

Biobank Data Standard

**Collecting, storing and sharing data
describing human biological material
for research**

Version 1



Introduction

This data standard has been developed to allow biobanks across the UK to share information about the samples they hold. This will allow a central portal to be created and made available to researchers so as to reduce the barriers to finding and accessing samples for research.

Wherever possible, existing standards have been used. These are indicated within the Source column¹.

In some tables, there is a field described as 'Disease Specific Data'. This is to recognise there may be places where additional data may be required for certain diseases. The aim of this standard however is to be independent of any specific disease, whilst recognising extensions may be desirable to capture disease-specific information.

Minimum Standard:

As a minimum the biobank should supply the data specified marked as Mandatory for All and Mandatory for Minimum Standard. No entry should be left empty. Those that are unknown or unavailable should be marked as such.

Best Practice:

The Biobank should provide data using the Patient hierarchy rather than the Collection hierarchy, allowing greater information to be made available to the researcher. No entry should be left empty. Those that are unknown or unavailable should be marked as such.

List of tables

Table Name	Requirement
Biobank	Mandatory for All
AvailableData	Mandatory for All
DiagnosisCode	Mandatory for All
OrganCode	Mandatory for All
Collection	Mandatory for Minimum Standard
SampleDetails	Mandatory for Minimum Standard
Collection	Mandatory for Minimum Standard
Collection_AvailableData	Mandatory for Minimum Standard
SampleDetail	Mandatory for Minimum Standard
Patient	Mandatory for Best Practice
PatientDiagnosis	Mandatory for Best Practice
SampleGroup	Mandatory for Best Practice
SolidSpecimen	Mandatory for Best Practice
FluidSample	Mandatory for Best Practice
TissueSample	Mandatory for Best Practice
Sample_AvailableData	Mandatory for Best Practice

¹ MIABIS <http://bbmri-wiki.wikidot.com/en:dataset>; BCCTB <http://www.breastcancertissuebank.org/>; STRATUM <http://stratumbiobanking.org/>

Table: Biobank		This table is to capture key information about a biobank, in particular the owner, contact details and the collections it holds.	
Field Name	Format	Description	Source
BiobankID	Free Text	Textual string of letters starting with the country code (according to standard ISO1366 alpha2) followed by the underscore “_” and post-fixed by a biobank ID or name specified by its juristic person (nationally specific)	MIABIS-01
Name of biobank	Free text	Textual string of letters denoting the name of the biobank in the local language.	MIABIS-02
Juristic person	Free text	Textual string of letters denoting the juristic person e.g. a university, concern, county council etc. for the biobank	MIABIS-03
URL	Free text	Textual string of letters with the complete http-address for the biobank URL	MIABIS-04
Country code	ISO-standard (3166 alpha2), two letter code.	Textual string of letters of the two letter code for the country of the biobank according to ISO-standard 3166 alpha2.	MIABIS-05
Contact person	Free text	Textual string of letters denoting the name of the contact person for the biobank	MIABIS-07
Contact phone	Free text	Phone to the "Contact person", including international call prefix	MIABIS-08
Contact email	Free text	Email address of the "Contact person"	MIABIS-09
Contact department	Free text	Department, or corresponding (e.g., division), of affiliation of the "Contact person"	MIABIS-10
Contact address	Free text	Street name and street number or PO Box of the "Contact person"	MIABIS-11
Contact ZIP / Post Code	Free text	ZIP/Post Code of the "Contact person"	MIABIS-12
Contact city	Free text	City of the "Contact person"	MIABIS-13
Contact country	ISO-standard (3166 alpha2), two letter code	Country of the "Contact person"	MIABIS-14
Last Updated	ISO-standard (8601) time format (YYYY-MM-DD)	Date in ISO-standard (8601) time format when data about the biobank was last updated in a database.	MIABIS-17

Table: AvailableData		Linked from the Collections and Specimen tables, it details what other data is available on the samples/collections	
Field Name	Format	Description	Source
Data Type	Free text	Textual string of letters detailing the type of the data e.g (Comorbidity, Clinical, Pathology, Omics)	New
Data Availability	Free text	Textual String of one of the following: Available to all / Available only through collaboration / Other	New

Table: DiagnosisCode		This table is used to capture the diagnosis coding that is used by the biobank.	
Field Name	Format	Description	Source
Code	Free Text	Textual representation of the code	New
Description	Free Text	Textual description of the diagnosis	New
Version	Free Text	Textual representation of the version (e.g. ICD-10)	New

Table: OrganCode		This table is used to capture the organ coding that is used by the biobank.	
Field Name	Format	Description	Source
Code	Free Text	Textual representation of the code	New
Description	Free Text	Textual description of the diagnosis	New
Version	Free Text	Textual representation of the version	New

Table: Collection		Where a biobank does not have the ability to supply information on patients and associated samples, it can use this table to give aggregated data on the collections they hold.	
Field Name	Format	Description	Source
Sample Collection/ Study ID	Free Text	Textual string depicting the unique ID or acronym for the sample collection or study	MIABIS-18
Study name	Free Text	Textual string of letters denoting the name of the study in English.	MIABIS-19
Description	Free Text	Textual string of letters describing the sample collection or study aim (max 200 characters).	MIABIS-20
Type of Collection	Free Text	Case-control, Cohort, Cross-sectional, Longitudinal, Twin-study, Quality control, Population-based, Disease specific, Other	MIABIS-30
Collection start	ISO-standard (8601) time format (YYYY-MM-DD)	Date in ISO-standard (8601) time format specifying when the sample collection starts	MIABIS-31
Collection end	ISO-standard (8601) time format (YYYY-MM-DD)	Date in ISO-standard (8601) time format specifying when the sample collection ends, if applicable	MIABIS-32
Sex	Free Text	Textual string of letters denoting the sex of the sample donors. Can be several values	MIABIS-35
Current sampled individuals	Integer	Number of individuals with biological samples in the sample collection/study at the date of Last updated.	MIABIS-48
Last updated	ISO-standard (8601) time format (YYYY-MM-DD)	Date in ISO-standard (8601) time format when data about the sample collection was last updated in a database	MIABIS-51
Access Conditions	Free Text	Textual String describing the access conditions to the collection	New
Collection Status	Free Text	Textual String describing the collection status of the collection	New
Hosting biobank	BiobankID	Textual string of letters of the biobank storing the biological samples that are part of the sample collection.	MIABIS-52

Table: Collection_AvailableData		This table is used to indicate what data is available for the collection.	
Field Name	Format	Description	Source
Collection.StudyID	Free Text	The identifier used in the Collection table	New
Data Type	Free Text	The data type, must match an entry in the available data table	New

Table: SampleDetail		Linked from the Collections table, this allows the aggregated data about samples to be recorded	
Field Name	Format	Description	Source
Material Type	Free Text	Most commonly abundant biological samples in biobanks. Can be several values.	MIABIS-41
Macroscopic Assessment	Free Text	Where applicable a textual string of letters detailing the macroscopic assessment of the sample e.g (Tumour, Normal)	BCCTB
Current sampled individuals	Integer	Number of individuals with biological samples in the sample collection/study at the date of Last updated (also see Planned sampled individuals)	MIABIS-48
OrganCode	Free Text	The code of the Organ. It must match an entry in the Organ table.	New
DiagnosisCode	Free Text	The code of the Diagnosis (if applicable). It must match an entry in the Diagnosis table.	New
FromCollection	Sample Collection/ Study ID	Textual string depicting the unique ID or acronym for the sample collection or study	New

Table: Patient		This captures the core information about any patient within a biobank	
Field Name	Format	Description	Source
AnonymisedID	Free Text	An identifier, made of either letters, numbers or both, as an anonymised identifier for the patient	BCCTB
Sex	Free Text	Textual string of letters denoting the sex of the sample donors.	MIABIS-35
DiseaseSpecificData	To be defined	An expectation that some diseases will extend the patient table.	NA

Table: PatientDiagnosis		This table is used to indicate the diagnosis for the patient.	
Field Name	Format	Description	Source
DiagnosisID	Free Text	An identifier, made of either letters, numbers or both, as an anonymised identifier for the diagnosis	BCCTB
DiagnosisDate	ISO-standard (8601) time format (YYYY-MM-DD)	Date in ISO-standard (8601) time format when the diagnosis was made.	BCCTB
Patient AnonymisedID	Free Text	The identifier used in the Patient table	BCCTB
DiagnosisCode	Free Text	Textual description of the diagnosis	BCCTB

Table: SampleGroup		A Sample Group is defined as samples collected from the same patient on the same day	
Field Name	Format	Description	Source
Identifier	Free Text	Textual string of letters and/or numbers denoting the identifier of the sample group.	BCCTB
Date Collected	ISO-standard (8601) time format (YYYY-MM-DD)	Date in ISO-standard (8601) time format when the samples were banked	BCCTB
Consent Details	Free Text	Textual string of letters denoting the consent details of the collection.	BCCTB
PatientDiagnosis DiagnosisID	Free Text	The ID of the diagnosis that is most relevant to the samples, it much match one in the PatientDiagnosis table	BCCTB
Age	Free Text	The age of the patient at the time of the sample. In groups: 0-5, 6-12, 13-17, 18-20, 21-30, 31-40, 41-50, 51-60, 61-70, 71-80, 81-90, 91-100, 101-110, 111-120.	STRATUM
DiseaseSpecificData	To be defined	An expectation that some diseases will extend the sample group table.	NA
Patient. AnonymisedID	Free Text	The anonymisedID of the patient that the SampleGroup belongs	BCCTB

Table: SolidSpecimen		A SolidSpecimen captures the information specific to the master specimen that is then sub-divided into smaller aliquots/samples	
Field Name	Format	Description	Source
Identifier	Free Text	Textual string of letters and/or numbers denoting the identifier of the sample group.	BCCTB
OrganCode	Free Text	The code of the Organ. It must match an entry in the Organ table.	BCCTB
Location	Free Text	Where applicable, the textual string of letters denoting the location of the organ (e.g. Left/Right)	EUROC-30
Disease Specific Data	NA	See other data standards	Multiple
SampleGroup identifier	Free Text	The identifier of the SampleGroup that the SolidSpecimen belongs. This must match an ID in the SampleGroup table.	BCCTB

Table: FluidSample		Captures information about fluid samples	
Field Name	Format	Description	Source
Barcode	Free Text	An identifier, made of either letters, numbers or both, as an anonymised identifier for the sample	BCCTB
Material Type	Free Text	Textual string of letters detailing the type of the sample e.g (Whole Blood, Serum)	MIABIS41
Availability	Free Text	Textual string to represent the availability of the sample	BCCTB
TimeToFreeze	Time (in minutes)	The number of minutes from the blood being taken to frozen.	New
StorageTemperature	Temperature (in Celsius)	The storage temperature of the samples in Celsius.	MIABIS-47
FreezeMethod	Free Text	Textual description describing how the samples were frozen (e.g. snap frozen)	New
Volume	Volume in ml	The volume of the sample.	BCCTB
ExtractionMethod	Free Text	Textual string of letters detailing how the blood was taken.	BCCTB
ProcessingMethod	FreeText	Textual description describing any processing performed.	New
StorageMedium	FreeText	Textual description of any storage medium added to the sample	new
Disease Specific Data	NA	See other data standards	Multiple
SampleGroup identifier	Free Text	The identifier of the SampleGroup that the SolidSpecimen belongs. This must match an ID in the SampleGroup table.	BCCTB

Table: TissueSample		A Tissue Sample captures the information about a tissue sample	
Field Name		Description	Source
Barcode	Free Text	An identifier, made of either letters, numbers or both, as an anonymised identifier for the sample	BCCTB
Material Type	Free Text	Textual string of letters detailing the type of the sample e.g (Whole Blood, Serum)	MIABIS41
MacroscopicAssesment	Free Text	Textual string of letters detailing the macroscopic assessment of the sample where applicable e.g. (normal, tumour)	
Availability	Free Text	Textual string to represent the availability of the sample	BCCTB
TimeToFreezeFromExcision	Time (in minutes)	The number of minutes from excision to freezing.	New
TimeToFreezeFromCeaseBloodFlow	Time (in minutes)	The number of minutes from the cut-off of blood flow to the sample to freezing.	New
StorageTemperature	Temperature (in Celsius)	The storage temperature of the samples in Celsius.	MIABIS-47
FreezeMethod	Free Text	Textual description describing how the samples were frozen (e.g snap frozen)	New
StorageMedium	FreeText	Textual description of any storage medium added to the sample (e.g. RNALater).	New
Disease Specific Data	NA	See other data standards	Multiple
SolidSpecimen identifier	Free Text	The identifier of the SolidSpecimen that the TissueSample belongs. This must match an ID in the SolidSpecimen table.	BCCTB

Table: Sample_AvailableData		This table is used to indicate what data is available for a tissue or fluid sample.	
Field Name	Format	Description	Source
barcode	Free Text	The identifier used in the Tissuesample or FluidSample table.	New
Data Type	Free Text	The data type, must match an entry in the available data table	New