

University of Groningen

## **Toward Global Biobank Integration by Implementation of the Minimum Information About Biobank Data Sharing (MIABIS 2.0 Core)**

Merino-Martinez, Roxana; Norlin, Loreana; van Enckevort, David; Anton, Gabriele; Schuffenhauer, Simone; Silander, Kaisa; Mook, Linda; Holub, Petr; Bild, Raffael; Swertz, Morris

*Published in:*

Biopreservation and biobanking

*DOI:*

[10.1089/bio.2015.0070](https://doi.org/10.1089/bio.2015.0070)

**IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.**

*Document Version*

Publisher's PDF, also known as Version of record

*Publication date:*

2016

[Link to publication in University of Groningen/UMCG research database](#)

*Citation for published version (APA):*

Merino-Martinez, R., Norlin, L., van Enckevort, D., Anton, G., Schuffenhauer, S., Silander, K., Mook, L., Holub, P., Bild, R., Swertz, M., & Litton, J-E. (2016). Toward Global Biobank Integration by Implementation of the Minimum Information About Biobank Data Sharing (MIABIS 2.0 Core). *Biopreservation and biobanking*, 14(4), 298-306. <https://doi.org/10.1089/bio.2015.0070>

### **Copyright**

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

### **Take-down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

# Toward Global Biobank Integration by Implementation of the Minimum Information About Biobank Data Sharing (MIABIS 2.0 Core)

Roxana Merino-Martinez,<sup>1</sup> Loreana Norlin,<sup>1</sup> David van Enckevort,<sup>2</sup> Gabriele Anton,<sup>3</sup> Simone Schuffenhauer,<sup>4</sup> Kaisa Silander,<sup>5</sup> Linda Mook,<sup>2</sup> Petr Holub,<sup>6</sup> Raffael Bild,<sup>7</sup> Morris Swertz,<sup>2</sup> and Jan-Eric Litton<sup>6</sup>

Biobanks are the biological back end of data-driven medicine, but lack standards and generic solutions for interoperability and information harmonization. The move toward a global information infrastructure for biobanking demands semantic interoperability through harmonized services and common ontologies. To tackle this issue, the Minimum Information About Biobank data Sharing (MIABIS) was developed in 2012 by the Biobanking and BioMolecular Resources Research Infrastructure of Sweden (BBMRI.se). The wide acceptance of the first version of MIABIS encouraged evolving it to a more structured and descriptive standard. In 2013 a working group was formed under the largest infrastructure for health in Europe, Biobanking and BioMolecular Resources Research Infrastructure (BBMRI-ERIC), with the remit to continue the development of MIABIS (version 2.0) through a multicountry governance process. MIABIS 2.0 Core has been developed with 22 attributes describing Biobanks, Sample Collections, and Studies according to a modular structure that makes it easier to adhere to and to extend the standard. This integration standard will make a great contribution to the discovery and exploitation of biobank resources and lead to a wider and more efficient use of valuable bioresources, thereby speeding up the research on human diseases. Many within the European Union have accepted MIABIS 2.0 Core as the “de facto” biobank information standard.

## Introduction

STORED HUMAN BIOLOGICAL MATERIAL and associated data suitable for large-scale molecular analyses are the most critical resources for translation of advances in molecular biology and medical technology into improved human health. To simplify the exchange of data between biobanks and biomedical research, the guideline on Minimum Information About Biobank data Sharing (MIABIS) was created in 2012 based on the work that had been done during the Biobanking and BioMolecular Resources Research Infrastructure (BBMRI) preparatory phase.<sup>1</sup> The aim was to standardize data elements used to describe any bio-

bank in the world, research on samples, and the associated data.<sup>2</sup> It proposed a simple structure with two main components: “Biobank” and “Sample Collection/Study” with a sum of 52 attributes. The sample donors and samples were represented as aggregated and meta-data using terms that were already part of existing ontologies and standard vocabularies in the biomedical field.

MIABIS was ontologized (OMIABIS) by Brochhausen et al.,<sup>3</sup> becoming the first ever biobank ontology representing the administrative entities in biobanking and their relations.

MIABIS 2.0 is a collection of components with associated attributes representing relevant concepts from biobanking

<sup>1</sup>BBMRI.se, Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden.

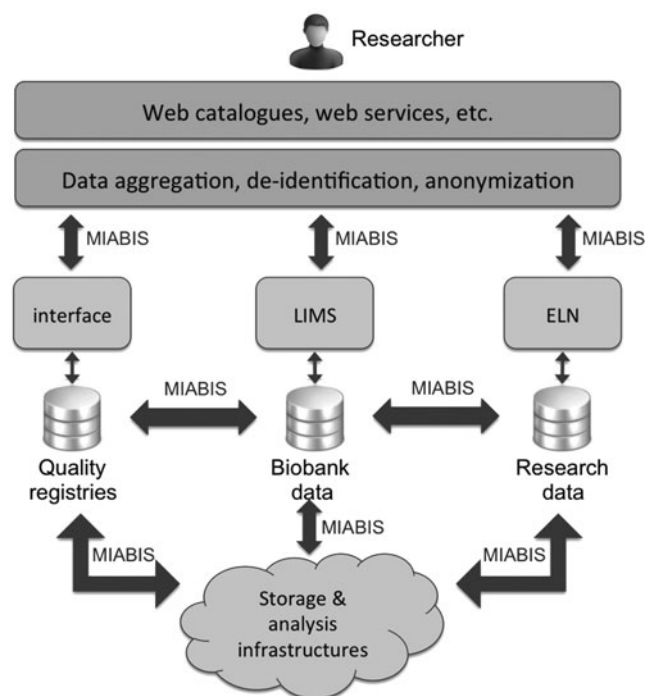
<sup>2</sup>BBMRI-NL, Department of Genetics, Genomics Coordination Center, University Medical Center Groningen, University of Groningen, Groningen, the Netherlands.

Institutes of <sup>3</sup>Epidemiology II and <sup>4</sup>Human Genetics, Helmholtz Center Munich, Munich, Germany.

<sup>5</sup>BBMRI.fi, National Institute for Health and Welfare, Genomics and Biomarkers Unit, Helsinki, Finland.

<sup>6</sup>BBMRI-ERIC, Graz, Austria.

<sup>7</sup>IMSE, TUM, Munich, Germany.



**FIG. 1.** Minimum Information About Biobank data Sharing (MIABIS) is a recommendation about what information should be stored in biobank and research information systems to be able to easily exchange information and data. The aim is to facilitate the reuse of bioresources and associated data by harmonizing the information in the most relevant components of the biomedical research ecosystem. LIMS, Laboratory Information Management System; ELN, Electronic Lab Notebook.

and biomedical research and can be used to integrate the most relevant building blocks of the biomedical research ecosystem (Fig. 1). This article describes MIABIS 2.0 Core, an update of MIABIS jointly prepared by a working group composed of representatives of BBMRI-ERIC<sup>4</sup> member countries. In this new version, certain attributes have been newly defined and structured and the component *Sample Collection/Study* has been separated into two distinct components that, together with the *Biobank* component, form the Core module of MIABIS. Guidelines on how to implement this standard in informatics systems and examples of use cases are also provided.

## Materials and Methods

### Governance

An international MIABIS Working Group was established under the umbrella of BBMRI-ERIC with representatives from national BBMRI nodes, with the aim to continue the development of MIABIS.

MIABIS is maintained through an extensive procedure where new components and attributes can be proposed and voted on by each country. We have been following the MoSCoW<sup>5</sup> prioritization to reach consensus on the MIABIS 2.0 components and attributes.

Each participating member country was asked to provide one or more contacts to participate in several

MIABIS review sessions, most of which were done through teleconference and web meetings. During and in between the meetings, each of the MIABIS data elements was reviewed by all and changes were proposed and subsequently accepted or rejected. We adhered to a consensus model for decision-making with the following characteristics:

- (1) Participatory and Collaborative: All group members were included and encouraged to participate and collaborate on the basis of their availability, expert knowledge, and their countries' special needs or area of interest.
- (2) Agreement Seeking: Seeking widespread or full agreement where possible. We have committed ourselves to the goal of generating as much agreement as possible, but "agree to disagree" on items and on a case-by-case basis discussed if attributes would be excluded from MIABIS for the short term or that inclusion and agreement were essential.
- (3) Process Oriented: We have carefully considered the process of making decisions, not just the result but also with a view to maintaining MIABIS in the future in a similar way. All contributions were welcome. The facilitator and the structure of the discussions specifically discouraged power leverage, adversarial positioning, and other group manipulation tactics.

### Changes from MIABIS 1.0

MIABIS was implemented in several use cases as biobank registers and catalogs and also as part of data models for biobank and research management systems. These experiences provided inputs to the MIABIS team regarding limitations and improvements to facilitate its easy adoption. The main suggestions were related to the large number of attributes, the need of a more structured design, and the incorporation of components that could cover other relevant actors in the biomedical research process.

The major update of MIABIS 2.0 from the previous MIABIS proposal<sup>2</sup> has been the introduction of a more modular and generic architecture. The "Biobank" component has been retained, while the "Sample Collection/Study" component was separated into two components: "Sample Collection" and "Study." These three components make up the MIABIS 2.0 Core. The data element list consists of a total of 22 attributes divided over these core components, presented in Tables 1–3 respectively. To ease adoption, none of the 22 attributes is mandatory or required because MIABIS is only a guideline highly recommended to be used when applicable.

The main steps for creating MIABIS 2.0 included the following:

- (1) The terms Biobank, Sample Collection, and Study have been defined.
- (2) The concept of "Sample Collection/Study" was separated into two distinct components: "Sample Collection" and "Study" as one study can be conducted using samples from more than one sample collection. A

“Sample Collection” is not necessarily associated with a specific “Study.”

- (3) MIABIS 2.0 has been structured in a more effective way. The definition of lists and structured data facilitates its implementation in informatics systems. For instance, the *Contact Information* is a structured data element that can be reused by several components in MIABIS.
- (4) The attribute *Biobank Type* was removed from the component “Biobank” because this concept is not consistently used by the biobank community
- (5) The attribute *Hosted Studies* was removed from the component “Biobank” because it reflects a relationship, not a real value.
- (6) The attributes *Average Age*, *Current Sampled Individuals*, *Current Total Individuals* were removed from the “Sample Collection” and “Study” components because they represent calculated values that can be derived from other components or data elements such as participant and sample.
- (7) The attributes comorbidity and medical records were removed from the “Sample Collection” and “Study” components because they are only associated to specific sample collections or studies.
- (8) *Type of collection* was renamed to *Collection Type* in the “Sample Collection” component and *Study Design* in the “Study” component.
- (9) We have made a modular structure of MIABIS, allowing additional components to be added, making MIABIS flexible and applicable for a variety of use cases and domains of biobanking and research studies. The proposed additional components currently being drafted in MIABIS 2.0 are “Biological Experiment,” “Participant,” “Rare Disease,” “Sample,” and “Quality.”
- (10) All definitions and value sets for the attributes have been reviewed and fine-tuned.

## Results

### *MIABIS 2.0 Core: components and attributes*

Before defining the minimum attributes of MIABIS 2.0 Core, it is important to clarify and define its three main components: “Biobank,” “Sample Collection,” and “Study.” Unanimous and unequivocal definitions for these terms are hard to find in literature, as described by Fransson et al.<sup>6</sup> Thus, we introduce the following general definitions aiming to facilitate the representation of MIABIS 2.0 Core as part of informatics data models:

Definition 1: *Biobank* represents an organization or an organizational unit that stores samples and data related to the samples.

Remarks: In MIABIS 2.0 Core, biobanks do not contain samples directly, but they are hosts of sample collections. On the level of biobanks, only attributes related to the organizational aspect of the biobanks are represented.

Definition 2: *Sample Collection* represents a set of samples with at least one common characteristic.

Definition 3: *Study* represents a set of samples brought together in the context of a research study.

Remarks: A study can combine samples from several sample collections and from several biobanks. One sample can participate in multiple studies.

Tables 1–6 provide the attribute list for each MIABIS 2.0 Core component.

The attributes *Contact Information* and *Disease* are classified as structured data as part of the modular approach of MIABIS. These attributes are reused by several components, for example, *Contact Information* is relevant to all three of the core components. The structures of *Contact Information* and *Disease* are presented in Tables 4–6. *Contact Information* consists of general “Contact Information,” for example, a biobank and “Research Information” to describe, for example, a Principal Investigator (PI).

TABLE 1. ATTRIBUTE DEFINITIONS OF THE COMPONENT “BIOBANK”

<i>Attribute code</i>	<i>Attribute name</i>	<i>Allowed values</i>	<i>Description</i>
MIABIS-2.0-01	ID	Text	Textual string of letters starting with the country code (according to standard ISO3166 alpha2) followed by underscore “_” and post-fixed by a biobank ID or name specified by its legal entity (nation specific)
MIABIS-2.0-02	Acronym	Text	Textual string of short name in use for the biobank. If applicable
MIABIS-2.0-03	Name	Text	Textual string of letters denoting the name of the biobank in English
MIABIS-2.0-04	URL	Text	Textual string of letters with the complete http-address for the biobank
MIABIS-2.0-05	Juristic Person	Text	Textual string of letters denoting the legal entity, for example, a university, county council, or other host organization for the biobank
MIABIS-2.0-06	Country	Text	ISO-standard (3166 alpha2), two-letter code for the country of the biobank
MIABIS-2.0-07	Contact Information	Structured data	Contact information for the contact person of the biobank MIABIS-2.0-07 ( <a href="https://github.com/MIABIS/miabis/wiki/Structured-data-and-lists#contact-information">https://github.com/MIABIS/miabis/wiki/Structured-data-and-lists#contact-information</a> )
MIABIS-2.0-08	Description	Text	Textual string of letters with a description about the biobank in English.

TABLE 2. ATTRIBUTE DEFINITIONS OF THE COMPONENT “SAMPLE COLLECTION”

<i>Attribute code</i>	<i>Attribute name</i>	<i>Allowed values</i>	<i>Description</i>
MIABIS-2.0-01	ID	Text	Sample Collection ID that also links the sample collection to the hosting biobank or study
MIABIS-2.0-02	Acronym	Text	Short name in use for the sample collection
MIABIS-2.0-03	Name	Text	The name of the sample collection in English
MIABIS-2.0-08	Description	Text	A description of the sample collection. Recommendation maximum 2000 characters
MIABIS-2.0-09	Sex <sup>a</sup>	Text list	The sex of the sample donor in the sample collection. Can be one or more of the following values: Male, Female, Unknown, Undifferentiated MIABIS-2.0-09 ( <a href="https://github.com/MIABIS/miabis/wiki/Structured-data-and-lists#sex">https://github.com/MIABIS/miabis/wiki/Structured-data-and-lists#sex</a> )
MIABIS-2.0-10	Age Low <sup>a</sup>	Integer	Age of youngest sample donor at time of sample donation
MIABIS-2.0-11	Age High <sup>a</sup>	Integer	Age of oldest sample donor at time of sample donation
MIABIS-2.0-12	Age Unit	Text list	Unit defining Age Low and Age High. Can be one of the following values: years, months, weeks, days
MIABIS-2.0-13	Data Categories <sup>a</sup>	Text list	The data categories from which data are available. Can be several values MIABIS-2.0-3 ( <a href="https://github.com/MIABIS/miabis/wiki/Structured-data-and-lists#data-categories">https://github.com/MIABIS/miabis/wiki/Structured-data-and-lists#data-categories</a> )
MIABIS-2.0-14	Material Type <sup>a</sup>	Text list	The biospecimen saved from a biological entity for propagation, for example, testing, diagnostics, treatment, or research purposes. Can be several values: MIABIS-2.0-14 ( <a href="https://github.com/MIABIS/miabis/wiki/Structured-data-and-lists#material-type">https://github.com/MIABIS/miabis/wiki/Structured-data-and-lists#material-type</a> )
MIABIS-2.0-15	Storage Temperature <sup>a</sup>	Text list	The long-term storage temperature at which the samples are stored after preparation. Can be several values MIABIS-2.0-15 ( <a href="https://github.com/MIABIS/miabis/wiki/Structured-data-and-lists#storage-temperature">https://github.com/MIABIS/miabis/wiki/Structured-data-and-lists#storage-temperature</a> )
MIABIS-2.0-16	Collection Type	Text list	The type of the sample collection. Can be several values MIABIS-2.0-16 ( <a href="https://github.com/MIABIS/miabis/wiki/Structured-data-and-lists#collection-type">https://github.com/MIABIS/miabis/wiki/Structured-data-and-lists#collection-type</a> )
MIABIS-2.0-17	Disease	Structured data	The disease of main interest in the sample collection, if any. Can be several values MIABIS-2.0-17 ( <a href="https://github.com/MIABIS/miabis/wiki/Structured-data-and-lists#disease">https://github.com/MIABIS/miabis/wiki/Structured-data-and-lists#disease</a> )
MIABIS-2.0-07	Contact Information	Structured data	Contact information for the contact person of the sample collection MIABIS-2.0-07 and MIABIS-2.0-23 ( <a href="https://github.com/MIABIS/miabis/wiki/Structured-data-and-lists#contact-information">https://github.com/MIABIS/miabis/wiki/Structured-data-and-lists#contact-information</a> )

<sup>a</sup>Attribute can be aggregated from data stored on the sample and subject levels in local management systems of the biobanks.

TABLE 3. ATTRIBUTE DEFINITIONS OF THE COMPONENT “STUDY”

<i>Attribute code</i>	<i>Attribute name</i>	<i>Allowed values</i>	<i>Description</i>
MIABIS-2.0-01	ID	Text	The unique ID or acronym of the study
MIABIS-2.0-02	Name	Text	The name of the study in English
MIABIS-2.0-08	Description	Text	A description of the study aim. Recommendation maximum 2000 characters
MIABIS-2.0-18	Principal Investigator	Text	The name of the person responsible for the study, for example, the principal investigator (PI)
MIABIS-2.0-07	Contact Information	Structured data	Contact information for the contact person of the study MIABIS-2.0-07 and MIABIS-2.0-23 ( <a href="https://github.com/MIABIS/miabis/wiki/Structured-data-and-lists#contact-information">https://github.com/MIABIS/miabis/wiki/Structured-data-and-lists#contact-information</a> )
MIABIS-2.0-19	Study Design	Text list	The design of the study. Can be one or several values MIABIS-2.0-19 ( <a href="https://github.com/MIABIS/miabis/wiki/Structured-data-and-lists#collection-type">https://github.com/MIABIS/miabis/wiki/Structured-data-and-lists#collection-type</a> )
MIABIS-2.0-09	Sex <sup>a</sup>	Text list	The sex of the study participants. Can be one or more of the following values: Male, Female, Unknown, Undifferentiated MIABIS-2.0-09 ( <a href="https://github.com/MIABIS/miabis/wiki/Structured-data-and-lists#sex">https://github.com/MIABIS/miabis/wiki/Structured-data-and-lists#sex</a> )

(continued)

TABLE 3. (CONTINUED)

<i>Attribute code</i>	<i>Attribute name</i>	<i>Allowed values</i>	<i>Description</i>
MIABIS-2.0-10	Age Low <sup>a</sup>	Integer	Age of the youngest study participant at time of inclusion
MIABIS-2.0-11	Age High <sup>a</sup>	Integer	Age of the oldest study participant at time of inclusion
MIABIS-2.0-12	Age Unit	Text list	Unit defining Age Low and Age High. Can be one of the following values: years, months, weeks, days
MIABIS-2.0-13	Data Categories <sup>a</sup>	Text list	The type of data that are associated with the samples in the study. Can be several values MIABIS-2.0-13 ( <a href="https://github.com/MIABIS/miabis/wiki/Structured-data-and-lists#data-categories">https://github.com/MIABIS/miabis/wiki/Structured-data-and-lists#data-categories</a> )
MIABIS-2.0-14	Material Type <sup>a</sup>	Text list	The biospecimen saved from a biological entity for propagation, for example, testing, diagnostics, treatment, or research purposes. Can be several values MIABIS-2.0-14 ( <a href="https://github.com/MIABIS/miabis/wiki/Structured-data-and-lists#material-type">https://github.com/MIABIS/miabis/wiki/Structured-data-and-lists#material-type</a> )
MIABIS-2.0-20	Total number of participants	Integer	Total number of individuals recruited to the study
MIABIS-2.0-21	Total number of sample donors	Integer	Total number of individuals with biological samples in the study
MIABIS-2.0-22	Inclusion Criteria	Text list	Information on type of parameters that determine which individuals will become study participants. Can be several values MIABIS-2.0-22 ( <a href="https://github.com/MIABIS/miabis/wiki/Structured-data-and-lists#inclusion-criteria">https://github.com/MIABIS/miabis/wiki/Structured-data-and-lists#inclusion-criteria</a> )

<sup>a</sup>Attribute can be aggregated from data stored on the sample and subject levels in local management systems of the biobanks.

TABLE 4. DEFINITION OF THE STRUCTURED DATA ATTRIBUTE “CONTACT INFORMATION”

<i>Attribute Code</i>	<i>Attribute name</i>	<i>Allowed values</i>	<i>Description</i>
MIABIS-2.0-07-A	First Name	Text	Textual string of letters denoting the first name of the Contact
MIABIS-2.0-07-B	Last Name	Text	Textual string of letters denoting the last name of the Contact
MIABIS-2.0-07-C	Phone	Formatted Text	Phone to the Contact, including international call prefix
MIABIS-2.0-07-D	E-mail	Text	E-mail address of the Contact person
MIABIS-2.0-07-E	Address	Text	Street name and street number or PO Box of the Contact person
MIABIS-2.0-07-F	ZIP	Text	ZIP code of the Contact person
MIABIS-2.0-07-G	City	Text	City of the Contact person
MIABIS-2.0-07-H	Country	Text	Country of the Contact person

TABLE 5. DEFINITION OF THE STRUCTURED DATA ATTRIBUTE “RESEARCH INFORMATION”

<i>Attribute code</i>	<i>Attribute name</i>	<i>Allowed values</i>	<i>Description</i>
MIABIS-2.0-23A	Juristic Person	Text	Textual string of letters denoting the legal entity, for example, a university, county council, or other host organization
MIABIS-2.0-23B	Department	Text	Department (e.g., division) or corresponding affiliation of the Contact person
MIABIS-2.0-23C	ORCID	Text	The persistent digital identifier that distinguishes researchers. As defined by ORCID <sup>a</sup>

<sup>a</sup>ORCID<sup>7</sup> website: <http://orcid.org/>

TABLE 6. DEFINITION OF THE STRUCTURED DATA ATTRIBUTE “DISEASE”

<i>Attribute code</i>	<i>Attribute name</i>	<i>Allowed values</i>	<i>Description</i>
MIABIS-2.0-17A	Disease Ontology	Text	Name of ontology used for disease. Can be several values, for example, ICD, SNOMED
MIABIS-2.0-17B	Disease Ontology Version	Text	Version of selected ontology for disease, for example, ICD-9, ICD-10, SNOMED-CT
MIABIS-2.0-17C	Disease Ontology Code	Text	Disease code from the selected Disease Ontology Version, for example, C61
MIABIS-2.0-17D	Disease Ontology Description	Text	Description from the selected Disease Ontology Code, for example, malignant neoplasm of prostate
MIABIS-2.0-17E	Disease-Free Text	Text	Explanation about disease or symptom in case of unknown disease or insufficient information

### *Guidelines for implementing MIABIS 2.0 Core in biobank and research informatics systems*

To be MIABIS compliant means (1) MIABIS concepts and attributes are part of the data model of biobank and research informatics systems or (2) data from biobank and research are mapped to MIABIS for sharing purposes.

Each component of MIABIS 2.0 Core can be modeled as a class in a logical model or as an entity in an Entity Relationship Diagram (ERD). The semantic of MIABIS 2.0 Core allows for the implementation of use cases for managing and sharing biobank and research data at the aggregated and metadata levels.

MIABIS 2.0 has two main parts: MIABIS 2.0 Core that represents a high level of information of biobank and biomedical research and *additional components* that describe particular subdomains derived from concepts that drive everyday biomedical research processes. This way, data from every component of the biomedical research ecosystem can be represented in a standardized manner. It will help create interfaces and tools to facilitate interoperability and reusability of data and knowledge. Currently, five additional components dealing with the concepts of “Biological Experiment,” “Participant,” “Rare Diseases,” “Sample,” and “Quality” have been proposed within the MIABIS Working Group. They are not presented in this article as they have not gone through the whole governance procedure and are still under the status of “proposed.” All updates and proposals can be found on MIABIS Wiki page (<https://github.com/MIABIS/miabis/wiki>).

### *Formal representation of MIABIS*

Figure 2 provides a logical representation of MIABIS 2.0 Core for a use case that could be a biobank catalog, part of a Laboratory Information Management System (LIMS) or a research management system, such as an Electronic Lab Notebook (ELN). It captures the semantics proposed for the core components.

When expressing the example logical model (Fig. 2) in an ERD, it introduces auxiliary entities as contact information, researcher information, references to ontologies, diseases, and so on. The attributes expressed as lists (*Material Type*, *Data Categories*, etc.) can be included as separated entities or in a unified table for all the lists. Examples of the use of MIABIS 2.0 Core in ERDs are as follows: the BBMRI-NL<sup>8</sup>

catalog, which can be downloaded as open source MOLGENIS software,<sup>9</sup> the data model for BBMRI-LPC<sup>10</sup> and for the Biobank Catalog for the BCNet network from the WHO organization.<sup>11</sup> Examples are also provided in the MIABIS wiki:

<https://github.com/MIABIS/miabis/wiki/Database-implementation>

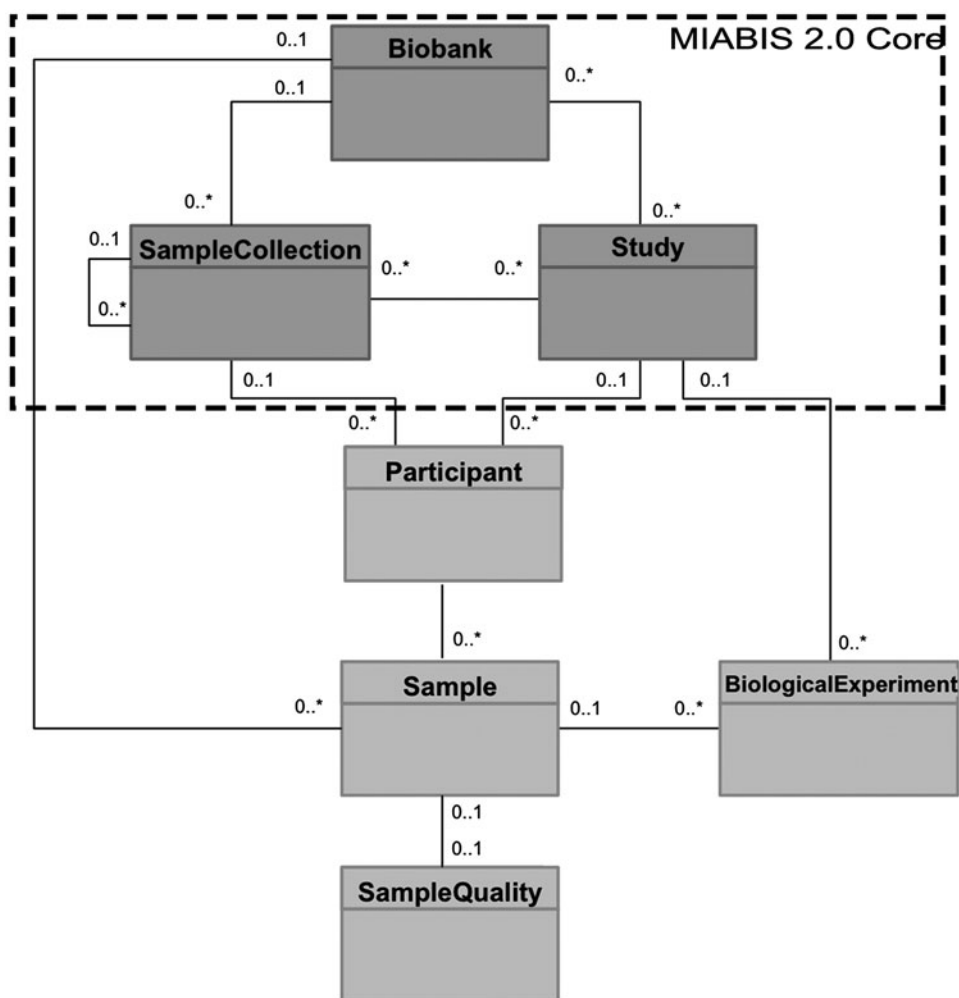
### *Data sharing*

MIABIS can be used to share data in a federated network of biobanks or in a virtual environment (e.g., cloud computing, business intelligence, service-oriented architecture). In a federated environment, as a first step, data from the biobanks should be mapped to the MIABIS data elements. This can be done manually or through a program or script, but it is essentially a manual job that can be substantially simplified if the biobank management systems already are MIABIS compliant. The mapping process needs to be done every time the data model is changed in the biobank database. Once the biobank data are mapped to MIABIS, an interface should be in place as a communication channel between the biobank and external services. In a federated solution, a uniform user interface can be used to distribute queries among the biobanks. A virtual environment applies data abstraction and data transformation techniques to deal with the different semantics in the biobanks. In this case, the mapping should produce linked data to a data layer that semantically connects the data with the data elements in MIABIS. This approach has been demonstrated as a proof of concept in the RD-Connect project<sup>12</sup> and in the BioMedBridges project.<sup>13</sup>

### *MIABIS 2.0 Core XML Schema*

A machine-readable version of MIABIS 2.0 Core allows for the implementation of several data sharing use cases such as the following:

- (1) A federation of catalogs where searches can be expanded to include data from another catalog.
- (2) Annotation of data sets in a data repository with the provenance of the samples.
- (3) Exchanging data about samples between a catalog and a track and trace system for sample requests, shipments, and so on.



**FIG. 2.** Logical data model of a use case that implements the MIABIS 2.0 Core and some additional components (Participant, Biological Experiment, Sample, and Sample Quality). Multiplicities in the diagram represent cardinality among the different classes for the specific use case.

An XML Schema Definition for MIABIS 2.0 Core to facilitate these use cases has been designed. The schema's guiding principles are as follows:

- (1) It must be possible to use each component individually.
- (2) Duplication of data should be avoided.
- (3) The resulting schema should be suitable for streaming access.
- (4) The XML file should be human-readable.

All relationships between components have been marked as optional to allow an implementation to cherry pick only the items desired, and we have used recognizable names for the fields.

To test the schema, a reference implementation has been created that generates Java classes that can be used to read and write a schema file. The XML schema can be found on the MIABIS Wiki (<https://github.com/MIABIS/miabis-xml>).

### *The case of the BBMRI-ERIC Directory*

MIABIS 2.0 Core is implemented in the BBMRI-ERIC Directory (<http://bbmri-eric.eu/bbmri-eric-directory>), which integrates biobank catalogs from the member countries. To interface with the Directory, sample collections should be modeled using partitioning of a parent set,<sup>14</sup> which is either a

biobank or another collection, where each collection is a partition. For the top-level collections, the parent set is the biobank, while each collection can be further partitioned into subcollections. This implies (1) each sample of a parent set belongs to exactly one (sub) collection; (2) there are no empty collections. Use of partitioning of the collection to obtain insight in the overall composition of the parent collections, and ultimately, the content of the whole biobank. If a biobank is not using the concept “*Sample Collection*,” one virtual sample collection for the biobank should contain all the samples. For clinical biobanks that lack standardized structuring, the software developers are advised to introduce sample collections that aggregate samples by, for instance, the material type.

### *Showcases of MIABIS*

Since 2012, several projects and biobanks have adopted the MIABIS standard in different use cases. Below is a short summary of these projects.

*Confederation of Cancer Biobanks.* The Confederation of Cancer Biobank from National Cancer Research Institute (NCRI) created the Biobank Data Standard for collection, storing, and sharing data describing human biological material. The standard includes components and attributes from MIABIS (<http://ccb.ncri.org.uk>).



*BBMRI-ERIC Directory.* The first version of the BBMRI-ERIC Directory (<http://bbmri-eric.eu/bbmri-eric-directory>) provides an aggregated view of the BBMRI-ERIC biobank infrastructure. Currently, MIABIS 2.0 is implemented on a common catalog platform that integrates biobank and research applications from most of the European countries that are members/observers of BBMRI-ERIC.

*BBMRI.se Sample Collection Register.* The Swedish BBMRI node created a MIABIS-compliant national catalog (<http://bbmriregister.se>) in 2012 with descriptive data about studies and sample collections using human biological samples in Sweden. The aim of the register is to raise awareness about sample accessibility for potential scientific collaborations.

*BBMRI.fi's KITE availability tool.* This is an open web-based database that contains information on available Finnish sample collections using MIABIS attributes. KITE also allows users to browse and search standardized lists of variables available in each sample collection. KITE also includes information on Finnish biobanks in MIABIS 2.0 format (<https://kite.fimm.fi>).

*BBMRI-NL national catalog.* The Dutch BBMRI node has published a national catalog since 2012 (<https://catalogue.bbmri.nl/>), which has been updated to MIABIS 2.0 using BBMRI-NL Rainbow Project 11. Currently, work is progressing to support fully automated updates of the data without human interaction from several large Dutch biobanks.

*BiobankCloud.* The BiobankCloud is an EU FP7-funded project aiming to create a cloud-computing platform as a service (PaaS) for storage and analysis of genomic data. MIABIS 2.0 is being used as part of the data model for representing metadata related to the biobank, collections, research study, and also metadata representing omics data ([www.biobankcloud.com/](http://www.biobankcloud.com/)).

*RD-Connect Databases.* Biobanks and clinical bioinformatics hub for rare diseases, or RD-Connect, is an EU FP7-funded project with the aim to provide an integrated platform connecting databases, registries, biobanks, and clinical bioinformatics for rare disease research. The project has 27 full partners, also outside the European Union such as in the United States (NIH) and Australia and 26 associated. RD-Connect also collaborates with a wide range of other people and organizations worldwide (<http://rd-connect.eu/>).

*EUDAT.* The MIABIS component study is used in B2SHARE, a user-friendly, reliable, and trustworthy way for researchers, scientific communities, and citizen scientists to store and share small-scale research data from diverse contexts ([www.eudat.eu/](http://www.eudat.eu/)).

*BioMedBridges.* Within the EC FP7-funded cluster project for Life Sciences, BioMedBridges, and a joint effort of the 10 biomedical sciences research infrastructures on the ES-FRI roadmap. “Biobank” and “Sample Collection” components are used in the data model of the *BioSample* service ([www.biomedbridges.eu/](http://www.biomedbridges.eu/)).

*International Agency for Research on Cancer.* In the global biobank network for International Agency for Research on Cancer (IARC), BBMRI.se is developing an international biobank catalog at the sample level for BCNet. The data model is fully based on MIABIS 2.0 (<http://ibb.iarc.fr/>).

## Discussion

MIABIS 2.0 Core is now simplified so that the components have smaller attribute lists. We foresee that the Core

will not be subject to much additional change. Modularization will support different kinds of use and will make it easier to add future extensions. As the MIABIS 2.0 Core attributes are often already represented in the databases of commercial and in-house management systems (e.g., LIMS, ELN, registries, and catalogs), little effort is needed to adopt the standard for data sharing.

MIABIS 2.0 will cover most of the relevant concepts in biobanking and biomedical research represented in MIABIS 2.0 Core and other additional components. Future plans, beyond the additional components already in progress, include adding components to represent personal data protection and clinical data.

As MIABIS is a conceptual information model, it does not define mandatory attributes. However, when implementing MIABIS in an informatics system, mandatory attributes are relevant according to the system requirements. When using MIABIS for biobank interoperability, the question of mandatory attributes gains more relevance.

Making informatics systems MIABIS compliant does not necessarily mean that the data model should follow MIABIS strictly. The informatics systems can keep their own idiosyncratic semantics and map their data to MIABIS for sharing purposes. Nevertheless, one of the aims of MIABIS is to shift the biobank and biomedical research community toward the use of the controlled vocabulary represented in this standard.

Another relevant issue is quality assessment for biobanks, sample collections, samples, and data. Sharing of bioresources requires consistent and standardized tools to evaluate the quality of the bioresource. This has a significant impact on the research process and will be addressed by the MIABIS component “Quality,” which is currently in preparation and will be associated with other components.

While a consensus procedure for proposing and accepting new components and attributes is in place, there is still some work to be done on the governance of MIABIS. It is envisaged that the maintenance and further development of MIABIS to include more countries will be aided by a content management platform that allows any interested countries, experts, and other interested parties to be informed, join, and contribute to the “under construction” components. Workgroup coordinators will post and display most current components, and country representatives will actively seek deliberation with their stakeholders. The voting process will be more formal but, by definition, transparent for all.

The biomedical research community is moving toward a more open-data and open-science environment. However, there is still a lack of dependable informatics platforms for researchers to share data in an easy and secure way. Even when several efforts are being made to facilitate sharing of biomedical data,<sup>15–17</sup> better information and dissemination of results in that direction are needed. At the moment, several informatics infrastructures providing services for storage, analysis, and integration of scientific data<sup>8,18,19</sup> are already implementing MIABIS 2.0 Core to facilitate biobank and research data sharing in trustable and secure environments.

We have developed a standard that can be used by biobanks, biomedical researchers, and software developers around the world, as the information recommended in MIABIS 2.0 Core is often a natural part of the data models for the biobank and research management systems. To

further promote the use of MIABIS and exemplify how it can be used once implemented, specifying scientific use cases will be beneficial.

The MIABIS 2.0 Core is the stable part of MIABIS 2.0, the de facto standard for sharing data in the BBMRI-ERIC community, and has also been adopted by other biobank networks. MIABIS 2.0 will shortly be finalized, including additional components that represent the most relevant building blocks of the biomedical research.

### Acknowledgments

We thank our fellow members in the MIABIS Working Group: Martin Fransson, Roman Siddiqui, Heimo Müller, Klaus Kuhn, Linda Zaharenko, Helmut Spengler, Araceli Diezfraile, Joakim Geeraert, Ondřej Vojtíšek, Anita Nieminen, Kristjan Metsalu, Murat Sariyar, Michael Hummel, and Cathleen Ploetzand. We also thank the European Commission for financially supporting the BBMRI preparatory phase (grant agreement 212111), the Swedish Research Council for granting the BBMRI.se project (grant agreement 829–2009-6285), BBMRI-NL, a research infrastructure financed by the Netherlands Organization for Scientific Research (NWO project 184.021.007), the Center for Translational Molecular Medicine, Translational IT project (CTMM-TraIT), and the Academy of Finland for financially supporting the BBMRI.fi network (grant no. 273527).

### Author Disclosure Statement

No conflicting financial interests exist.

### References

1. Yuille M, van Ommen GJ, Brechot C, et al. Biobanking for Europe. *Brief Bioinform* 2008;9:14–24.
2. Norlin L, Fransson MN, Eriksson M, et al. A Minimum Data Set for Sharing Biobank Samples, Information and Data: MIABIS. *Biopres Biobank* 2012;10:343–348.
3. Brochhausen M, Fransson MN, Kanaskar NV, et al. Developing a semantically rich ontology for the biobank-administration domain. *Journal of Biomedical Semantics* 2013;4:23.
4. BBMRI-ERIC. <http://bbmri-eric.eu/>
5. MoSCoW Analysis (6.1.5.2). *A Guide to the Business Analysis Body of Knowledge* (second ed.). Toronto, Ontario, Canada: International Institute of Business Analysis; 2009. ISBN 978-0-9811292-1-1.
6. Fransson MN, Rial-Sebbag E, Brochhausen M, et al. Toward a common language for biobanking. *Eur J Hum Genet* 2015;23:22–28.
7. Editorial. Credit where credit is due. *Nature* 2009;462:825.
8. BBMRI.nl catalogue: <http://catalogue.bbmri.nl>
9. Swertz MA, et al. The MOLGENIS toolkit: Rapid prototyping of biosoftware at the push of a button. *BMC Bioinform* 2010;11 Suppl 12:S12
10. BBMRI-LPC. [www.bbmri-lpc.org/](http://www.bbmri-lpc.org/)
11. BCNet. <http://bcnet.iarc.fr>
12. RD-Connect. <http://rd-connect.eu>
13. BioMedBridges. [www.biomedbridges.eu](http://www.biomedbridges.eu)
14. Halmos PR. *Naive Set Theory*. NY: Springer; 1960:28. ISBN 9780387900926.
15. Lehmann S, Guadagni F, Moore H, et al. Standard pre-analytical coding for biospecimens: Review and implementation of the Sample PREanalytical Code (SPREC). *Biopres Biobank* 2012;10:366–374.
16. Harris JR, Burton P, Knoppers BM, et al. Toward a roadmap in global biobanking for health. *Eur J Hum Genet* 2012;20:1105–1111.
17. Wichmann HE, Kuhn KA, Waldenberger M, et al. Comprehensive catalog of European biobanks. *Nat Biotechnol* 2011;29:795–797.
18. Yuille M, Dixon K, Platt A, et al. The UK DNA banking network: A 'fair access' biobank. *Cell Tissue Bank* 2010; 11:241–251.
19. BiobankCloud: [www.biobankcloud.com](http://www.biobankcloud.com)

Address correspondence to:  
 Roxana Merino-Martinez, MS  
 BBMRI.se  
 Department of Medical Epidemiology and Biostatistics  
 Karolinska Institutet  
 Nobels väg 12A  
 Stockholm 171 77  
 Sweden

E-mail: [roxana.martinez@ki.se](mailto:roxana.martinez@ki.se)