BIOPRESERVATION AND BIOBANKING
Volume 00, Number 00, 2017

DOI: 10.1089/bio.2016.0113

ORIGINAL ARTICLE

Conception and Implementation of an Austrian Biobank Directory Integration Framework

Philipp Hofer-Picout,[1,*] Horst Pichler,[2,*] Johann Eder,[2] Sabrina B. Neururer,[1] Heimo Müller,[3] Robert Reihs,[3] Petr Holub,[4,5] Thomas Insam,[6] and Georg Goebel[1]

Introduction: Sample collections and data are hosted within different biobanks at diverse institutions across Europe. Our data integration framework aims at incorporating data about sample collections from different biobanks into a common research infrastructure, facilitating researchers' abilities to obtain high-quality samples to conduct their research. The resulting information must be locally gathered and distributed to searchable higher level information biobank directories to maximize the visibility on the national and European levels. Therefore, biobanks and sample collections must be clearly described and unambiguously identified. We describe how to tackle the challenges of integrating biobank-related data between biobank directories using heterogeneous data schemas and different technical environments.
Methods: To establish a data exchange infrastructure between all biobank directories involved, we propose the following steps: (A) identification of core entities, terminology, and semantic relationships, (B) harmonization of heterogeneous data schemas of different Biobanking and Biomolecular Resources Research Infrastructure (BBMRI) directories, and (C) formulation of technical core principles for biobank data exchange between directories.
Results: (A) We identified the major core elements to describe biobanks in biobank directories. Since all directory data models were partially based on Minimum Information About BIobank Data Sharing (MIABIS) 2.0, the MIABIS 2.0 core model was used for compatibility. (B) Different projection scenarios were elaborated in collaboration with all BBMRI.at partners. A minimum set of mandatory and optional core entities and data items was defined for mapping across all directory levels. (C) Major core data exchange principles were formulated and data interfaces implemented by all biobank directories involved.
Discussion: We agreed on a MIABIS 2.0-based core set of harmonized biobank attributes and established a list of data exchange core principles for integrating biobank directories on different levels. This generic approach and the data exchange core principles proposed herein can also be applied in related tasks like integration and harmonization of biobank data on the individual sample and patient levels.

Keywords: BBMRI, biobank directory, data integration, EDI interface, MIABIS, RESTful

Introduction

To conduct (bio-)medical studies with significant results, researchers need to gather an appropriate number of samples and data, a challenging process often requiring cross-institutional collaboration, especially for rare diseases where individuals and samples are generally limited.[1] Consequently, a key question is where to get proper samples to perform a research project.[2] Unfortunately, researchers are frequently unaware of suitable samples outside their own institution or department—a problem which can be solved by the establishment of (inter-)nationally searchable biobank directories for harmonization and integration of biobank resources.[3]

Biobank directories summarize aggregated meta-level information and give a first insight into which kinds of sample collections and data (sample-related data or personal donor information comprising medical history and therapy

[1]Department of Medical Statistics, Informatics and Health Economics, Medical University of Innsbruck, Innsbruck, Austria.
[2]Department of Information and Communication Systems, University of Klagenfurt, Klagenfurt, Austria.
[3]Institute of Pathology, Medical University Graz, Graz, Austria.
[4]Institute of Computer Science, Masaryk University, Brno, Czech Republic.
[5]Biobanking and BioMolecular Resources Research Infrastructure (BBMRI-ERIC), Graz, Austria.
[6]Department of Obstetrics and Gynecology, Medical University of Innsbruck, Innsbruck, Austria.
*Both these authors contributed equally to this work.

records, metabolomics, proteomics, genomics data, or lifestyle data from questionnaires) are gathered and provided by different medical institutions in individual countries across Europe. Local biobank directories maintained inside a hospital would permit researchers to refer to sample collections outside their own departments.

The integration of existing biobanks into a common European research infrastructure is one key initiative of the Biobanking and Biomolecular Resources Research Infrastructure (BBMRI) project.[4] Several independent biobank directories on different regional levels were established over the past 2 years, such as the Biobank Registry of the Medical University of Innsbruck (MUI) (http://biobankregister.i-med.ac.at)[5] and the Austrian (http://catalog.bbmri.at) and European Biobank Directory (http://bbmri-eric.eu/bbmri-eric-directory-2.0).[6]

One important task of the BBMRI European Research Infrastructure Consortium (BBMRI-ERIC) is to harmonize available biobank data and distribute information from the local to the European level. Figure 1 shows a scenario where a collection owner at the MUI feeds collection data into the local directory of the Biobank Innsbruck, which is automatically aggregated and distributed to directories of higher regional levels by defined electronic data interchange (EDI) interfaces. In addition, each directory must provide a searchable Web portal where researchers and other interested persons are enabled to search for collections, samples, and contacts they might be interested in, for example, by specifying how many samples of a specific material type with a specific diagnosis are required.

The development of biobank directories is an ongoing parallel effort on all regional levels, developed in evolutionary steps and versions. Requirements are frequently added, changed, and completed whenever new features and problems arise from experiences gathered in previous development iterations or advances in research, which results in multiple, heterogeneous data schemas and technical implementations. To exchange data between such heterogeneous environments, we had to identify and solve several data interoperability problems.[7,8]

Schema integration and harmonization

Semantic and structural heterogeneity occur when biobank data are harmonized and integrated from different data schemas. Schema heterogeneity involves missing core concepts, naming differences, or incompatible structuring of semantically equal entities in multiple data schemas developed independently from each other.[9]

Despite harmonization attempts by the European research community,[10] ambiguities in the exact interpretation of, aggregation of, and the relationship between biobanks, sample collections, subcollections, samples, and aliquots still exist, often depending on national laws or other regulatory frameworks and specific use cases.

Some of the problems we encountered, when comparing the various databases and questions arising therefrom, were as follows: (1) what exactly is the difference between a collection and subcollection in a biobank? (2) Can one specific sample be part of several collections or subcollections? (3) A diagnosis could be described by a free text definition in one collection and by a coding standard, like the International Classification of Diseases (ICD) version 10 or Systematized Nomenclature of Medicine Clinical Terms (SNOMED-CT), in another one; (4) the body site is explicitly stated in one collection and included in an ICD-10 diagnosis code in another one; and (5) different study types require different types of samples (e.g., blood, serum, plasma, tissue, urine, and isolated biomolecules, such as DNA, RNA, or proteins), which are usually described and therefore structured in different ways.[2]

Challenges of EDI in a biobank directory integration framework

The establishment of a common infrastructure for EDI across distributed and disparate systems sharing (bio-) medical data leads to different ethical, legal, and technological challenges and needs, which have been addressed and dealt with by developers and researchers in related bioinformatics projects[11] and literature.[12] In this problem area, several challenges must be tackled, for example, (1) which data format and transport protocol can be used to exchange and synchronize data in heterogeneous technical environments? (2) How is one specific biobank or collection uniquely identified over all synchronized directories? (3) Who triggers updates and how are they propagated, and how can conflicts from interfering updates be avoided? (4) How can security, privacy, and authorization issues be dealt with?

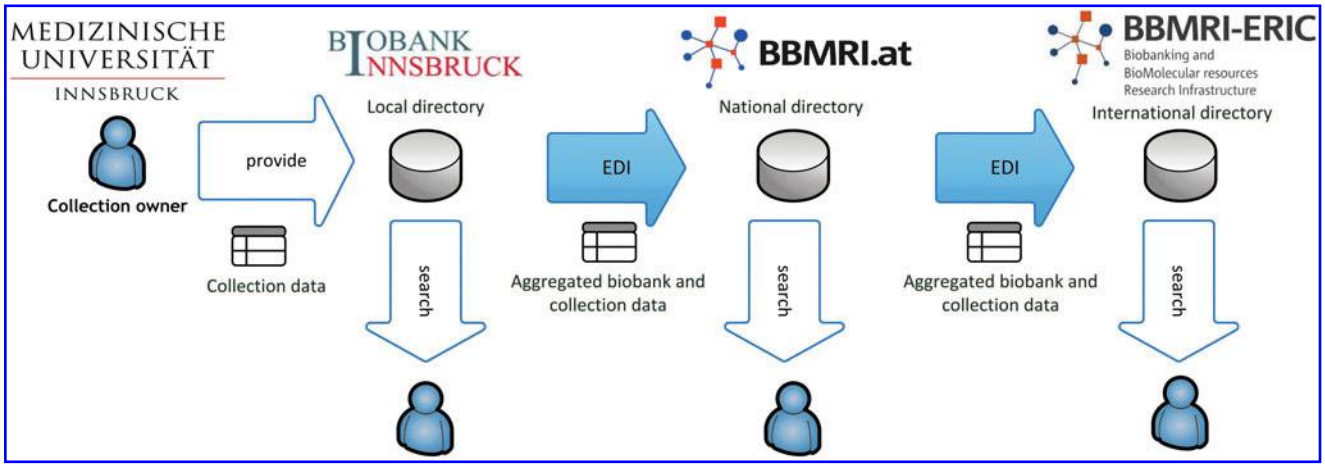

FIG. 1. Data integration scenario involving three biobank directories on local, national, and European level.

(5) How can interoperability be guaranteed when schemas and interfaces of partners change over time?

To meet the above-mentioned challenges, we proposed principles for and realized a data integration framework to facilitate the exchange of aggregated biobank information between synchronized biobank directories on institutional, national, and international levels.

Materials and Methods

To establish a framework for biobank directory data integration, the following strategies were elaborated to address each of the problem areas within BBMRI.at.

Identification of core entities, terminology, and semantic relationships

The initial versions of each directory were implemented mainly in parallel with, but independently from, one another. It was thus necessary to establish a joint understanding of the essential biobank concepts and meta-attributes used to describe different sample collection types. The aim was to achieve an agreement on how to efficiently integrate data from different collections into a common infrastructure.

First, an overall identification of the main concepts and relationships was carried out by an analysis and alignment of the respective data models from different biobank directories as well as biobank metadata provided by different BBMRI.at-partners. In the next step, *schema matching* was performed to cope with semantic equivalence and semantic relationships between objects in two different data schemas.[13] When matching two biobank directory data schemas, we were confronted with variations of entity and attribute nomenclature, data types and field values, and terminologies used, such as the granularity of aggregation of biobanks and sample collections. We relied on different schema matching heuristics proposed by Halevy.[14] Heuristics denote different matching decision criteria, for example, different attribute names used by two data schemas. For instance, the attributes ''BodySite'' and ''AnatomicalPartName'' in two different data schemas are synonyms both describing the anatomical origin of a sample, for example, blood or a specific organ.

Based on the different heuristics obtained from national directories and organizations, it was necessary to agree on a semantically equivalent representation of core entities and attributes represented on all levels. Standardized schema definition in all biobank directories is an ongoing work. Standardization efforts like the Minimum Information About BIobank Data Sharing (MIABIS)[15] were used as a starting point to reach a harmonization of biobank data by creating compatible collection registries on the institutional, national, and European levels to search, discover, and share common information between interested parties. During the last years, the MIABIS standard has been established as a standard within the BBMRI-ERIC directory.

Schema integration and harmonization

As the primary objective was to integrate national biobanks into a common European infrastructure, the above-mentioned steps were performed top-down starting with an agreement on a minimum set of exchange attributes between the European and national-level directories. Based on the identification of core concepts available in all different directories, it was necessary to agree on a common subset of deliverable MIABIS attributes for data exchange between local BBMRI.at-partners, the national and European biobank directory.

Schema mapping was used to transform an object from one schema to another[13] and deal with the heterogeneity of the existing data schemas. To map semantically equivalent attributes between different directories, it was necessary to cope with different representations of core entities, attributes, and data types, as well as different granularities of data representations. Data instances of semantically equivalent catalogue attributes were considered to find commonly used terminologies, selection lists, unit representations, sample counts, and disease or body site codes. Schema mapping was performed between the BBMRI-ERIC and BBMRI.at biobank directories by following a top-down strategy. Next, commonly deliverable attributes identified in all Austrian biobanks were mapped to the national BBMRI.at biobank directory.

Formulation of technical core principles for biobank data exchange

Apart from harmonization of heterogeneous directory data models, we were additionally confronted with the data interchange between different technical implementations. We therefore began to develop a foundation for data interchange between biobanks and our directories, which we started by defining several principles, derived from state-of-the art concepts usually applied in EDI scenarios.[11] A list of technical core principles for electronic data exchange between biobank directories was defined, taking into consideration the following aspects identified and derived from related information exchange challenges dealt with in existing (bio-)informatics literature: (1) privacy,[16] (2) biobank data exposition, (3) update responsibility,[11](4) data security,[16] (5) synchronization and (6) versioning conflicts,[17] (7) traceability of biobank resources/unique ID mapping across different directories,[18] and (8) data format and transport protocol.[19]

Results

Identification of core entities, terminology, and semantic relationships

We considered three directory data models containing information about biobanks and sample collections on different regional levels: (1) BBMRI-ERIC version 2, (2) BBMRI.at version 1, and (3) MUI Biobank directory version 1 (Table 1). The European biobank directory provides an overview about the national BBMRI nodes and contains meta-level attributes about biobanks and collections, including ID, biobank type, textual description, material types and diagnoses, contact information, data access policies, and the approximate number of samples or categories of data available. Information within the BBMRI-ERIC directory is organized into a general biobank-related information part and a second part containing a list of aggregated information about the collections associated therewith. The BBMRI.at directory contains data of four BBMRI.at-associated biobanks and some of their collections: the Biobank Graz, the

Table 1. Exemplar Results of the Identification of Core Entities, Such as Data Types and Semantics of Data Items Used to Represent Biospecimen Repositories for the Medical University of Innsbruck Biobank Directory (Extract)

Name	*Type*	*Mandatory*	*Definition*
Sample collection			
SampleCollectionID	String	Yes	Unique free text name of a sample collection.
Description	String	Yes	Free text description of a sample collection.
CategoryOfData	[Formatted string]	Yes	Types of data stored along with the sample collection.
PlannedSampledIndividuals	Int	No	Total number of donors involved in this sample collection.
CollectionStart	Date	No	Date when the collection of samples started.
CollectionEnd	Date	No	Date when the collection of samples finished.
…	…	…	…
Samples			
SamplesID	Autogenerated string	Yes	Main ID of a sample group within a sample collection.
MaterialTypeName	MaterialType	Yes	Material type, for example, Whole Blood, Serum, FFPE, Frozen tissue, and DNA.
StorageTemperature	[Formatted string]	Yes	The long-term storage temperature of all samples in a sample group after preparation.
AnatomicalPartName	String	No	Anatomical location/organ of the human body of all samples in the group.
Diagnosis	[DiagnosisType]	No	One or more diseases annotated to all samples in this group. Diagnosis strings are either specified as ICD-10 codes or free text. Specification of disease annotations is not mandatory, for example, for health status ''healthy.''
NumberofSamples	Int	No	Exact number of samples in the sample set, if this can be specified.
Dimensions	[Formatted string]	Yes	Approximate size of the sample set, for example, ''1 to 10'' or ''1000 to 10,000.''
…	…	…	…

Samples are subgroups of a specific material type and optional diagnosis, and belong to one sample collection. Attributes belonging to the MIABIS core dataset are displayed in a separate column where applicable.

ICD, International Classification of Diseases; MIABIS, Minimum Information About BIobank Data Sharing.

Biobank Innsbruck, the Medical University of Vienna (MedUni Wien) Biobank, and the Vienna Veterinary Biobank (VetBioBank). Collections within the BBMRI.at directory are aggregated into sample groups sharing one specific property, for example, same material type, body site, or disease. The MUI biobank directory stores aggregated data about different in-house sample collections distributed at different MUI departments. Like in the BBMRI.at directory, sample collections are further aggregated into homogeneous material type groups with optional body site or diagnosis specifications.

Members of BBMRI.at and BBMRI-ERIC increased their cooperation efforts with MIABIS. The MIABIS core model has been partially used, extended, or restructured by national- and European-level partners. As a result, all biobank directory data models considered herein are, at least, partially based on MIABIS 2.0. Free text or the ICD-10 definitions are partially provided within the BBRMI.at, MUI, and BBMRI-ERIC directory as annotations of disease-oriented collections.

Integration and harmonization of heterogeneous data schemas

We elaborated different aggregation scenarios in collaboration with the BBMRI.at partners providing biobank resources. We decided to use the MIABIS core entities[20] ''Biobanks,'' ''Sample Collections,'' and ''Samples'' as a basis for harmonization. The MIABIS entity ''Samples'' was used to describe subgroups within sample collections. They share at least one common property, for example, material type or diagnosis, but have no individual identifier. Mappings were performed between (1) the MUI biobank directory and the BBMRI.at directory and (2) the BBMRI.at and BBMRI-ERIC directory.

We found that sample collections within the BBMRI.at and MUI biobank directory are minimally comparable regarding their (1) collection description, (2) contact information, and (3) available data categories (Table 2). The (1) material type, the (2) disease status, as well as the (3) approximate number of samples, and (4) aliquots is the minimum information that is required for the aggregation of collection subgroups within the BBMRI.at directory (Table 3). Partners are encouraged, but not forced, to provide additional information to describe material groups, such as diagnoses specifications or anatomical sites, within their collections. ICD-10 and SNOMED-CT are provided specifying diseases in the BBMRI.at directory. Partners are free to decide which terminologies to use or whether they allow free text input fields. We agreed on terms and definitions of the most important biobank-related concepts, based upon standards (ISO/TC 276,

Table 2. Mapping of Sample Collections on BBMRI.at-Level

BBMRI.at directory				*MUI biobank directory*			
M/O	*Name*	*Type*	*MIABIS*	*M/O*	*Name*	*Type*	*MIABIS*
M	Id	String	Yes	M	SampleCollectionID	String	Yes
M	Name	String	Yes				
M	Contact	Contact	Yes	M	ContactID	ContactType	Yes
O	People	[Contact + Role]	No	O	JuristicPerson	JuristicPerson Type	Yes
O	Acronym	String	Yes				
O	Description	String	Yes	M	Description	String	Yes
O	CollectionStarted	Date	No	O	CollectionStart	Date	No
O	CollectionEnded	Date	No	O	CollectionEnd	Date	No
O	SurveyData	[SurveyDataType]	No				
O	DataCategories	[DataCategories Type]	Yes	M	CategoryOfData	[DataCategories Type]	Yes
O	NumberOfDonors	Int	Yes	O	PlannedSampledIndividuals	Int	Yes
				O	PlannedTotalIndividuals	Int	Yes
				O	LIMS	String	No

M/O denotes mandatory and optional (O) attributes. Attributes belonging to the MIABIS core dataset are displayed in a separate column where applicable. Lines in the table body are used to group semantically corresponding attributes between the two directories.
MUI, Medical University of Innsbruck.

Table 3. Mapping of Material Groups Within Sample Collections Between Austrian Biobank Directories, BBMRI.at, and Medical University of Innsbruck

BBMRI.at directory				*MUI biobank directory*			
M/O	*Name*	*Type*	*MIABIS*	*M/O*	*Name*	*Type*	*MIABIS*
M	Id	Autogenerated String	Yes	M	SamplesID	Autogenerated String	Yes
O	Description	String	No				
M	MaterialType	MaterialType	Yes	M	MaterialTypeName	Material Type	Yes
M	NumberOfSamples	Int	No	O	NumberOfSamples	Int	No
				M	Dimension	[Formatted string]	No
M	TotalNumber OfAliquots	Int	No	O	AliquotsOrBlocksPer Sample	Int	No
				O	StoredSampleAmount	Decimal	No
				O	StoredSampleAmount Unit	[Formatted string]	No
O	ICD10Diagnosis	[Formatted string]	No	O	Diagnosis	[DiagnosisType]	Yes
M	isDiseased	Boolean	Yes				
O	SNOMED-CT	[Formatted String]	No				
O	Orphanet	[Formatted String]	No				
O	StorageTemperature	StorageTemperature Type	Yes	O	StorageTemperature	[Formatted string]	Yes
O	BodySite	String	No	O	AnatomicalPartName	String	Yes
O	SPRECCode	[Formatted string]	No	O	SprecCode	String	No
				O	SampleHandling	String	No

M/O denotes mandatory (M) and optional (O) attributes. Attributes belonging to the MIABIS core dataset are displayed in a separate column, where applicable. Lines in the table body are used to group semantically corresponding attributes between the two directories.
SNOMED-CT, Systematized Nomenclature of Medicine Clinical Terms.

CEN/TC 140) by the Internal Organization for Standardization (ISO) and European Committee for Standardization (CEN), and complemented, where necessary, since some ambiguity still exists.

Technical core principles for biobank data exchange

We defined a list of technical core principles for data exchange:

(i) Protection of sensitive biobank data: Privacy protection demands increase as the data aggregation level decreases. When a biobank directory provides data on samples from a particular individual, appropriate IT-based methods for the protection of personal information must be implemented. Based on an evaluation of current approaches in the literature, we propose that sensitive data items (name, social security number, etc.) should be either sufficiently deidentified,[16] while preserving data usability,[21] or stored and linked in a separate registry.[22] Information about individual samples and donors in all biobank directories must be only accessible to authorized researchers and removed after the revocation of an informed consent.

(ii) Biobank data exposition: Distribution and access of information about biobanks, collections, and individual samples must be handled by the biobank proprietors and be in conformity with donor-informed consents. Biobank proprietors have the legal responsibility for samples and data and must ensure that only authorized and correct information are published; any modification or removal of data in a local biobank information system or directory (e.g., due to the revocation of an informed consent) should be immediately reflected within higher level biobank directories. Consequently, we favor a push-based data transmission strategy where it is the responsibility of the biobank owners to push data up to the next level to implement connectors to the directory interfaces.

(iii) Update responsibility: Only the node nearest to the data source is authorized to modify data. Nearest nodes refer to a biobank management system or biobank directory run by an institution or department owning a biobank, for example, the MUI biobank directory. Thus, national and international directories are allowed to receive, but not to change, data.

(iv) Security: The above three core principles require that write-access to biobank resources must be authenticated and authorized to prevent unauthorized data manipulation. For this, a role-based authorization is necessary to ensure that a partner cannot change a resource it does not own.

(v) Avoid partial synchronizations: Only complete sample collection data sets can be submitted to target systems for insertion or change.

(vi) Versioning of interfaces: Data exchange interfaces must be immune to changes in the underlying directory data structures. To achieve this, we propose to use interface versioning concepts. Each schema update may result in a new version of the interface being made available alongside older versions. To request a specific version of an interface, users can specify which data model version they wish to receive.

(vii) Identification of resources/ID mapping: Each partner institution, biobank, and sample collection must be distinctly identifiable in the Austrian and European biobank directory.

(viii) Data format and transport protocol: We evaluated diverse technologies and decided on a Representational State Transfer (REST)ful architecture[23] to exchange biobank and sample collection data in JSON format (JavaScript Object Notation; www.json.org). The JSON dataset, which is used for data exchange between the National and the MUI biobank directory, comprises three basic resources, namely ''Biobank,'' ''Sample Collection,'' and ''Samples,'' with all mandatory and optional meta-attributes according to the data exchange core model. Exemplar JSON documents describing biobanks, collection lists, and individual collections with subgroups can be retrieved by the following URLs:

- http://catalog.bbmri.at/directory/biobanks/MUI the Biobank Innsbruck
- http://catalog.bbmri.at/directory/biobanks/MUI/collections all BB Innsbruck Collections
- http://catalog.bbmri.at/directory/biobanks/MUI/collections/45 BB Innsbruck Collection with id 45

As there are different Web frameworks and technologies used in the registries (PHP, Java), the REST endpoints were implemented individually with suitable plugins and software libraries, widely available for most programming languages and platforms.

In the following, we describe the scenario of data exchange between the MUI biobank directory, the BBMRI.at directory, and the BBMRI-ERIC directory:

Local data capturing at MUI. Currently, additions and changes to data therein are provided manually by a Web-based user interface. Specific data interchange connectors between proprietary software of each local collection and the MUI biobank directory are going to be implemented in the upcoming versions.

Sending data from MUI to BBMRI.at. Data stored in the MUI biobank directory are periodically sent to the REST-interface of the BBMRI.at directory for insertion or update (Fig. 2). The communication is HTTPS secured and inserting POST- and PUT-operations requires authentication and authorization by username and password. The GET interfaces to request information about biobanks and sample collections of the BBMRI.at directory are currently open for everybody as the aggregation of sample information contains no privacy-breaching data that may lead to individuals. Uniform Resource Identifiers (URIs) are used to identify, request, and manipulate biobanks and collections in the parent directory, such as BBMRI.at. Each biobank or collection is assigned an own unique URI, for example, http://catalog.bbmri.at/directory/biobanks/MUI identifies the MUI biobank and http://catalog.bbmri.at/directory/biobanks/MUI/collections/75 identifies the Human Genome Collection in Innsbruck.

Sending data from BBMRI.at to BBMRI-ERIC. Currently, data exchange between the BBMRI.at directory and BBMRI-ERIC is handled by Lightweight Directory Access Protocol (LDAP) data records in LDAP Data Interchange Format (LDIF) containing aggregated information about Austrian biobanks and sample collections. LDAP data records are sent and inserted (overwrite if exists) into the ERIC directory through file transfer. Quite recently, BBMRI-ERIC

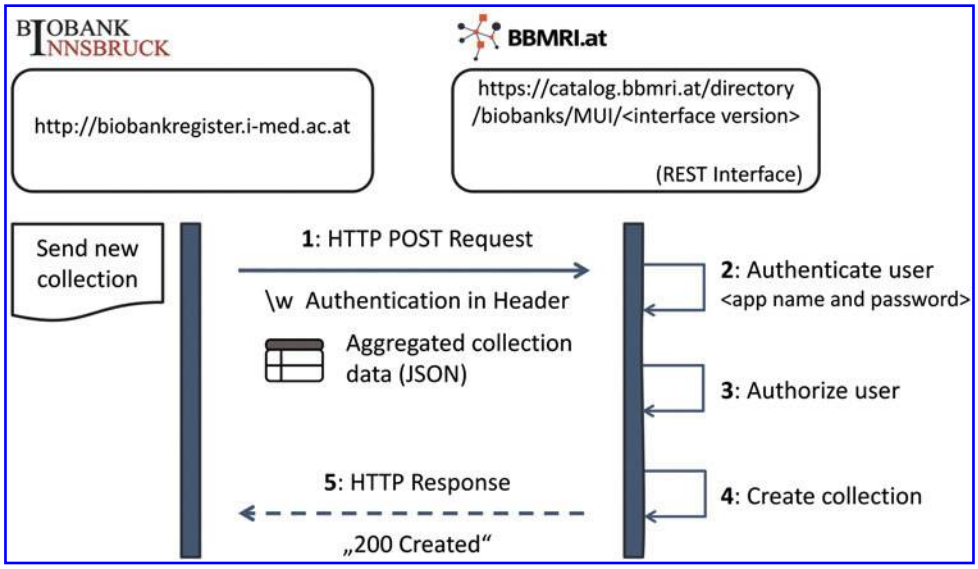

FIG. 2. Transmission of a new sample collection (POST request) from the MUI biobank directory to BBMRI.at directory using the Representational State Transfer protocol. MUI, Medical University of Innsbruck.

published their Interface Version 2.5 based on a MOLGENIS[24] directory, which features a REST-based approach to send and receive data about biobanks and collections. The corresponding BBMRI.at connectors are currently implemented.

Discussion

We established a core set of biobank attributes and formulated essential data exchange principles for sharing aggregate biobank data within an (inter-) national environment of biobank directories. Such a framework facilitates biobank custodians and researchers to integrate and find suitable sample collections from several biobank institutions within searchable (inter-) national biobank directories. Consequently, various Austrian biobank structures and directory data models were analyzed to obtain the major core entities, terminologies, and semantic relationships. We agreed on standard terminologies and a minimum set of core data items that can be provided by all BBMRI.at partner biobanks, such as the MUI biobank directory. As all data models were partially based on MIABIS, the same has been selected as a core data model to align with. However, MIABIS does not provide any personal- and sample-level information, allowing deeper inquiries and thus reidentification of individuals. Information on individual samples and participants, which would allow deeper inquiries or the reidentification of individuals, are, among others, most likely going to be included in upcoming versions.

The formulation of core principles serves as a foundation for EDI between local biobanks and (inter-)national directories. The assignment of unique identifiers to each biobank or sample collection is tackled by the national BBMRI.at directory using URIs allowing reidentification of a biobank or sample collection in subsequent interface calls. We propose to allow only institutions or departments owing biobank and sample collection to add, modify, and distribute data within their local biobank directory or information system, which has the following advantages: first, there is no need for synchronization mechanisms to solve conflicts between interfering updates from different levels. Second, allowing changes on any level would result in the necessity to implement conflict resolution mechanisms and, most certainly, confusion. Appropriate authentication and authorization mechanisms were implemented to ensure that a partner cannot change a resource it does not own. We agreed on using only atomic entities for data exchange, that is, whole biobank or sample collection data sets. This strategy simplified the architecture and transfer protocol tremendously, as we did not have to deal with partial updates and there was no need for a versioning mechanism to keep track of changes within biobank resources over time. Interface versioning allowed coping with version changes in the underlying data models. However, the interface owner must ensure that data sent to older interface versions are correctly mapped to the new schema. We decided to use REST as it is supported by a wide range of technical frameworks and programming languages and is the most frequently used throughout the BBMRI-ERIC community.[4] Several widely used biobank information systems[25] provide REST-based Web services. Furthermore, the Health Level 7 (HL7) organization proposes the use of REST in their Fast Healthcare Interoperability Resources (FHIR) standard, a next generation standards framework for harmonizing Electronic Health Records (EHR).[26] HL7 FHIR defines a set of predefined resources (e.g., ''Patients,'' ''Specimen,'' or ''BodySite''), which are machine-readable data formats and elements for electronic health information exchange. In the future, REST interfaces and resources provided by FHIR could facilitate sending and mapping clinical and specimen data from healthcare IT systems to MIABIS-based local, national, and European biobank directories, as they all rely on the same technical standard and similar core data items.

As can be seen from our technical principles, developing, running, and maintaining the required infrastructure are resource intensive. In the long run, these tasks should be performed by a designated maintenance team of IT specialists and biobank administrators who will be responsible for long-term servicing and further development. Although within BBMRI.at these tasks are still accomplished by BBMRI.at research staff members, BBMRI-ERIC has

already set up the Common Service IT to implement and maintain core internal and external services.[27]

The evolution of data schemas and interfaces by versions is currently driven by BBMRI-ERIC and then distributed in a top-down manner to national nodes and regional biobanks. Fields are ''optional,'' ''recommended,'' or ''required.'' Therefore, each new interface version aims at increasing the data level and quality of described biobanks, collections, and samples. For example, a field like ''Diagnosis'' is most likely to be promoted from ''recommended'' to ''required'' in one of the next versions, as it is frequently required in directory search queries.

Currently, data input at the MUI biobank directory or within local biobanks is mostly performed manually. Since manual upgrades of changes of sample collections and material groups are performed only in certain time intervals, local changes within material groups of sample collections, like the number of samples of a material group within a specific sample collection of a biobank, are frequently, not immediately, reflected within higher level directories. For the MUI biobank directory, such manual updates will be replaced by an automated data aggregation from a hospital-wide biobank information system in the future. Efforts in collecting and integrating biobanks and sample collections hosted at different Austrian institutions within national and European biobank directories are still ongoing.

The practical usability of such a data integration framework for gathering appropriate samples in research projects and medical studies becomes more apparent with an increasing number of national and European sample collections, which can be found in national or international biobank directories for respective research purposes. For the future, one major challenge will be to raise awareness of existing directories and encourage biobank custodians to share biobank resources in this context.

The harmonization approach as well as the data exchange core principles proposed in this article can be applied in other fields, such as the integration and harmonization of biobank data on individual samples and at the patient level.

Acknowledgment

This work was supported by the Austrian Biobanking and BioMolecular Resources Research Infrastructure (BBMRI.at) funded by the Austrian Federal Ministry of Science, Research and Economy (BMWFW GZ 10.470/0016-II/3/2013).

Author Disclosure Statement

No conflicting financial interests exist.

References

1. Mora M, Angelini C, Bignami F, et al. The EuroBioBank Network: 10 Years of hands-on experience of collaborative, transnational biobanking for rare diseases. Eur J Hum Genet 2015;23:1116–1123.
2. Müller H, Reihs R, Zatloukal K, et al. State-of-the-art and future challenges in the integration of biobank catalogues. In: Holzinger A, Röcker C, Ziefle M (eds). *Smart Health Open Problems and Future Challenges.* Cham, Switzerland: Springer International Publishing; 2015; 8700:261–273.
3. Chabannon C, Honstettre A, Daufresne L-M, et al. Publication of biological samples collections catalogues by tumor banks. Bull Cancer 2010;97:181–189.
4. Holub P, Litton J, et al. RDA data fabric IG (DFIG): BBMRI-ERIC IT. Zenodo, 2015. Available at: https://zenodo.org/record/51593#.WNzxHWdAolo.
5. Hofer P, Fiegl H, Angerer J, et al. A concept of a MIABIS based register of biosample collections at the Medical University of Innsbruck. Stud Health Technol Inform 2014; 205:293–297.
6. Mayrhofer MT, Holub P, Wutte A, et al. BBMRI-ERIC: The novel gateway to biobanks. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2016;59:379–384.
7. Kormeier B. Data warehouses in bioinformatics In: Chen M, Hofestädt R (eds). *Approaches in Integrative Bioinformatics*. Heidelberg: Springer Berlin; 2014: 113–114.
8. Embley DW, Thalheim B. *Handbook of Conceptual Modeling: Theory, Practice, and Research Challenges.* Heidelberg: Springer Science & Business Media Berlin; 2012:441–442.
9. Eder J, Dabringer C, Schicho M, Stark K. Information systems for federated biobanks. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); 2009: 156–190.
10. Quinlan PR, Mistry G, Bullbeck H, et al. A Data standard for sourcing fit-for-purpose biological samples in an integrated virtual network of biobanks. Biopreserv Biobank 2014;12:184–191.
11. Data exchange principles. Available at: https://nbn.org.uk/national-biodiversity-network/archive-information/data-exchange-principles/ (accessed January 16, 2017).
12. Mascalzoni D, Dove ES, Rubinstein Y, et al. International charter of principles for sharing bio-specimens and data. Eur J Hum Genet 2005;23:721–728.
13. Doan A, Halevy A, Ives Z. Schema matching and mapping In: Doan A (ed). *Principles of Data Integration*. Amsterdam: Elsevier; 2012: 121–160.
14. Halevy AY. Why your data won't mix. Queue-Semi-Structured Data 2005;3:50–58.
15. Norlin L, Fransson MN, Eriksson M, et al. A minimum data set for sharing biobank samples, information, and data: MIABIS. Biopreserv Biobank 2012;10:343–348.
16. Lablans M, Borg A, Ueckert F. A RESTful interface to pseudonymization services in modern web applications. BMC Med Inform 2015;15:1–10.
17. Varanasi B, Belida S. Versioning, paging, and sorting. In: Varanasi B, Belida S (eds). *Spring REST.* New York: Apress; 2015: 105–119.
18. Bravo E, Calzolari A, De Castro P, et al. Quantifying the use of bioresources for promoting their sharing in scientific research. BMC Med 2015;13:1–12.
19. Stockinger H, Attwood T, Chohan SH. Experience using web services for biological sequence analysis. Brief Bioinf 2008;9:493–505.
20. Merino-Martinez R, Norlin L, van Enckevort D, et al. Toward global biobank integration by implementation of the minimum information about BIobank Data Sharing (MIABIS 2.0 Core). Biopreserv Biobank 2016;14:298–306.
21. Khokhara RH, Chen R, Fung BCM, et al. Quantifying the costs and benefits of privacy-preserving health data publishing. J Biomed Inform 2014;50:107–121.

22. Suomen Biopankit: Data protection. Available at: www.biopankki.fi/en/data-protection (accessed June 5, 2017).
23. Guinard D, Ion I, Mayer S. In search of an internet of things service architecture: REST or WS-*? A developers' perspective. In: Puiatti, A, Gu T (eds). *Mobile and Ubiquitous Systems: Computing, Networking, and Services*. Heidelberg: Springer Berlin; 2012;104:326–337.
24. Pang C, van Enckevort D, de Haan M. MOLGENIS/connect: A system for semi-automatic integration of heterogeneous phenotype data with applications in biobanks. Bioinformatics 2016;32:2176–2183.
25. Butters OW, Issa S, Lusted J, et al. The Biomedical Research Infrastructure Software as a Service Kit (BRISSKit): Technical description [version 1; referees: 2 approved with reservations] F1000Research. 2016; 1905.
26. Lòpez JD, Moreno LG, Anzola D, et al. Standardization of clinical documents through HL7-FHIR for Colombia. Int J Comput Sci Inf Technol 2016;8:15–27.
27. Holub P. BBMRI-ERIC common service IT. Available at: www.bbmri-eric.eu/BBMRI-ERIC/common-service-it/ (accessed June 28, 2016).

Address correspondence to:
Philipp Hofer-Picout, MSc
Department of Medical Statistics, Informatics and Health Economics
Medical University of Innsbruck
Schöpfstraße 41/1
A-6020 Innsbruck
Austria
E-mail: philipp.hofer@i-med.ac.at