Tool for assessment of regulatory and ethical requirements

Contents

Tables

Figures

1 Executive Summary

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.

The online tool described here assists the user in the assessment of ethical and legal requirements connected to sharing sensitive data. The tool assesses the regulatory requirements established in BioMedBridges deliverable 5.1[1] and presents relevant information and contractual templates in a user-friendly way.

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[2]. 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

[1] http://www.biomedbridges.eu/deliverables/52
[2] www.biomedbridges.eu

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.

3 General Approach

3.1 Legal and ethical background

With respect to data sharing, a number of the research infrastructures involved in BioMedBridges 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 when making data or samples available, for example in data bridges.

3.2 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).

3.3 Features of the tool

The tool provides 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 making data (or samples) available, for example in data bridges, regarding the ethical and legal requirements. 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.

4 Knowledge Base - Requirements Matrix & Assessment

The ethical and legal requirements underlying the tool are captured in a matrix that serves as the knowledge base. In order to develop this matrix, the following three main tasks had to be performed:

1. Identify the dimensions of the matrix and organising all requirements (identified in deliverable 5.1) within the matrix (see section 4.1)
2. Workflow definition for assessment and guidance based on the matrix and requirements (see section 4.2)
3. Determine the tool output (i.e. recommendations, see section 4.3).

4.1 Matrix Structure

The matrix structures all 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 analysed regarding relevant and redundant/irrelevant information contained in them. Based on this analysis, the dimensions of the matrix were determined (in the following, the term “data bridge” should be read as “new context for the data”):

- 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 bride. 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 1.

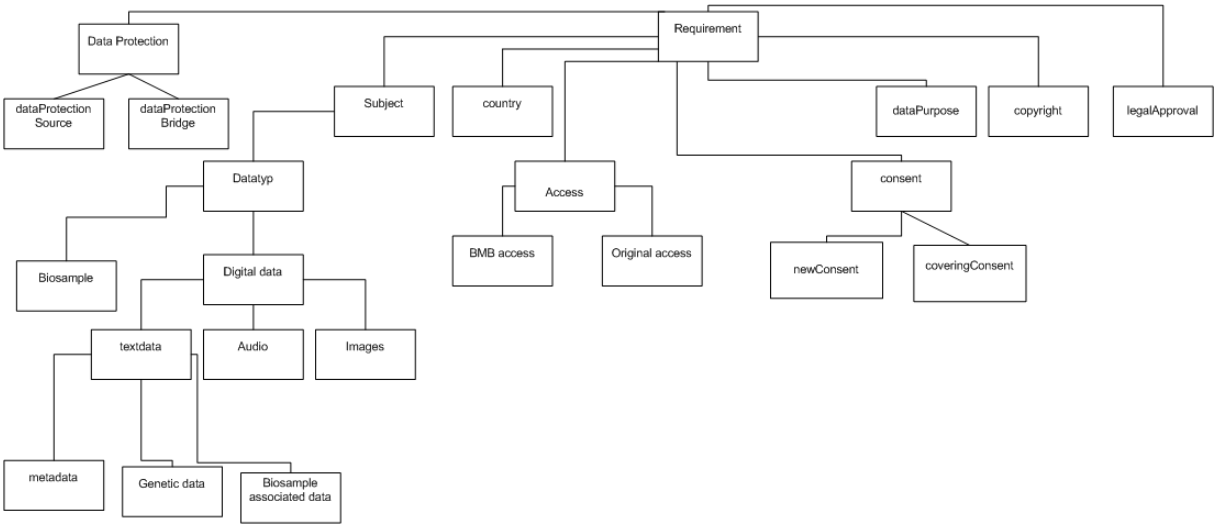

Figure 1 *All dimensions in relation to a requirement including the dependencies of the data type categories*

4.2 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 2).

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 1 *Example of the requirements matrix*

id	rqNr	subject	dataType	dataProtection Source	dataPurpose	legalAp proval	copyright	dataProtection Bridge	covering Consent	new Consent	bmb Access	original Access	country
1	DP1.1	non-human	metadata	pseudonym	any	any	any	anonym	any	any	any	any	eu
2	DP1.1	non-human	metadata	identifying	any	any	any	anonym	any	any	any	any	eu
3	DP1.2	non-human	metadata	pseudonym	any	any	any	anonym	any	any	any	any	eu
4	DP1.2	non-human	metadata	identifying	any	any	any	anonym	any	any	any	any	eu
5	DP1.3	non-human	metadata	pseudonym	any	no	any	pseudonym	no	yes	any	any	eu
6	DP1.3	non-human	metadata	identifying	any	no	any	pseudonym	no	yes	any	any	eu
7	DP1.3	non-human	metadata	pseudonym	any	no	any	identifying	no	yes	any	any	eu
8	DP1.3	non-human	metadata	identifying	any	no	any	identifying	no	yes	any	any	eu
9	DP1.3.0	non-human	metadata	identifying	any	no	any	identifying	no	no	any	any	eu
10	DP1.3.0	non-human	metadata	identifying	any	no	any	pseudonym	no	no	any	any	eu
11	DP1.3.0	non-human	metadata	pseudonym	any	no	any	identifying	no	no	any	any	eu
12	DP1.3.0	non-human	metadata	pseudonym	any	no	any	pseudonym	no	no	any	any	eu
13	DP1.4	human	metadata	any	any	any	any	any	any	any	any	any	eu
14	DP2.1	human	image	identifying	any	any	any	any	any	any	any	any	eu
15	DP2.2	human	image	identifying	any	any	any	any	any	any	any	any	eu
16	DP2.3	human	image	pseudonym	data linking	any	any	any	any	any	any	any	eu
17	DP2.3	human	image	identifying	data linking	any	any	any	any	any	any	any	eu
18	DP2.3	human	image	any	any	any	any	any	any	any	open	restricted	eu

4.3 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

data type =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 meta-data 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.

4.4 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:

title=*Attributevalue*(**parentID**= *pID*, **id**=*id,* **requirements list** = *yes/no*)

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.

5 Tool Development

5.1 Workflow: Perspective of the Researcher

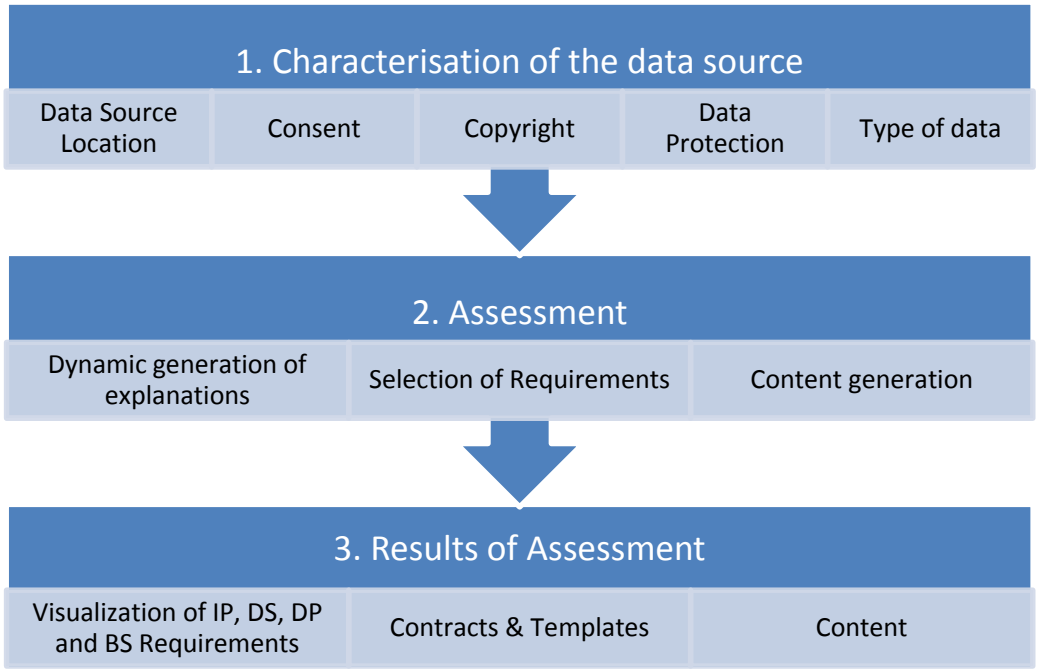

Figure 2 *Wizard (Questionnaire)*

On the basis of the questionnaire provided in Appendix 3 to this document in order to characterise a data source, the assessment tool selects requirements from the different requirement clusters. Each question relates to a column in the requirements matrix (see Figure 1) and the answers to these questions lead to a selection of one or more rows (requirements) from the requirements matrix. The results of the assessment consist of structured listing of requirements, explanations and related information (content), and contracts&templates. 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).

5.2 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.

5.3 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.

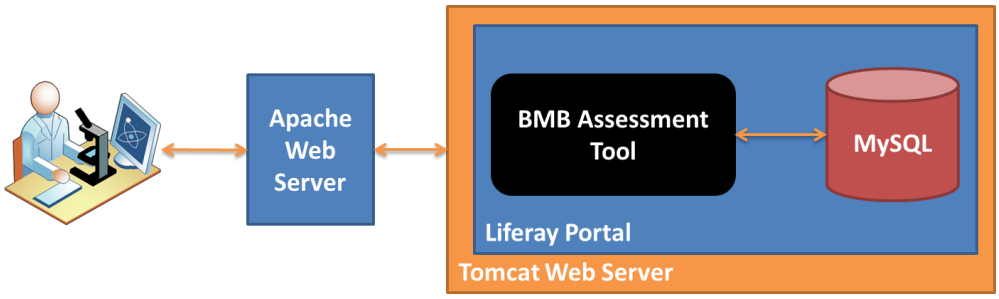

Figure 3 *High-Level Architecture Assessment Tool*

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.

5.4 Concepts and Data Base Structure

In Figure 4, 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

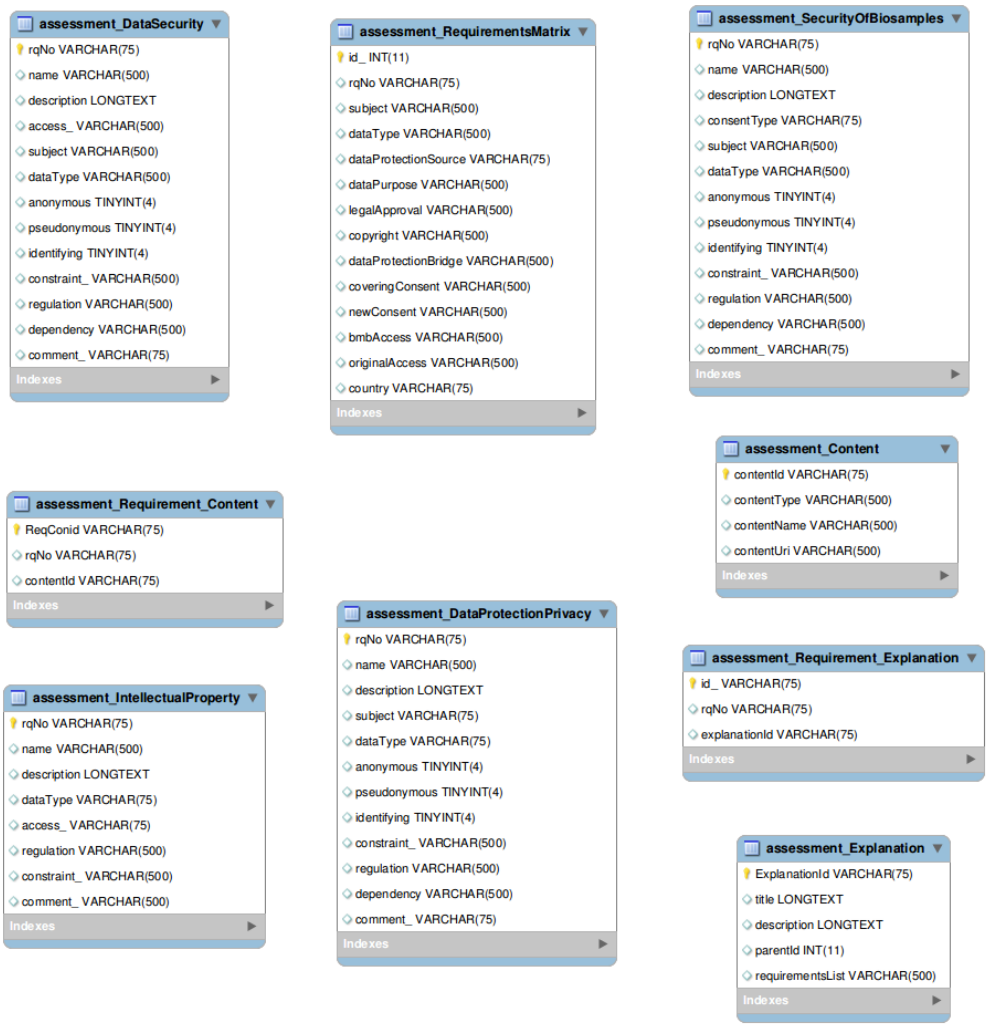

Figure 4 *BioMedBridges assessment tool Enhanced Entity Relationship Model*

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.

6 Templates & Contracts

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.

6.1 The BioMedBridges 'Ethical Governance Framework'

The bridge-building within the BioMedBridges project is primarily governed by the "BioMedBridges Ethical Governance Framework (EGF)[3]". 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

[3] BioMedBridges Ethical Governance Framework Version 1.1, April 2013

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

6.2 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 a bridge, which indicates in this context a platform for making data and samples available. 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 counterperfomance provisions. These fields are indicated in Figure 5 as greyscale blocks.

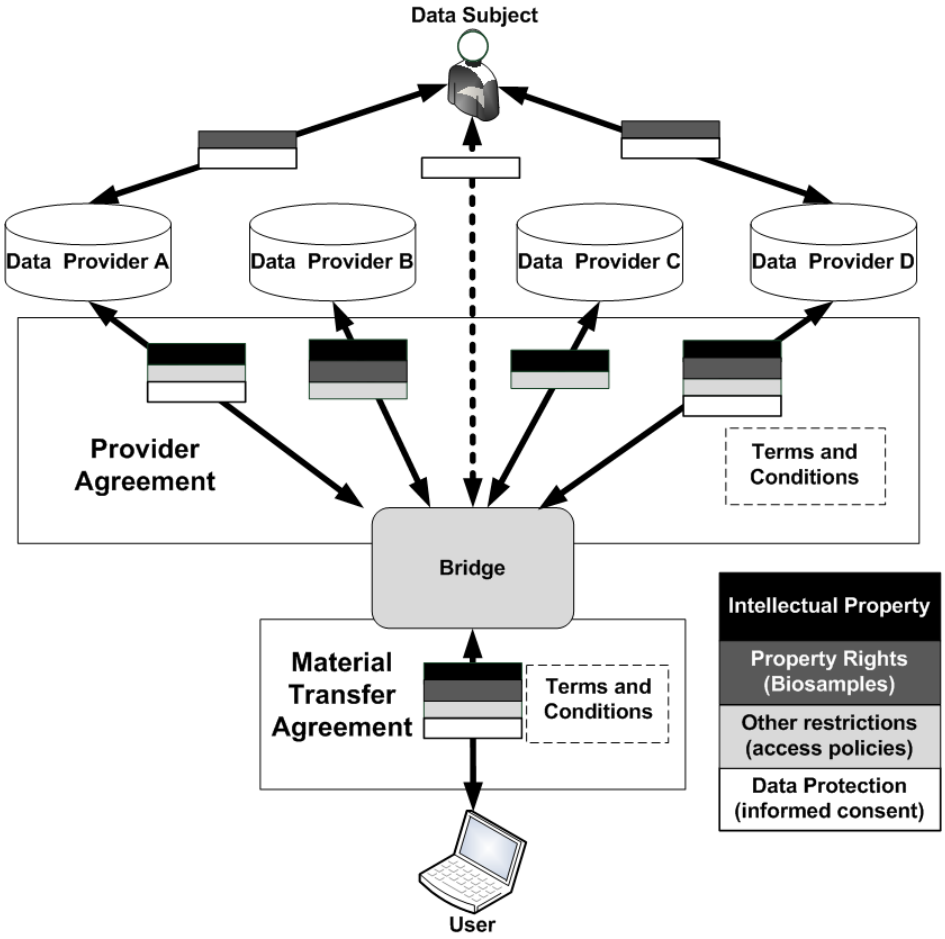

Figure 5 *Contractual Relations*

6.3 Informed Consent Templates

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 affordable[4]. 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

[4] 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[5]. 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[6], 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

[5] 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", http://www.ak-med-ethik-komm.de/aktuelles.html

[6] 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[7].

6.4 Data or sample 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[8]*",* 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

[7] See e.g. D.M. Gitter, "The Challenge of Achieving Open-Source Sharing of Biobank Data", in: Giovanni Pascuzzi, Umberto Izzo and Matteo Macilotti (Ed.): Comparative Issues in the Governance of Research Biobanks, 2013, chapter 10, 165 pp.

[8] 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).

6.5 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.

7 Testing performed

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.

7.1 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 6 shows a screenshot of the

spreadsheet that was used for testing. Full test scenarios and results are provided in the supplement.

	id	rqNr	subject	dataType	dataProtectionSource	dataPurpose	legalApproval	copyright	dataProtectionBridge
2	1	DP1.1	non-huma	metadata	pseudonym	any	any	any	anonym
3	2	DP1.1	non-huma	metadata	identifying	any	any	any	anonym
4	3	DP1.2	non-huma	metadata	pseudonym	any	any	any	anonym
5	4	DP1.2	non-huma	metadata	identifying	any	any	any	anonym
6	5	DP1.3	non-huma	metadata	pseudonym	any	no	any	pseudonym
7	6	DP1.3	non-huma	metadata	identifying	any	no	any	pseudonym
8	7	DP1.3	non-huma	metadata	pseudonym	any	no	any	identifying
9	8	DP1.3	non-huma	metadata	identifying	any	no	any	identifying
10	9	DP1.3.0	non-huma	metadata	identifying	any	no	any	identifying
11	10	DP1.3.0	non-huma	metadata	identifying	any	no	any	pseudonym
12	11	DP1.3.0	non-huma	metadata	pseudonym	any	no	any	identifying
13	12	DP1.3.0	non-huma	metadata	pseudonym	any	no	any	pseudonym
14	13	DP1.4	human	metadata	any	any	any	any	any
15	14	DP2.1	human	image	identifying	any	any	any	any

Figure 6 *Screenshot of the requirement matrix with filters*

7.2 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.

8 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). 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 (see Appendix 4).

However, there is potential for improvements and extensions. For example, a graphical wizard for entering new requirements into the tool is highly desirable; this would also allow customizing previously listed requirements. Further, it should be possible to specify more t, 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.

9 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