

# BBMRI-ERIC DIRECTORY

*MANUAL FOR DATA MANAGERS*

Version 3.6.12

Authors: Dieuwke Roelofs-Prins, Aneas Hodselmans, Esther van Enkevort

# TABLE OF CONTENTS

<b>Table of Contents</b>	<b>2</b>
<b>Introduction</b>	<b>2</b>
<b>Data integration process</b>	<b>2</b>
Manual data entry	4
Manual data upload of Excel or CSV files	6
Scheduled file ingest of CSV files	11
Structure of the CSV files	12
Programmatic updates through the RESTful API's	12
<b>Data model</b>	<b>13</b>
Download the latest model and data	13
Structure of the identifiers	13
Biobank	16
Collection	17
Network	20
Contact information	21
Also known in	22
Star Model	23
<b>Code lists</b>	<b>24</b>
Disease types	25
<b>Quality</b>	<b>25</b>
Quality checks data input	26
Warnings data input	26
Quality marks	26

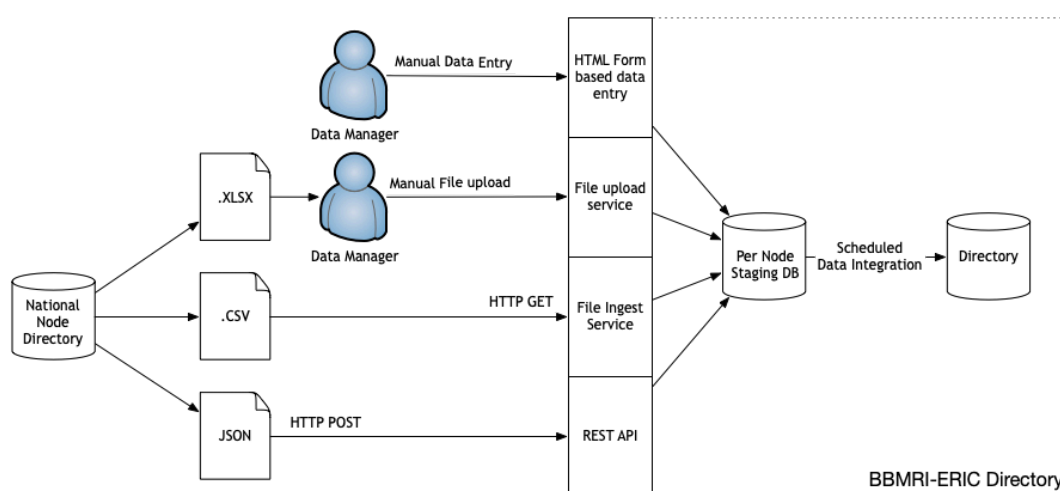
## INTRODUCTION

This manual is a technical description of the BBMRI-ERIC Directory data model and the process of updating data in the Directory. The intended audience of this manual is the National Node Data Manager. Throughout this document the BBMRI-ERIC Directory is used for the examples. However, this manual also applies to local BBMRI instances. Although the model remains the same, minor differences between the local and ERIC Directories exist, these will be indicated in the footnotes.

## DATA INTEGRATION PROCESS

The BBMRI-ERIC Directory has a federated process of updating the data, where each National Node is responsible for updating the data for the biobanks in the node. This is done in a staging area that gives the national node exclusive access to update the data.<sup>1</sup> Data in the Directory can be managed in four different ways:

1. Manual data entry if the National Node does not host a National Directory
2. Manual upload of Excel or CSV files exported from the National Directory
3. Scheduled file ingest of CSV files from the National Directory
4. Programmatic updates initiated by the National Directory (using the Directory's RESTful API's, see the Interoperability on <https://molgenis.gitbooks.io/molgenis/content/>)



Regardless of the method used to update the staging area the data from the staging area is integrated into the Directory through a nightly scheduled job. This means that it takes one day before changes are visible for the outside world. In the meantime, the data manager of the National Node can access and verify the data in the National Node's staging area.

Next to the data that is provided by the National Node the Directory holds quality marks that are based upon the self-assessment filled in by the biobanks. These parameters are managed by BBMRI-ERIC's quality management team and cannot be updated by the National Node. However, for a smooth process of application for the quality marks it is paramount that the biobank and the collections are registered in the Directory before the self-assessment is filled in.

### MANUAL DATA ENTRY

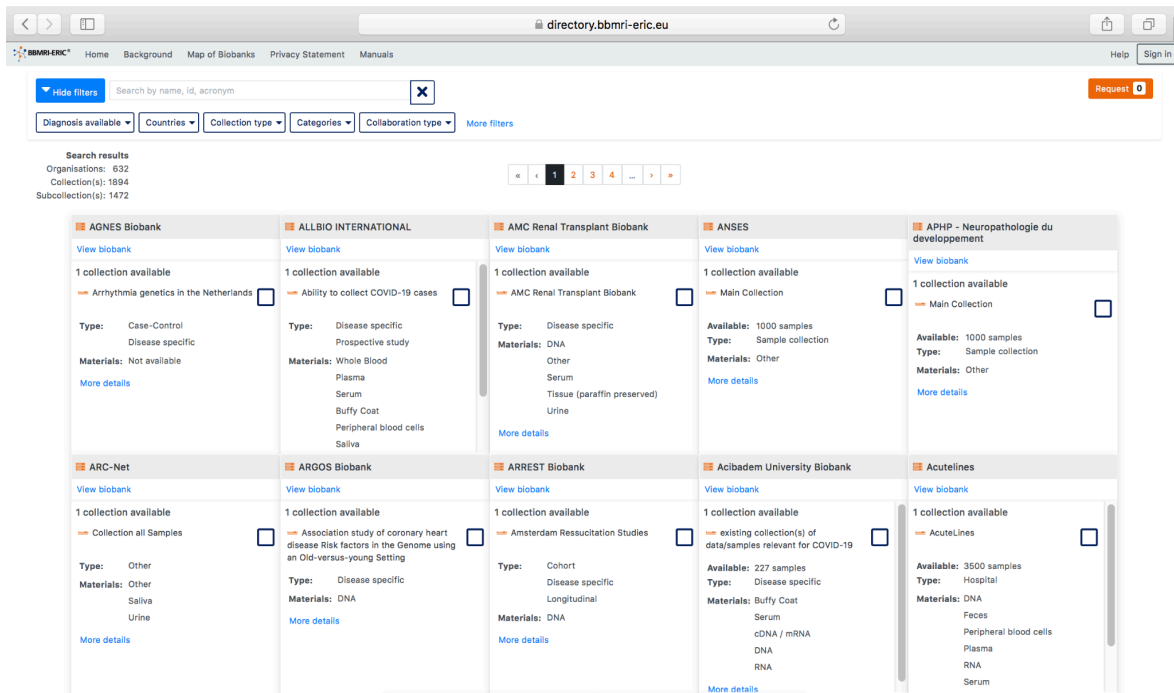
The manual data entry option is to be used only as a last resort option if no National Directory or database is available. Data entry is done in the staging area for the National Node and changes will

---

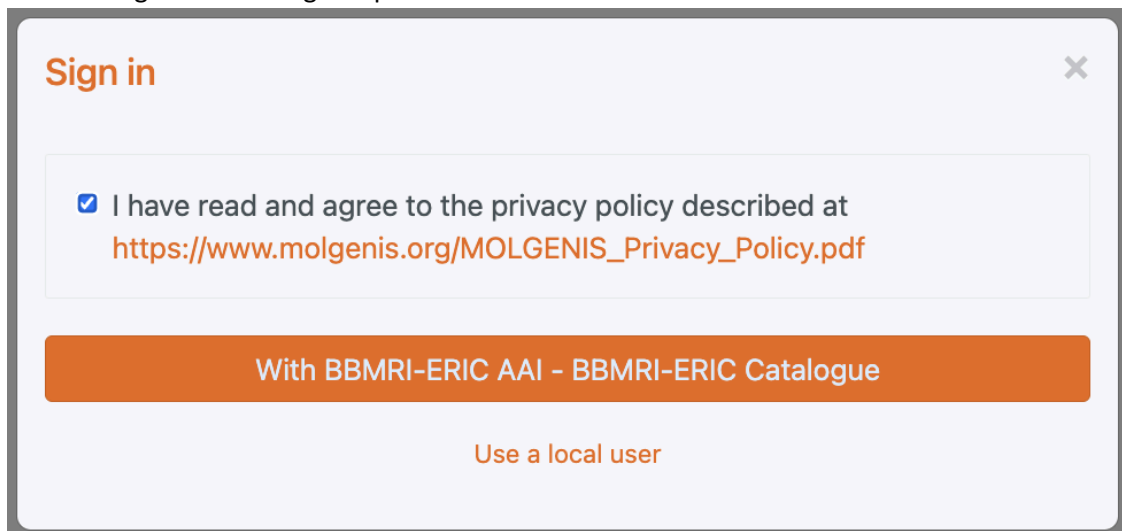
<sup>1</sup> This part only applies to the BBMRI-ERIC Directory, the local directory instances do not contain staging areas and can upload the data directly to the directory model.

be published to the production version on a daily basis based on a scheduled job that runs during the night.

1. Go to <https://directory.bbmri-eric.eu/>

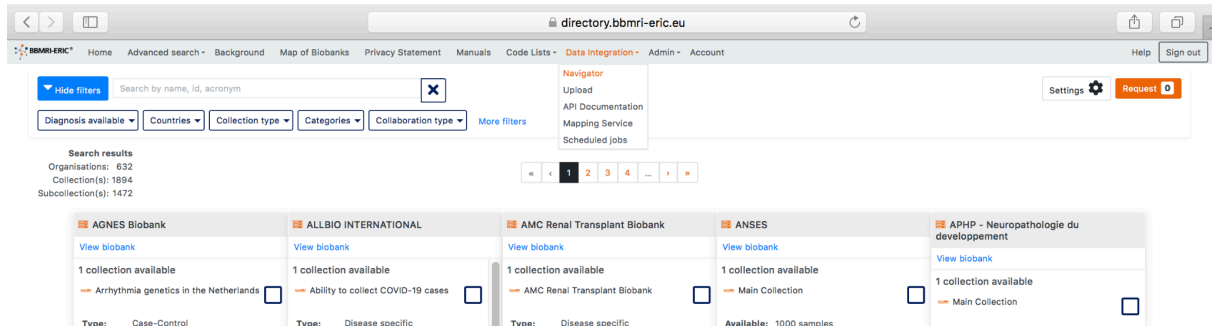


2. Click on "Sign in" in the right top corner of the screen.

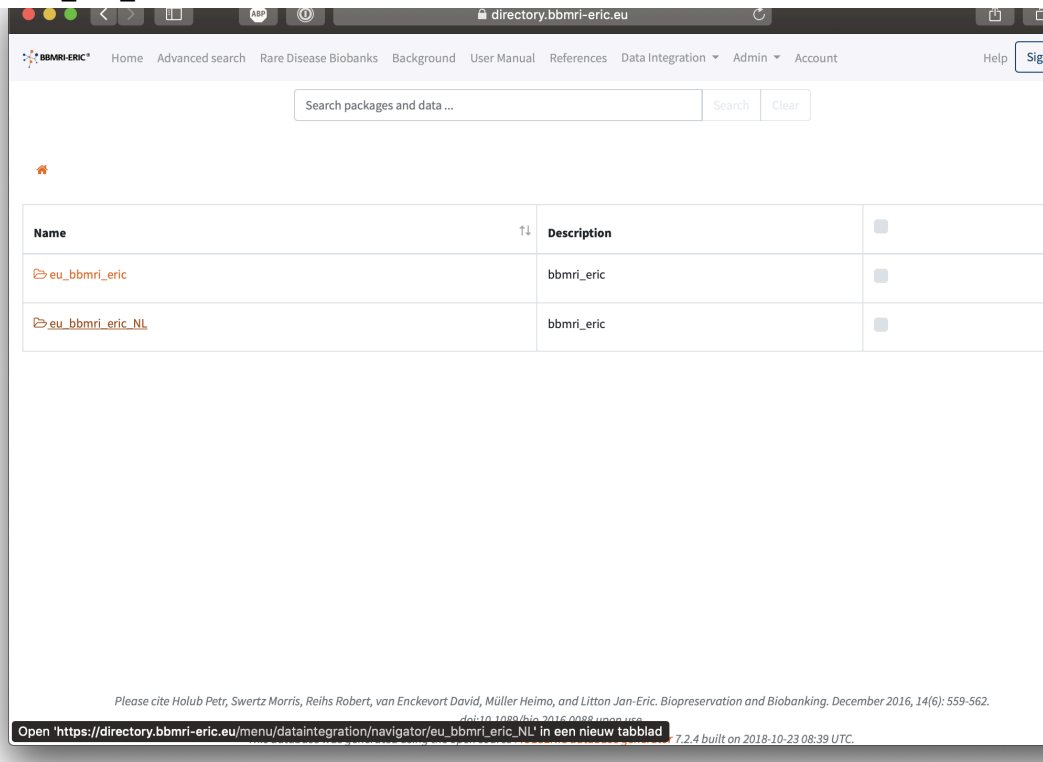


This work is part of the ADOPT BBMRI-ERIC project, funded by the European Commission, topic H2020- INFRADEV-3-2015, Grant Agreement Nr 676550

3. Sign in “With BBMRI-ERIC AAI”. If you haven’t an account yet, create it (see [manual](#)). After creating it, send a mail to [molgenis-support@umcg.nl](mailto:molgenis-support@umcg.nl) to get permission(PI in CC). If you have an account, sign in.
4. Click on “Data integration” in the menu. and Click on “Navigator” in the menu.



5. Click in the following screen on the link to go to your national node’s staging area. All countries are denoted by their `eu_bbmri_eric_<CC>`, where `<CC>` is your country code, e.g.<sup>2</sup> `eu_bbmri_eric_NL`.



<sup>2</sup> This only applies to the BBMRI-ERIC Directory, local instances can use the model without country code: `eu_bbmri_eric`.

This work is part of the ADOPT BBMRI-ERIC project, funded by the European Commission, topic H2020- INFRADEV-3-2015, Grant Agreement Nr 676550

6. Click on the link for the data table you want to edit.

BBMRI-ERIC Direc...

BBMRI-ERIC Home Advanced search Background Map of Biobanks Privacy Statement Manuals Code Lists Data Integration Admin Account Help Sign out

Search packages and data ...

bbmri\_NL

Name	Description
NL: Biobank, Collection or Network also known in	Network, biobank or collection also exists in ...
NL: biobanks	Description of the biobank organisation, like name, location, network, contact persons, collaboration opportunities and quality assessments
NL: collections	Description of the data and samples collected within a biobank, collections may be divided in sub-collections
NL: networks	Description of the biobank networks, like name, contact information and all kinds of common agreements (access policies, charters, sops)
NL: persons	Information on the contact persons of a biobank, collection of network, like name, address and e-mail address

Please cite Holub Petr, Swertz Morris, Reihls Robert, van Enckevort David, Müller Helmo, and Litton Jan-Eric. Biopreservation and Biobanking. December 2016, 14(6): 559-562. doi:10.1089/bio.2016.0088 upon use.  
 This database was created using the open source MQLGENIS software 10.1.0 built on 2023-01-30 11:41 UTC.  
 Please cite Van der Velde et al (2018), Swertz et al (2010) or Swertz & Jansen(2007) on use.

7. Edit the data using the edit and delete buttons.

BBMRI-ERIC Direc...

BBMRI-ERIC Advanced search Background Map of Biobanks Privacy Statement Manuals Code Lists Data Integration Admin Account Help Sign out

Collections Description of the data and samples collected within a biobank, collections may be divided in sub-collections

Search data values

Data item filters

Data item selection

Name	Acronym	Description
Bio-UV 2017: Photo(chemo)therapy in skin diseases		In this study, the molecular mechanisms of ac
BioPersMed Cohort: cardiovascular and metabolic diseases		This cohort comprises a huge number of sam
Blood And Tissue Samples From Gynaecological Tumours / Gynaecological Cancers		Blood as well as fresh frozen tissue samples t
Blood Bank Collection: healthy subjects		The sample collection of the Department of B
Blood and Tears: affective disorders, schizophrenic disorders, anxiety		Blood Collection for future biomarker search j
COVAC-DM Study: diabetes mellitus type 1, diabetes mellitus type 2, COVID-19, COVID-19 vaccine		It is unknown, if people with established diabk
COVID-19 Convalescent Cohort: SARS-CoV-2, COVID-19 recovered		Biobank Graz is looking for COVID-19 recover
Ability to collect COVID-19 cases		Ability to prospectively collect COVID-19 cas
Clinibil Study: post cardiac-surgery, ICU		Biological samples were as part of the Clinibil
Colorectal Cancer (CRC) Whole Slide Image - Clinical Annotation Cohort		Inclusion criteria: Patients with colorectal ca...
Colorectal Cancer (CRC) Whole Slide Image - Survival Cohort		Inclusion criteria: Patients with colorectal ca...
Colorectal Cancer (CRC) Whole Slide Image - Vascular Invasion Cohort		Inclusion criteria: Patients with colorectal ca...
Colorectal Cancer Collection		
Colorectal Cancer Collection - Colorectal Tissue		
Colorectal Cancer Collection - Colorectal Tissue		
Colorectal Cancer Collection - Colorectal Tissue		
Colorectal Cancer Collection - Colorectal Tissue		

8. When you press the edit button you get a popup to edit the data of one record.

This work is part of the ADOPT BBMRI-ERIC project, funded by the European Commission, topic H2020- INFRADEV-3-2015, Grant Agreement Nr 676550

## Collections

Images

Samples  
 Select all Deselect all  
 Denotes whether access to samples, data or images may be obtained on fee-based basis.

**Access via join projects to**

Data  
 Images  
 Samples  
 Select all Deselect all  
 Denotes whether access to samples, data or images may be obtained on joint project basis.

**Access description**

Short description of access rules.

**Access uri**

URI describing the access policy.

**PD/SOP**

Data processing PD/SOP  
 Data storage PD/SOP  
 Data transport PD/SOP  
 Sample processing PD/SOP  
 Sample storage PD/SOP  
 Sample transport PD/SOP  
 Select all Deselect all  
 Availability of Process Descriptions (PDs) and/or Standard Operating Procedures (SOPs)

Cancel Save

9. To save the record you might have to scroll down to reveal the save button.

BBMRI-ERIC Direc...  
 Home Advanced search Background Map of Biobanks Privacy Statement Manuals Code Lists Data Integration Admin Account Help Sign out

Images

Samples  
 Select all Deselect all  
 Denotes whether access to samples, data or images may be obtained on fee-based basis.

**Access via join projects to**

Data  
 Images  
 Samples  
 Select all Deselect all  
 Denotes whether access to samples, data or images may be obtained on joint project basis.

**Access description**

Short description of access rules.

**Access uri**

URI describing the access policy.

**PD/SOP**

Data processing PD/SOP  
 Data storage PD/SOP  
 Data transport PD/SOP  
 Sample processing PD/SOP  
 Sample storage PD/SOP  
 Sample transport PD/SOP  
 Select all Deselect all  
 Availability of Process Descriptions (PDs) and/or Standard Operating Procedures (SOPs)

Cancel Save

Please cite Holub Petr, Swertz Morris, Reihls Robert, van Enckevort David, Müller Helmo, and Litton Jan-Eric. *Biopreservation and Biobanking*, December 2016, 14(6): 559-562. doi:10.1089/bio.2016.0088 upon use.  
 This database was created using the open source MOLOGENIS software  
 Please cite Van der Velde et al (2018), Swertz et al (2010) or Swertz & Jansen(2007) on use.

10. All changes are automatically published from staging to production during the night.

## MANUAL DATA UPLOAD OF EXCEL OR CSV FILES

Manual upload is done in the staging area for the National Node<sup>3</sup> and changes will be published to the production version on a daily basis based on a scheduled job that runs during the night. The Excel

<sup>3</sup> Local BBMRI instances can directly upload to the directory model without using the staging area and two letter country code in the table names. For example: *eu\_bbmri\_collections*.

This work is part of the ADOPT BBMRI-ERIC project, funded by the European Commission, topic H2020- INFRADEV-3-2015, Grant Agreement Nr 676550



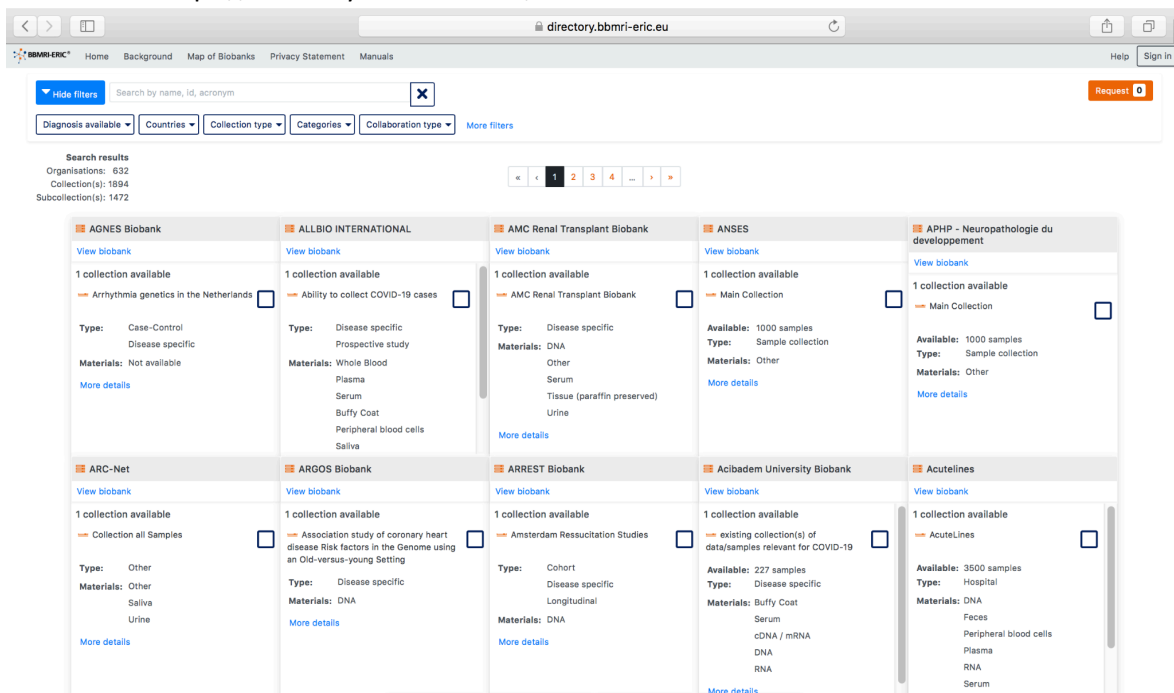
file or CSV files have to adhere to a specific structure. CSV files can be bundled together in a ZIP file, where each file has the name of the table and Excel files can have multiple sheets, where each sheet has the name of the table. The names are as follows:

- Collection: eu\_bbmri\_eric\_<CC>\_collections, e.g. eu\_bbmri\_eric\_NL\_collections
- Biobanks: eu\_bbmri\_eric\_<CC>\_biobanks, e.g. eu\_bbmri\_eric\_NL\_biobanks
- Networks: eu\_bbmri\_eric\_<CC>\_networks, e.g. eu\_bbmri\_eric\_NL\_networks
- Contact information: eu\_bbmri\_eric\_<CC>\_persons, e.g. eu\_bbmri\_eric\_NL\_persons
- Facts: eu\_bbmri\_eric\_<CC>\_facts, e.g. eu\_bbmri\_eric\_NL\_facts
- Also known in: eu\_bbmri\_eric\_<CC>\_also\_known\_in, e.g. eu\_bbmri\_eric\_NL\_also\_known\_in

Our data management team will provide you with a template upon request, or see section “Download the latest model and data” to retrieve a template yourself.

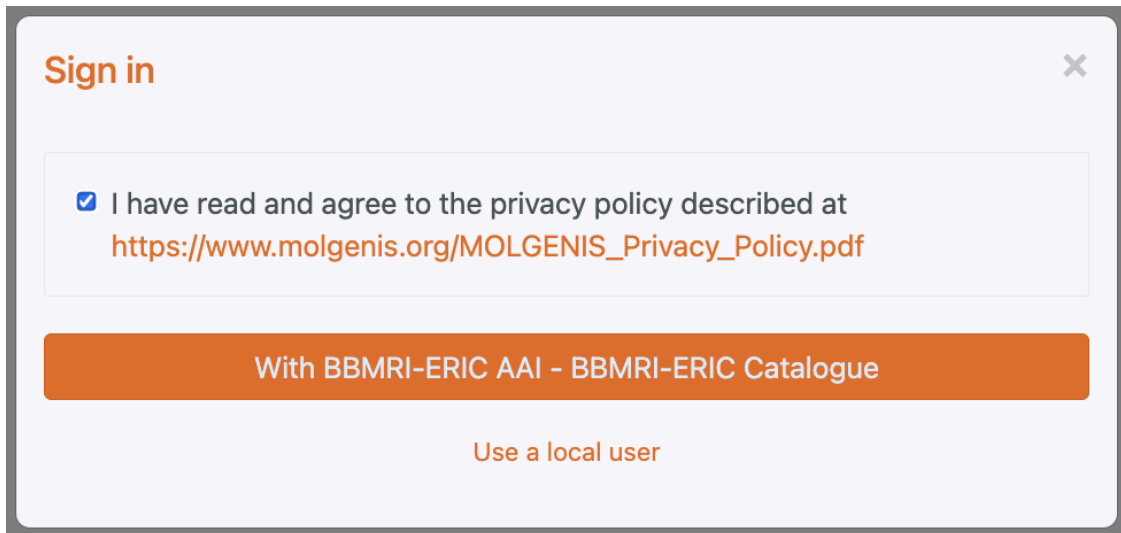
The steps to upload the files are as follows:

1. Go to <https://directory.bbmri-eric.eu/>

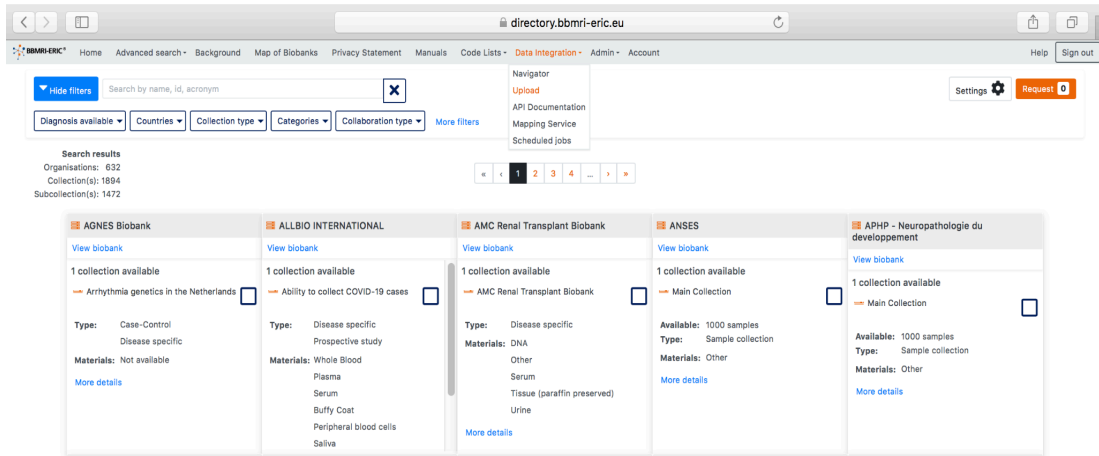


2. Click on “Sign in” in the right top corner of the screen.

This work is part of the ADOPT BBMRI-ERIC project, funded by the European Commission, topic H2020- INFRADEV-3-2015, Grant Agreement Nr 676550



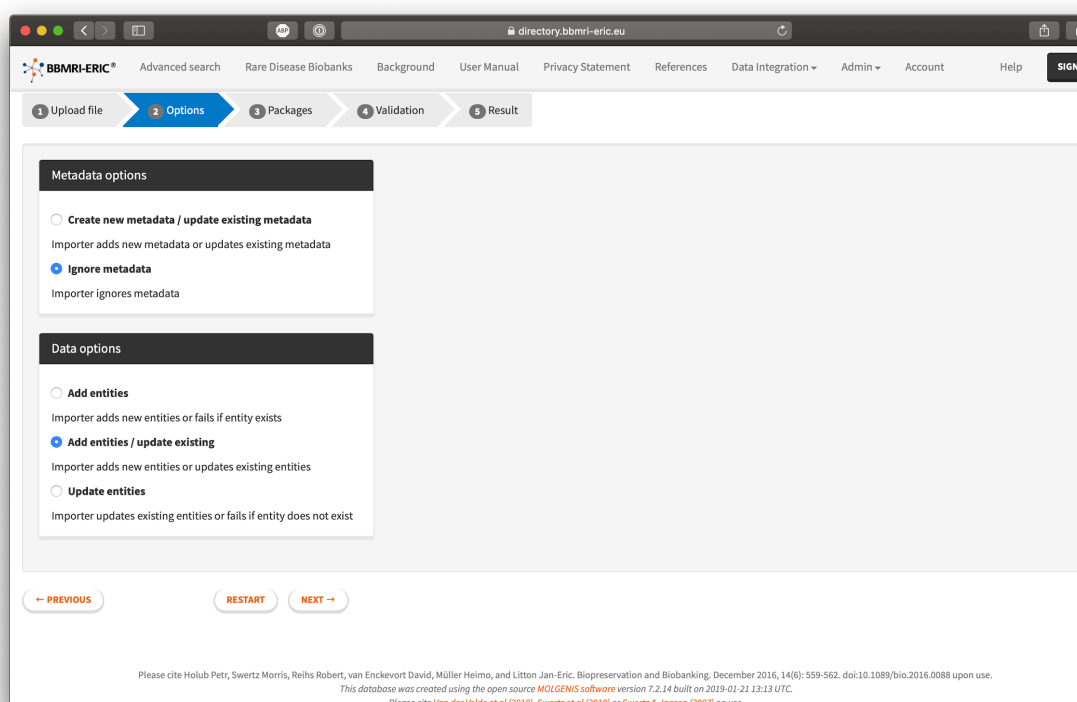
3. Sign in “With BBMRI-ERIC AAI”. If you haven’t an account yet, create it (see [manual](#)). After creating it, send an email to [molgenis-support@umcg.nl](mailto:molgenis-support@umcg.nl) to get permission(PI in CC). If you have an account, sign in.
4. Click on “Data integration”, “Upload” to go to the data upload wizard.



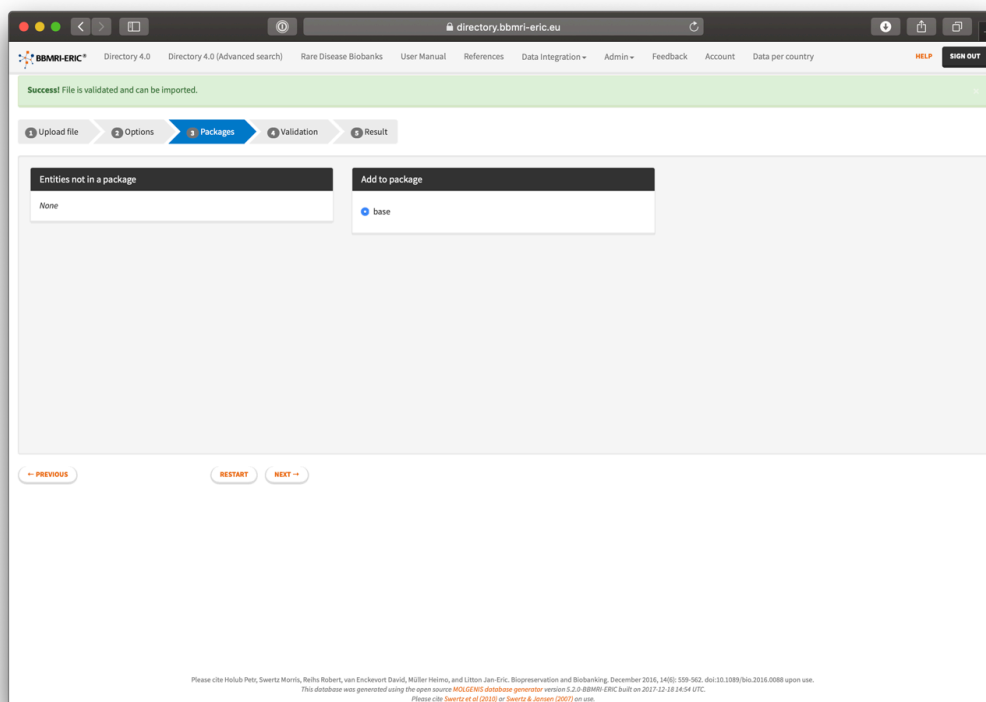
5. Select your Excel file to upload and press next.

This work is part of the ADOPT BBMRI-ERIC project, funded by the European Commission, topic H2020- INFRADEV-3-2015, Grant Agreement Nr 676550

6. Select the following options: “Ignore/metadata” and “Add entities / update existing” and press next.



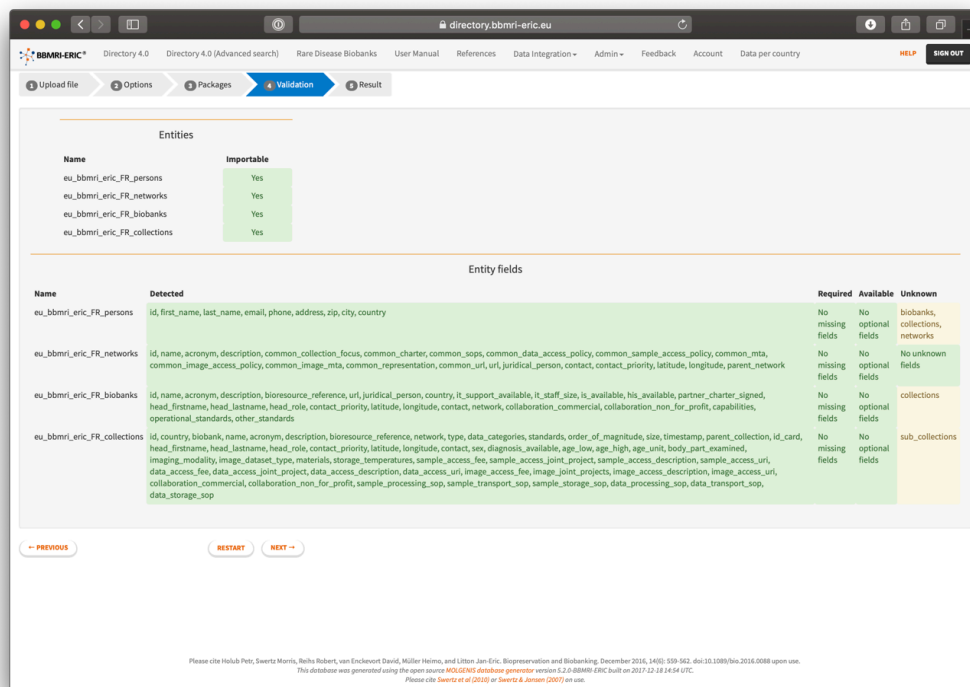
7. Accept the default options and select next.



8. Verify that there are no Errors (denoted with a red background colour) indicated during the validation and press “Next”. There might be some warnings (denoted with a yellow colour)

This work is part of the ADOPT BBMRI-ERIC project, funded by the European Commission, topic H2020- INFRADEV-3-2015, Grant Agreement Nr 676550

indicated during the validation process when your file contains additional columns or does not provide data for optional columns.



- The actual updating will now start, and at the end give a report with the number of updated records. At the end you can press “Finish” to complete the process. If the upload fails, please take a screenshot of the error report and contact the helpdesk.

## SCHEDULED FILE INGEST OF CSV FILES

To set up a scheduled file ingest of CSV files you will have to provide separate files for the five different entities (biobanks, collections, networks, contact information and also known in) and they should be made available for download by the system using HTTP or HTTPS. We currently do not support authentication, but if necessary you can limit the access to your files to the IP address of the Directory. To report on failures you need to provide us with an e-mail address that will receive error notifications. You should file a ticket in the Directory queue of the BBMRI-ERIC at <https://helpdesk.bbmri-eric.eu/> to get the system setup.

## STRUCTURE OF THE CSV FILES

The CSV files should be well structured and adhere to a few specific rules:

- No embedded newlines
- Text should be enclosed with double quotes
- Lists should be enclosed with double quotes and use commas as separator for the list elements
- Dates should be ISO8601 formatted
- When there is no value for a certain field you should leave it completely empty

This work is part of the ADOPT BBMRI-ERIC project, funded by the European Commission, topic H2020- INFRADEV-3-2015, Grant Agreement Nr 676550

## PROGRAMMATIC UPDATES THROUGH THE RESTFUL API'S

The Directory is built using the MOLGENIS Software. MOLGENIS offers a RESTful + JSON API to perform basic Create, Read, Update, Delete (CRUD) operations on the data in the Directory. Examples of these commands can be found in the accompanying Postman configuration (<https://doi.org/10.5281/zenodo.3367102>). An extensive manual for the RESTful API can be found in our Swagger developer documentation, which you will find in the 'Data Integration' menu. When using the REST-API, please make sure you use the staging area's endpoint for your national node, by using the country code (CC) according to the valid ISO-3166-1 alpha 2 standards<sup>4</sup>.

- For API V2 use: `/api/v2/eu_bbmri_eric_<CC>_<Entity>`
- For the Data API use: `/api/data/eu_bbmri_eric_<CC>_<Entity>`

The entity is either *collections*, *biobanks*, *networks*, *persons*, *also\_known\_in* or *facts*.

## DATA MODEL

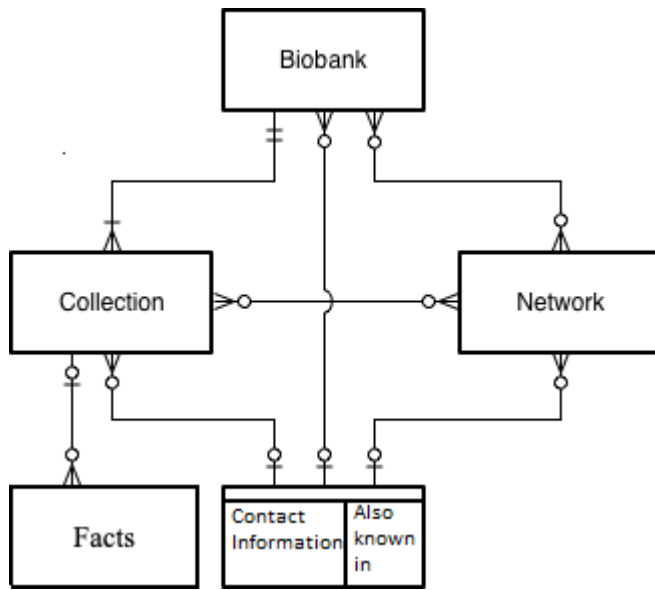
The data model of the directory consists of 6 entities that have to be managed by the national nodes and several look-up lists for the acceptable values for the fields.

1. Contact information: Describes contact information for biobanks, collections or networks
2. Also known in: Describes the source in which the biobank, collection or networks also exists
3. Network: Describes networks of collaborating biobanks
4. Biobank: Describes the biobank organisation
5. Collection: Describes the biobank collections and sub collections
6. Facts: Describes aggregates of samples of the collections

For simplicity, the supporting entities that describe the look-up lists are not described in this manual. The model can be visually represented as follows.

---

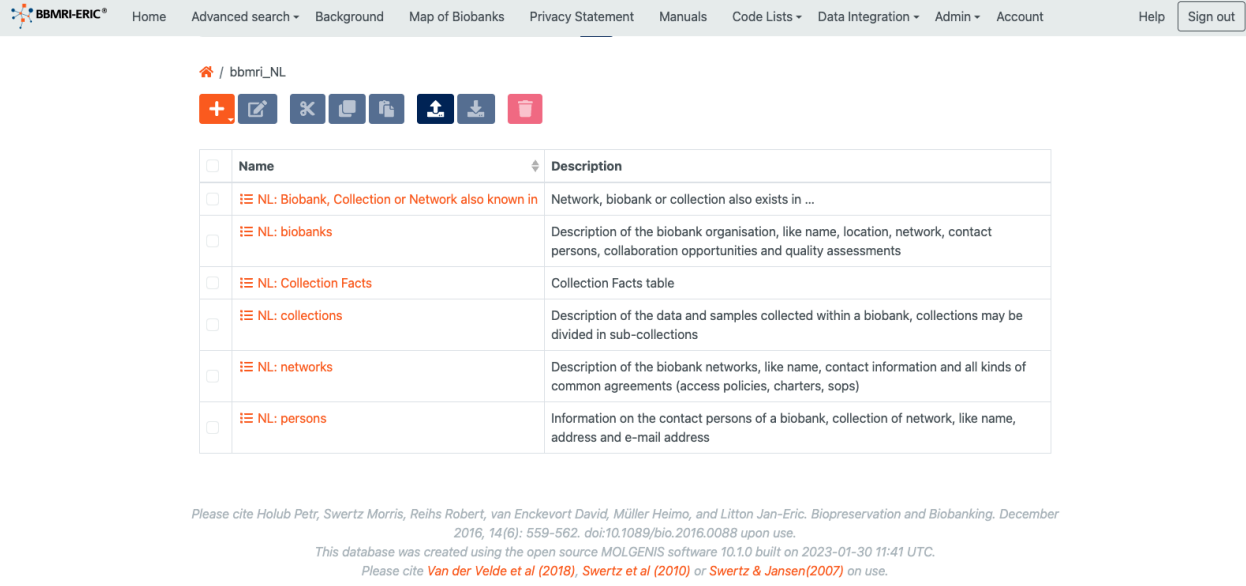
<sup>4</sup> Note: for BBMRI-UK, this code is "UK" instead of the ISO standards prescribed "GB".



This work is part of the ADOPT BBMRI-ERIC project, funded by the European Commission, topic H2020- INFRADEV-3-2015, Grant Agreement Nr 676550

## DOWNLOAD THE LATEST MODEL AND DATA

The latest model and data from your national node staging area can be obtained using the Navigator.



The screenshot shows the BBMRI-ERIC Navigator interface. At the top, there is a navigation bar with links for Home, Advanced search, Background, Map of Biobanks, Privacy Statement, Manuals, Code Lists, Data Integration, Admin, Account, Help, and Sign out. Below the navigation bar, the breadcrumb path is "/ bbmri\_NL". A toolbar contains icons for adding, editing, deleting, and downloading. The main content area displays a table with the following data:

<input type="checkbox"/>	Name	Description
<input type="checkbox"/>	NL: Biobank, Collection or Network also known in	Network, biobank or collection also exists in ...
<input type="checkbox"/>	NL: biobanks	Description of the biobank organisation, like name, location, network, contact persons, collaboration opportunities and quality assessments
<input type="checkbox"/>	NL: Collection Facts	Collection Facts table
<input type="checkbox"/>	NL: collections	Description of the data and samples collected within a biobank, collections may be divided in sub-collections
<input type="checkbox"/>	NL: networks	Description of the biobank networks, like name, contact information and all kinds of common agreements (access policies, charters, sops)
<input type="checkbox"/>	NL: persons	Information on the contact persons of a biobank, collection of network, like name, address and e-mail address

Below the table, there is a citation notice: "Please cite Holub Petr, Swertz Morris, Reihls Robert, van Enckevort David, Müller Helmo, and Litton Jan-Eric. Biopreservation and Biobanking. December 2016, 14(6): 559-562. doi:10.1089/bio.2016.0088 upon use. This database was created using the open source MOLGENIS software 10.1.0 built on 2023-01-30 11:41 UTC. Please cite Van der Velde et al (2018), Swertz et al (2010) or Swertz & Jansen(2007) on use."

1. Click on “Navigator” in the menu.
2. Go to your national node staging area
3. Select all the entities you want to download by clicking the checkboxes left to the entity names
4. Click the download button

This provides you with the latest model including the currently available data from your national node.

This work is part of the ADOPT BBMRI-ERIC project, funded by the European Commission, topic H2020- INFRADEV-3-2015, Grant Agreement Nr 676550

## STRUCTURE OF THE IDENTIFIERS

The Directory is a federated infrastructure, and to prevent collisions in identifiers we have defined a specific structure for the identifiers in the Directory.

Biobank ID: `bbmri-eric:ID:<CC>_<local id>`

Collection ID: `<Biobank ID>:collection:<local id>`

Network ID: `bbmri-eric:networkID:<CC>_<local id>`

Contact ID: `bbmri-eric:contactID:<CC>_<local id>`

Also Known In ID: `bbmri-eric:akiID:<CC>_<local id>`

Facts ID: `bbmri-eric:factID:<CC>_<local id>`

Where <CC> has to be replaced by the valid ISO-3166-1 alpha 2 country code for the country of residence and <local id> with a local ID generated by the national node to be unique within its context. We advise to use the ID of the entity in the national directory if one is present. All local parts should be limited to roman letters and numbers (a-z, A-Z, 0-9).

## BIOBANK

The biobank entity (`eu_bbmri_eric_biobanks`) describes the biobank organisation.

Attribute * = Mandatory	Description	Cardinality	Acceptable values
<b>id *</b>	Unique identifier of this record	1	[a-z][A-Z][0-9][:-_] starting with the prefix <code>bbmri-eric:ID: + ISO-3166-1 alpha 2 <a href="#">country</a> code</code>
<b>name *</b>	Name of the biobank	1	Text, maximum 255 characters, recommended to be less than 60 characters
acronym	Short name or acronym of the biobank if applicable	0..1	Text, maximum 255 characters, recommended to be less than 20 characters
<b>description *</b>	Description of the biobank in English	1	Text
url	URL of the website of the biobank	0..1	URL
location	The city where the biobank is located	0..1	Text, maximum 255 characters, recommended to be less than 60 characters
<b>country *</b>	The country in which the biobank resides	1	ISO-3166-1 alpha 2 <a href="#">country</a> code
latitude	Latitude of the location of the primary site of the biobank	0..1	WGS84 coordinate
longitude	Longitude of the location of the primary site of the biobank	0..1	WGS84 coordinate
head	Person in charge of the biobank.	0..1	Unique identifier of the referenced contact information record
<b>contact *</b>	Contact information of the primary external contact of the biobank	1	Unique identifier of the referenced contact information record
juridical_person	Name of the organisation (legal entity) of the biobank	0..1	Text, maximum 255 characters, recommended to be less than

This work is part of the ADOPT BBMRI-ERIC project, funded by the European Commission, topic H2020- INFRADEV-3-2015, Grant Agreement Nr 676550



network	List of networks in which the biobank participants	0..n	60 characters Comma separated list of unique identifiers of the referenced network information records
also_known	List of also known in which are linked to collection participates	0..n	Comma separated list of unique identifiers of the referenced also known information records
capabilities	List of the capabilities that the biobank can offer to a researcher as a service	0..n	Comma separated list of unique identifiers of the <a href="#">capabilities</a> .
collaboration_commercial	Indication if the biobank is able to participate in commercial collaborations	0..1	true, false
collaboration_non_for_profit	Indication if the biobank is able to participate in collaborations with not-for-profit organisations	0..1	true, false
<i>withdrawn</i>	This biobank is withdrawn from the Directory	0..1	true, false

The biobank entity in the published production version contains some extra attributes, which are filled automatically or by external parties:

Attribute	Description	Cardinality	Acceptable values
<b>pid *</b>	persistent identifier (read only)	1	Text, maximum 255 characters, recommended to be less than 60 characters
quality	Biobank quality assessment (read only)	0..1	Comma separated list of quality assessment marks
<b><i>national_node *</i></b>	The biobanks originates from this national node	1	ISO-3166-1 alpha 2 <a href="#">country</a> code

## COLLECTION

The collection entity (eu\_bbmri\_eric\_collections) describes the data and samples collected in the biobank at the (sub)collection level. The collection can be described with sub collections to provide detailed information on the available materials, diseases or other attributes. A collection can be subdivided on any distinct criterion, but should always maintain strict partitioning (i.e. there should not be overlap between sub collections. Each sample should be represented in only one sub collection of the collection).

*N.B.: In order for COVID-19 collections to be found, the collection must be a member of the COVID-19 network (see network attribute)*

Attribute * = Mandatory	Description	Cardinality	Acceptable values
<b>Descriptives</b>			
<b>id *</b>	The unique identifier of the record	1	[a-z][A-Z][0-9][:-_] starting with the ID of the biobank in which the collection resides + :collection:
<b>name *</b>	Name of the collection	1	Text, maximum 255 characters, recommended to be less than

This work is part of the ADOPT BBMRI-ERIC project, funded by the European Commission, topic H2020- INFRADEV-3-2015, Grant Agreement Nr 676550

acronym	Short name or acronym of the collection if applicable	0..1	60 characters Text, maximum 255 characters, recommended to be less than 20 characters
<b>description *</b>	Description of the collection in English	1	Text
url	URL of the website of the collection	0..1	URL
location	The city where the collection is located	0..1	Text, maximum 255 characters, recommended to be less than 60 characters
<b>country *</b>	The country in which the collection resides	1	ISO-3166-1 alpha 2 <a href="#">country</a> code
latitude	Latitude of the location of the primary site of the collection	0..1	WGS84 coordinate
longitude	Longitude of the location of the primary site of the collection	0..1	WGS84 coordinate
head	Person in charge of the collection.	0..1	Unique identifier of the referenced contact information record
<b>contact *</b>	Contact information of the primary external contact of the collection	1	Unique identifier of the referenced contact information record
withdrawn	This collection is withdrawn from the Directory	0..1	true, false
<b>Belongs to</b>			
parent_collection	Parent collection of which the collection is a part	0..1	Unique identifier of the referenced collection
sub_collection	<i>sub collection of which the collection is a part</i>	0..1	<i>(automatically filled field)</i>
<b>biobank *</b>	The biobank that hosts the collection	1	Unique identifier of the referenced biobank
network	List of networks in which the collection participates	0..n	Comma separated list of unique identifiers of the referenced network information records
also_known	List of also known in which are linked to collection participates	0..n	Comma separated list of unique identifiers of the referenced also known information records
<b>Characteristics</b>			
<b>type *</b>	The type of the sample collection	1	Comma separated list of <a href="#">collection types</a>
<b>data_categories *</b>	The categories of data that are available as part of the collection	1..n	Comma separated list of <a href="#">data types</a>
<b>order_of_magnitude *</b>	Number of samples in the collection expressed as orders of magnitude	1	<a href="#">Integer n</a> , where 10 <sup>n</sup> is the best approximation of the number of samples in the collection (range 0 - 8)
size	Exact size of collection in number of unique sample ID's at the point in time given in the specified timestamp	0..1	Integer ≥ 0
timestamp	Exact timestamp at which the size of the collection as specified in size was determined	0..1	Timestamp in ISO 8601 format (yyyy-mm-ddThh:mm:ss+nnnn), e.g. 2016-11-15T09:53:13+0100, it is acceptable to set the time component to 00:00:00 when not known.
<b>Donor data</b>			
number of donors	Exact number of donors for	0..1	Integer ≥ 0

This work is part of the ADOPT BBMRI-ERIC project, funded by the European Commission, topic H2020- INFRADEV-3-2015, Grant Agreement Nr 676550

	whom there are samples and/or data in the collection		
order_of_magnitude_donors	Number of donors for whom there are samples and/or data in the collection expressed as orders of magnitude	0..n	<a href="#">Integer n</a> , where 10 <sup>n</sup> is the best approximation of the size of the collection (range 0 - 8)
sex	The sex of the individuals whose samples are part of the collection	0..n	Comma separated list of <a href="#">sex</a>
diagnosis_available	Diagnosis available in the collection	0..n	ICD-10 with the prefix <i>urn:miriam:icd:</i> Available <a href="#">diagnosis</a> can denoted by entire chapters (e.g. <i>urn:miriam:icd:l</i> ), blocks (e.g. <i>urn:miriam:icd:A00-A09</i> ), or individual codes. (e.g. <i>urn:miriam:icd:A09</i> . OR Orphanet codes with the prefix <i>ORPHA:</i> (e.g. <i>ORPHA:10</i> ).
age_low	Age of the youngest individual in the collection at the time the sample was taken	0..1	Integer $\geq 0$ and $\leq$ age_high
age_high	Age of the oldest individual in the collection at the time the sample was taken	0..1	Integer $\geq 0$ and $\geq$ age_low,
age_unit	Unit defining age low and age high	0..1	Comma separated list of MIABIS <a href="#">age_units</a> .
<b>Sample data</b>			
materials	The types of biological material present in the collection	0..n	Comma separated list of MIABIS <a href="#">material types</a> . When specified the collection type must include Sample Collection.
storage_temperatures	The temperature at which the samples are stored	0..n	Comma separated list of <a href="#">storage temperatures</a> according to the MIABIS classification
<b>Imaging Data</b>			
body_part_examined	The body part that was the target of examination for the image taken	0..n	Comma separated list of DICOM <a href="#">body part</a> codes (based on SNOMED-CT codes). When specified the collection type must include Image collection.
imaging_modality	The imaging modality used for generating the image	0..n	Comma separated list of DICOM <a href="#">image modality</a> codes. When specified the collection type must include Image collection.
image_dataset_type	The datatype of the images in the collection	0..n	Comma separated list of DICOM image <a href="#">dataset types</a> . When specified the collection type must include Image collection.
<b>Policies</b>			
collaboration_commercial	Indication if the material in the collection is available for use in a commercial context	0..1	true, false
collaboration_non_for_profit	Indication if the material in the collection is available for use in a not-for-profit context	0..1	true, false
data_use	Data access policy/policies in data use ontology format.	0..n	Comma separated list of unique identifiers of the

This work is part of the ADOPT BBMRI-ERIC project, funded by the European Commission, topic H2020- INFRADEV-3-2015, Grant Agreement Nr 676550

commercial_use	Indication if commercial users can request access to samples and/or data	0..1	referenced <a href="#">data use</a> ontology true, false
access_fee	Indication if an access fee is required for access to samples,data or imaging	0..1	Comma separated list of f unique identifiers of <a href="#">access types</a>
access_joint_project	Indication if a joint project is required for access to samples,data or imaging	0..1	Comma separated list of f unique identifiers of <a href="#">access types</a>
access_description	Short description of the conditions for access to samples,data or imaging in English	0..1	Text
access_uri	URL to a detailed description of the access conditions for access to samples,data or imaging	0..1	URL
PD/SOP	Indication if the samples and/or data does has sop in a way off processing, shipment or storage	0..1	Comma separated list of f unique identifiers of <a href="#">SOPs</a>

The collection entity in the published production version contains some extra attributes, which are filled automatically or by external parties:

Attribute	Description	Cardinality	Acceptable values
<b>biobank_label *</b>	The label of the biobank that hosts the collection	1	Text, maximum 255 characters, recommended to be less than 60 characters
combined_network	List of networks in which the collection and biobanks participates (read only)	0..n	Comma separated list of unique identifiers of the referenced network information records
categories	Collection categories based on diseases (read only)	0..1	Comma separated list of unique identifiers of the referenced categories information records
quality	Collection quality assessment (read only)	0..1	Comma separated list of quality assessment marks
combined_quality	Collection and Biobank quality assessment (read only)	0..1	Comma separated list of quality assessment marks
<b>national_node *</b>	The collection originates from this national node	1	ISO-3166-1 alpha 2 <a href="#">country</a> code

This work is part of the ADOPT BBMRI-ERIC project, funded by the European Commission, topic H2020- INFRADEV-3-2015, Grant Agreement Nr 676550

## NETWORK

The network entity (eu\_bbmri\_eric\_networks) describes a biobank network.

Attribute * = Mandatory	attribute	Cardinality	Acceptable values
<b>id *</b>	The unique identifier of the record	1	[a-z][A-Z][0-9][:-_] starting with the prefix bbmri-eric:networkID: + ISO-3166-1 alpha 2 <a href="#">country</a> code
<b>name *</b>	The name of the network	1	Text, maximum 255 characters, recommended to be less than 60 characters
acronym	Short name or acronym of the network, if applicable	0..1	Text, maximum 255 characters, recommended to be less than 20 characters
<b>description *</b>	Description of the network, in English	1	Text
location	The city where the network is located	0..1	Text, maximum 255 characters, recommended to be less than 60 characters
latitude	Latitude of the location of the primary site of the network	0..1	WGS84 coordinate
longitude	Longitude of the location of the primary site of the network	0..1	WGS84 coordinate
also_known	List of also known in which are linked to collection participates	0..n	Comma separated list of unique identifiers of the referenced also known information records
url	URL of the website of the network	0..1	URL,
juridical_person	Name of the organisation (legal entity) of the network	0..1	Text
<b>contact *</b>	Contact information of the primary external contact of the network	1	Unique identifier of the referenced contact information record
parent_network	Parent or larger network that this network is a part of	0..1	Unique identifier of the parent network
common_network_elements	elements who are the same for all biobanks/collections within the network	0..1	Comma separated list of f unique identifiers of <a href="#">common network</a>
withdrawn	This network is withdrawn from the Directory	0..1	true, false

The network entity in the published production version contains some extra attributes, which are filled automatically or by external parties:

Attribute	Description	Cardinality	Acceptable values
<b>national_node *</b>	The network originates from this national node	1	ISO-3166-1 alpha 2 <a href="#">country</a> code

This work is part of the ADOPT BBMRI-ERIC project, funded by the European Commission, topic H2020- INFRADEV-3-2015, Grant Agreement Nr 676550

## CONTACT INFORMATION

The contact information entity (eu\_bbmri\_eric\_persons) describes the contact information for a biobank, collection or network.

Attribute * = Mandatory	Description	Cardinality	Acceptable values
<b>id *</b>	Unique identifier of this record	1	[a-z][A-Z][0-9][:-_], starting with bbmri-eric:contactID: + ISO-3166-1 alpha 2 <a href="#">country</a> code.
title_before_name	Titles that are prefixed to the name	0..1	Text, maximum 255 characters, recommended to be less than 60 characters
first_name	First name of the person to contact	0..1	Text, maximum 255 characters, recommended to be less than 60 characters
last_name	Last name of the person to contact	0..1	Text, maximum 255 characters, recommended to be less than 60 characters
title_after_name	Titles that are appended to the name (e.g. MD, PhD)	0..1	Text, maximum 255 characters, recommended to be less than 60 characters
<b>email *</b>	E-mail address to contact	1	text
phone	Telephone number including international prefix	0..1	Compliant to the E.123 norm, international notation including the international access number and using spaces to visually separate groups of numbers, e.g. +31 20 1234567
address	Address including routing information where necessary	0..1	text
zip	ZIP or postal code	0..1	text
city	City	0..1	text
<b>country*</b>	Country	1	ISO-3166-1 alpha 2 <a href="#">country</a> code
role	Official role of the person	0..1	text
<i>biobanks</i>	<i>Person is linked to this biobank</i>	<i>0..1</i>	<i>(automatically filled field)</i>
<i>collections</i>	<i>Person is linked to this collection</i>	<i>0..1</i>	<i>(automatically filled field)</i>
<i>network</i>	<i>Person is linked to this network</i>	<i>0..1</i>	<i>(automatically filled field)</i>

The “contact-information” entity in the published production version contains some extra attributes, which are filled automatically or by external parties:

Attribute	Description	Cardinality	Acceptable values
<b>national_node *</b>	The contact originates from this national node	1	ISO-3166-1 alpha 2 <a href="#">country</a> code

This work is part of the ADOPT BBMRI-ERIC project, funded by the European Commission, topic H2020- INFRADEV-3-2015, Grant Agreement Nr 676550

## ALSO KNOWN IN

The “also known in” entity (`eu_bbmri_eric_also_known_in`) describes in where the network, biobank or collection also exists.

Attribute * = Mandatory	Description	Cardinality	Acceptable values
<code>id</code> *	<i>Unique identifier of this record</i>	1	<i>[a-z][A-Z][0-9][:-_], starting with <code>bbmri-eric:akid</code>: + ISO-3166-1 alpha 2 <a href="#">country</a> code. (automatically generated field)</i>
<code>name_system</code> *	Name of the source in which the biobank, collection or network also exists	1	Text, maximum 255 characters, recommended to be less than 60 characters
<code>pid</code>	Persistent Identifier of the biobank, collection or network in the other source	0..1	Text, maximum 255 characters, recommended to be less than 60 characters
<code>url</code> *	Link to the biobank, collection or network in the other source	1	URL

The “also known in ” entity in the published production version contains some extra attributes, which are filled automatically or by external parties:

Attribute	Description	Cardinality	Acceptable values
<code>national_node</code> *	The “also known in” originates from this national node	1	ISO-3166-1 alpha 2 <a href="#">country</a> code

This work is part of the ADOPT BBMRI-ERIC project, funded by the European Commission, topic H2020- INFRADEV-3-2015, Grant Agreement Nr 676550

## STAR MODEL

In order to generate aggregated data from individual level data the “fact table/ star model” can be used. Which is a special form of a *Dimensional Data Model* developed.

The star schema is the simplest form or building block of a dimensional data model organizing data into facts and dimensions. A fact table contains the numeric measures produced by an operational measurement event. Within the Star Model data schema, a fact is an itemized entity that is countable or measureable and is characterized by a number of attributes which are grouped into so-called dimensions(attributes in the table below). The MIABIS Star Model selects only the most important data elements from the MIABIS core components (see table below).

The fact table provides a tool for researchers to run feasibility queries and find suitable donors, samples, and/or data for their research question.

Queries like:

- the number of donors with certain characteristics
- the number of samples with certain characteristics
- the number of data sets with certain characteristics
- the number of collections with certain characteristics

or arbitrary combinations of these can be answered with the implemented data model at MIABIS core level.

*REMARK*“Because of the adopted method of data creation and collection the number of donors presented in the table below should not be added as it may give the wrong sums”

The Star Model entity “eu\_bbmri\_eric\_facts” describes the individual level data of the collection

Attribute * = Mandatory	Description	Cardinality	Acceptable values
<i>id</i> *	Unique identifier of this record	1	[a-z][A-Z][0-9][:-_], starting with <i>bbmri-eric:factID:</i> + ISO-3166-1 alpha 2 <a href="#">country code</a> .
<b>collection</b> *	The collection that hosts the facts	1	Unique identifier of the referenced collection
sex	The sex of the individuals whose samples are part of the collection	0..1	List of <a href="#">sex</a>
age	Unit defining age of sample donor at time of sample donation	0..1	List of <a href="#">age range</a> .
sample_type	The types of biological material	0..1	list of MIABIS <a href="#">material types</a> . When specified the facts type must include Sample Collection.
disease	Diagnosis available in the collection	0..1	ICD-10 with the prefix <i>urn:miriam:icd:</i> Available <a href="#">diagnosis</a> Use only the individual codes. (e.g. urn:miriam:icd:A09. OR Orphanet codes with the prefix <i>ORPHA:</i> (e.g. ORPHA:10). Do not use entire chapters (e.g. urn:miriam:icd:!), blocks

This work is part of the ADOPT BBMRI-ERIC project, funded by the European Commission, topic H2020- INFRADEV-3-2015, Grant Agreement Nr 676550



			(e.g. urn:miriam:icd:A00-A09)
number_of_samples	Exact number of samples and/or data in the collection	0..1	Integer ≥ 0
number_of_donors	Exact number of donors for whom there are samples and/or data in the collection	0..1	Integer ≥ 0
last_update*	Date the fact information was last updated in the source system	1	YYYY-MM-DD

## CODE LISTS

The Directory uses code lists to specify the acceptable options where there is a limited choice. Options can be referenced by their identifier and if an attribute can contain only one value the attribute value will be directly set to this identifier. [Code lists](#) themselves are maintained by BBMRI-ERIC and cannot be updated by a National Node's data manager.

The tables in this document are for reference only. For the actual values, please refer to the specification at <https://directory.bbmri-eric.eu/> and click the right table below the menu item CODE LISTS

## DISEASE TYPES

The [disease type](#) is designed to be an extensible, ontology based code list. Currently it comprises the ICD-10 and Orphanet ontologies, but in the future other disease ontologies such as SNOMED CT might be included. To support this extensibility the ICD-10 codes have been prefixed with the URI scheme urn:miriam:icd: (e.g. urn:miriam:icd:C19, urn:miriam:icd:C00-C97) as specified by the Miriam registry team EMBL-EBI. The disease types table contains entries for all levels of the ICD-10 ontology, chapters, blocks and codes and sub codes. Data integrators should refrain from creating other blocks of contiguous codes that are not specified in ICD-10. For sources that contain disease codes in another ontology data integrators should convert these to the corresponding ICD-10 codes. Both SNOMED-CT and Orphanet maintain mappings for their ontologies. It is strongly advised to provide individual codes instead of blocks of codes. For the Orphanet codes the prefix ORPHA: can be used (e.g. ORPHA:10).

## QUALITY

To secure the data quality, checks are performed [within the system](#). During the upload error messages will help correct your data input if for example ID's are not correct. Frequently BBMRI will also give [warnings about the completeness](#) of the data entry. This will be done by a python script and send afterwards to the National nodes. Next to the quality of the data input BBMRI provides also information about the quality of biobanks, collections. BBMRI head office can assign certificates to give an indication about [qualities and standards](#).

### QUALITY CHECKS DATA INPUT

The table below shows the checks

Tables	Attributes	Description of the check	Check time/phase	Error message
persons	id	check if the ID is correct like the <a href="#">structure of Identity</a>	When entering data in the database (either manually, through a file or via API) and in the process of combining the data from National Nodes into the BBMRI-ERIC Directory	change ID
networks	id	check if the ID is correct like the <a href="#">structure of Identity</a>	When entering data in the database (either manually, through a file or via API) and in the process of combining the data from National Nodes into the BBMRI-ERIC Director	change ID
biobank	id	check if the ID is correct like the <a href="#">structure of Identity</a>	When entering data in the database (either manually, through a file or via API) and in the process of combining the data from National Nodes into the BBMRI-ERIC Director	change ID
collections	id	check if the ID is correct like the <a href="#">structure of Identity</a>	When entering data in the database (either manually, through a file or via API) and in the process of combining the data from National Nodes into the BBMRI-ERIC Director	change ID
collections	age_unit	If there is only 1 age unit selected	between migration National nodes and directory	mail to NN, contact biobank about more than one age_unit, data can not published
collections	age high / age low	if the age high is higher than the age low	In the process of combining the data from National Nodes into the BBMRI-ERIC Director	mail to NN, contact biobank about more than one age_unit, data can not published

### WARNINGS DATA INPUT

The quality checks /warnings available through python scripts pay attention to attributes at different levels (biobank, collection, etc) taking care of different aspects (access policies, geolocation, etc) and raising the warnings (in *italics*) when the data provided are not compliant with the data model. Please find here the [complete list of warning messages](#).

This work is part of the ADOPT BBMRI-ERIC project, funded by the European Commission, topic H2020- INFRADEV-3-2015, Grant Agreement Nr 676550

## QUALITY MARKS

The collection entity (eu\_bbmri\_eric\_collections) describes the data and samples collected in the biobank at the (sub)collection level. The collection can be described with sub collections to provide detailed information on the available materials, diseases or other attributes. A collection can be subdivided on any distinct criterion, but should always ma

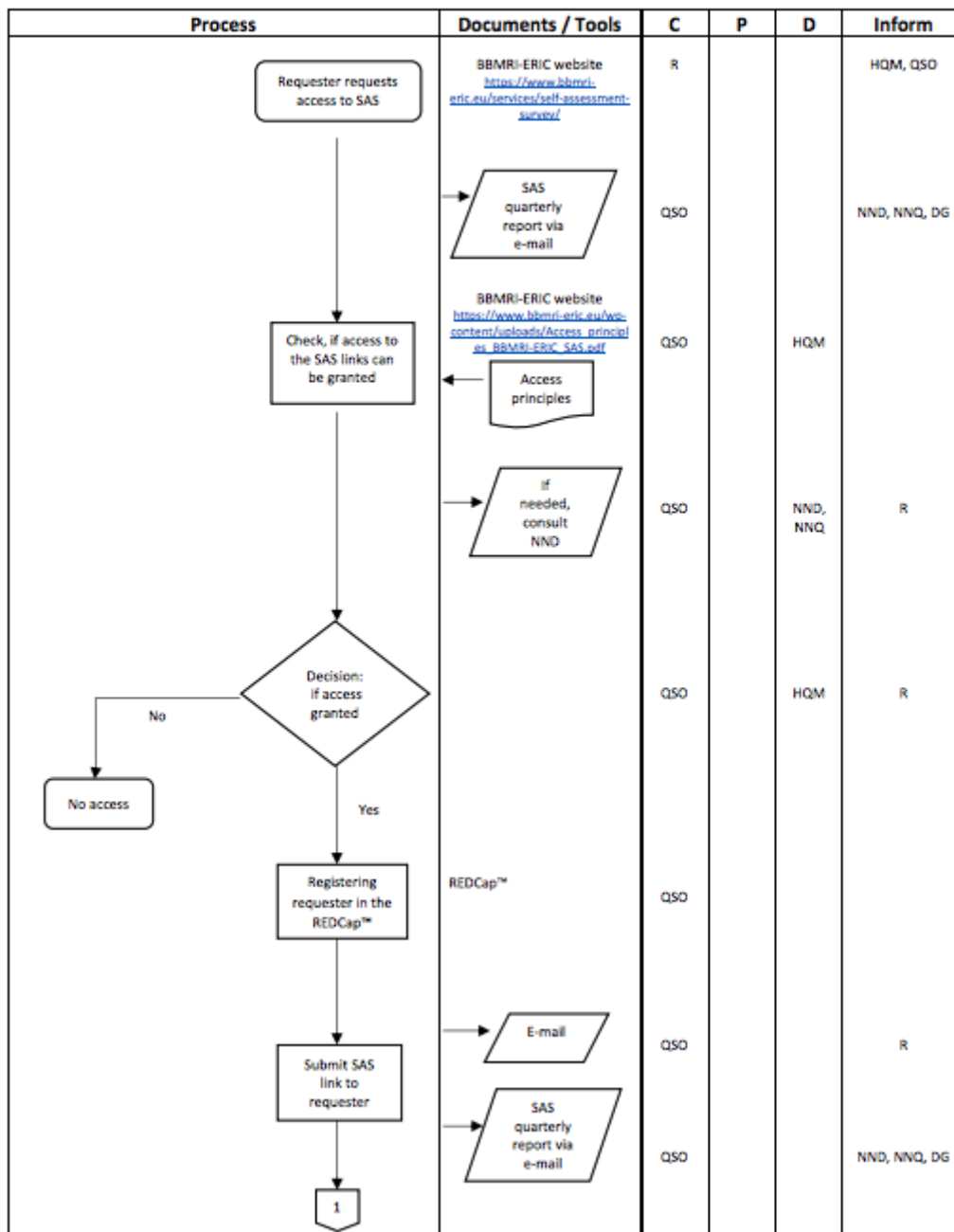
The results of the BBMRI-ERIC Self Assessment Survey, third-party certifications and Audits, shown as Q-marks in the Directory are managed by the BBMRI.QM team. You will not be able to update these values yourself. request a quality mark via

<https://www.bbmri-eric.eu/services/self-assessment-survey/>

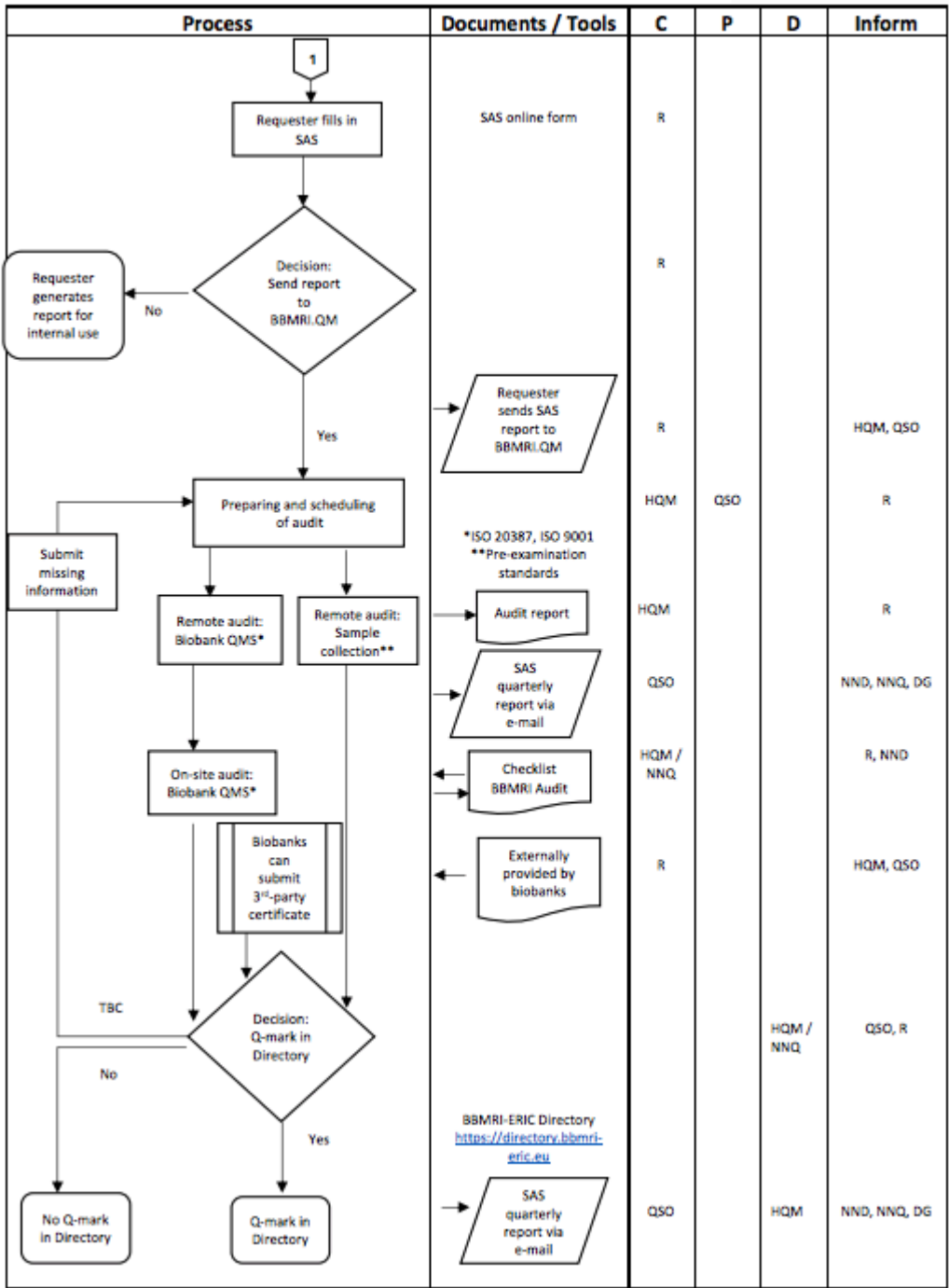
The document below shows the workflow.

Follow the workflow below to add quality marks.

### Process Flow: Q-Assessment Scheme for Biobanks and Sample Collections



This work is part of the ADOPT BBMRI-ERIC project, funded by the European Commission, topic H2020- INFRADEV-3-2015, Grant Agreement Nr 676550



This work is part of the ADOPT BBMRI-ERIC project, funded by the European Commission, topic H2020- INFRADEV-3-2015, Grant Agreement Nr 676550








**Legend:**

R – Requester	C – Conduct
DG – Director General	D – Decide
NND – National Node Director	P – Participate
NNQ – National Node Quality Officer	Q-mark – Quality mark
HQM – Head of Quality Management	SAS – Self-Assessment Survey
QSO – Quality Management Service Officer	REDCap™ – online software for creation of Self-Assessment Survey
	TBC – To be clarified

\* ISO 20387:2018 Biotechnology – Biobanking- General requirements for biobanking;  
ISO 9001: 2015 Quality managements systems – Requirements;

- \*\* Specifications for pre-examinations processes for:
- frozen tissue (FT) Part 1: Isolated RNA (ISO 20184-1:2018)
  - frozen tissue (FT) Part 2: Isolated proteins (ISO 20184-2:2018)
  - FFPE tissue Part 1: Isolated RNA (ISO 20166-1:2018)
  - FFPE tissue Part 2: Isolated proteins (ISO 20166-2:2018)
  - FFPE tissue Part 3: Isolated DNA (ISO 20166-3:2018)
  - venous whole blood Part 1: Isolated cellular RNA (ISO 20186-1:2019)
  - venous whole blood Part 2: Isolated genomic DNA (ISO 20186-2:2019)
  - venous whole blood Part 3: Isolated circulating cell free DNA from plasma (ISO 20186-3:2019)
  - metabolomics in urine, venous blood serum and plasma (CEN 16945:2016)
  - will be continuously updated

**Description process flow symbols:**

Process Flow Symbols	Meaning
	Start / End
	Activity
	Decision
	Connection to another process
	Document
	Verbal information / information via telephone, e-mail ....
	Connector

### Latest document revisions

#### **Version 3.5.6**

**Date:** 20/04/2020

**Author:** Marije van der Geest

**Edits:** Added COVID-19 items (*covid19biobank* and *COVID-19*) for biobank and collections table. Added code lists for *Services provided by the biobank* and *Relevant data and products*.

**Date:** 29/04/2020

**Author:** Marije van der Geest

**Edits:** Updated lists *Sex*, *Material types*, *Lab standards* & *Operational standards*.

#### **Version 3.6**

**Date:** 02/07/2020

**Author:** Marije van der Geest

**Edits:** Added section: *Download latest model and data*.

#### **Version 3.6.1**

**Date:** 07/07/2020

**Author:** Marije van der Geest

**Edits:** Added information for national nodes and local directory instances.

#### **Version 3.6.2**

**Date:** 28/07/2020

**Author:** Marije van der Geest

**Edits:** Added Orphanet codes description for *diagnosis\_available*, COVID-19 collection information and information about NN endpoints for using the REST-API

#### **Version 3.6.3**

**Date:** 02/12/2020

**Author:** Brenda Hijmans

**Edits:** Added DUO code description, changed order of some attributes in collections to match order in Molgenis

#### **Version 3.6.4**

**Date:** 26/4/2021

**Author:** Aneas Hodselmans

**Edits:** Delete "contact priority" in biobanks, collections and networks, Deleted standard in collections and biobanks. Added quality workflow and changed the list of lab/operational standards.

#### **Version 3.6.5**

**Date:** 11/5/2021

**Author:** Esther van Enckevort

**Edits:** Updated documentation for the *order\_of\_magnitude* and *order\_of\_magnitude\_donor* fields to specify the acceptable range.

#### **Version 3.6.6**

**Date:** 7/7/2022

**Author:** Aneas Hodselmans

**Edits:** New application pictures and reference to the codelist tabel via de menu

#### **Version 3.6.7**

**Date:** 14/2/2023

**Author:** Aneas Hodselmans

**Edits:** qualities and warning messages, codelists via menu, [proposal 1](#) and [proposal 2](#)

#### **Version 3.6.8**

**Date:** 22/5/2023

This work is part of the ADOPT BBMRI-ERIC project, funded by the European Commission, topic H2020- INFRADEV-3-2015, Grant Agreement Nr 676550

**Author:** Aneas Hodselmans  
**Edits:** Add starmodel & network improvements  
**Version 3.6.9**  
**Date:** 18/10/2023  
**Author:** Aneas Hodselmans  
**Edits:** Cardinality

**Version 3.6.10**  
**Date:** 1/11/2023  
**Author:** Aneas Hodselmans  
**Edits:** Cardinality

**Version 3.6.11**  
**Date:** 22/01/2024  
**Author:** Aneas Hodselmans and Dieuwke Roelofs-Prins  
**Edits:** Withdrawn adjustments and update of some pictures

**Version 3.6.12**  
**Date:** 29/03/2024  
**Author:** Aneas Hodselmans and Dieuwke Roelofs-Prins  
**Edits:** Pagenumbers