Sulfonate]

\Samples\Primary_Specimen\Fluids\Bone Marrow Aspirate\	Bone Marrow Aspirate
\Samples\Primary_Specimen\Fluids\CSF\	CSF
\Samples\Primary_Specimen\Fluids\Synovial_Fluid\	Synovial Fluid
\Samples\Primary_Specimen\Fluids\Peritoneal_Fluid\	Peritoneal Fluid
\Samples\Primary_Specimen\Fluids\Saliva\	Saliva
\Samples\Primary_Specimen\Fluids\Sputum\	Sputum
\Samples\Primary_Specimen\Fluids\Stool\	Stool
\Samples\Primary_Specimen\Swab\	Swabs
\Samples\Primary_Specimen\Swab\Type\	Type
\Samples\Primary_Specimen\Swab\Type\Cotton\	Cotton
\Samples\Primary_Specimen\Swab\Type\Dacron\	Dacron
\Samples\Primary_Specimen\Swab\Type\Polyester\	Polyester
\Samples\Primary_Specimen\Swab\Source\	Source
\Samples\Primary_Specimen\Swab\Source\Buccal\	Buccal
\Samples\Primary_Specimen\Swab\Source\Cervical\	Cervical
\Samples\Primary_Specimen\Swab\Source\MRSA\	MRSA
\Samples\Primary_Specimen\Swab\Source\Nasopharyngeal\	Nasopharyngeal
\Samples\Primary_Specimen\Swab\Source\Penile\	Penile
\Samples\Primary_Specimen\Swab\Source\Rectal\	Rectal
\Samples\Primary_Specimen\Swab\Source\Throat\	Throat
\Samples\Primary_Specimen\Swab\Source\Other_Site\	Other Site
\Samples\Primary_Specimen\Fluids\Thin_Prep\	Thin Prep
\Samples\Primary_Specimen\Fluids\Urine\	Urine
\Samples\Primary_Specimen\Fluids\Urine\Random Urine	Random Urine
\Samples\Primary_Specimen\Fluids\Urine\24h_Urine\	24h Urine
\Samples\Primary_Specimen\Fluids\Urine\24h_Urine\10g_Boric_Acid\	10g Boric Acid
\Samples\Primary_Specimen\Fluids\Urine\24h_Urine\25mL_6N_HCl\	25mL 6N HCl
\Samples\Primary_Specimen\Fluids\Urine\24h_Urine\10mL_50%_Sulfuric_Acid\	10mL 50% Sulfuric Acid
\Samples\Primary_Specimen\Fluids\Urine\24h_Urine\5g_Sodium_Carbonate\	5g Sodium Carbonate
\Samples\Primary_Specimen\Fluids\Urine\24h_Urine\No_additive\	No additive
\Samples\Derivative_Specimen\	Sample Derivatives
\Samples\Derivative_Specimen\Fluids\	Fluids
\Samples\Derivative_Specimen\Fluids\Blood_Derivatives\	Blood Derivatives
\Samples\Derivative_Specimen\Fluids\Blood_Derivatives\Buffy_Coat\	Buffy Coat
\Samples\Derivative_Specimen\Fluids\Blood_Derivatives\Packed_RBC_Pellet\	Packed RBC Pellet

\Samples\Derivative_Specimen\Fluids\Blood_Derivatives\Plasma\	Plasma
\Samples\Derivative_Specimen\Fluids\Blood_Derivatives\Platelet_Preparations\	Platelet Preparations
\Samples\Derivative_Specimen\Fluids\Blood_Derivatives\Platelet_Preparations\Platelet_Rich_Plasma\	Platelet Rich Plasma
\Samples\Derivative_Specimen\Fluids\Blood_Derivatives\Platelet_Preparations\Pheresis_Platelet_Unit\	Pheresis Platelet Unit
\Samples\Derivative_Specimen\Fluids\Blood_Derivatives\Platelet_Preparations\Platelet_Concentrate_Unit\	Platelet Concentrate Unit
\Samples\Derivative_Specimen\Fluids\Blood_Derivatives\Serum\	Serum
\Samples\Derivative_Specimen\Fluids\Other_Fluid_Derivatives\	Other Fluid Derivatives
\Samples\Derivative_Specimen\Fluids\Other_Fluid_Derivatives\Cell_Pellet\	Cell Pellet
\Samples\Derivative_Specimen\Fluids\Other_Fluid_Derivatives\Supernatant\	Supernatant
\Samples\Derivative_Specimen\Nucleic_Acids\	Nucleic Acids
\Samples\Derivative_Specimen\Nucleic_Acids\DNA\	DNA
\Samples\Derivative_Specimen\Nucleic_Acids\RNA\	RNA
\Samples\Derivative_Specimen\Nucleic_Acids\Total Nucleic Acid\	Total Nucleic Acid
\Samples\Derivative_Specimen\Microbial_Isolates\	Microbial Isolates
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\	Bacteria
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Anaerobes-Gram_Positive\	Anaerobes - Gram Positive
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Anaerobes-Gram_Positive\Cocci\	Cocci
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Anaerobes-Gram_Positive\Cocci\Peptostreptococcus\	Peptostreptococcus
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Anaerobes-Gram_Positive\Cocci\Peptostreptococcus\Peptostreptococcus_micros\	Peptostreptococcus micros
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Anaerobes-Gram_Positive\Rods\	Rods
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Anaerobes-Gram_Positive\Rods\Actinomyces\	Actinomyces
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Anaerobes-Gram_Positive\Rods\Actinomyces\Actinomyces_dentalis	Actinomyces dentalis
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Anaerobes-Gram_Positive\Rods\Actinomyces\Actinomyces_israelii\	Actinomyces israelii
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Anaerobes-Gram_Positive\Rods\Actinomyces\Actinomyces_odontolyticus\	Actinomyces odontolyticus
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Anaerobes-Gram_Positive\Rods\Clostridium\	Clostridium
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Anaerobes-Gram_Positive\Rods\Clostridium\Clostridium_perfringens\	Clostridium perfringens
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Anaerobes-Gram_Positive\Rods\Clostridium\Clostridium_difficile\	Clostridium difficile
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Anaerobes-Gram_Positive\Rods\Clostridium\Clostridium_tertium\	Clostridium tertium
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Anaerobes-Gram_Positive\Rods\Clostridium\Clostridium_histolytica\	Clostridium histolytica
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Anaerobes-Gram_Positive\Rods\Proprionibacterium\	Proprionibacterium
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Anaerobes-Gram_Positive\Rods\Proprionibacterium\Propionibacterium_acnes\	Proprionibacterium acnes
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Anaerobes-Gram_Negative\	Anaerobes - Gram Negative
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Anaerobes-Gram_Negative\Cocci\	Cocci
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Anaerobes-Gram_Negative\Cocci\Veillonella\	Veillonella
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Anaerobes-Gram_Negative\Cocci\Veillonella\Veillonella_parvula\	Veillonella parvula

\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Anaerobes-Gram_Negative\Rods\	Rods
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Anaerobes-Gram_Negative\Rods\Bacteroides\	Bacteroides
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Anaerobes-Gram_Negative\Rods\Bacteroides\Bacteroides_fragilis\	Bacteroides fragilis
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Anaerobes-Gram_Negative\Rods\Bacteroides\Bacteroides_ovatus\	Bacteroides ovatus
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Anaerobes-Gram_Negative\Rods\Bacteroides\Bacteroides_pyogenes\	Bacteroides pyogenes
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Anaerobes-Gram_Negative\Rods\Bacteroides\Bacteroides_thetaiotaomicron\	Bacteroides thetaiotaomicron
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Anaerobes-Gram_Negative\Rods\Bacteroides\Bacteroides_vulgatus\	Bacteroides vulgatus
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Anaerobes-Gram_Negative\Rods\Fusobacterium\	Fusobacterium
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Anaerobes-Gram_Negative\Rods\Fusobacterium\Fusobacterium_nucletum\	Fusobacterium nucleatum
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Anaerobes-Gram_Negative\Rods\Fusobacterium\Fusobacterium_necrophorum\	Fusobacterium necrophorum
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Anaerobes-Gram_Negative\Rods\Fusobacterium\Fusobacterium_polymorphum\	Fusobacterium polymorphum
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Anaerobes-Gram_Negative\Rods\Fusobacterium\Fusobacterium_novum\	Fusobacterium novum
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Anaerobes-Gram_Negative\Rods\Porphyromonas\	Porphyromonas
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Anaerobes-Gram_Negative\Rods\Porphyromonas\Porphyromonas_asaccharolytica\	Porphyromonas asaccharolytica
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Anaerobes-Gram_Negative\Rods\Porphyromonas\Porphyromonas_gingivalis\	Porphyromonas gingivalis
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Anaerobes-Gram_Negative\Rods\Porphyromonas\Porphyromonas endodontalis\	Porphyromonas endodontalis
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Anaerobes-Gram_Negative\Rods\Prevotella\	Prevotella
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Anaerobes-Gram_Negative\Rods\Prevotella\Prevotella_dentalis\	Prevotella dentalis
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Anaerobes-Gram_Negative\Rods\Prevotella\Prevotella_oralis\	Prevotella oralis
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Microaerophilic_Species\	Microaerophilic Species
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Microaerophilic_Species\Campylobacter\	Campylobacter
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Microaerophilic_Species\Campylobacter\Campylobacter_jejuni\	Campylobacter jejuni
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Microaerophilic_Species\Campylobacter\Campylobacter_coli\	Campylobacter coli
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Microaerophilic_Species\Campylobacter\Campylobacter_fetus\	Campylobacter fetus
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Microaerophilic_Species\Helicobacter\	Helicobacter
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Microaerophilic_Species\Helicobacter\Helicobacter_pylori\	Helicobacter pylori
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Positive\	Aerotolerant - Gram Positive
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Positive\Cocci\	Cocci
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Positive\Cocci\Staphylococcus\	Staphylococcus
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Positive\Cocci\Staphylococcus\Staphylococcus_aureus\	Staphylococcus aureus
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Positive\Cocci\Staphylococcus\Staphylococcus_epidermidis\	Staphylococcus epidermidis
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Positive\Cocci\Staphylococcus\coagulase-_Staphylococci\	coagulase- Staphylococci
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Positive\Cocci\Staphylococcus\Staphylococcus saprophyticus\	Staphylococcus saprophyticus
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Positive\Cocci\Staphylococcus\Staphylococcus_lugdenensis\	Staphylococcus lugdenensis
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Positive\Cocci\Streptococcus\	Streptococcus

\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Positive\Cocci\Streptococcus\Lancefield_Grouping\	Lancefield Antigen Grouping
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Positive\Cocci\Streptococcus\Lancefield_Grouping\Group_A\	Group A
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Positive\Cocci\Streptococcus\Lancefield_Grouping\Group_B\	Group B
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Positive\Cocci\Streptococcus\Lancefield_Grouping\Group_C\	Group C
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Positive\Cocci\Streptococcus\Lancefield_Grouping\Group_D\	Group D
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Positive\Cocci\Streptococcus\Lancefield_Grouping\Group_E\	Group E
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Positive\Cocci\Streptococcus\Lancefield_Grouping\Group_F\	Group F
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Positive\Cocci\Streptococcus\Streptococcus_pyogenes\	Streptococcus pyogenes
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Positive\Cocci\Streptococcus\Streptococcus_anginosus\	Streptococcus anginosus
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Positive\Cocci\Streptococcus\Streptococcus_equi\	Streptococcus equi
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Positive\Cocci\Streptococcus\Streptococcus_pneumoniae\	Streptococcus pneumoniae
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Positive\Cocci\Streptococcus\Streptococcus_equisimilis\	Streptococcus equisimilis
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Positive\Cocci\Streptococcus\Streptococcus bovis\	Streptococcus bovis
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Positive\Cocci\Streptococcus\Streptococcus_intermedius\	Streptococcus intermedius
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Positive\Cocci\Streptococcus\Stretptococcus_milleri\	Streptococcus milleri
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Positive\Cocci\Streptococcus\Streptococcus_agalactiae\	Streptococcus agalactiae
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Positive\Cocci\Streptococcus\Streptococcus_dysgalactiae\	Streptococcus dysgalactiae
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Positive\Cocci\Enterococcus\	Enterococcus
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Positive\Cocci\Enterococcus\Enterococcus_faecium\	Enterococcus faecium
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Positive\Cocci\Enterococcus\Enterococcus_faecalis\	Enterococcus faecalis
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Positive\Cocci\Enterococcus\Enterococcus_casseliflavus\	Enterococcus casseliflavus
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Positive\Cocci\Enterococcus\Enterococcus_gallinarum\	Enterococcus gallinarum
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Positive\Cocci\Micrococcus\	Micrococcus
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Positive\Cocci\Micrococcus\Micrococcus_luteus\	Micrococcus_luteus
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Positive\Cocci\Gemella\	Gemella
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Positive\Cocci\Gemella\Gemella_morbillorum\	Gemella morbillorum
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Positive\Cocci\Gemella\Gemella_sanguinis\	Gemella sanguinis
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Positive\Rods\	Rods
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Positive\Rods\Bacillus\	Bacillus
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Positive\Rods\Bacillus\Bacillus_cereus\	Becillus cereus
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Positive\Rods\Bacillus\Bacillus_subtilis\	Bacillus subtilis
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Positive\Rods\Lactobacillus\	Lactobacillus
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Positive\Rods\Listeria\	Listeria
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Positive\Rods\Listeria\Listeria_monocytogenes\	Listeria monocytogenes
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Positive\Rods\Corynebacterium\	Corynebacterium

\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Negative\	Aerotolerant - Gram Negative
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Negative\Rods\	Rods
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Negative\Rods\Enterobacteriaceae\	Enterobacteriacaea
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Negative\Rods\Enterobacteriaceae\Escherichia\	Escherichia
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Negative\Rods\Enterobacteriaceae\Escherichia\Escherichia coli\	Escherichia coli
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Negative\Rods\Enterobacteriaceae\Salmonella\	Salmonella
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Negative\Rods\Enterobacteriaceae\Salmonella\Salmonella_typhhimurium\	Salmonella typhimurium
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Negative\Rods\Enterobacteriaceae\Salmonella\Salmonella_enteriditis\	Salmonella enteriditis
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Negative\Rods\Enterobacteriaceae\Salmonella\Salmonella_cholerasuis\	Salmonella cholerasuis
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Negative\Rods\Enterobacteriaceae\Shigella\	Shigella
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Negative\Rods\Enterobacteriaceae\Shigella\Shigella_flexneri\	Shigella flexneri
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Negative\Rods\Enterobacteriaceae\Shigella\Shigella_dysenteriae\	Shigella dysenteriae
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Negative\Rods\Enterobacteriaceae\Shigella\Shigella_sonnei\	Shigella sonnei
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Negative\Rods\Enterobacteriaceae\Proteus\	Proteus
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Negative\Rods\Enterobacteriaceae\Proteus\Proteus mirabilis\	Proteus mirabilis
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Negative\Rods\Enterobacteriaceae\Proteus\Proteus_vulgaris\	Proteus vulgaris
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Negative\Rods\Enterobacteriaceae\Citrobacter\	Citrobacter
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Negative\Rods\Enterobacteriaceae\Citrobacter\Citrobacter_freundii\	Citrobacter freundii
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Negative\Rods\Enterobacteriaceae\Citrobacter\Citrobacter_rodentium\	Citrobacter rodentium
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Negative\Rods\Enterobacteriaceae\Citrobacter\Citrobacter_koseri\	Citrobacter koseri
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Negative\Rods\Enterobacteriaceae\Citrobacter\Citrobacter_braakii\	Citrobacter braakii
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Negative\Rods\Enterobacteriaceae\Klebsiella\	Klebsiella
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Negative\Rods\Enterobacteriaceae\Klebsiella\Klebsiella_pneumoniae\	Klebsiella pneumoniae
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Negative\Rods\Enterobacteriaceae\Klebsiella\Klebsiella_oxytoca\	Klebsiella oxytoca
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Negative\Rods\Enterobacteriaceae\Morganella\	Morganella
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Negative\Rods\Enterobacteriaceae\Morganella\Morganella morganii\	Morganella morganii
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Negative\Rods\Enterobacteriaceae\Serratia\	Serratia
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Negative\Rods\Enterobacteriaceae\Serratia\Seratia_marcescens\	Serratia marcescens
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Negative\Rods\Enterobacteriaceae\Serratia\Serratia_odorifera\	Serratia odorifera
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Negative\Rods\Enterobacteriaceae\Enterobacter\	Enterobacter
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Negative\Rods\Enterobacteriaceae\Enterobacter\Enterobacter_aerogenes\	Enterobacter aerogenes
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Negative\Rods\Enterobacteriaceae\Enterobacter\Enterobacter_sakazakii\	Enterobacter sakazakii

\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Negative\Rods\Enterobacteriaceae\Enterobacter\Enterobacter_cloacae\	Enterobacter cloacae
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Negative\Rods\Enterobacteriaceae\Enterobacter\Enterobacter_agglomerans\	Enterobacter agglomerans
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Negative\Rods\Enterobacteriaceae\Yersinia\	Yersinia
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Negative\Rods\Enterobacteriaceae\Yersinia\Yersinia_enterocolitica\	Yersinia enterocolitica
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Negative\Rods\Enterobacteriaceae\Yersinia\Yersinia_pseudotuberculosis\	Yersinia pseudotuberculosis
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Negative\Rods\Pseudomonales\	Pseudomonales
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Negative\Rods\Pseudomonales\Pseudomonas\	Pseudomonas
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Negative\Rods\Pseudomonales\Pseudomonas\Pseudomonas_aeruginosa\	Pseudomonas aeruginosa
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Negative\Rods\Pseudomonales\Pseudomonas\Pseudomonas_fluorescens\	Pseudomonas fluorescens
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Negative\Rods\Pseudomonales\Pseudomonas\Pseudomonas_putida\	Pseudomonas putida
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Negative\Rods\Pseudomonales\Pseudomonas\Pseudomonas_stutzeri\	Pseudomonas stutzeri
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Negative\Rods\Pseudomonales\Pseudomonas\Pseudomonas_syringae\	Pseudomonas syringae
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Negative\Rods\Pasteurellacea\	Pasteurellacea
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Negative\Rods\Pasteurellacea\Haemophilus\	Haemophilus
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Negative\Rods\Pasteurellacea\Haemophilus\Haemophilus_influenzae\	Haemophilus influenzae
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Negative\Rods\Pasteurellacea\Haemophilus\Haemophilus_parainfluenzae\	Haemophilus parainfluenze
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Negative\Rods\Pasteurellacea\Haemophilus\Haemophilus_aphrophilus\	Haemophilus aphrophilus
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Negative\Rods\Pasteurellacea\Haemophilus\Haemophilus_hemolyticus\	Haemophilus hemolyticus
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Negative\Rods\Pasteurellacea\Haemophilus\Haemophilus_parahemolyticus\	Haemophilus parahemolyticus
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Negative\Rods\Pasteurellacea\Pasteurella\	Pasteurella
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Negative\Rods\Pasteurellacea\Pasteurella\Pasteurella_multocida\	Pasteurella multocida
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Negative\Rods\Burkholderales\	Burkholderales
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Negative\Rods\Burkholderales\Burkholderia\	Burkholderia
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Negative\Rods\Burkholderales\Burkholderia\Burkholderia_cepacia\	Burkholderia cepacia
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Negative\Rods\Burkholderales\Burkholderia\Burkholderia_cenocepacia\	Burkholderia cenocepacia
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Negative\Cocci\	Cocci
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Negative\Cocci\Acinetobacter\	Acinetobacter
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Negative\Cocci\Acinetobacter\Acinetobacter_baumanii\	Acinetobacter baumanii
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Negative\Cocci\Acinetobacter\Acinetobacter_calcoaceticus\	Acinetobacter calcoaceticus

\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Negative\Cocci\Acinetobacter\Acinetobacter_lwoffii\	Acinetobacter lwoffii
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Negative\Cocci\Neisseria\	Neisseria
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Negative\Cocci\Neisseria\Neisseia_gonorrhea\	Neisseria gonorrhea
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Negative\Cocci\Neisseria\Neisseria_meningiditis\	Neisseria meningiditis
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Negative\Cocci\Moraxella\	Moraxella
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Aerotolerant-Gram_Negative\Cocci\Moraxella\Moraxella_catarrhalis\	Moraxella catarrhalis
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Acid_Fast\	Acid Fast
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Acid_Fast\Mycobacterium\	Mycobacterium
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Acid_Fast\Mycobacterium\Mycobacterium_abscessus\	Mycobacterium abscessus
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Acid_Fast\Mycobacterium\Mycobacterium_avium\	Mycobacterium avium
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Acid_Fast\Mycobacterium\Mycobacterium_africanum\	Mycobacterium africanum
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Acid_Fast\Mycobacterium\Mycobacterium_bovis\	Mycobacterium bovis
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Acid_Fast\Mycobacterium\Mycobacterium_kansasii\	Mycobacterium kansasii
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Acid_Fast\Mycobacterium\Mycobacterium_leprae\	Mycobacterium leprae
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Acid_Fast\Mycobacterium\Mycobacterium_marinum\	Mycobacterium marinum
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Acid_Fast\Mycobacterium\Mycobacterium_tuberculosis\	Mycobacterium tuberculosis
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Acid_Fast\Mycobacterium\Mycobacterium_intracellularae\	Mycobacterium intracellularae
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Acid_Fast\Nocardia\	Nocardia
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Acid_Fast\Nocardia\Nocardia_asteroides\	Nocardia asteroides
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Acid_Fast\Nocardia\Nocardia_brasilensis\	Nocardia brasilensis
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Acid_Fast\Rhodococcus\	Rhodococcus
\Samples\Derivative_Specimen\Microbial_Isolates\Bacteria\Acid_Fast\Rhodococcus\Rhodococcus_equi\	Rhodococcus equi
\Samples\Derivative_Specimen\Microbial_Isolates\Fungi\	Fungi
\Samples\Derivative_Specimen\Microbial_Isolates\Fungi\Yeasts\	Yeasts
\Samples\Derivative_Specimen\Microbial_Isolates\Fungi\Yeasts\Candida\	Candida
\Samples\Derivative_Specimen\Microbial_Isolates\Fungi\Yeasts\Candida\Candida_albicans\	Candida albicans
\Samples\Derivative_Specimen\Microbial_Isolates\Fungi\Yeasts\Candida\Candida_dublinensis\	Candida dublinensis
\Samples\Derivative_Specimen\Microbial_Isolates\Fungi\Yeasts\Candida\Candida_glabrata\	Candida glabrata
\Samples\Derivative_Specimen\Microbial_Isolates\Fungi\Yeasts\Candida\Candida_guilliermondii	Candida guillermondii
\Samples\Derivative_Specimen\Microbial_Isolates\Fungi\Yeasts\Candida\Candida_krusei\	Candida krusei
\Samples\Derivative_Specimen\Microbial_Isolates\Fungi\Yeasts\Candida\Candida_parasilopsis\	Candida parasilopsis
\Samples\Derivative_Specimen\Microbial_Isolates\Fungi\Yeasts\Candida\Candida_tropicalis\	Candida tropicalis
\Samples\Derivative_Specimen\Microbial_Isolates\Fungi\Yeasts\Cryptococcus\	Cryptococcus
\Samples\Derivative_Specimen\Microbial_Isolates\Fungi\Yeasts\Cryptococcus\Cryptococcus_neoformans\	Cryptococcus neoformans
\Samples\Derivative_Specimen\Microbial_Isolates\Fungi\Yeasts\Saccharomyces\	Saccharomyces

\Samples\Derivative_Specimen\Microbial_Isolates\Fungi\Yeasts\Saccharomyces\Saccharomyces_boulardii\	Saccharomyces boulardii
\Samples\Derivative_Specimen\Microbial_Isolates\Fungi\Yeasts\Saccharomyces\Sacharomyces_cerevisciae\	Saccharomyces cerevisciae
\Samples\Derivative_Specimen\Microbial_Isolates\Fungi\Yeasts\Malessezia\	Malessezia
\Samples\Derivative_Specimen\Microbial_Isolates\Fungi\Yeasts\Malessezia\Malessezia_furfur\	Malessezia furfur
\Samples\Derivative_Specimen\Microbial_Isolates\Parasites\	Parasites
\Samples\Derivative_Specimen\Microbial_Isolates\Parasites\Blood-borne\	Blood borne
\Samples\Derivative_Specimen\Microbial_Isolates\Parasites\Blood-borne\Babesia\	Babesia
\Samples\Derivative_Specimen\Microbial_Isolates\Parasites\Blood-borne\Babesia\Babesia_microti\	Babesia microti
\Samples\Derivative_Specimen\Microbial_Isolates\Parasites\Blood-borne\Babesia\Babesia_divergens\	Babesia divergens
\Samples\Derivative_Specimen\Microbial_Isolates\Parasites\Blood-borne\Plasmodium\	Plasmodium
\Samples\Derivative_Specimen\Microbial_Isolates\Parasites\Blood-borne\Plasmodium\Plasmodium_falciparum\	Plasmodium falciparum
\Samples\Derivative_Specimen\Microbial_Isolates\Parasites\Blood-borne\Plasmodium\Plasmodium_malariae\	Plasmodium malariae
\Samples\Derivative_Specimen\Microbial_Isolates\Parasites\Blood-borne\Plasmodium\Plasmodium_ovalae\	Plasmodium ovale
\Samples\Derivative_Specimen\Microbial_Isolates\Parasites\Blood-borne\Plasmodium\Plasmodium_vivax\	Plasmodium vivax
\Samples\Derivative_Specimen\Microbial_Isolates\Parasites\Blood-borne\Filariases\	Filariases
\Samples\Derivative_Specimen\Microbial_Isolates\Parasites\Blood-borne\Filariases\Loa_loa\	Loa loa
\Samples\Derivative_Specimen\Microbial_Isolates\Parasites\Blood-borne\Filariases\Wuchereria_bancrofti\	Wuchereria bancrofti
\Samples\Derivative_Specimen\Microbial_Isolates\Parasites\Blood-borne\Filariases\Brugia_malay\	Brugia malay
\Samples\Derivative_Specimen\Microbial_Isolates\Parasites\Blood-borne\Filariases\Brugia_timori\	Brugia timori
\Samples\Derivative_Specimen\Microbial_Isolates\Parasites\Blood-borne\Trypanosoma\	Trypanosoma
\Samples\Derivative_Specimen\Microbial_Isolates\Parasites\Blood-borne\Trypanosoma\Trypanosoma_cruzi\	Trypanosoma cruzi
\Samples\Derivative_Specimen\Microbial_Isolates\Parasites\Blood-borne\Trypanosoma\Trypanosoma_brucei\	Trypanosoma brucei
\Samples\Derivative_Specimen\Microbial_Isolates\Parasites\Blood-borne\Trypanosoma\Trypanosoma_rhodesiense\	Trypanosoma rhodesiense
\Samples\Derivative_Specimen\Microbial_Isolates\Parasites\Blood-borne\Trypanosoma\Trypanosoma_gambiensis\	Trypanosoma gambiensis
\Samples\Derivative_Specimen\Microbial_Isolates\Parasites\Intestinal\	Intestinal
\Samples\Derivative_Specimen\Microbial_Isolates\Parasites\Intestinal\Cestodes\	Cestodes
\Samples\Derivative_Specimen\Microbial_Isolates\Parasites\Intestinal\Cestodes\Diphyllobothrium_latum\	Diphyllobothrium latum
\Samples\Derivative_Specimen\Microbial_Isolates\Parasites\Intestinal\Cestodes\Taenia_solium\	Taenia solium
\Samples\Derivative_Specimen\Microbial_Isolates\Parasites\Intestinal\Cestodes\Taenia_saginatta\	Taenia saginatta
\Samples\Derivative_Specimen\Microbial_Isolates\Parasites\Intestinal\Cestodes\Hymenolepsis_diminuta\	Hymenolepsis diminuta
\Samples\Derivative_Specimen\Microbial_Isolates\Parasites\Intestinal\Cestodes\Hymenolepsis_nana\	Hymenolepsis nana
\Samples\Derivative_Specimen\Microbial_Isolates\Parasites\Intestinal\Nematodes\	Nematodes
\Samples\Derivative_Specimen\Microbial_Isolates\Parasites\Intestinal\Nematodes\Ascaris_lumbricoides\	Ascaris_lumbricoides
\Samples\Derivative_Specimen\Microbial_Isolates\Parasites\Intestinal\Nematodes\Strongyloides_stercoralis\	Strongyloides stercoralis
\Samples\Derivative_Specimen\Microbial_Isolates\Parasites\Intestinal\Nematodes\Necator_americanum\	Necator americanum

\Samples\Derivative_Specimen\Microbial_Isolates\Parasites\Intestinal\Nematodes\Enterobius_vermicularis\	Enterobium vermicularis
\Samples\Derivative_Specimen\Microbial_Isolates\Parasites\Intestinal\Nematodes\Trichuris_trichuria\	Trichuris trichuria
\Samples\Derivative_Specimen\Microbial_Isolates\Parasites\Intestinal\Protozoans\	Protozoans
\Samples\Derivative_Specimen\Microbial_Isolates\Parasites\Intestinal\Protozoans\Entamoeba_histolytica\	Entamoeba histolytica
\Samples\Derivative_Specimen\Microbial_Isolates\Parasites\Intestinal\Protozoans\Cryptosporidium_parvum\	Cryptosporidium parvum
\Samples\Derivative_Specimen\Microbial_Isolates\Parasites\Intestinal\Protozoans\Giardia_lamblia\	Giardia lamblia
\Samples\Derivative_Specimen\Microbial_Isolates\Parasites\Intestinal\Protozoans\Toxoplasma_gondii\	Toxoplasma gondii
\Samples\Derivative_Specimen\Microbial_Isolates\Parasites\Intestinal\Trematodes\	Trematodes
\Samples\Derivative_Specimen\Microbial_Isolates\Parasites\Intestinal\Trematodes\Fasciola\	Fasciola
\Samples\Derivative_Specimen\Microbial_Isolates\Parasites\Intestinal\Trematodes\Fasciola\Fasciola_buski\	Fasciola buski
\Samples\Derivative_Specimen\Microbial_Isolates\Parasites\Intestinal\Trematodes\Fasciola\Fasciola_hepatica\	Fasciola hepatica
\Samples\Derivative_Specimen\Microbial_Isolates\Parasites\Intestinal\Trematodes\Schistosoma\	Schistosoma
\Samples\Derivative_Specimen\Microbial_Isolates\Parasites\Intestinal\Trematodes\Schistosoma\Schistosoma_japonicum\	Schistosoma japonicum
\Samples\Derivative_Specimen\Microbial_Isolates\Parasites\Intestinal\Trematodes\Schistosoma\Schistosoma_hematobium\	Schistosoma_hematobium
\Samples\Derivative_Specimen\Microbial_Isolates\Parasites\Intestinal\Trematodes\Schistosoma\Schistosoma_mansoni\	Schistosoma_mansoni
\Samples\Derivative_Specimen\Microbial_Isolates\Parasites\Intestinal\Trematodes\Schistosoma\Schistosoma_mekongi\	Schistosoma_mekongi
\Samples\Derivative_Specimen\Microbial_Isolates\Parasites\Other_Sites\	Other Sites
\Samples\Derivative_Specimen\Microbial_Isolates\Parasites\Other_Sites\Nematodes\	Nematodes
\Samples\Derivative_Specimen\Microbial_Isolates\Parasites\Other_Sites\Nematodes\Trichinella_spiralis\	Trchinella spiralis
\Samples\Derivative_Specimen\Microbial_Isolates\Viruses\	Viruses
\Samples\Derivative_Specimen\Microbial_Isolates\Viruses\dsDNA_Viruses\	dsDNA Viruses
\Samples\Derivative_Specimen\Microbial_Isolates\Viruses\dsDNA_Viruses\Adenoviruses\	Adenovirus
\Samples\Derivative_Specimen\Microbial_Isolates\Viruses\dsDNA_Viruses\Herpesviruses\	Herpesviruses
\Samples\Derivative_Specimen\Microbial_Isolates\Viruses\dsDNA_Viruses\Herpesviruses\Cytomegalovirus\	Cytomegalovirus
\Samples\Derivative_Specimen\Microbial_Isolates\Viruses\dsDNA_Viruses\Herpesviruses\Epstein-Barr_Virus\	Epstein-Barr Virus
\Samples\Derivative_Specimen\Microbial_Isolates\Viruses\dsDNA_Viruses\Herpesviruses\HSV-1\	HSV 1
\Samples\Derivative_Specimen\Microbial_Isolates\Viruses\dsDNA_Viruses\Herpesviruses\HSV-2\	HSV 2
\Samples\Derivative_Specimen\Microbial_Isolates\Viruses\dsDNA_Viruses\Papillomaviruses\	Papillomavirus
\Samples\Derivative_Specimen\Microbial_Isolates\Viruses\ssDNA_Viruses\	ssDNA Viruses
\Samples\Derivative_Specimen\Microbial_Isolates\Viruses\ssDNA_Viruses\Parvovirus_B19\	Parvovirus B19
\Samples\Derivative_Specimen\Microbial_Isolates\Viruses\dsRNA_Viruses\	dsRNA Viruses
\Samples\Derivative_Specimen\Microbial_Isolates\Viruses\(+)ssRNA_Viruses\Reovirus\	Reovirus
\Samples\Derivative_Specimen\Microbial_Isolates\Viruses\(+)ssRNA_Viruses\Rotavirus\	Rotavirus
\Samples\Derivative_Specimen\Microbial_Isolates\Viruses\(+)ssRNA_Viruses\	(+)ssRNA Viruses
\Samples\Derivative_Specimen\Microbial_Isolates\Viruses\(+)ssRNA_Viruses\Picornaviruses\	Picornaviruses

\Samples\Derivative_Specimen\Microbial_Isolates\Viruses\(+)ssRNA_Viruses\Picornaviruses\Enteroviruses\	Enteroviruses
\Samples\Derivative_Specimen\Microbial_Isolates\Viruses\(+)ssRNA_Viruses\Picornaviruses\Hepatitis_A\	Hepatitis A
\Samples\Derivative_Specimen\Microbial_Isolates\Viruses\(+)ssRNA_Viruses\Picornaviruses\Hepatitis_C\	Hepatitis C
\Samples\Derivative_Specimen\Microbial_Isolates\Viruses\(+)ssRNA_Viruses\Picornaviruses\Rhinoviruses\	Rhinoviruses
\Samples\Derivative_Specimen\Microbial_Isolates\Viruses\(+)ssRNA_Viruses\Picornaviruses\Rubella_Virus\	Rubella Virus
\Samples\Derivative_Specimen\Microbial_Isolates\Viruses\(-)ssRNA_Viruses\	(-)ssRNA Viruses
\Samples\Derivative_Specimen\Microbial_Isolates\Viruses\(-)ssRNA_Viruses\Influenza_Virus\	Influenza Virus
\Samples\Derivative_Specimen\Microbial_Isolates\Viruses\(-)ssRNA_Viruses\Influenza_Virus\Influenza_A\	Influenza A
\Samples\Derivative_Specimen\Microbial_Isolates\Viruses\(-)ssRNA_Viruses\Influenza_Virus\Influenza_B\	Influenza B
\Samples\Derivative_Specimen\Microbial_Isolates\Viruses\ssRNA-RT_Viruses\	ssRNA-RT Viruses
\Samples\Derivative_Specimen\Microbial_Isolates\Viruses\dsRNA-RT_Viruses\	dsRNA-RT Viruses
\Samples\Derivative_Specimen\Microbial_Isolates\Viruses\dsRNA-RT_Viruses\HIV\	HIV
\Samples\Derivative_Specimen\Microbial_Isolates\Viruses\dsRNA-RT_Viruses\HIV\HIV_I\	HIV I
\Samples\Derivative_Specimen\Microbial_Isolates\Viruses\dsRNA-RT_Viruses\HIV\HIV_II\	HIV II
\Samples\Derivative_Specimen\Microbial_Isolates\Viruses\dsDNA-RT_Viruses\	dsDNA-RT Viruses
\Samples\Derivative_Specimen\Microbial_Isolates\Viruses\dsDNA-RT_Viruses\Hepatitis_B_Virus\	Hepatitis B Virus
\Samples\Derivative_Specimen\Solid_Tissues\	Solid Tissues
\Samples\Derivative_Specimen\Solid_Tissues\Paraffin_Block\	Paraffin Block
\Samples\Derivative_Specimen\Solid_Tissues\Paraffin_Section\	Paraffin Section
\Samples\Derivative_Specimen\Solid_Tissues\Frozen_Block\	Frozen Block
\Samples\Derivative_Specimen\Solid_Tissues\Frozen_Section\	Frozen Section
\Samples\Derivative_Specimen\Solid_Tissues\Epon_Block\	Epon Block
\Samples\Derivative_Specimen\Solid_Tissues\Epon_Section\	Epon Section
\Samples\Primary_Specimen\Solid Tissues\	Solid Tissues
\Samples\Primary_Specimen\Solid Tissues\Source\	Source
\Samples\Primary_Specimen\Solid Tissues\Source\Genitourinary System\	Genitourinary System
\Samples\Primary_Specimen\Solid Tissues\Source\Genitourinary System\Kidney\	Kidney
\Samples\Primary_Specimen\Solid Tissues\Source\Genitourinary System\Ureter\	Ureter
\Samples\Primary_Specimen\Solid Tissues\Source\Genitourinary System\Bladder\	Bladder
\Samples\Primary_Specimen\Solid Tissues\Source\Genitourinary System\Urethra\	Urethra
\Samples\Primary_Specimen\Solid Tissues\Source\Circulatory System\	Circulatory System
\Samples\Primary_Specimen\Solid Tissues\Source\Respiratory System\	Respiratory System
\Samples\Primary_Specimen\Solid Tissues\Source\Lymphoid System\	Lymphoid System
\Samples\Primary_Specimen\Solid Tissues\Source\Lymphoid System\Lymph_Node\	Lymph Node
\Samples\Primary_Specimen\Solid Tissues\Source\Lymphoid System\Spleen\	Spleen

\Samples\Primary_Specimen\Solid Tissues\Source\GI System\	Gastrointestinal System
\Samples\Primary_Specimen\Solid Tissues\Source\Endocrine System\	Endocrine System
\Samples\Primary_Specimen\Solid Tissues\Source\Endocrine System\Adrenal\	Adrenal
\Samples\Primary_Specimen\Solid Tissues\Source\Endocrine System\Thyroid\	Thyroid
\Samples\Primary_Specimen\Solid Tissues\Source\Endocrine System\Parathyroid\	Parathyroid
\Samples\Primary_Specimen\Solid Tissues\Source\Endocrine System\Pituitary\	Pituitary
\Samples\Primary_Specimen\Solid Tissues\Source\Endocrine System\Hypothalamus\	Hypothalamus
\Samples\Primary_Specimen\Solid Tissues\Source\Nervous System\	Nervous System
\Samples\Primary_Specimen\Solid Tissues\Source\Reproductive System\	Reproductive System
\Samples\Primary_Specimen\Solid Tissues\Source\Muskuloskeletal System\	Muskuloskeletal System
\Samples\Primary_Specimen\Solid Tissues\Source\Skin and Connective Tissues\	Skin and Connective Tissues
\Samples\Primary_Specimen\Solid Tissues\Collection Method\	Collection Method
\Samples\Primary_Specimen\Solid Tissues\Collection Method\Biopsy\	Biopsy
\Samples\Primary_Specimen\Solid Tissues\Collection Method\Autopsy\	Autopsy
\Samples\Primary_Specimen\Solid Tissues\Collection Method\Fine Needle Aspirate\	Fine Needle Aspiration
\Samples\Primary_Specimen\Solid Tissues\Collection Method\Surgical Excision\	Surgical Excision
\Samples\Primary_Specimen\Solid Tissues\Collection Method\Wedge Resection\	Wedge Resection
\Samples\Primary_Specimen\Solid Tissues\Fixative\	Fixative
\Samples\Primary_Specimen\Solid Tissues\Fixative\Formalin\	Formalin
\Samples\Primary_Specimen\Solid Tissues\Fixative\Bouins\	Bouins
\Samples\Primary_Specimen\Solid Tissues\Fixative\Frozen-OCT\	Frozen-OCT
\Samples\Primary_Specimen\Solid Tissues\Fixative\Frozen\	Frozen
\Samples\Primary_Specimen\Solid Tissues\Fixative\Fresh/Unfixed\	Fresh/Unfixed