# Deliverable D5.2

<table>
<thead>
<tr>
<th>Project Title:</th>
<th>Building data bridges between biological and medical infrastructures in Europe</th>
</tr>
</thead>
<tbody>
<tr>
<td>Project Acronym:</td>
<td>BioMedBridges</td>
</tr>
<tr>
<td>Grant agreement no.:</td>
<td>284209</td>
</tr>
<tr>
<td>Research Infrastructures, FP7 Capacities Specific Programme; [INFRA-2011-2.3.2.] “Implementation of common solutions for a cluster of ESFRI infrastructures in the field of &quot;Life sciences&quot;</td>
<td></td>
</tr>
<tr>
<td>Deliverable title:</td>
<td>Tool for assessment of regulatory and ethical requirements; including supportive documents</td>
</tr>
<tr>
<td>WP No.</td>
<td>5</td>
</tr>
<tr>
<td>Lead Beneficiary:</td>
<td>10: Technologie- und Methodenplattform für die vernetzte medizinische Forschung e.V. (TMF)</td>
</tr>
<tr>
<td>WP Title</td>
<td>Secure access</td>
</tr>
<tr>
<td>Contractual delivery date:</td>
<td>31 December 2013</td>
</tr>
<tr>
<td>Actual delivery date:</td>
<td>23 December 2013</td>
</tr>
<tr>
<td>WP leader:</td>
<td>Christian Ohmann, Klaus Kuhn 5: UDUS, 7:TUM-MED</td>
</tr>
</tbody>
</table>
# List of Contributors

<table>
<thead>
<tr>
<th>Field of Activity</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chair WP5</td>
<td>WP5 Christian Ohmann (UDUS)</td>
</tr>
<tr>
<td>Co-Chair</td>
<td>WP5 Klaus Kuhn (TUM-MED)</td>
</tr>
<tr>
<td>Editorial team</td>
<td>Krister Helbing (TMF), Murat Sariyar (TMF), Irene Schlünder (TMF)</td>
</tr>
<tr>
<td>General approach and methodology</td>
<td>Krister Helbing(TMF), Töresin Karakoyun(UDUS)</td>
</tr>
<tr>
<td>Task Forces</td>
<td><strong>Knowledge Base - Requirements Matrix &amp; Assessment:</strong> Krister Helbing(Lead, TMF), Sabine Brunner (TUM-MED), Töresin Karakoyun (UDUS), Stefan Klein (Erasmus MC), Murat Sariyar (TMF), Sebastian Seufert (UDUS), Astrid Woollard (STFC)</td>
</tr>
<tr>
<td></td>
<td><strong>Tool Development:</strong> Töresin Karakoyun (Lead, UDUS), Benjamin Braasch (UDUS), Martin Eckert (UDUS), Krister Helbing (TMF), Christian Krauth (UDUS), Murat Sariyar (TMF), Sebastian Seufert (UDUS), Astrid Woollard (STFC)</td>
</tr>
<tr>
<td></td>
<td><strong>Templates &amp; Contracts:</strong> Irene Schlünder (Lead, TMF), Krister Helbing(TMF), Roland Krause, Ugis Sarkans (EMBL-EBI), Astrid Woollard (STFC)</td>
</tr>
<tr>
<td></td>
<td><strong>Ontology:</strong> Wolfgang Kuchinke (Lead, UDUS), Christian Ohmann (UDUS)</td>
</tr>
<tr>
<td>Report</td>
<td><strong>Executive Summary and Introduction:</strong> Krister Helbing (TMF), Murat Sariyar (TMF)</td>
</tr>
<tr>
<td></td>
<td><strong>General Approach:</strong> Krister Helbing(TMF)</td>
</tr>
<tr>
<td></td>
<td><strong>Research process description that defines data bridges within BioMedBridges:</strong> Wolfgang Kuchinke(UDUS), Christian Ohmann (UDUS)</td>
</tr>
<tr>
<td></td>
<td><strong>Knowledge Base - Requirements Matrix &amp; Assessment:</strong> Krister Helbing (TMF), Murat Sariyar (TMF)</td>
</tr>
<tr>
<td></td>
<td><strong>Tool Development:</strong> Töresin Karakoyun (UDUS), Christian Krauth (UDUS)</td>
</tr>
<tr>
<td><strong>Templates &amp; Contracts:</strong></td>
<td>Irene Schlünder (TMF)</td>
</tr>
<tr>
<td>---------------------------</td>
<td>-----------------------</td>
</tr>
<tr>
<td><strong>Ontology:</strong></td>
<td>Wolfgang Kuchinke (UDUS)</td>
</tr>
<tr>
<td><strong>Testing and Outlook:</strong></td>
<td>Murat Sariyar (TMF)</td>
</tr>
</tbody>
</table>

| **Review and commenting** | Caroline Carbonnelle (Inserm), Philipp Gormanns (HMGU), Stefan Klein (Erasmus MC), Chris Morris (STFC), Christian Ohmann (UDUS), Michael Raess (HMGU), Ugis Sarkans (EMBL-EBI), Stephanie Suhr (EMBL-EBI) |

| **Initial workshop (Goals and concept)** | Sabine Brunner (TUM-MED), Tom Hancocks (EMBL-EBI), Krister Helbing (TMF), Töresin Karakoyun (UDUS), Stefan Klein (Erasmus MC), Roland Krause (TMF), Torsten Meiners (EU-OPENSSCREEN), Christian Ohmann (UDUS), UGIS Sarkans (EMBL-EBI), Sebastian Semler (TMF), Sebastian Seufert (UDUS), Astrid Woollard (STFC) |
Contents

1 Executive Summary .......................................................... 6
2 Introduction ............................................................................ 7
3 General Approach ..................................................................... 8
3.1 Summary of Deliverable D5.1 ............................................... 8
   3.1.1 Goals and Approach ...................................................... 8
   3.1.2 Results and conclusions for legal assessment tool ............ 9
   3.1.3 Desired features of the tool ........................................... 10
3.2 Task forces .......................................................................... 10
4 Research process description that defines data bridges within BioMedBridges .................................................................. 12
   4.1 Data flow for general research with focus on database use .... 12
   4.2 Data flow consolidated with decision workflow ................. 15
5 Knowledge Base-Requirements Matrix & Assessment (Task Force 1) ................................................................. 17
   5.1 Matrix .............................................................................. 18
      5.1.1 Defining the matrix structure ....................................... 18
      5.1.2 Organising all requirements within the matrix ............. 20
   5.3 Workflow ........................................................................... 21
   5.4 Tool output ......................................................................... 22
6 Tool Development (Task Force 2) ............................................ 23
   6.1 Aim .................................................................................. 23
   6.2 Assessment Tool .................................................................. 23
      6.2.1 Main Use Cases ........................................................... 23
      6.2.2 Workflow of the tool: Perspective of the Researcher .... 25
      6.2.3 Output Screen .............................................................. 25
      6.2.4 Technology/Architecture ............................................. 26
      6.2.5 Concepts and Data Base Structure ............................... 26
7 Templates & Contracts (Task Force 3) ...................................... 29
   7.1 The BioMedBridges ‘Ethical Governance Framework’ .......... 29
   7.2 Contractual Relationships between persons involved in the exchange of biosamples and/or data ....................... 30
   7.3 Informed Consent Document ............................................. 31
   7.4 Provider Agreement .......................................................... 34
   7.5 Material Transfer Agreement (MTA) ................................. 35
8 Ontology (Task Force 4) .......................................................... 37
   8.1 Necessity of an ontology ..................................................... 37
   8.2 Legal ontology as basis for legal queries ............................. 38
9 Testing of the prototype ........................................................... 48
   9.1 Matrix Tests ...................................................................... 48
   9.2 Tool Tests .......................................................................... 49
10 Outlook ................................................................................... 50
11 Supplemental material (online) .............................................. 50
12 Delivery and schedule ............................................................ 51
13 Adjustments made ................................................................. 51
14 Background information ....................................................... 51
Tables

Table 1 Example of the requirements matrix .......................................................... 21
Table 2 Use case: See Results .................................................................................. 23
Table 3 Use case: Assessment Wizard ...................................................................... 24
Table 4 Relation between questions of the decision workflow and concepts of the ontology .......................................................... 41
Table 5 Mapping of workflow concepts to the Data Bridge Ontology ...................... 45

Figures

Figure 1 Taskforces and their interdependencies ...................................................... 12
Figure 2 Basic research model (from16) .................................................................. 13
Figure 3 Data flow in BioMedBridges Data Bridges (DB=database) ....................... 15
Figure 4 Decision flow (Evaluation of data sources). The arrows show how the researcher is led through the different questions of the decision workflow module to assess the necessary degree of data protection ........................................ 16
Figure 5 Consolidated form of the data flow in BioMedBridges Data Bridges (DB=database) ................................................................................................. 17
Figure 6 All dimensions in relation to a requirement including the dependencies of the data type categories .................................................................................. 20
Figure 7 Wizard (Questionnaire) ........................................................................... 25
Figure 8 High-Level Architecture Assessment Tool .................................................. 26
Figure 9 BioMedBridges assessment tool Enhanced Entity Relationship Model ....... 28
Figure 10 Contractual Relations ........................................................................... 31
Figure 11 Basic concepts of the Data Bridge Ontology ............................................ 41
Figure 12 Complete Data Bridge Ontology .............................................................. 47
Figure 13 Screenshot of the requirement matrix with filters ..................................... 49
1 Executive Summary

This report describes and supplements the online ‘Tool for the assessment of ethical and legal requirements’ developed as deliverable 5.2 of the BioMedBridges project. The tool is available at http://hhu3.at.xencon.de/.

BioMedBridges is constructing the e-infrastructure to allow interoperability between data and services in the biological, medical, translational and clinical domains. For this purpose, BioMedBridges provides data bridges between biomedical sciences research infrastructures. By enabling interoperability and cooperation between infrastructures, data protection and data security challenges are lifted to a new and higher level. To address this problem, work package 5 ‘Secure Access’ is developing a security framework which will ensure that services provided by BioMedBridges are compliant with national and European data protection regulations, privacy rules and access policies.

The first output of the work package was a ‘Report on regulations, privacy and security requirements’ (D5.1)\(^1\). The second deliverable, described in this report, provides an online tool for the assessment of ethical and legal requirements (D5.2). The tool assesses the regulatory requirements established in D5.1 for specific use cases and presents relevant information and contractual templates in a user-friendly way.

In order to achieve the goals of D5.2, the following task forces were established: requirements matrix & assessment, tool development, contracts & template and ontology. The first task force defined a matrix as the knowledge base for the assessment tool and a workflow for querying the tool. The second task force developed templates and contracts for informed consent, provider agreements and material transfer agreements. In task force 3, the assessment was implemented via a java-based, open-source framework. The requirement matrix and the tool were tested extensively in a first prototype of the online tool. Task force 4 identified and assessed existing ontologies for all relevant concepts in D5.2 and created a tool-specific ontology based on this work.

\(^1\) BioMedBridges Deliverable 5.1: ‘Report on regulations, privacy and security requirements’
For the near future, three extensions of the tool should be considered: a more comprehensive and detailed ontology in order to be able to cover more requirements than implemented at present, a graphics wizard for entering new requirements into the tool, and the option to enable a data provider to specify more than two data sources.

2 Introduction

The general aim of BioMedBridges is to develop harmonised solutions for ESFRI research infrastructures in the field of life sciences. These harmonised solutions consist of registries, standards and interoperability tools, covering data security and restricted access for sensitive data. On this basis, BioMedBridges will implement the technical infrastructure to allow interoperability between data and services in the biological, medical, translational and clinical domains, especially with respect to so-called data bridges between research infrastructures. These data bridges connect very heterogeneous data sources (e.g. sequencing technologies, biosamples data, clinical data, spectroscopy and imaging and even synchrotron data for structure determination) in order to address research questions on a broader data basis.

It is the general aim of WP5 to support legal interoperability of data sharing and to develop a security framework which will ensure that services provided by the project are compliant with local, national and European regulations and privacy rules. The specific aim of this deliverable has been the development of a tool for the assessment of ethical and general legal requirements based on an analysis of the complex issues resulting from international data and biomaterial sharing between different research infrastructures. In addition, contractual templates and generic texts have been developed to support legally sound cooperation for data exchange.

It is important to note that, while WP5 spends the majority of its effort and resources on addressing challenges connected to the data bridges constructed within the BioMedBridges project, the first two deliverables concerning the

__________________________

2 www.biomedbridges.eu
ethical and legal requirements also address issues connected to sharing of biosamples. As such, deliverable 5.1 and 5.2 go beyond the requirements of the project itself and support legally sound collaboration within the wider scientific community.

The tool is based on the results of the previous Working Tasks of WP5: Regulations and privacy requirements for using the data bridges (WT1), rules and regulations for accessing databases of e-Infrastructures (WT2), regulations and security issues regarding security of biosamples (WT3.1), regulations and security issues regarding animal protection (WT3.2) and rules and regulations regarding data connected to intellectual property and licences in e-Infrastructures (WT3.3). The results of these WTs, which can be found in the D5.1 report\(^3\), have been summarized in requirement clusters, each related to the special topic of each WT (data protection, data security, security of biosamples, animal protection and intellectual property). The challenge of this deliverable is the development of a practical, online, interactive tool for researchers to assess ethical, intellectual property, data protection and security requirements based on the planned data bridges.

\section{General Approach}

\subsection{Summary of Deliverable D5.1}

\subsection{Goals and Approach}

With respect to data sharing, a number of the research infrastructures involved in the project already have some form of data curation and data protection in place or are in the process of developing this.

However, by enabling interoperability and cooperation between infrastructures, the project highlights new and more complex data protection and data security challenges. For example, openly accessible biomolecular and structural data may become linked to data that is and has to remain protected, such as patient data. The combination of different sources of open data may, in some cases, ...
result in an increased risk of identification. Thus, data protection and data security are of essential relevance for this project. The basis for the security framework developed to address this is provided by a systematic in-depth analysis of legal and ethical rules for sharing data and information between infrastructures on a European level. The resulting report from D5.1 compiles the necessary requirements to ensure legal interoperability with respect to data protection, privacy and security of the envisioned data bridges.

During the development of the deliverable 5.1 report, several concepts from computer science were employed. First, legal interoperability is defined as a specification of the general interoperability concept, which describes the ability of diverse systems to work together and to exchange information. Second, data bridges are interpreted as interfaces, as common boundaries where a direct contact between two different entities or systems occurs and where information is exchanged. Third, concepts from requirements engineering are used, such as the Domain scenario, usage scenarios and requirement clusters, to derive the necessary requirements for the data bridges.

### 3.1.2 Results and conclusions for legal assessment tool

The main results for D5.2 are the requirements clusters of the following domains: Data Protection/Privacy, Data Security, Intellectual Property and Licenses and Security of Biosamples. These requirements were described and structured in tables (an example can be found in Appendix 1). The requirement clusters will be the basis for the development of the security framework of BioMedBridges, but they were also used to develop legally compliant use cases and incorporated into the legal assessment tool of BioMedBridges.

In summary, data sharing and interoperability between the research infrastructures is restricted considerably on the legal level by the lack of a common data protection framework and legal standards, a variety of access rules, different ethical and privacy requirements, and different intellectual property/licensing rules and barriers to maintain data security. The requirements clusters identify interfaces to allow legal interoperability by defining conditions for the combination of open access, restricted access, pseudonymisation and anonymisation of data, and decisions of Ethics Committees (e.g. based on informed consent). They are the first step to achieve legal interoperability between research infrastructures. However, to
automate the assignments of applicable rules and policies, a common legal ontology that is connected to domain specific ontologies was implemented as a component of the interoperability interface. For this task, common legal ontologies were investigated with respect to their ability to support the development of the security framework and the legal assessment tool.

3.1.3 Desired features of the tool

The tool for the assessment of ethical and legal requirements should provide the following features:

— **Cover many use cases**: The tool is designed for researchers intending to exchange research data across borders, be it an internal user from within the BioMedBridges community or an external user. Consequently, the requirements of the tool and the templates are not limited to the use within the BioMedBridges project.

— **Provide concrete benefits to the user**: The user of the tool can assess the feasibility of a research project regarding the ethical and data protection requirements based on the intended use of the data bridge. The tool presents the requirements established in D5.1 in a user-friendly way so the user easily gains an understanding of the specific regulatory requirements applicable to the planned research project. The tool raises awareness of ethical and data protection requirements and suggests solutions to fulfil applicable requirements.

— **The content must be adjustable**: On the one hand, the requirement cluster in D5.1 focuses on the EU level while neglecting the national level. Thus the tool has to offer the opportunity to be extended in order to reflect national jurisdiction. On the other hand, the tool should allow permanent updating of rules and regulations. Therefore the content of the tool has to be generic. These requirements are fulfilled through a design that allows constant enrichment and updating of the content without altering the implementation of the tool.

3.2 Task forces

In order to achieve the abovementioned goals, four task forces were established (see also Figure 1):
The task force requirements matrix & assessment (TF1) constructed a matrix as a knowledge-based structure for the online interactive legal & ethical guidance and assessment tool. This was done by determining the columns of the matrix and by incorporating all requirements (identified in D5.1) within the matrix. This TF also defined a workflow for user queries and determined the tool output based on the selected requirements. The results from the work of this were used as basis for the development of the tool.

The task force tool development (TF2) was responsible for the development of an online interactive tool for legal & ethical guidance and assessment. Its main tasks were the implementation of a database based on the matrix and the development of the tool with a user interface based on workflow descriptions.

The task force contracts & template (TF3) identifies and generates the templates to support the user to fulfil the legal & ethical requirements. For this task, the contractual relations between the user, the patient, the data provider and the data bridge had to be defined. Template categories based on the requirements were defined and corresponding templates generated. These templates or parts of the templates are linked to the requirements.

The task force ontology (TF4) identified and assessed existing ontologies in the field and generated a tool-specific ontology based on available ontologies. This ontology supports the definition of terms used within and by the tool and the description of the relation between questions of the decision workflow and concepts of the ontology.
4 Research process description that defines data bridges within BioMedBridges

4.1 Data flow for general research with focus on database use

Because the data bridges will be used for all kinds of research, determined by the individual research question, the data bride has to be integrated to the general research process on the conceptual level. In general, research is a systematic inquiry to describe, explain, predict and control observed phenomena. In research, inductive methods are used to analyze the observed phenomenon and identify the general principles and processes underlying it; deductive methods verify the hypothesized principles through observations. In the conventional approach, the research process covers the creation of a research study. The process for a research study can be divided into the following stages: (1) identification of a research problem (research questions that are significant and feasible to study), (2) preparation of a research proposal (including literature review, research design, research method like experiment, data analysis and protection of human subject), (3) conduction of a pilot study (application of data-collection method and other procedures), (4)

conduction of the main study, and (5) preparation of a report. Such research can be done for two kinds of inquiry representing methodological approaches: qualitative and quantitative Research\textsuperscript{6}. For quantitative research the aim is a complete description, whereas for qualitative research the aim is to classify features, count them, and construct statistical models to be able to explain the observations. Both of these methods use data: quantitative research uses experiments and tools, such as questionnaires or laboratory equipment, to collect numerical data, while on the other hand the qualitative researcher analyses data from interviews, pictures, or artifacts. But research is not only the collection of data; research is the process that turns data into information that addresses the research question. It can be said that the formulation of an effective research question is the most important step in the research process.

The workflow of research is based on a general model for research\textsuperscript{7}. In general, the researcher begins with a question, collects data, synthesizes it (put the different data together in some coherent form), and analyse it in light of the question. This data analysis turns data into information and the research comes up with conclusions (Figure 2).

\textit{Figure 2 Basic research model (from\textsuperscript{16})}

\begin{itemize}
\item \textsuperscript{6} J. Neill (2007): Qualitative versus Quantitative Research: Key Points in a Classic Debate. 28 Feb 2007
\item \textsuperscript{7} W. Badke: Research Strategies: Finding Your Way through the Information Fog. iUniverse, Inc, 4th edition (2011)
\end{itemize}
Not shown in the figure, there is also a flow back from conclusions to new questions. Based on this basic research model, which describes the core of database-focused research, the BioMedBridges research process for data bridges was developed (Figure 3). Research steps common to all use cases were identified and set in relation.

The researcher needs some knowledge to generate an idea and to formulate a question. Based on this prior knowledge and the data the researcher already has ("Own data"), the researcher generates a hypothesis and formulates a research question to obtain meaningful data ("Formulation of research question"). The formulation of the research questions leads to the need for data and possibly for building a Data Bridge between "Own data" and a new database ("Identification of new DB", "Access to new data"). The use case description of D5.1 shows that data bridges do not only relate to the exchange or transfer of data to the researcher, but also to the linking and analysis of data. In every bridge, a new data set is accessed or even generated and this data set is integrated and harmonised with the existing data (Own data) to be analysed in view of the research question. Where not enough data was used to answer the research question, a new data gathering round is initiated, beginning with a "Search for new data". In case sufficient data to answer the research question was obtained, the analysis leads to a description of results (conclusions).

A data flow that includes different forms of access and processing of data (extraction, analysis, linking) is given in Figure 3. The step to ensure legal interoperability comes before access to new data and processing (e.g. linking) of data. The researcher always operates within a “knowledge domain”, i.e. the valid knowledge used to refer to a research area, to generate a hypothesis or idea to formulate the research question that triggers the search for new data. Access to new data may include linking, processing or extraction of data. The linking process may cover access to additional databases such as databases with genomics data (new DB). The analysis of the use cases in Del. 5.1 showed that data is often distributed over several databases and has to be integrated. The extraction of data leads to a new data set that is incorporated with the own data of the researcher.
4.2 Data flow consolidated with decision workflow

The answers to several questions in the tool specify the data source and are used for the evaluation of the degree of data protection. Here, an internal “own” data source, an external data source and what we refer to as a new data source are considered. The latter accounts for new datasets that originate from linking data from two different existing data sources. For all three data sources, specifications exists that the Evaluation module (blue area) (Figure 4) uses to determine the data protection risk and the necessary data protection measures. Sometimes, this will lead automatically to the conclusion that the data access can be permitted; sometimes to the conclusion that data access is impermissible; and sometimes to the conclusion that an application must be made to an Ethical Oversight Committee. The specifications consist of the following components:

- Purpose of data usage
- Characterization of data source: internal, external, etc.
- Human / Non-human data

Figure 3 Data flow in BioMedBridges Data Bridges (DB=database)
- Type of data or biomaterial
- Type of data protection relevant for data source
- Type of access to data source: open or restricted
- Policy for data source
- Type of consent necessary
- Type of commercial value of data (e.g. intellectual property issues, pharmaceutical industry users)

These different topics concerning the characterization of the data sources can be depicted as a decision flow (sequence of questions of the decision workflow) (Figure 4) with the aim to evaluate the degree of data protection necessary for the intended purpose (e.g. access, extraction, linkage).

*Figure 4 Decision flow (Evaluation of data sources). The arrows show how the researcher is led through the different questions of the decision workflow module to assess the necessary degree of data protection*

The decision flow (Appendix 3) can be integrated with the data flow of the data bridges, resulting in a joint diagram (Figure 5).
5 Knowledge Base - Requirements Matrix & Assessment (Task Force 1)

The aim of this task force was the definition of a matrix as a knowledge base for the online interactive legal and ethical guidance and assessment tool. In order to achieve this goal, the following three main tasks had to be performed:

1. Identifying the dimensions of the matrix and organising all requirements (identified in deliverable 5.1) within the matrix (see section 5.1)

2. Workflow definition for assessment and guidance based on the matrix and requirements (see section 5.2)

3. Determining the Tool output (see section 5.3).
5.1 Matrix

5.1.1 Defining the matrix structure

The matrix should be able to structure all the requirements of the different requirement clusters identified in deliverable 5.1. First, the requirements tables of Data Protection, Intellectual Property, Data Security and Security of Biosamples were analyzed regarding relevant and redundant/irrelevant information contained in them. Based on this analysis, the dimensions of the matrix were determined:

— The column rqNR describes the requirement number. The first two digits of the number represent the requirement cluster (DP = Data Protection, DS = Data Security, IP = Intellectual Property, BS = Security of biosamples). In order to describe additional information besides these requirement clusters, the requirement cluster ‘additional information’ (AI) was introduced. After the two characters follows an incremental number. For example, DP1 is the first requirement related to data protection.

— The column subject describes the data subject on which the data was collected. The dimension has two values: human and non-human (e.g. animal, plant, etc.).

— The column dataType describes the type of data (e.g. text data, image) related to the requirement. The dimension has the following values: metadata = overview description of a dataset, or a set of biological samples, biosample = any biological sample, image = any kind of image (e.g. DICOM images), genetic data = genetic data, and biosample associated data = any kind of data that is stored with (a link to) the biosample. Besides these shapes there are the following generic values: digital = any kind of digital information, such as images, video or audio files including any data type but biosamples; textdata = all types of text data including metadata, biosample associated data and genetic data.

— The column dataPurpose describes the purpose of data usage and contains the values data sharing/transferring, data linking/merging/combining and others.
The column **dataProtectionSource** describes the kind of data protection of the data source. Its values are anonymous = this requirement has to be fulfilled for an anonymous data source, pseudonymous = this requirement has to be fulfilled for a pseudonymous data source, and identifying = this requirement has to be fulfilled for an identifying data source.

The column **dataProtectionBridge** describes the kind of data protection of the data bridge. Its values are anonymous = this requirement has to be fulfilled for an anonymous data bridge, pseudonymous = this requirement has to be fulfilled for a pseudonymous data bridge, and identifying = this requirement has to be fulfilled for identifying data bridge (this kind of data protection is only allowed for metadata containing identifying data of a researcher).

The column **originalAccess** describes access to the data source used. Its values are: open = open access with no restrictions such as authentication and restricted = access restricted by authentication/registration.

The column **bmbAccess** describes the desired type of access for the new data bridge. Its values are: open = open access with no restrictions such as authentication, restricted = access restricted by authentication/registration, and combined = different access tiers.

The column **coveringConsent** describes whether the new research project is covered by an existing consent given by the data subject. Its values are yes and no.

The column **newConsent** describes whether it is feasible to get consent from the data subject for the new data bridge. Its values are yes and no.

The column **legalApproval** describes whether there is a legal basis for using the data source without consent from the data subject. Its values are yes and no.

The column **copyright** describes whether the data are subject to copyright or intellectual property restrictions. Its values are yes and no.

The dimension **country** describes the country in which the data source is located. Its values are the respective country's acronyms (e.g. GB/UK, DE, NL, etc).
The dimensions described, their relation to a requirement and the hierarchy of the data type categories are illustrated in Figure 6.

![Figure 6 All dimensions in relation to a requirement including the dependencies of the data type categories](image)

**5.1.2 Organising all requirements within the matrix**

The matrix is represented in a two-dimensional table. Each column represents one dimension of the matrix (e.g. country); each row represents one possible requirement description with respect to all matrix dimensions. There can be more than one row per requirement, if one requirement can assume more than one value of a requirement dimension (e.g. DP1.1 for dataProtectionSource=identifying and pseudonym) (see Figure 7).

The following rules were defined for the matrix input:

- The matrix contains at least one entry per requirement.
- All dimensions have to be specified for a requirement. If the dimension is not relevant for the requirement (e.g. for a data protection requirement, copyright is not relevant), the value for this dimension has to be “any”. The value “any” means that the requirement is valid for all values of that dimension; hence, the tool has to consider all of them.
Only one value of a dimension can be specified for a requirement (e.g. dataProtectionSource = identifying). If one requirement has more than one value per dimension it has to be entered for each value.

Table 1 illustrates the organisation of different requirements in the matrix.

<table>
<thead>
<tr>
<th>id</th>
<th>rqNr</th>
<th>subject</th>
<th>dataType</th>
<th>dataProtection</th>
<th>dataPurpose</th>
<th>legalAp</th>
<th>copyright</th>
<th>dataProtection</th>
<th>covering</th>
<th>new</th>
<th>Consent</th>
<th>legalAp</th>
<th>bmb</th>
<th>Access</th>
<th>original</th>
<th>Access</th>
<th>country</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>DP1.1</td>
<td>non-human</td>
<td>metadata</td>
<td>pseudonym</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>eu</td>
</tr>
<tr>
<td>2</td>
<td>DP1.1</td>
<td>non-human</td>
<td>metadata</td>
<td>identifying</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>eu</td>
</tr>
<tr>
<td>3</td>
<td>DP1.2</td>
<td>non-human</td>
<td>metadata</td>
<td>identifying</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>eu</td>
</tr>
<tr>
<td>4</td>
<td>DP1.2</td>
<td>non-human</td>
<td>metadata</td>
<td>identifying</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>eu</td>
</tr>
<tr>
<td>5</td>
<td>DP1.3</td>
<td>non-human</td>
<td>metadata</td>
<td>pseudonym</td>
<td>any</td>
<td>no</td>
<td>any</td>
<td>pseudonym</td>
<td>no</td>
<td>yes</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>eu</td>
</tr>
<tr>
<td>6</td>
<td>DP1.3</td>
<td>non-human</td>
<td>metadata</td>
<td>identifying</td>
<td>any</td>
<td>no</td>
<td>any</td>
<td>pseudonym</td>
<td>no</td>
<td>yes</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>eu</td>
</tr>
<tr>
<td>7</td>
<td>DP1.3</td>
<td>non-human</td>
<td>metadata</td>
<td>pseudonym</td>
<td>any</td>
<td>no</td>
<td>any</td>
<td>identifying</td>
<td>no</td>
<td>yes</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>eu</td>
</tr>
<tr>
<td>8</td>
<td>DP1.3</td>
<td>non-human</td>
<td>metadata</td>
<td>identifying</td>
<td>any</td>
<td>no</td>
<td>any</td>
<td>identifying</td>
<td>no</td>
<td>yes</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>eu</td>
</tr>
<tr>
<td>9</td>
<td>DP1.3.0</td>
<td>non-human</td>
<td>metadata</td>
<td>identifying</td>
<td>any</td>
<td>no</td>
<td>any</td>
<td>identifying</td>
<td>no</td>
<td>no</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>eu</td>
</tr>
<tr>
<td>10</td>
<td>DP1.3.0</td>
<td>non-human</td>
<td>metadata</td>
<td>identifying</td>
<td>any</td>
<td>no</td>
<td>any</td>
<td>identifying</td>
<td>no</td>
<td>no</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>eu</td>
</tr>
<tr>
<td>11</td>
<td>DP1.3.0</td>
<td>non-human</td>
<td>metadata</td>
<td>pseudonym</td>
<td>any</td>
<td>no</td>
<td>any</td>
<td>identifying</td>
<td>no</td>
<td>no</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>eu</td>
</tr>
<tr>
<td>12</td>
<td>DP1.3.0</td>
<td>non-human</td>
<td>metadata</td>
<td>pseudonym</td>
<td>any</td>
<td>no</td>
<td>any</td>
<td>identifying</td>
<td>no</td>
<td>no</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>eu</td>
</tr>
<tr>
<td>13</td>
<td>DP1.3.0</td>
<td>non-human</td>
<td>metadata</td>
<td>pseudonym</td>
<td>any</td>
<td>no</td>
<td>any</td>
<td>identifying</td>
<td>no</td>
<td>no</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>eu</td>
</tr>
<tr>
<td>14</td>
<td>DP1.3.0</td>
<td>non-human</td>
<td>metadata</td>
<td>pseudonym</td>
<td>any</td>
<td>no</td>
<td>any</td>
<td>identifying</td>
<td>no</td>
<td>no</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>eu</td>
</tr>
<tr>
<td>15</td>
<td>DP2.1</td>
<td>human</td>
<td>image</td>
<td>identifying</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>eu</td>
</tr>
<tr>
<td>16</td>
<td>DP2.1</td>
<td>human</td>
<td>image</td>
<td>identifying</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>eu</td>
</tr>
<tr>
<td>17</td>
<td>DP2.2</td>
<td>human</td>
<td>image</td>
<td>pseudonym</td>
<td>data linking</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>eu</td>
</tr>
<tr>
<td>18</td>
<td>DP2.2</td>
<td>human</td>
<td>image</td>
<td>identifying</td>
<td>data linking</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>eu</td>
</tr>
</tbody>
</table>

5.3 Workflow

A workflow for user queries was pre-defined in such a manner that usability is considered and many non-feasible selections are avoided (see Appendix 3). The questions and possible selections are given by the attributes gained by the work described in section 5.1. Only one answer is possible for each question and the selection is processed internally according to the text after the selection options. For example, the selection of Metadata as the data type (question 4 in Appendix 3) leads to the assignment

\[ \text{data type} = \text{Metadata, Textdata and digital data and any}, \]

and this means that requirements having the data type ‘metadata’ or ‘text data’ or ‘digital’, or where the data type is non-specified, are affected by this selection.

There are three leaps (goto-statements), i.e. skipping of questions, in the workflow to avoid possible pitfalls and guide the user efficiently:
— After question 4: the user is guided to question 10 if non-human data is selected without metadata information, because in this case only metadata information concerning the researcher would be relevant for questions 5-9.

— After question 5: the user is guided to question 10 if the data source is anonymous, because the bridge can only be anonymous and questions 7-9 concerning pseudonymous are irrelevant.

— After question 6: the user plans an anonymous bridge, hence, questions 7-9 concerning pseudonymous data are again irrelevant.

5.4 Tool output

A user query results in a dynamically generated list of requirements (tables) and corresponding explanations (text). The requirement tables are polished versions of tables provided by deliverable 5.1. Explanations are pre-defined for all requirements and produced by referring to a mapping table that associates explanations and requirements in an extensible fashion. Requirement tables, explanations and the mapping table are provided in the supplement of this report. The content of an explanation is related to relevant attribute values, i.e. those that lead to several similar requirements and/or require special attention. For example, the dimension “data protection” has one general explanation independent of any specifications, and a number of different explanations according to the values “Identifying Metadata”, “Human Data”, “Data linking or merging”, “ Biosample”, “Genetic data”, “Text data”, “Pseudonymised Data” and “Anonymous Data”. An explanation has the following structure:

\[
\text{title} = \text{AttributeValue}(\text{parentID} = pID, \text{id} = id, \text{requirements list} = \text{yes/no})
\]

\[
\text{text} = \text{explanationtext}
\]

Mapping of requirements to explanations is accomplished by a table that assigns the explanation ID to all requirements that are associated with the value `AttributeValue`. The parent ID gives the ID of the explanation that is hierarchically one abstraction level above the explanation considered, which enables a structured listing of all explanations that are mapped to the requirements selected by the matrix. The argument “requirement list” is set to ‘yes’ if tables are directly associated with the explanation.
Important additional information such as non-feasibility of research plans can also be considered by the matrix and mapped to an explanation, e.g. if the user plans a pseudonymous bridge without a legal basis or consent, an error message appears as the outcome.

6 Tool Development (Task Force 2)

6.1 Aim

Development of an online interactive legal and ethical guidance and assessment tool to encourage researchers to take data privacy, data protection, intellectual property and security of biosamples aspects into consideration when building data bridges. The tool aids the researchers with explanations, contract templates and basic requirements.

6.2 Assessment Tool

6.2.1 Main Use Cases

Main use cases describe the interaction between a researcher and the assessment tool and which prerequisites have to be fulfilled. The use case “Assessment Wizard” describes step by step how to use the wizard in general. The use case “See Results” describes how to handle the results overview.

<table>
<thead>
<tr>
<th>Use Case ID</th>
<th>UC_AT.AssessDataBridge_01</th>
</tr>
</thead>
<tbody>
<tr>
<td>Use Case Name</td>
<td>Assess Data bridge consisting of two data sources</td>
</tr>
<tr>
<td>Summary</td>
<td>Assess Data bridge consisting of two data sources</td>
</tr>
<tr>
<td>Priority</td>
<td>Essential</td>
</tr>
<tr>
<td>Use Frequency</td>
<td>Always</td>
</tr>
<tr>
<td>Direct Actors</td>
<td>Researcher</td>
</tr>
<tr>
<td>Prereq./Trigger</td>
<td>Researcher has access to the tool</td>
</tr>
</tbody>
</table>

**Main Success Use Case Scenario**

1. Researcher accesses Assessment Tool web page
2. Researcher answers questions by using the wizard
3. Researcher defines envisaged data bridge purpose
4. Researchers answer questions related to data source 1
5. Researchers answer questions related to data source 2
6. Research checks assessment overview
### Use case: Assessment Wizard

<table>
<thead>
<tr>
<th>Use Case ID</th>
<th>UC_AT.SeeResults_01</th>
</tr>
</thead>
<tbody>
<tr>
<td>Use Case Name</td>
<td>See Results</td>
</tr>
<tr>
<td>Summary</td>
<td>Access results of assessment</td>
</tr>
<tr>
<td>Priority</td>
<td>Essential</td>
</tr>
<tr>
<td>Use Frequency</td>
<td>Always</td>
</tr>
<tr>
<td>Direct Actors</td>
<td>Researcher</td>
</tr>
<tr>
<td>Prereq./Trigger</td>
<td>Wizard has been completed</td>
</tr>
<tr>
<td>Main Success Use Case Scenario</td>
<td></td>
</tr>
<tr>
<td>1.</td>
<td>Researcher completes assessment</td>
</tr>
<tr>
<td>3.</td>
<td>For each requirements section tables are shown with links to additional content such as documents, tools or web pages</td>
</tr>
<tr>
<td>4.</td>
<td>For each section explanations are generated based on the assessment</td>
</tr>
</tbody>
</table>

### Notes and Questions

- All given answers will live as long as you’re at the same view (ViewScoped). Partly completed data entries won’t be saved.
6.2.2 Workflow of the tool: Perspective of the Researcher

1. Purpose of data bridge
   Based on research question

2. Characterisation of data source 1
   Data Source Location  Consent  Copyright  Data Protection  Type of data

3. Characterisation of data source 2
   Data Source Location  Consent  Copyright  Data Protection  Type of data

4. Assessment
   Dynamic generation of explanations  Selection of Requirements  Content generation

5. Results of Assessment
   Visualization of IP, DS, DP and BS Requirements  Contracts & Templates  Content

Figure 7 Wizard (Questionnaire)

On the basis of the questionnaire provided in Appendix 3 to this document, the assessment tool selects requirements from the different domains. Each question relates to a column in the requirements matrix (see Figure 6) and the answers to these questions lead to a selection of one or more rows (requirements) from the requirements matrix. To enforce semantic plausibility, leaps (goto-statements) were implemented in the wizard. Screenshots of the tool can be found in the user manual (see Appendix 2).

6.2.3 Output Screen

The output screen is divided into:

- Intellectual Property,
- Security of Biosamples,
- Data Security and
- Data Protection/Privacy.

Each section comprises requirements, explanations and links to content.
6.2.4 Technology/Architecture

The assessment tool is based on a typical web environment, consisting of Liferay, Apache and Tomcat. The assessment tool itself (Portlet JSR 286) was written in Java using Managed Beans and Primefaces and JSF as frontend technology. The Liferay theme was adopted to meet BioMedBridges corporate design as far as possible.

![High-Level Architecture Assessment Tool](image)

**Figure 8 High-Level Architecture Assessment Tool**

Liferay Portal is an open-source software that is used primarily in businesses as an employee and business process oriented enterprise portal. Liferay enables an employee to see and personalise information, data and applications in a unified user interface in his/her own web browser. It is developed in Java. There are three segments in Liferay Portal. The core application by the same name is the base for the other two segments. Liferay CMS is a content managing system building on Liferay Portal. Liferay Collaboration is used for web-based teamwork and social networks. Liferay Portal is based on a service-oriented architecture (SOA). This enables further Liferay or self-written components to be added via portlets or access to existing applications respectively.

Primefaces is a Java Framework based on Java Server Faces (JSF). JSF is a Java specification for building user interfaces for web applications.

6.2.5 Concepts and Data Base Structure

In Figure 9, a high-level or conceptual data model used in database design has been used to visualize the used data structures and data model of the assessment tool. All content is generated automatically based on the
assessment of the researcher. It is possible to change the knowledge basis in order to keep the knowledge actual to legislations and regulations. Moreover, additional content, explanations or new requirements can be added via the database interface.

The assessment tool comprises the following main concepts:

- Requirements matrix
- Requirements
- Explanations
- Content

In addition, the following concepts are used to enable relationships/dependencies between the concepts:

- Requirements-to-content mapping
- Requirements-to-explanation mapping
Figure 9 BioMedBridges assessment tool Enhanced Entity Relationship Model

The tables are described as followed:

— **Requirements Matrix**: The requirements matrix is basis for the assessment. It functions as the decision matrix and is based on the answers of the researcher (each column of the requirements matrix mapped to one question).

— **Explanation**: The explanations are generated on the basis of the requirements. The explanations are dynamically loaded as a tree structure from the database in order to generate text similar to a document structure, compromising of chapters, sub-chapters and so on.
— **Content**: Requirements can be linked to additional or external content such as documents, web pages, images, and so on.

— **Requirements (IP, DP, DS, BS)**: For each requirement type, different tables or concepts were implemented. All requirements sections reflect the different demands and data types needed for the specific subsections, such as intellectual property.

### 7 Templates & Contracts (Task Force 3)

In order to support the legally sound cooperation within and beyond the BioMedBridges project, contractual templates and generic texts were developed and integrated into the tool. These templates and texts reflect some major issues concerning trans-border biomedical research and exchange. The legal background, however, is far from being clear and reliable. In contrast, the legal landscape is undergoing a process of being established, changed and harmonized throughout Europe. The tool can only reflect, but not change, this unsatisfactory situation. Therefore, the templates are based on a survey of a number of European and national contractual models and forms; however, these are in no way the only possible solution but suggest certain policies and assume certain legal opinions that seem to be currently prevailing. The scope of the templates and the main legal assumptions are outlined in this chapter. All templates are provided in the supplementary material to this report.

#### 7.1 The BioMedBridges ‘Ethical Governance Framework’

The bridge-building within the BioMedBridges project is primarily governed by the “BioMedBridges Ethical Governance Framework (EGF)”. It addresses the Research Infrastructures involved in the BioMedBridges project in order to ensure that principal rules concerning animal protection and data protection are implemented and establishes a governing and monitoring system and assigns responsibilities. The templates of this tool can be used in order to supplement the EGF. They are harmonised with the content of the EGF and provide

---

8 [BioMedBridges Ethical Governance Framework Version 1.1](#), April 2013

BioMedBridges Deliverable D5.2
complementary contractual templates, such as an informed consent form and a data provider agreement.

7.2 Contractual Relationships between persons involved in the exchange of biosamples and/or data

The BioMedBridges project is characterized by a wide range of participating infrastructures and a respective range of goals and policies; this is all the more true for other initiatives that involve the exchange of biological research data and biospecimens. Research activities in the wider community outside of BioMedBridges may vary from offering biosamples from animals to those from human beings; some services include the provision of data, some do not. Some have a very restrictive access policy while others offer open source data and material. This variety of services entails a respective variety of legal constraints. Thus, there will be many samples of templates; every use case requires a specified contract model. On the other hand, the cases have a sort of hierarchy regarding legal issues that can arise. Therefore, the templates are designed in a generic fashion such that the full set of available provisions cover all legal issues that can reasonably arise. The result is a contractual model for a maximum case. The maximum case will be: a biobank providing human biosamples with data, be it metadata or other types of data, and having a restrictive access policy. Lighter versions are derived from this model: by indicating their requirements, the user selects the legal provisions that are necessary to regulate his bridge. The tool will then provide the appropriate template. In the maximum model there are three layers of legal relationships: data subject – data provider, data provider – bridge, and finally bridge – end user.

Four fields of law can be identified that have to be taken into account: data protection/privacy, intellectual property, material property rights, and a pool of remaining issues that is called “miscellaneous” and can include liability restrictions, charges or any other counterperformance provisions. These fields are indicated in Figure 10 as greyscale blocks.
Wherever human material or personal data are used, the first document that is needed is usually an "informed consent": the person whose biomaterial or personal data have been collected and stored in a biobank and/or database has the fundamental right to prohibit the storage and further use of the biomaterial and/or any existing data linked to the material. The template "Information Sheet and Consent Form" intends to help in cases where the initial consent given by the donor/data subject at the initial collection of the material/data does not cover the intended use within a new research project.

Data protection issues do not arise in cases of animal biosamples or data. Since a data subject does not exist in such a case (animals are not data subjects with data protection concerns), no data protection issue arises. Therefore, no "informed consent" in the sense of data protection legislation is necessary.
Similarly, data protection issues do not arise where data related to humans - be it data linked to or derived from human biosamples, be it other data - are anonymised in the sense that the data contain no information that could reasonably be used by anyone to identify the individuals who donated them or to whom they relate. Anonymisation is therefore still widely seen as remedy to avoid any data protection concerns, since anonymised data are no longer considered to be personal data and thus do not fall within the scope of data protection legislation. It has to be noted that anonymisation rendering personal data into non-personal data does not comprise pseudonymisation.

On the other hand, the requirement for proper de-identification in the sense that the donor/data subject is no longer identifiable by reasonable means imposes increasing requirements for technical safeguards. Especially where biospecimens or genetic data are used, constant technical developments challenge the concept of anonymisation due to the fact that sequencing is getting more and more feasible and affordable⁹. Moreover, anonymisation deprives the donor of the right to withdraw their consent and claim the remaining material. Therefore, data protection agencies are beginning to recommend that anonymisation should rather be avoided than to be used as means to protect the donor. It is therefore strongly recommended to seek consent from the donor instead of trying to address this issue by anonymisation.

Asking for further consent assumes that the donor of the biospecimen or data in question has given their prior consent to be contacted again by the institution now seeking further consent to use the biospecimen/data. This consent to be contacted is an indispensable prerequisite to use this template. If this requirement is not met, it would be a breach of privacy as such to seek consent for further use of the biospecimen/data.

Valid consent for the use of personal data, including the use of human biosamples, requires that the following conditions are met:

The consent must be freely given. This means that potential donors must be sure that the decision will not affect any present or future health care or entail

⁹ See e.g. E. Ayday, E. De Cristofaro, J.-P. Hubaux, G. Tsudik, “The Chills and Thrills of Whole Genome Sequencing” (2013)
any other disadvantages. It is generally recognised that consent can only be
given on the basis of proper information on the intended use and the rights of
the donor, e.g. to withdraw the given consent at any time (“informed consent”).
Therefore, the Information Sheet provided should be a visible part of the
consent declaration in order to underline that the participant has been properly
informed before signing the declaration.

Informed consent includes the explicit indication of a specified research
purpose. This requirement is one of the great challenges for biomedical
research and even more for biobanks. A purpose is without doubt properly
specified in cases where the use of biospecimens/data is limited to one single
study with a clearly defined study purpose outlined in a study protocol. Beyond
this, it is widely recognised that the indication of a certain disease that the
intended research addresses can be seen as a sufficient specification. The
concept of “broad consent to medical research”, which would be most helpful
for research with regard to data protection concerns in cases of the use of
human biospecimens through biobanks, is gaining ground within the EU\textsuperscript{10}. It
can, however, still not be seen as prevailing and is not in all cases a reliable
form of consent. It is therefore recommended to indicate the research purpose
as precisely as possible.

Incidental findings regarding undetected health risks or diseases of the donor
raise the question whether there is an obligation to inform the donor. On the
other hand, the “right not to know” must be respected. This conflict cannot
completely be resolved, but at least duly handled through the implementation of
an explicit declaration regarding incidental findings.

To take into account varying research purposes and remaining legal
uncertainties, in particular in the context of international cooperation of
biobanks\textsuperscript{11}, the template of the consent declaration is generic. It contains
several paragraphs from which the donor/data subject can choose to consent
to or not. Some are indispensable to any valid consent whereas others extend

\textsuperscript{10} See e.g. the consent form template “Mustertext zur Spende, Einlagerung und
Nutzung von Biomaterialien sowie zur Erhebung, Verarbeitung und Nutzung von Daten
in Biobanken” very recently issued by the „Arbeitskreis Medizinischer Ethik-
Kommissionen“, \url{http://www.ak-med-ethik-komm.de/aktuelles.html}

\textsuperscript{11} See the very good overview of B.S. Elger and A.L. Caplan, “Consent and
anonymization in research involving biobanks”, EMBO reports Vol 7, No 7, 2006, p. 661
the scope of the consent. As discussed above, broad consent and consent to anonymisation are not unquestioned and may be invalid under certain jurisdictions or in the future. However, considering the fact that the concept of broad consent for medical research is gaining ground, it is recommended that consent is collected in this sense to enable future use of the specimen/data. It is, however, important to ensure that the currently intended use is covered by the scope of the consent under present applicable law.

The biosamples and data in question have to be specified as precisely as possible. Great care should be taken to ensure that the indication is correct and complete. Otherwise, the risk that the consent does not cover the intended use is not properly addressed.

In order to be safely within the law, open-access to human data in the ideal sense of the concept should only be granted in cases where duly anonymised (not pseudonymised) data are offered. The data should only bear a minor risk of re-identification; biosamples and full or broad genetic data should be excluded, at least as long as the data subject has not unambiguously consented that their data is published. With regard to personal data, the concept of “fair access” is a choice, since it combines confidentiality of personal data with the possibility of making data available for research purposes\textsuperscript{12}.

### 7.4 Provider Agreement

The second relationship is that between the data/material provider and the bridge. This relationship is regulated by the Provider Agreement. This agreement determines the conditions under which the material/data is provided to the bridge and how it can be further used and distributed.

First, within BioMedBridges, this relation is governed by the “BioMedBridges Ethical Governance Framework\textsuperscript{13}”, which provides fundamental rules for the process of bridge building and sets out policies for the BioMedBridges project that specifically relate to ethical and regulatory issues with regard to the storage and access of data. It addresses the research infrastructures involved


\textsuperscript{13} BioMedBridges Ethical Governance Framework Version 1.1, April 2013
in the BioMedBridges project in order to ensure that principal rules concerning animal protection and data protection are implemented. It establishes a governing and monitoring system and assigns responsibilities. Appendix 1 of the Ethical Governance Framework contains a “data provider form” that helps to identify the legal constraints of the provided data.

The BioMedBridges Ethical Governance Framework does not provide contractual clauses. Taking into account that the bridge is being built to serve research interests and, therefore, prepares the governance of the relation to the end-user (i.e. the researcher who uses the data/material for their research), provisions regulating issues such as liability, intellectual and material property rights etc. are useful. Since bridge building will at least entail a disclosure of data including those that affect the rights of third parties (e.g. rights of donors/data subjects falling into the scope of data protection) to the bridge or the partner institute, it is recommended to use the provider agreement template. This helps to avoid future legal issues concerning data protection, intellectual property, patentability, liability and perhaps material property rights between the collaborating participants within BioMedBridges.

The template “provider agreement” covers the “maximum case”: the bridge-building infrastructures provide human personal data or even biosamples to the partner institute or to the bridge. In all other cases, e.g. the case that the provider provides only meta- and/or assay data in order to allow the enhanced search for the data/material at the site, the respective provisions can be left out. In simple cases it might be sufficient to use the “data provider form” in Appendix 1 of the BioMedBridges Ethical Governance Framework. This form serves the determination of basic constraints of the provided material/data that limit the use of the material/data for further research and must be observed in the relation of the Bridge to the end-user (researcher).

7.5 Material Transfer Agreement (MTA)

The last document aims at governing the relation to the researcher who intends to use the data and/or biosamples – first of the bridge and finally of the participating database and/or biobank. BioMedBridges participants are not bound to a common or uniform access policy with regard to end-
users/researchers. The various access policies depend instead on the material/data in question.

Restricted access will be granted in cases of the provision of human personal data or biosamples, since it has to be ensured that the research project is covered by the consent of the donor/data subject and by the approval of the appropriate ethics committee. The MTA does not replace the access procedure of the data base and/or biobank since it will be signed after access has been granted. Access procedures have to be established separately to ensure that the material/data are only offered to authorised researchers.

Wherever “open access” is the policy of the participating institute, the MTA might serve as “Terms and Conditions” that form the contractual basis for the usage of the service offered. Open access implies that any researcher, irrespective of their research project, is allowed to get data from the resource in question without any access procedure. It does, however, not necessarily mean that the service is provided without any contractual constraints such as observing intellectual property rights, credits, and limitation of liability. The delivery of data/material can therefore well be conditional on the acceptance of the “Terms and Conditions”.

The conditions set out in the Provider Agreement (above) have to be observed for shaping the transfer agreement with the end user since it cannot be transferred what has not been provided. Therefore, all restrictions rooted in the Provider Agreement must be included in the end user transfer agreement. Special attention should be paid to the scope of consent in case of human personal data/biosamples. In cases of doubt, the provider should be contacted in order to clarify whether new consent is needed and whether the donor/data subject can be contacted again.

The MTA can be used by the bridge in cases where the research data is directly provided by the bridge or by the Research Infrastructure holding the data/biosamples.

The MTA template is based on a survey of a number European and national database and biobank initiatives and a review of MTAs already in place. A certain standard of contractual framework has emerged that contains typical clauses, e.g. to limit liability and to choose the applicable jurisdiction. A most
useful model MTA is provided by BBMRI; this has served as guideline for the
template implemented in the online tool.

The MTA template is generic, similar to the informed consent document and
the Provider Agreement. Clauses not needed for the context of the bridge in
question can be omitted.

8 Ontology (Task Force 4)

8.1 Necessity of an ontology

Increasingly, ontologies are used to enable data integration in order to provide
the ability to manipulate data transparently across heterogeneous data
sources\textsuperscript{14}. Ontologies therefore have become relevant to a number of
applications, including medical information management and biomolecular
information systems. Semantic data integration uses a conceptual
representation of the data and of their relationships; thus, it employs at its core
the concept of ontology. Ontologies were developed by the Artificial Intelligence
community to enable knowledge sharing\textsuperscript{15} and a common use of ontologies is
data standardization and conceptualization. In the context of this tool, an
ontology is used to structure the querying process of the assessment tool.
Enabling some kind of ontology-based querying\textsuperscript{16} in the tool was necessary
because the data protection and data security concepts are difficult and their
relation complex, which prevents the use of an easy assessment step. Based
on the ontology, the decision workflow is structured with the questions
concerning the purpose of data usage and the character of the data source at
the beginning, leading to questions about the identifiability of data and the type
of consent.

\textsuperscript{14} I.F. Cruz, H. Xiao: The Role of Ontologies in Data Integration. Journal of Engineering
Intelligent Systems, 13 (4), 1-8, (2005)
\textsuperscript{15} N. Guarino: Formal Ontology and Information Systems. In: Proceedings of the 1st
International Conference on Formal Ontologies in Information Systems, FOIS 1998, 3–
15, (1998)
\textsuperscript{16} F. Di Pinto, D. Lembo, M. Lenzerini, R. Mancini et al.: Optimizing Query Rewriting in
Ontology-Based Data Access. EDBT/ICDT ‘13, March 18-22 2013, Genoa, Italy (2013)
Thus, an ontology covering personal data but also non-personal data requirements had to be incorporated into the tool. As mentioned above, an ontology defines a common vocabulary for researchers who need to share information in a domain. It includes definitions of basic concepts in the domain and relations among them\cite{Noy2014}. Using an ontology it becomes easier to share a common understanding of the structure of information, enable reuse of domain knowledge and make domain assumptions explicit.

We used an ontology as a method of organizing the concepts in the data privacy domain and connect them with concepts for data sharing in such a way that the concepts can be mapped to the various practices that occur in the data protection domain (e.g. anonymisation, access restrictions). For this purpose, we considered the data flow described in the specific cases of the proposed data bridges in order to get as output a list of concepts that data provider and data user must consider. These concepts are connected to privacy requirements that must be fulfilled by all parties involved in the process of sharing of sensitive patient data. These requirements cover the understanding of conceptual information about rights, obligations and actions, including patient consent, privacy preserving techniques such as anonymization or pseudonymization and additional rights, such as those of the data subject to withdraw consent or to be notified of incidental findings. This complexity leads naturally to the necessity for a high-level, ontology-based model as basis for the assessment tool\cite{Rahmouni2011}.

### 8.2 Legal ontology as basis for legal queries

We reviewed the existence of available legal and privacy ontologies for use in our tool. There are several legal ontologies concerning data privacy, confidentiality, and intellectual property which were evaluated for usage.

---

\textsuperscript{17} N.F. Noy, D.L. McGuinness: Ontology Development 101: A Guide to Creating Your First Ontology. \url{http://protege.stanford.edu/publications}

\textsuperscript{18} H. Boussi Rahmouni, T. Solomonides, M. Casassa Mont, et al. (2011): \textit{A Model-driven Privacy Compliance Decision Support for Medical Data Sharing in Europe}
The **NEURONA** ontology\(^\text{19}\) is an application-oriented ontology and model for the knowledge necessary for the development of data protection compliance applications that offers reports regarding the correct use of security measures for personal data. The project is financed by the Spanish Ministry of Industry, Tourism and Commerce. The project’s general goal is the development of techniques and systems to incorporate intelligence in the three main areas of corporate security: legal, organizational and technological. The construction of the ontology was focused on the acquisition of conceptual domain knowledge extracted from the legal and related documents and the interaction with the legal experts. The first level of concepts includes activity, country and consent.

**OntoPrivacy** is a modification of a glossary of keywords from the Italian personal data protection code to support search and allow retrieval and visualization\(^\text{20}\). It has been created for the creation of a glossary of keywords extracted from the Italian Personal Data Protection Code.\(^\text{21}\) This glossary has been manually created by an expert in the legal field and is made up of both specific terms from the Public Administration domain (e.g. atto amministrativo/administrative act) and generic words (e.g. parere/opinion). OntoPrivacy has been created as part of the Law Making Environment project in order to support the metaSearch (mS) software, a tool that allows querying profiles of legislative data.

**Legl.OPD** ontology modeled concepts from the Spanish Data Protection act. The LRI-Core Ontology covers the main concepts that are common to all legal domains. It has five main categories: Physical-Entity (e.g. substances and physical objects), Mental-Entity (e.g. emotions), Abstract-Entity (e.g. numbers), Role (e.g. social and communication roles), and Occurrence (e.g. events). In addition, the **BDSG** ontology formalizes the German Personal Data Protection Act.

In general, the diversity and complexity of the rules governing data protection in Europe results in the need for a modeling approach that is able to abstract from


\(^{20}\) V. Cappelli et al.: Modelization of Domain Concepts Extracted from the Italian Privacy Legislation

these complexities rules that can be used for the automation of queries. Privacy requirements are the obligations that must be fulfilled for a compliant sharing and processing of sensitive patient data. This covers consent, anonymity, and the rights of the data subject, including their right to be notified of incidental findings. Using the OWL language, it is easy to model the conceptual privacy domain of “data sharing” and its components as hierarchies of classes and of properties to represent the relationships between these classes. In this way, privacy requirements (e.g. consent) and practices (e.g. anonymisation) may be modeled as OWL classes and assigned to the “Data Sharing” resource as object properties. But an evaluation of the existing privacy ontologies showed that these approaches did not fit to the research data flow-based approach for the assessment tool. They are too specific (including PIN, patient address, etc.) or focus on the legal domain without considering research purposes. Therefore, none of the available ontologies was used. Instead, a small ontology was created based on elements of these existing ones.

The first step was the creation of a glossary of the workflow processes and the legal terms identified in the use cases of deliverable 5.1. The center of the ontology forms the data bridge (Figure 11) connecting data provider with data user. The data bridge enables data sharing. Most often these are non-human data with little data protection requirements. Where identifying or sensible data are involved in data sharing an identification risk arises and the identifying data becomes part of the personal data area (Figure 11, blue area). Part of this area was adopted from Rahmouni et al., who modeled the legal domain of the EU data protection directive. They omitted concepts of ethics committee decisions or approvals by data protection committee. Enabling the system to simulate the role of ethics committees or data protection boards will be part of the further development of their ontology. However, because of the importance for the use of human data for research and the involvement of informed consent, the concept of ethics committees had to be included in the system.

22 H. Boussi Rahmouni; T. Solomonides; M. Casassa Mont, et al. (2011): A Model-driven Privacy Compliance Decision Support for Medical Data Sharing in Europe
23 H.B. Rahmouni; T. Solomonides; M. Casassa Mont, et al. (2011): A Model-driven Privacy Compliance Decision Support for Medical Data Sharing in Europe
All concepts of the workflow were mapped to the ontology developed as part of this deliverable to guarantee comprehensive coverage of the information flow associated with the workflow (Table 2 Use case: See Results Table 2 and Table 3). The listing shows that the mapping is complete with the exceptions of data processing terms (e.g. DataAnalysis, DataExtraction), which are treated in the workflow in a generic way. Data processing terms were included because the analysis of use cases in deliverable 5.1 showed that the Data Bridges are not only tools for transferring data but often include some kind of data service like data annotation, data analysis, data extraction.

Table 4 Relation between questions of the decision workflow and concepts of the ontology

<table>
<thead>
<tr>
<th>Decision workflow</th>
<th>Ontology concepts</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Purpose for data usage</td>
<td>SharingPurpose</td>
<td></td>
</tr>
<tr>
<td>Own (internal) data source or external data source</td>
<td>DataUser, DataProvider</td>
<td></td>
</tr>
<tr>
<td>Question</td>
<td>Answer</td>
<td></td>
</tr>
<tr>
<td>-------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Where is your data source located</td>
<td>MemberState</td>
<td></td>
</tr>
<tr>
<td>What is the original species of the data</td>
<td>Human, Animal</td>
<td></td>
</tr>
<tr>
<td>What kind of data or biomaterial will be used (data classification)</td>
<td>Metadata</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Text data</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Image data</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Genetic data</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Biomaterial</td>
<td></td>
</tr>
<tr>
<td>What kind of data protection has the source data</td>
<td>PrivacyStatus, IdentificationRisk, Personal, IdentifyingData</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Identifying (none)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pseudonym Anonym</td>
<td></td>
</tr>
<tr>
<td>What kind of data protection do you plan for your bridge?</td>
<td>Identifying (Only metadata), e.g. identity of researcher</td>
<td></td>
</tr>
<tr>
<td>Consent of your research subject</td>
<td>InformedConsent, ConsentNecessity</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Different types of consent exist</td>
<td></td>
</tr>
<tr>
<td>Use the data source without consent</td>
<td>ConsentNecessity</td>
<td></td>
</tr>
<tr>
<td>Access to the used data open or restricted</td>
<td>OpenData, NonidentifyingData</td>
<td></td>
</tr>
<tr>
<td>Type of access</td>
<td>OpenData, Deidentification, anonymised</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Open Restricted</td>
<td></td>
</tr>
<tr>
<td>Data subject to copyright</td>
<td>IP</td>
<td></td>
</tr>
</tbody>
</table>

Data producers or data providers generate research data or make data available. The data user (data consumer) is the researcher that has a research question, their own preliminary data and uses the data bridge to obtain additional data. The quality of the research data that is provided via different research data repositories is in part determined by the degree to which data can be used without limitation for scientific research while at the same time complying with applicable rules and regulations. Not only the use of human personal data, but also the open and free use of research data takes place within a legal framework and policy guidelines.
The OECD ‘Principles and Guidelines for Access to Research Data’ provide guidance regarding access to research data. The principles of “Open Access” are listed in the ‘Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities’. Based on the Usage Scenarios developed in Deliverable 5.1, the ontology does not use the term Data access but the more general term Data processing, because data access only plays a preliminary role in the sharing of data via the data bridges. The assessment tool developed here relates to the legal interface resulting from the act of data sharing. In general, the BioMedBridges data bridge represents an interface between two or more data sources provided by different research infrastructures. These data are provided under certain legal conditions or policies such as open access, restricted access, copyright, explicit consent, and/or positive vote of an Ethics Committee. The requirements clusters that were developed (see Del 5.1) defined the conditions under which data sources underlying different data protection rules and access rules can be shared, and the assessment tool translates data access and protection rules and provides the rules for all possible combinations of data sources. In this way, the legal interface set up by the tool may provide the rules to allow sharing data from an open access database with anonymised human data in a restricted access database.

Interoperability is indispensable on a technical, semantic and legal level. From the legal point of view, harmonized licences and policies have to be developed and employed to support data sharing. However, a number of barriers need to be overcome to ensure harmonized licensing conditions. An example of such barriers is formed by the possible different conditions imposed by national legislations.

The ontology introduces the concept of IdentificationRisk for SharedData. IdentifyingData is the core concept of the personal data area of the ontology. The ontology takes does not take non-identifying and animal data into account.

---

26 See: http://oa.mp.de/openaccess-berlin/berlindeclaration.html
See: http://english.vsnu.nl/web/show/id=88938/langid=42
28 K. Janssen: Legal interoperability – barriers to the harmonization of licences. Share PSI workshop Position Paper

BioMedBridges Deliverable D5.2
as data protection requirements are only brought up with the emergence of identifying data (PrivacyStatus, SharingPurpose). Following the EU Data Protection Directive, the exhaustive legal definition that “processing of personal data (‘processing’) shall mean any operation or set of operations which is performed upon personal data, whether or not by automatic means, such as collection, recording, organization, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, blocking, erasure or destruction” applies\(^29\). Whether the anonymisation of personal data may fall under this definition as a special form of processing of personal data is currently discussed controversially\(^30\).


Based on the assessment of the identification risk, the existence of identifying data (= personal data) results in data protection rules coming into effect. In contrast to previously developed privacy ontologies, the concepts EthicsCommittee and Committee (Data Protection Committee) were included in the ontology. Although Ethics Committees and Data Protection Committees are subject to national regulations and there exists a wide variety of number, structure and policies in Europe, approval by an ethics committee is essential for research with many types of human data. Consequently, it was added to the ontology and Research Ethics Committee Approval as well as Data Access Committee Approval have been included in the decision flow (matrix, part data security) (Appendix 3).

---

**Table 5 Mapping of workflow concepts to the Data Bridge Ontology**

<table>
<thead>
<tr>
<th>No.</th>
<th>workflow</th>
<th>ontology</th>
<th>comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Purpose of data use</td>
<td>ConsentNecessity</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Research question</td>
<td>ConsentNecessity</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Data sharing</td>
<td>Shared data, SharingPurpose</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Data transfer</td>
<td>DataTransfer</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Data linking</td>
<td>DataLinkage</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Merging</td>
<td>DataEnrichment</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Combining</td>
<td>DataProcessing</td>
<td>DataLinkage</td>
</tr>
<tr>
<td>8</td>
<td>Internal data source</td>
<td>DataUser</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>External data source</td>
<td>DataProvider</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Location</td>
<td>MemberState</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Country</td>
<td>MemberState</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Human species</td>
<td>Human</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Non-human</td>
<td>Animal, NonLiving</td>
<td>NonIdentifyingData</td>
</tr>
<tr>
<td>14</td>
<td>Biomaterial</td>
<td>Biomaterial</td>
<td>Biomaterial not cons.</td>
</tr>
<tr>
<td>15</td>
<td>Genomic sample</td>
<td>GenomicSample</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Gene sequence sample</td>
<td>DataType</td>
<td>Biomaterial not cons.</td>
</tr>
<tr>
<td>17</td>
<td>Protein sample</td>
<td>DataType</td>
<td>Biomaterial not cons.</td>
</tr>
<tr>
<td>18</td>
<td>Metadata</td>
<td>DataElement</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>Text data</td>
<td>DataElement</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>Image data</td>
<td>DataElement</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>Genetic data</td>
<td>DataElement</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>Data protection</td>
<td>Deidentification</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>Identifying</td>
<td>PersonalIdentifyingData</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>Pseudonym</td>
<td>Pseudonymised, PrivacyStatus</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>Anonym</td>
<td>anonymised, PrivacyStatus</td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>Open access</td>
<td>OpenData</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>Restricted access</td>
<td>RestrictedAccess</td>
<td>IdentificationRisk</td>
</tr>
<tr>
<td>28</td>
<td>Combined access</td>
<td>CombinedAccess</td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>Specific consent</td>
<td>InformedConsent</td>
<td>ConsentFormat</td>
</tr>
<tr>
<td>30</td>
<td>Consent fee type of research</td>
<td>ConsentFee</td>
<td>ConsentFormat</td>
</tr>
<tr>
<td>31</td>
<td>Consent</td>
<td>InformedConsent</td>
<td>ConsentFormat</td>
</tr>
<tr>
<td>32</td>
<td>Ethics approval</td>
<td>EthicsApproval</td>
<td>InformedConsent</td>
</tr>
<tr>
<td>33</td>
<td>Commercial value</td>
<td>CommercialValue</td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>Patent</td>
<td>Patent</td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>Non-commercial</td>
<td>NonCommercial</td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>Academic</td>
<td>Academic</td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>Genotype data</td>
<td>GenotypeData</td>
<td></td>
</tr>
<tr>
<td>38</td>
<td>Gene sequence data</td>
<td>GeneSequenceData</td>
<td></td>
</tr>
</tbody>
</table>

---

**Diagram:**

- DataBridge
- ReadData
- EncryptedData
- DataExtraction
- DataTransformation
- DataAnalysis
- DataFiltering
- DataNormalisation

---

BioMedBridges Deliverable D5.2
SharedData that has no identification risk becomes NonidentifyingData and connects to OpenData in the ontology. Open data\(^{31}\) is the idea that certain data should be freely available to everyone to use and republish, without restrictions imposed by copyright, patents or other means of control. Problems may arise because the data are commercially valuable or can be aggregated into products of potential value. Control may be exerted through access restrictions, licenses, copyright, patents and charges for access or re-use. For example, the Reactome database\(^{32}\) is an open-source, open access, manually curated and peer-reviewed pathway database. Pathway annotations are authored by expert biologists in collaboration with Reactome editorial staff and cross-referenced to many bioinformatics databases: Entrez Gene, Ensembl and UniProt, the UCSC and HapMap Genome Browsers, the KEGG Compound and ChEBI small molecule databases, PubMed, and Gene Ontology.

The tool ontology provided the structuring concepts and their relationships for the core terms of the matrix (e.g. ConsentSpecificity, SharingPurpose, MemberState,) (Appendix 5). Additional terms were needed for the technical realization of the matrix. Consequently, in addition to the basic concepts of data privacy, the complete ontology (Figure 12) included an area of data processing operations and also connections to the intellectual property domain, De-identification and the DataType concept. The DataType concept connects the ontology to the different data types considered in the workflow such as metadata, text data, and image data. The figure highlights one difference of the tool ontology compared with other privacy ontologies such as, for example, the Ontology model of data-sharing contexts\(^{33}\). SharedData can have a risk of identification of a person. Such a risk may be enhanced in cases where non-identifying human data are linked with genetic data. The concept of IdentificationRisk connects the SharedData with the IdentifyingData (Figure 12).

\(^{31}\) http://en.wikipedia.org/wiki/Open_data  
\(^{32}\) http://www.reactome.org/ReactomeGWT/entrypoint.html  
Because of the inclusion of the concept of IdentificationRisk, which relates to IdentifyingData for human personal data but also to potentially identifying genetic data (e.g. gene sequences) and NonidentifyingData (e.g. all animal data) that can be easily shared, our ontology-driven privacy compliance assessment allows for a much wider range of rule sets than the available privacy ontologies that focus only on human personal data and often on national data protection regulations (e.g. NEURONA for Spain). The available ontologies focus on legal concepts, whereas we confine the legal concepts to the domain of data protection of identifying and personal data. This enables us to create rule sets for shared NonidentifyingData that constitute the predominance of data sources in ESFRI infrastructures. On the other hand, although our assessment tool considers the heterogeneity of available data sources, in its current state, it omits the national data protection specifics that still exist in Europe.

To confirm the research-focused approach and to get initial information about the usability of the tool, interviews with potential users at several research
infrastructures were conducted. Three interviews were conducted involving researchers from FIMM (WP8, potential use case personalized medicine in prostate cancer), a Finnish institution (WP10, population cohort studies), and INFRAFRONTIER and HMGU (WP7). In summary, the tool was judged to be very useful concerning data protection and/or ethical issues. The support provided by the tool in the form of assistance and guidance in the case of uploading of potentially identifying data to an open access data source was recognised (Appendix 4).

9 Testing of the prototype

Due to the complexity of the requirement matrix with respect to possible queries, testing of the assessment tool necessitated a structured approach and careful selection of relevant use cases. The assessment tool was tested on two different levels:

- semantic and logic integrity of the requirement matrix
- correctness of the implementation (tool).

Based on the first level, use cases for the tool test were developed.

9.1 Matrix Tests

To test and, if necessary, correct the requirement matrix, two test scenarios were developed:

- Non-contradictory requirement sets: for all requirements, it was separately tested whether other requirements that are assessed as non-contradictory by the matrix were indeed compatible with the requirement considered and whether all compatible requirements were selected.

- Correct requirement selection: it was checked whether queries corresponding to the requirements result in the requirements expected.

Two raters worked separately and exchanged their tests in order to achieve cross-validated results and increase the quality of the data. In case of divergent results, the raters discussed the concrete test, resolved the problem and
performed the test again. The concrete tests were performed by using the filter tool in MS Excel spreadsheets. Figure 13 shows a screenshot of the spreadsheet that was used for testing. Full test scenarios and results are provided in the supplement.

<table>
<thead>
<tr>
<th>id</th>
<th>dp</th>
<th>dataType</th>
<th>dataProtectionSource</th>
<th>dataPurpose</th>
<th>legalApproval</th>
<th>copyright</th>
<th>dataProtectionBridge</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>DP1.1</td>
<td>non-human metadata</td>
<td>pseudonym</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
</tr>
<tr>
<td>3</td>
<td>DP2.1</td>
<td>non-human metadata</td>
<td>identifying</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
</tr>
<tr>
<td>4</td>
<td>DP1.2</td>
<td>non-human metadata</td>
<td>pseudonym</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
</tr>
<tr>
<td>5</td>
<td>DP1.2</td>
<td>non-human metadata</td>
<td>identifying</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
</tr>
<tr>
<td>6</td>
<td>DP1.3</td>
<td>non-human metadata</td>
<td>pseudonym</td>
<td>any</td>
<td>no</td>
<td>any</td>
<td>pseudonym</td>
</tr>
<tr>
<td>7</td>
<td>DP1.3</td>
<td>non-human metadata</td>
<td>identifying</td>
<td>any</td>
<td>no</td>
<td>any</td>
<td>pseudonym</td>
</tr>
<tr>
<td>8</td>
<td>DP1.3</td>
<td>non-human metadata</td>
<td>pseudonym</td>
<td>any</td>
<td>no</td>
<td>any</td>
<td>identifying</td>
</tr>
<tr>
<td>9</td>
<td>DP1.3</td>
<td>non-human metadata</td>
<td>identifying</td>
<td>any</td>
<td>no</td>
<td>any</td>
<td>identifying</td>
</tr>
<tr>
<td>10</td>
<td>DP3.0</td>
<td>non-human metadata</td>
<td>identifying</td>
<td>any</td>
<td>no</td>
<td>any</td>
<td>identifying</td>
</tr>
<tr>
<td>11</td>
<td>DP3.0</td>
<td>non-human metadata</td>
<td>identifying</td>
<td>any</td>
<td>no</td>
<td>any</td>
<td>pseudonym</td>
</tr>
<tr>
<td>12</td>
<td>DP3.0</td>
<td>non-human metadata</td>
<td>pseudonym</td>
<td>any</td>
<td>no</td>
<td>any</td>
<td>identifying</td>
</tr>
<tr>
<td>13</td>
<td>DP3.0</td>
<td>non-human metadata</td>
<td>pseudonym</td>
<td>any</td>
<td>no</td>
<td>any</td>
<td>pseudonym</td>
</tr>
<tr>
<td>14</td>
<td>DP4.0</td>
<td>human</td>
<td>metadata</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
</tr>
<tr>
<td>15</td>
<td>DP5.1</td>
<td>human</td>
<td>image</td>
<td>identifying</td>
<td>any</td>
<td>any</td>
<td>any</td>
</tr>
</tbody>
</table>

**Figure 13** Screenshot of the requirement matrix with filters

### 9.2 Tool Tests

After completion of the matrix test, the matrix was frozen and assumed to be correct. As the workflow in 5.3 precludes some non-feasible selections, this had to be considered in the construction of use case scenarios for the tool test. For all requirements, a use case was developed based on the workflow. Then, expected requirements were calculated by the matrix – again, with the filter tool of MS Excel and in a cross-validation context with two raters. Finally, the concrete queries were entered into the tool and the results were compared with the expected ones. Overall, the tests showed that the core functionality, i.e. the requirement matrix, was implemented correctly. The image requirements were not selected because of a typo (using “images” instead of “image” as the data type), and the requirement tables were displayed incorrectly because of a missing escaping mechanism related to commas. In some instances, non-repeatable output was generated, e.g. a requirement was not shown; this could be due to random errors in reading out the database. For all use case scenarios and the results, we refer to the corresponding spreadsheets provided in the supplement.
10 Outlook

The assessment tool has several features, such as extensibility or usability (e.g. guiding the user through the questions and preventing non-feasible queries). However, there is potential for improvements and extensions. First, a more comprehensive and detailed ontology with relevant terms related to biomedical research and associated legal issues is necessary in order to cover a wider range of requirements than currently implemented. Second, a graphical wizard for entering new requirements into the tool is highly desirable; this would also allow customizing previously listed requirements. Third, as a data provider it should be possible to enter data from more than two data sources, which affects the workflow described in section 5.2. From our perspective, such improvements are essential for a broader acceptance and continuity of the assessment tool. First steps with respect to continuity were already taken by using an open-source solution, for which it is, for example, straightforward to change the database management system.

11 Supplemental material (online)

This report and all supplemental material are available at http://www.biomedbridges.eu/deliverables/52-0:

1. Tool content: Data protection/privacy requirements
2. Tool content: Data security requirements
3. Tool content: Explanations for requirements
4. Tool content: Intellectual property requirements
5. Tool content: Mapping of requirements to explanations
6. Tool content: Matrix tests
7. Tool content: Requirement matrix
8. Tool content: Security requirements of biosamples
9. Tool content: Tests performed
10. Template: Data transfer agreement – personal data
11. Template: Data transfer agreement – non-personal data
12. Template: Material transfer agreement – human biosamples
13. Template: Material transfer agreement – non-personal biosamples
14. Template: Provider agreement – human biosamples
15. Template: Provider agreement – non-personal biosamples
16. Template: Data provider agreement – personal data
17. Template: Data provider agreement – non-personal data
18. Template: Information Sheet and Consent Form
12 Delivery and schedule

The delivery is delayed:  ☐ Yes ☑ No

13 Adjustments made

No adjustments were made.

14 Background information

This deliverable relates to WP 5; background information on this WP as originally indicated in the description of work (DoW) is included below.

WP 5 Title: Secure access

Lead: Christian Ohmann (UDUS), Klaus Kuhn (TUM-MED)

Participants: EMBL, STFC, UDUS, TUM-MED, ErasmusMC, TMF, HMGU, INSERM

Data sharing between the different e-Infrastructures will lift ethical, legal and security concerns to a new level of complexity. A preliminary analysis of data security requirements of the e-Infrastructures ECRIN, EATRIS, BBMRI, ELIXIR, Euro-BioImaging, INSTRUCT, EMBL-EBI, and Inrafrontier showed that following e-Infrastructures are storing or processing patient related data (or biosamples): EATRIS, ECRIN, BBMRI, Euro-BioImaging and EMBL-EBI. In addition, INSTRUCT is interested in secure sample transport and in intellectual property rights; Infrafrontier stores high-throughput data from mice. BBMRI with its focus on the availability of biomaterials is currently emphasizing aspects like k-anonymity and metadata management for its data. Sharing of imaging data by Euro-BioImaging poses challenges with respect to anonymisation and intellectual property. Therefore, an ethical, regulatory and security framework for international data sharing that covers these diverse areas and different types of data (e.g. clinical trials data, mouse data, and human genotype and DNA sequence data) is of crucial importance for BioMedBridges.

<table>
<thead>
<tr>
<th>Work package number</th>
<th>WP5</th>
<th>Start date or starting event:</th>
<th>month 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Work package</td>
<td>Secure access</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

BioMedBridges Deliverable D5.2
<table>
<thead>
<tr>
<th>title</th>
<th>Activity Type</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RTD</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Person-months per participant</td>
<td>61</td>
<td>15</td>
<td>54</td>
<td>0</td>
<td>58</td>
<td>5</td>
<td>34</td>
<td>10</td>
<td>4</td>
</tr>
</tbody>
</table>

**Objectives**

Based on an analysis of the complex ethical, legal and regulatory issues resulting from international data and biomaterial sharing between different e-Infrastructures, WP5 will develop a security framework that will ensure that services provided by BioMedBridges are compliant with local, national and European regulations and privacy rules. Therefore the developed legal framework will allow the use of data bridges, that consider among other regulations the EU Directive 95/46/EC, EU Directive 2001/20/EC (GCP), national data protection acts, GLP rules, animal protection laws, laws about biobanking, laws concerning genetic data and stem cell research, data access approval rules (by informed consent), rules by Hospital Boards or Research / Ethics Committees as well as regulations for intellectual property and license rights.

The legal foundation will be applied for the development of a security framework employing security policies, account policies, consent, user agreements of the participating infrastructures and authentication and authorization services. Existing standards and concepts of European e-infrastructures (e.g. GÉANT / eduGAIN and TERENA) will be considered.

WT1: Regulations and privacy requirements for using the data bridges

WT 2: Rules and regulations for accessing databases of e-Infrastructures

WT 3.1: Regulations and security issues regarding security of biosamples

WT 3.2: Regulations and security issues regarding animal protection

WT 3.3: Rules and regulations regarding data connected to intellectual property and licences in e-Infrastructures

WT 4: Development of a tool for assessment of ethical and legal requirements

BioMedBridges Deliverable D5.2
WT 5: Security requirements for an e-infrastructure addressing the use cases

WT 6: Threat and risk analysis for sharing data or biomaterials

WT 7: Design of the security architecture and framework

WT 8: Implementation of a pilot for the security framework

Description of work and role of participants

In WT 1-4 regulations, requirements and design aspects; in WT 5-8 the security implementation are addressed.

In the first part, information collection will require extensive contacting and considerable travelling. In the second part, staff exchange will be an important way to coordinate activities. WT5 will be chaired by UDUS and TUM.

WT 1: Regulations and privacy requirements for using the data bridges (M1-M12)
(Leader: UDUS, Participants: EMBL-EBI, Erasmus MC, HMGU, STFC, TMF, TUM, FVB, INSERM)

This task will analyse the legal and ethical situation concerning the sharing and transfer of data and the access to data in a trans-European context for all e-Infrastructures. The legal implications and corresponding data exchange strategies will be analysed on European, national, regional (e.g. data protection law in Scotland) and local (e.g. hospital law) level. Legal implications for different types of data and the linking of data have to be considered, including biobank data, genetic data, stem cell research data, data of children and vulnerable will be paid to personal data (Directive 95/46/EC) and the roles of data controller and data processor for the data bridges. Subcontracting will be needed for lawyer support and translation of legal documents.

WT 2: Rules and regulations for accessing databases of e-Infrastructures (M6-M18)
(Leader: UDUS, Participants: EMBL-EBI, Erasmus MC, HMGU, STFC, TMF, TUM, FVB, INSERM)

This task will analyse the rules, regulations and associated practices and policies concerning the access to e-Infrastructure databases. A survey will analyse the situation and policies of all e-Infrastructure databases.

Special attention will be paid to the role of different types of informed consent, research exemptions, policies, and approvals by Hospital Biobanks Boards or Research and Ethics Committees.

WT 3.1: Regulations and security issues regarding security of biosamples (M1-
This task will analyse the rules and regulations that affect data protection and security of bio samples. Especially the physical transfer of samples may be restricted by national legislations.

WT 3.2: Regulations and security issues regarding animal protection (M1-M12)  
(Leader: TMF, Participants: EMBL-EBI, Erasmus MC, UDUS, HMGU, TUM, FVB, INSERM)

This task will analyse the rules, practices and regulations concerning data protection and the protection of animal welfare.

WT 3.3: Rules and regulations regarding data connected to intellectual property and licences in e-Infrastructures (M1-M12) (Leader: EMBL-EBI, Participants: Erasmus MC, UDUS, HMGU, STFC, TMF, TUM, FVB, INSERM)

This task will analyse the rules, practices and regulations concerning the access to databases and the sharing of data protected by intellectual property rights.

WT 4: Development of a tool for assessment of ethical and legal requirements and supporting documents (M13-24) (Leader: TMF, Participants: EMBL-EBI, Erasmus MC, UDUS, HMGU, STFC, TUM, FVB, INSERM)

In this WT all results of the previous WTs will be collected, integrated and interdependencies will be developed. The different dimensions of the developed requirements matrix will cover: (1) kind of data (patient data, molecular data, mouse data, phenotype data, etc.), kind of data protection (anonymisation, pseudonymisation, none), regulations and rules for secure access. A priority list of combinations of these dimensions that may happen during cooperation between different e-Infrastructures will be analysed and depicted. In addition, contractual templates and generic texts will be developed to support a legal sound cooperation for data exchange.

WT 5: Security requirements for an e-infrastructure addressing the use cases (M6-30). (Leader: TUM, Participants: EMBL-EBI, Erasmus MC, UDUS, HMGU, STFC, TMF, FVB, INSERM)

Utilizing results from the previous WTs and focussing on a priority list of use cases including WP8, WP7 and WP10, security requirements for aggregated or shared data or biomaterials will be identified, including confidentiality, integrity, and availability. These requirements will consider the different levels of integration (WP4), type and content of integrated data (including the specific risk of re-identification) or shared biomaterials, security policies and consent agreements of the participating infrastructures and European regulations. The
use of de-identification and (k-) anonymity will be specified.

Requirements for data access layers will be defined. Suggested tiers are: (1) Public access to meta and coarse grained data, where typical risks need to be considered (e.g. statistical inference of membership); (2) access to k-anonymous derived or summary data based on use agreements and user accounts, (3) access to de-identified microdata integrated / accessible across infrastructures which requires approval of a data access committee. Consent agreements and security policies of the participating infrastructures will be considered in these tiers.

WT 6: Threat and risk analysis for sharing data or biomaterials (M9-30) (Leader: TUM, Participants: EMBL-EBI, Erasmus MC, UDUS, HMGU, STFC, TMF, FVB, INSERM)

Based on the security requirements, a threat and risk analysis will be performed. Attacker models, origins of threats (e.g. trails), and possible points of attack will be identified, considering results from latest research. Following typical (risk) categories need to be considered: Membership disclosure, attribute disclosure and re-identification. The risk analysis will weigh the different threats, considering the interests of researchers, protection of research-related IP, and privacy of patients.

WT 7: Design of the security architecture and framework (M18-30) (Leader: EMBL-EBI, Participants: TUM)

 Derived from the requirements developed in previous WTs, a security framework will be designed, comprising authentication, authorization, and accounting services. Different security solutions will be evaluated, ranging from decentralized to tightly integrated authentication and authorisation. Access layers and corresponding approval workflows will be specified. Authentication mechanisms for the integrated databases need to be designed, using standards (e.g. OpenID, Shibboleth, Liberty Alliance) and utilizing concepts or solutions from European identity federation initiatives (GÉANT and TERENA). The security policies of BioMedBridges will comprise access policies and use agreements and will consider security policies of participating infrastructures and European laws and regulations (derived from WT 4). The security framework needs access to a repository of authorization rules as part of a metadata repository. These authorization rules will be based on consent and regulations of the participating infrastructures combined with rules and contracts for co-operation. Authorization policies have to be expressed in an appropriate format (e.g. XACML). The policy administration repository will be related to defined access tiers. Logging of user activities is used to ensure accountability.

WT 8: Implementation of a pilot for the security framework (M24-48) (Leader: EMBL-EBI, Participants: TUM, UDUS, STFC, TMF)

Implementation will need close collaboration with WP4 and WP3. Parallel to the implementation steps of the services provided by WP4, and for the same use
cases, the security framework developed in this WP will be implemented. The policy administration repository will be a central part of this implementation.

Subcontracting for legal costs: UDUS (partner 5) for legal costs associated with WP5 - Work Task 1 of WP5 will analyse the legal and ethical situation concerning the sharing and transfer of data and the access to data in a trans-European context for all e-Infrastructures. Subcontracting is required for legal advice and the translation of legal documents.
Appendices

Appendix 1: Example requirement table ................................................................. 2
Appendix 2: User Manual .......................................................................................... 3
Wizard ....................................................................................................................... 3
Appendix 3: Questionnaire and decision workflow ............................................... 7
Questions and leaps ................................................................................................. 7
Decision workflow .................................................................................................... 10
Appendix 4: Stakeholder and user identification .................................................... 12
Questions for the clarification of the approach ...................................................... 13
Results of potential user interviews ...................................................................... 13
Appendix 5: Mapping of matrix concepts to the ontology ....................................... 15

Figures

Figure 1 Wizard: purpose of data usage ................................................................. 3
Figure 2 FAQs ......................................................................................................... 3
Figure 3 Wizard: Data Provider No.1 (Example) ..................................................... 4
Figure 4 Wizard: Data Provider No.2 (Example) ..................................................... 4
Figure 5 Wizard: Required Data Questions ............................................................ 5
Figure 6 Wizard: Confirmation/ Summary ............................................................... 5
Figure 7 Assessment Results Screen ..................................................................... 6
Figure 8 Workflow questions 7) – end ................................................................. 11
### Appendix 1: Example requirement table (potential identification risk for human research subjects)

<table>
<thead>
<tr>
<th>No.</th>
<th>Name</th>
<th>Description</th>
<th>Research subject</th>
<th>Data type</th>
<th>Anonymously</th>
<th>Pseudonymously</th>
<th>Identifying</th>
<th>Identifyingconstraint</th>
<th>Regulation</th>
<th>dependency</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Removing the metadata</td>
<td>Identifying patient data in DICOM meta data have to be removed or replaced</td>
<td>human</td>
<td>image</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>Data linking, public access</td>
<td>Data Protection Directive 95/46/EC Art. 6 (e). No data shall be kept longer than necessary</td>
<td>none</td>
</tr>
<tr>
<td>2.2</td>
<td>Removing the data on the image</td>
<td>Identifying patient data on an image have to be removed to avoid re-identification of the patient</td>
<td>human</td>
<td>image</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>Data linking, public access</td>
<td>Data Protection Directive 95/46/EC Art. 6 (e). No data shall be kept longer than necessary</td>
<td>none</td>
</tr>
<tr>
<td>2.3</td>
<td>Altering the image</td>
<td>Images have to be altered to prevent re-identification. In neuroimaging there are for example methods (defacing or skull stripping) to prevent face reconstruction.</td>
<td>human</td>
<td>image</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>Data linking, public access</td>
<td>Data Protection Directive 95/46/EC Art. 6 (e) requires minimization of risk of de-identification.</td>
<td>none</td>
</tr>
</tbody>
</table>
Appendix 2: User Manual


**Wizard**

Enter the address into the web browser.

On the web page, select “Assessment Tool”.

You will be directed to the first tab of the wizard (see Figure 1).

![Figure 1 Wizard: purpose of data usage](image)

At the bottom of the assessment tool there are a few FAQs to assist the user and to give additional information to the tool (see Figure 2).

![Figure 2 FAQs](image)

Start answering the question and press the next button to get to the next tab.

On the second and third tab of the wizard all questions according to the data provider are given. Depending on your answers the tool will hide or show different questions (compare Figure 3 and Figure 4).
If you have forgotten to answer one or more questions the tool will gently remind you to answer all questions (see Figure 5) and indicate the missing ones (all mandatory questions are additionally marked with *).
Figure 5 Wizard: Required Data Questions

After completion of all questions the tools will summarize your answers. To confirm and to get the results of your assessment press the “Send Assessment” button. To go back press “Back” or “New Assessment” to reset your answers and start from the beginning (see Figure 6).

Figure 6 Wizard: Confirmation/ Summary

After sending the assessment an information message shows if the request was successful or not (see yellow box in Figure 7). The result page
compromises both data providers each with five sections with requirements and an overview of the entered answers (see Figure 7). Further there is a button to get back to the Confirmation/Summary section and one to start a new assessment.

Figure 7 Assessment Results Screen
Appendix 3: Questionnaire and decision workflow

Questions and leaps

1) What is the purpose for data usage?
   a) Redistribution
   b) Data linking
   c) Others
      • If a) is selected: dataPurpose="Redistribution" and any,
      • if b) is selected: dataPurpose="Data linking" and any,
      • if c) is selected: dataPurpose="Others" and any.
   ➔ Select and go to II)

2) Where is your data source located?
   a) DE
   b) GB / UK
   c) FR
   d) NL
   e) ES
   f) Other
      • If a-e) is selected: country=x and EU,
      • if f) is selected: country=EU.
   ➔ Select and go to 3).

3) What is the original species of the data?
   a) Human
   b) Non-human
      • If a) is selected: subject="Human" and any,
      • if b) is selected: subject="Non-human" and any.
   ➔ Select and go to 4).

4) What kind of data or biomaterial will be used (data classification)?
   a) Metadata
   b) Textdata
   c) Images
   d) Genetic data
   e) Biosample
   f) Biosample associated data
      • If a) is selected: dataType="Metadata", “Textdata", "Digitaldata" and any,
      • if b) is selected: dataType="Textdata", “Digitaldata” and any,
      • if c) is selected: dataType="Images", “Digitaldata" and any,
- if d) is selected: dataType="Genetic data", "Textdata", "Digitaldata" and any,
- if e) is selected: dataType="Biosample" and any,
- if f) is selected: dataType="Biosample associated data", "Textdata", "Digitaldata" and any.

→ If (subject = "Human") or (subject = "Non-human" and dataType = "Metadata") go to 5), else go to 10).

5) **What kind of data protection has the source data?**
   a) Identifying
   b) Pseudonym
   c) Anonym
      - If a) is selected: dataProtectionSource="Identifying" and any,
      - if b) is selected: dataProtectionSource="Pseudonym" and any,
      - if c) is selected: dataProtectionSource="Anonym" and any.

→ If a) or b) go to 6), else go to 10).

6) **What kind of data protection do you plan for your bridge?**
   a) Identifying
      → (Only available if subject="Non-human" and dataType="Metadata")
   b) Pseudonym
   c) Anonym
      - If a) is selected: dataProtectionBridge="Identifying" and any,
      - if b) is selected: dataProtectionBridge="Pseudonym" and any,
      - if c) is selected: dataProtectionBridge="Anonym" and any.

→ If a) or b) go to 7), else go to 10).

7) **Does your consent cover your new research project?**
   a) Yes
   b) no
      - If a) is selected: coveringConsent="Yes" and any,
      - if b) is selected: coveringConsent="No" and any.

→ If b) go to 8), else go to 10).

8) **Do you have a legal basis to use the data source without consent?**
   a) Yes
   b) No
      - If a) is selected: legalApproval="Yes" and any,
      - if b) is selected: legalApproval="No" and any.

→ If b) go to 9), else go to 10).

9) **Is it feasible for you to get a new consent of your research subject?**
   a) Yes
   b) no
      - if a) is selected: newConsent="Yes" and any,
      - if b) is selected: newConsent="No" and any.

→ Select and go to 10).
10) Is the access to the used data open or restricted?
   a) Open
   b) Restricted
      • If a) is selected: originalAccess="Open" and any,
      • if b) is selected: originalAccess="Restricted" and any.
   ➔ Select and go to 11).

11) What is the desired type of access?
   a) Open
   b) Restricted
   c) Combined (different access tiers)
      • If a) is selected: bmbAccess="Open" and any,
      • if b) is selected: bmbAccess="Restricted" and any,
      • if c) is selected: bmbAccess="Combined", "Open", "Restricted" and any.
   ➔ Select and go to 12).

12) Are the data subject to copyright or intellectual property?
   a) Yes
   b) No
      • If a) is selected: copyright="Yes" and any,
      • if b) is selected: copyright="No" and any.
   ➔ Select and go to next source.
Decision workflow

Figure 8 Workflow questions1) – 6)
Figure 9 Workflow questions 7) – end
Appendix 4: Stakeholder and user identification

The aim is to give potential users of data bridges some guidance on ethical, legal and data security issues when linking or sharing data and provide them with (contractual) templates (e.g. licensing agreements, material transfer agreements, consent, etc.). However, we were not sure how potential users would like to use this tool and how we should design the online version so that it is useful for them. To address these challenges, we asked all Use Case Coordinators to provide us with potential user contacts to conduct individual telephone interviews with them. We ended up with three interviews, which were regarded to be enough as point of reference for the potential user perspective. The potential user is regarded to be a typical researcher.

The following telephone interviews were conducted:

— 02 September 2013, WP8: WP8 coordinator: Henrik Edgren; potential user: Päivi Östling; researcher at FIMM (Finland); research field: personalized medicine in prostate cancer

— 18 September 2013, WP10: WP10 coordinator: Ugis Sarkans; potential user: Kaisa Silander; researcher in Finland; research field: diseases by means of population cohort studies

— 09 October 2013, WP7: WP7 leader and potential user: Michael Raess; researcher (biologist), project manager at INFRAFRONTIER and HMGU. Involved with organization tasks: constructional and financial issues of the research infrastructures

First, the potential users were provided with the outline of the project/ work task. The BMB project and the connection to the WT5.4 vision/ aim to develop a legal, ethical, data privacy/ protection/ security assessment tool was explained.
Questions for the clarification of the approach

If you begin a research project where data usage (sharing/transfer/linking etc.) is or may be a concern, how is your approach in concerns of ethical, legal and data protection/privacy and security issues?

1. Do you have expertised partners to contact? Legal-counselling in-house? Extern?
2. Do you have guiding (contractual) templates (e.g. licensing agreements, material transfer agreements, consent, etc.), documents?
3. Do you have experience with ethics committees?
4. Have there been any problems with ethical, legal aspects?

Results of potential user interviews

In summary, the tool was seen by all participants as very useful for the topics of data protection and/or ethics. The support of the tool by providing assistance and guidance in the case of uploading of identifying data to an open access data source is recognised. It was stated that the tool should be able to assist and guide the planning of clinical studies, e.g. by simplifying the choice of informed consent types for different settings. The users had different opinions about the guidance function of the tool.

A suggestion was that the tool should provide checklists of things to consider when entering new research collaborations (e.g. ethics or IP). In addition, support by the decision when or how data should be anonymised (e.g. sequence data) is needed, as well as information about potential restrictions to use data. The tool should provide templates for consent or material transfer agreements for new research collaborations. It is acknowledged that the problem lies in the differences of national laws.

As entrance to the tool, the selection of different data types was preferred over a selection of research fields, because different research fields use the same kind of data and thus have the same kinds of problems. Thus, the navigation by
data types is considered as easier and corresponds better to the research flow. The data type model is preferred over research fields as research fields all use the similar kind of data and thus have the same kinds of problems. Navigation by data type is considered easier. A researcher’s view model for the tool’s user interface which starts with the characterization of data sources the user wants to use is favored. For WP7 no issues of data protection and IP are foreseen; all data sources used, like ArrayExpress, Gene Expression Atlas or data import from University Graz (BBMRI) and CERM (Instruct) is freely accessible. Nonetheless, it is seen that the tool may be of value to support researchers by supporting researchers to prevent an upload of potentially identifying data. Data protection issues are dealt with in the personalised medicine use case.

The national infrastructure components of INFRAFRONTIER operate under tight legal requirements, and complicacy to those requirements is a prerequisite for becoming part of the INFRAFRONTIER Research Infrastructure. Therefore, the ability to provide legally compliant data lies in the responsibility of the national institutions. Thus the researchers at member and partner institutions are the potential users of the tool. For these researchers the tool should support the data upload of identifying data. The Finish researcher had interest in receiving material transfer agreements. In addition, the publication of some kinds of data, like raw data, and the selection of the right kind of informed consent need support. An added value would consist when the tool could be used to assist the planning of clinical studies, e.g. to provide the right informed consent for different settings.

It would be very useful if the tool could provide users with a checklist of things to think about when entering new research collaboration, e.g. ethics or IP. More specifically, it would be useful to know when or how data should be anonymised (e.g. sequence data). It would be great to have suggestions for tools for anonymisation or aggregation of data. It would be very useful to know more about the restrictions of use of data.

Templates for consent in new research collaborations were seen as a useful feature.
Appendix 5: Mapping of matrix concepts to the ontology

The following table shows the mapping of matrix concepts to ontology concepts of the Data Bridge Ontology, i.e. how the ontology provides the structuring concepts and their relationships for the core terms of the matrix (e.g. ConsentSpecifity, SharingPurpose, MemberState). Additional terms are needed for the technical realization of the matrix. (“technical” concerns specific terms included in the realization of the assessment tool, but with little importance for the decision workflow and the ontology. Biosampling was not considered in the ontology).

<table>
<thead>
<tr>
<th>No.</th>
<th>matrix</th>
<th>ontology</th>
<th>comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>legal basis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>subject</td>
<td></td>
<td>technical</td>
</tr>
<tr>
<td>3</td>
<td>data protection</td>
<td></td>
<td>technical</td>
</tr>
<tr>
<td>4</td>
<td>dependency</td>
<td></td>
<td>technical</td>
</tr>
<tr>
<td>5</td>
<td>original consent</td>
<td>ConsentSpecifity</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>desired type of research</td>
<td>ConsentSpecifity</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>access</td>
<td></td>
<td>technical</td>
</tr>
<tr>
<td>8</td>
<td>original access</td>
<td></td>
<td>technical</td>
</tr>
<tr>
<td>9</td>
<td>human</td>
<td>Human</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>image</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>pseudonym</td>
<td>pseudonymised</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>identifying</td>
<td>Deidentification</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>anonym</td>
<td>anonymised</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>informed consent</td>
<td>InformedConsent</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>data linking</td>
<td>DataLinkage</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>ethic committee approval</td>
<td>EthicsCommittee</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>biosample</td>
<td></td>
<td>not considered</td>
</tr>
<tr>
<td>18</td>
<td>textdata</td>
<td></td>
<td>DataType</td>
</tr>
<tr>
<td>19</td>
<td>genetic data</td>
<td></td>
<td>DataType</td>
</tr>
<tr>
<td>20</td>
<td>digital</td>
<td></td>
<td>DataType</td>
</tr>
<tr>
<td>21</td>
<td>metadata</td>
<td></td>
<td>DataType</td>
</tr>
<tr>
<td>22</td>
<td>non-commercial</td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>commercial</td>
<td>IP</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>national law</td>
<td>MemberState</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>open access</td>
<td>OpenData</td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>open</td>
<td>OpenData</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>restricted</td>
<td></td>
<td>technical</td>
</tr>
<tr>
<td>28</td>
<td>specific consent</td>
<td>ConsentSpecifity</td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>broad consent</td>
<td>ConsentSpecifity</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>legal basis</td>
<td></td>
<td>technical</td>
</tr>
<tr>
<td>31</td>
<td>Privacy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>Requirement</td>
<td></td>
<td>technical</td>
</tr>
<tr>
<td>33</td>
<td>Description</td>
<td></td>
<td>technical</td>
</tr>
<tr>
<td>34</td>
<td>Research subject</td>
<td></td>
<td>technical</td>
</tr>
<tr>
<td>35</td>
<td>Data type</td>
<td></td>
<td>DataType</td>
</tr>
<tr>
<td>36</td>
<td>Anonymous</td>
<td>anonymised</td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>Pseudonymous</td>
<td></td>
<td>(DataType)</td>
</tr>
<tr>
<td>38</td>
<td>Identifying</td>
<td>IdentifyingData</td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>Constraint</td>
<td></td>
<td>technical</td>
</tr>
<tr>
<td>40</td>
<td>Regulation</td>
<td></td>
<td>technical</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>48</td>
<td><strong>data security</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>49</td>
<td>Data Access Committee Committee</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>Approval Committee</td>
<td></td>
<td></td>
</tr>
<tr>
<td>51</td>
<td>Research Ethics Committee EthicsCommittee</td>
<td></td>
<td></td>
</tr>
<tr>
<td>52</td>
<td>secondary use SharingPurpose</td>
<td></td>
<td></td>
</tr>
<tr>
<td>53</td>
<td>Council of Europe’s Recommendations technical</td>
<td></td>
<td></td>
</tr>
<tr>
<td>54</td>
<td>Dec. of Helsinki EthicsCommittee technical</td>
<td></td>
<td></td>
</tr>
<tr>
<td>55</td>
<td>ESHG 2003 technical</td>
<td></td>
<td></td>
</tr>
<tr>
<td>56</td>
<td>UNESCO 2003 technical</td>
<td></td>
<td></td>
</tr>
<tr>
<td>57</td>
<td>CIOMS WHO 2002 technical</td>
<td></td>
<td></td>
</tr>
<tr>
<td>58</td>
<td>transfer of data DataTransfer technical</td>
<td></td>
<td></td>
</tr>
<tr>
<td>59</td>
<td>MTA technical</td>
<td></td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>share data SharedData technical</td>
<td></td>
<td></td>
</tr>
<tr>
<td>61</td>
<td>verifying identity Personal</td>
<td></td>
<td></td>
</tr>
<tr>
<td>62</td>
<td>permissions technical</td>
<td></td>
<td></td>
</tr>
<tr>
<td>63</td>
<td>authentication system technical</td>
<td></td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>authorization system technical</td>
<td></td>
<td></td>
</tr>
<tr>
<td>65</td>
<td>audit trail technical</td>
<td></td>
<td></td>
</tr>
<tr>
<td>66</td>
<td>Accountability technical</td>
<td></td>
<td></td>
</tr>
<tr>
<td>67</td>
<td>Approval for use DataUser technical</td>
<td></td>
<td></td>
</tr>
<tr>
<td>68</td>
<td>Approval for processing DataProcessing technical</td>
<td></td>
<td></td>
</tr>
<tr>
<td>69</td>
<td>Approval for redistribution DataTransfer technical</td>
<td></td>
<td></td>
</tr>
<tr>
<td>70</td>
<td>Identifying data Personal</td>
<td></td>
<td></td>
</tr>
<tr>
<td>71</td>
<td>Personal data Personal</td>
<td></td>
<td></td>
</tr>
<tr>
<td>72</td>
<td>usage of data SharingPurpose</td>
<td></td>
<td></td>
</tr>
<tr>
<td>73</td>
<td>Limitation of use SharingPurpose</td>
<td></td>
<td></td>
</tr>
<tr>
<td>74</td>
<td>Removal of data Deidentification</td>
<td></td>
<td></td>
</tr>
<tr>
<td>75</td>
<td>donor data not considered</td>
<td></td>
<td></td>
</tr>
<tr>
<td>76</td>
<td>data loss technical</td>
<td></td>
<td></td>
</tr>
<tr>
<td>77</td>
<td>backup technical</td>
<td></td>
<td></td>
</tr>
<tr>
<td>78</td>
<td>IP IP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>79</td>
<td>copyright material IP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>80</td>
<td>commercially sensitive IP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>81</td>
<td>copyright material IP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>82</td>
<td>open access IP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>83</td>
<td>restricted access IP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>84</td>
<td>Data encryption IP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>85</td>
<td>metadata encryption IP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>86</td>
<td>data access agreement IP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>87</td>
<td>licensing agreement IP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>88</td>
<td>quality control IP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>89</td>
<td>limited liability IP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>90</td>
<td>quality statement IP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>91</td>
<td>MTA IP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>92</td>
<td>biological samples IP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>93</td>
<td>Limited liability statement IP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>94</td>
<td>Material transfer agreement IP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>95</td>
<td>Non-disclosure agreement IP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>96</td>
<td>Disposal of biological samples IP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>97</td>
<td>Removal of certain metadata IP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Appendix</td>
<td>Description</td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>-------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>98</td>
<td>temporary embargo</td>
<td></td>
<td></td>
</tr>
<tr>
<td>99</td>
<td>User authentication</td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>Data labeled as 'restricted'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>101</td>
<td>Staff awareness</td>
<td></td>
<td></td>
</tr>
<tr>
<td>102</td>
<td>patentability</td>
<td></td>
<td></td>
</tr>
<tr>
<td>103</td>
<td>hashing of search queries</td>
<td></td>
<td></td>
</tr>
<tr>
<td>104</td>
<td>depositor</td>
<td></td>
<td></td>
</tr>
<tr>
<td>105</td>
<td>researcher</td>
<td></td>
<td></td>
</tr>
<tr>
<td>106</td>
<td>copyright material</td>
<td></td>
<td></td>
</tr>
<tr>
<td>107</td>
<td>commercially sensitive copyright material</td>
<td></td>
<td></td>
</tr>
<tr>
<td>108</td>
<td>security of biosamples</td>
<td></td>
<td></td>
</tr>
<tr>
<td>109</td>
<td>Deletion of personal data</td>
<td></td>
<td></td>
</tr>
<tr>
<td>110</td>
<td>Deletion of identifying data</td>
<td></td>
<td></td>
</tr>
<tr>
<td>111</td>
<td>Renewal of consent</td>
<td></td>
<td></td>
</tr>
<tr>
<td>112</td>
<td>specific consent</td>
<td></td>
<td></td>
</tr>
<tr>
<td>113</td>
<td>broad consent</td>
<td></td>
<td></td>
</tr>
<tr>
<td>114</td>
<td>Research Ethics Committee</td>
<td></td>
<td></td>
</tr>
<tr>
<td>115</td>
<td>Approval</td>
<td></td>
<td></td>
</tr>
<tr>
<td>116</td>
<td>relevant regulatory authority</td>
<td></td>
<td></td>
</tr>
<tr>
<td>117</td>
<td>Data Access request</td>
<td></td>
<td></td>
</tr>
<tr>
<td>118</td>
<td>Material Transfer Agreement</td>
<td></td>
<td></td>
</tr>
<tr>
<td>119</td>
<td>Anonymised data</td>
<td></td>
<td></td>
</tr>
<tr>
<td>120</td>
<td>researchers' personal data</td>
<td></td>
<td></td>
</tr>
<tr>
<td>121</td>
<td>publication of personal data</td>
<td></td>
<td></td>
</tr>
<tr>
<td>122</td>
<td>donor</td>
<td></td>
<td></td>
</tr>
<tr>
<td>123</td>
<td>additional approval</td>
<td></td>
<td></td>
</tr>
<tr>
<td>124</td>
<td>cross-border data transfer</td>
<td></td>
<td></td>
</tr>
<tr>
<td>125</td>
<td>Law of exporting country</td>
<td></td>
<td></td>
</tr>
<tr>
<td>126</td>
<td>Transfer of biosample</td>
<td></td>
<td></td>
</tr>
<tr>
<td>127</td>
<td>ISBER 2008</td>
<td></td>
<td></td>
</tr>
<tr>
<td>128</td>
<td>OECD Guidelines 2009</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Security of Biosamples**

- Deletion of personal data
- Deletion of identifying data
- Renewal of consent
- Specific consent
- Broad consent
- Research Ethics Committee
- Approval
- Relevant regulatory authority
- Data Access request
- Material Transfer Agreement
- Anonymised data
- Researchers' personal data
- Publication of personal data
- Donor
- Additional approval
- Cross-border data transfer
- Law of exporting country
- Transfer of biosample
- ISBER 2008
- OECD Guidelines 2009