

UPLOADING DATA TO SAMPLE CATALOGUE

(We) simply upload what Excel/CSV you have.

Then you will have your own private 'catalogue' area (as basis for conversion)

Mary Wang

Fondazione Telethon

mwang@telethon.it

David van Enkevort

University Medical Center Groningen

david.van.enkevort@umcg.nl

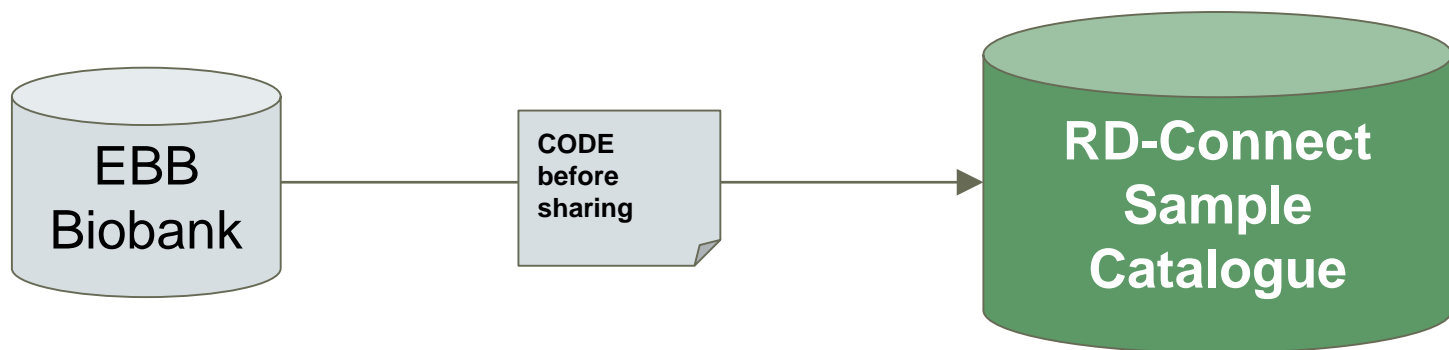


For each biobank we must:



2

1. **Upload** data to from local biobank to central search
2. **Encode** data in same structure so it can be searched/compared





1. Federate from biobank to catalogue

3

1. Simply take your sample list

PatientID	Birthdate	Residence	Sampletype	Status	Gender	Phenotype	Disease	Molecularanal	Karyotype	Diagnosis	Export
7790	13/11/83	GENOVA GE	DNA	Proband	M	Affected	CHROMOSOMAL	AVAILABLE	AVAILABLE	Molecular, Cytogenetic	Yes
7790	13/11/83	GENOVA GE	Leukocytes	Proband	M	Affected	CHROMOSOMAL	AVAILABLE	AVAILABLE	Molecular, Cytogenetic	Yes
7790	13/11/83	GENOVA GE	Lymphoblastoid	Proband	M	Affected	CHROMOSOMAL	AVAILABLE	AVAILABLE	Molecular, Cytogenetic	Yes
7791	29/09/82	GENOVA GE	DNA	Proband	M	Affected	CHROMOSOMAL	AVAILABLE	AVAILABLE	Molecular, Cytogenetic	Yes
7791	29/09/82	GENOVA GE	Leukocytes	Proband	M	Affected	CHROMOSOMAL	AVAILABLE	AVAILABLE	Molecular, Cytogenetic	Yes

2. Add description of your attributes (see MOLGENIS 'EMX' docs)

3. Upload the file

1 Upload file

2 Options

3 Packages

4 Validation

5 Result

Upload a file

RING_14_9sept2015_EMX.xlsx

← Previous

Restart

Next →



2. Encode samples in same standard so they be searched in a comparable way



4

Pseudonomized Sample ID *	Text
Material Type *	Value list
Anatomical Site *	Value list
Sex *	Male / Female
Disease *	Orphanet code
Diagnosis Type	Value list
Age at Sampling	Number
Age at Death	Number
Age at Diagnosis	Number
Age at Remission	Number
Affected	Yes / No / Empty

Family members available	Yes / No / Empty
Related samples available	Yes / No / Empty
Genotype data available	Yes / No / Empty
Registry	Yes / No / Empty
Sample Collection	hyperlink
Hosting Biobank *	Registry & Biobank Finder ID
Registry	Registry & Biobank Finder ID
Pseudonomized Patient ID *	Text

* Required

Template downloadable from:

<http://rd-connect.eu/biosamples-data/sample-catalogue/>

<https://samples.rd-connect.eu/>



Example: upload biobank

5

	A	B	C	D	E	F	G	H	I	J	K	L
1	ID	ParticipantID	SampleID	Sex	AgeAtSampling	MaterialType	Orpha	DiagnosisType	Affected	Family	Genotype	LastUpdated
2	patient1:sample-1	patient1	sample-1	M	42	serum	232	mol	1	1	1	11/04/17
3	patient1:sample-2	patient1	sample-2	M	43	DNA	232	mol	1	1	1	11/04/17
4	patient1:sample-3	patient1	sample-3	M	43	RNA	232	mol	1	1	1	11/04/17
5	patient2:sample-1	patient2	sample-1	M	30	serum	170	clin	1	1	1	11/04/17
6	patient2:sample-2	patient2	sample-2	M	30	plasma	170	clin	1	1	1	11/04/17
7	patient2:sample-3	patient2	sample-3	M	31	urine	170	clin	1	1	1	11/04/17
8	patient2:sample-4	patient2	sample-4	M	32	blood	170	clin	1	1	1	11/04/17
9	patient3:sample-1	patient3	sample-1	F	45	blood	310	clin	1	1	1	11/04/17
10	patient3:sample-2	patient3	sample-2	F	46	plasma	310	clin	1	1	1	11/04/17
11	patient3:sample-3	patient3	sample-3	F	46	serum	310	clin	1	1	1	11/04/17
12	patient3:sample-4	patient3	sample-4	F	47	urine	310	clin	1	1	1	11/04/17
13	patient4:sample-1	patient4	sample-1	M	42	blood	662	clin	1	0	0	11/04/17
14	patient4:sample-2	patient4	sample-2	M	45	blood	662	clin	1	0	0	11/04/17
15	patient4:sample-3	patient4	sample-3	M	46	blood	662	clin	1	0	0	11/04/17
16	patient4:sample-4	patient4	sample-4	M	47	blood	662	clin	1	0	0	11/04/17
17	patient5:sample-1	patient5	sample-1	F	34	serum	461	mol	0	1	1	11/04/17
18	patient5:sample-2	patient5	sample-2	F	35	serum	461	mol	0	1	1	11/04/17
19	patient5:sample-3	patient5	sample-3	F	35	plasma	461	mol	0	1	1	11/04/17
20	patient5:sample-4	patient5	sample-4	F	37	serum	461	mol	0	1	1	11/04/17
21	patient6:sample-1	patient6	sample-1	F	23	serum	586	clin	1	1	1	11/04/17
22	patient6:sample-2	patient6	sample-2	F	25	serum	586	clin	1	1	1	11/04/17

biobank_samples attributes entities packages +

Normal View Ready Sum=1



We need to tell MOLGENIS what to do



6

Each Excel sheet is an 'entity' and each column is an 'attribute'

	A	B	C	D	E	F	G	H	I
1	name	entity	label	dataType	nillable	idAttribute	visible		
2	ID	biobank_samples	ID		FALSE	TRUE	FALSE		
3	ParticipantID	biobank_samples	Participant ID	string	FALSE				
4	SampleID	biobank_samples	Sample ID	string	FALSE				
5	Sex	biobank_samples	Sex	string	TRUE				
6	AgeAtSampling	biobank_samples	Age at sampling	int	TRUE				
7	MaterialType	biobank_samples	Material type	string	FALSE				
8	AnatomicalSite	biobank_samples	Anatomical site	string	FALSE				
9	Disease	biobank_samples	Disease	string	FALSE				
10	DiagnosisType	biobank_samples	Diagnosis type	string	FALSE				
11	Affected	biobank_samples	Affected	string	FALSE				
12	Family	biobank_samples	Family members available	string	FALSE				
13	Genotype	biobank_samples	Genotype	string	FALSE				
14	LastUpdated	biobank_samples	Date of last update	string	TRUE				
15									
16									
17									
18									
19									
20									
21									
22									
23									
24									
25									



Upload wizard

If you want to upload data yourself, ask us for an account



7

RD Connect Sample Catalogue EuroBioBank Data Integration Admin Feedback Account Sign out

Welcome to the RD Connect Sample Catalogue

Upload
Catalogue
Scripts
Jobs
Automatic imports
ID-Card indexer
User Manager
Permission Manager
Menu Manager
Theme Manager
Index manager
Log manager
Settings

This catalogue is intended to facilitate the discovery of samples and samples data from Rare Diseases biobanks. It also provides information about sample collections and studies done on the registered samples.

Edit page

<https://samples.rd-connect.eu/>



Upload wizard



8

RD Connect

Sample Catalogue

EuroBioBank ▾

Data Integration ▾

Admin ▾

Feedback

Account

Sign out

1 Upload file

2 Options

3 Packages

4 Validation

5 Result

Upload a file

demo_model.xlsx

Choose file

← Previous

Restart

Next →

Click Next



Upload wizard



9

RD Connect

Sample Catalogue

EuroBioBank ▾

Data Integration ▾

Admin ▾

Feedback

Account

Sign out

1 Upload file

2 Options

3 Packages

4 Validation

5 Result

Entity options

☒ Add entities

Importer adds new entities or fails if entity exists

☐ Add entities / update existing

Importer adds new entities or updates existing entities

☐ Update entities

Importer updates existing entities or fails if entity does not exist

second time you might want update

← Previous

Restart

Next →

Click Next



Upload wizard



10

RD Connect

[Sample Catalogue](#)

[EuroBioBank](#)

[Data Integration](#)

[Admin](#)

[Feedback](#)

[Account](#)

[Sign out](#)

Success! File is validated and can be imported.

1 Upload file

2 Options

3 Packages

4 Validation

5 Result

Entities not in a package

None

Add to package

- ☒ base
- ☐ biobank

Advanced feature

[← Previous](#)

[Restart](#)

[Next →](#)

Click Next



Upload wizard



11

RD Connect Sample Catalogue EuroBioBank Data Integration Admin Feedback Account Sign out

1 Upload file 2 Options 3 Packages 4 Validation 5 Result

Entities

Name	Importable
biobank_samples	Yes

Entity fields

Name	Detected	Required	Available	Unknown
biobank_samples	ID, ParticipantID, SampleID, Sex, AgeAtSampling, MaterialType, Orpha, DiagnosisType, Affected, Family, Genotype, LastUpdated	No missing fields	No optional fields	No unknown fields

← Previous Restart Next →

MOLGENIS reports the plan (final check!)

Click Next



Upload wizard



12

RD Connect

Sample Catalogue

EuroBioBank

Data Integration

Admin

Feedback

Account

1 Upload file

2 Options

3 Packages

4 Validation

5 Result

Import success

Imported 34 biobank_samples entities.

Permissions

Select Group:

All Users

Entity Class	Edit metadata	Edit	View	Count	None
biobank_samples	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>

Save

← Previous

Restart

Finish →

Done!

Now you can
decide to share

Done!



View the results in 'data explorer'



13

RD Connect Sample Catalogue EuroBioBank Data Integration Admin Feedback Account Sign out

Data Explorer
Mapping Service
SORTA
Tag Wizard
Model registry

Welcome to the RD-Connect Sample Catalogue

This catalogue is intended to facilitate the discovery of samples and samples data from Rare Diseases biobanks. It also provides information about sample collections and studies done on the registered samples.



Choose your newly uploaded set



14

RD Connect Sample Catalogue EuroBioBank Data Integration Admin Feedback Account Sign out

Sample Catalogue Sample Information

Search data values

Data item filters

Data item selection

Select all Deselect all

Wizard

Data Aggregates Charts Annotators

Sample Catalogue Demo Biobank Samples Demo

Sample ID	Disease	MIABIS Material Type	Material Type	Anatomical Site	Sex	Diagnosis Type	Genotype data available
urn:rdconnect.tngb:aaaacwyqsmweb6qwhz3mx2aaae	Ring chromosome 14	Blood	Leukocyte	Blood	Female	Molecular,Cytogenetics	Yes
urn:rdconnect.tngb:aaaacwyqsmwed6qwhz3mx2aaae	Ring chromosome 14	Immortalized Cell Lines	Lymphoblast	Blood	Male	Molecular,Cytogenetics	Yes
urn:rdconnect.tngb:aaaacwyqsmwed6qwhz3mx2aaai	Ring chromosome 14	Blood	Leukocyte	Blood	Male	Molecular,Cytogenetics	Yes
urn:rdconnect.tngb:aaaacwyqsmwed6qwhz3mx2aaam	Ring chromosome 14	DNA	DNA	Blood	Male	Molecular,Cytogenetics	Yes
urn:rdconnect.tngb:aaaacwyqsmwef6qwhz3mx2aaae	Ring chromosome 14	DNA	DNA	Blood	Female	Molecular,Cytogenetics	Yes
urn:rdconnect.tngb:aaaacwyqsmwef6qwhz3mx2aaai	Ring chromosome 14	Blood	Leukocyte	Blood	Female	Molecular,Cytogenetics	Yes
urn:rdconnect.tngb:aaaacwyqsmwef6qwhz3mx2aaam	Ring chromosome 14	Immortalized Cell Lines	Lymphoblast	Blood	Female	Molecular,Cytogenetics	Yes
urn:rdconnect.tngb:aaaacwyqsmwef6qwhz3mx2aaaq	Ring chromosome 14	Immortalized Cell Lines	Lymphoblast	Blood	Female	Molecular,Cytogenetics	Yes
urn:rdconnect.tngb:aaaacwyqsmweb6qwhz3mx2aaae	Ring chromosome 14	DNA	DNA	Blood	Female	Molecular,Cytogenetics	Yes
urn:rdconnect.tngb:aaaacwyqsmweb6qwhz3mx2aaai	Ring chromosome 14	Blood	Leukocyte	Blood	Female	Molecular,Cytogenetics	Yes
urn:rdconnect.tngb:aaaacwyqsmweb6qwhz3mx2aaam	Ring chromosome 14	DNA	DNA	Blood	Female	Molecular,Cytogenetics	Yes
urn:rdconnect.tngb:aaaacwyqsmwef6qwhz3mx2aaae	Ring chromosome 14	Immortalized Cell Lines	Lymphoblast	Blood	Female	Molecular,Cytogenetics	Yes
urn:rdconnect.tngb:aaaacwyqsmwen6qwhz3mx2aaae	Ring chromosome 14	DNA	DNA	Blood	Male	Molecular,Cytogenetics	Yes
urn:rdconnect.tngb:aaaacwyqsmwen6qwhz3mx2aaai	Ring chromosome 14	Blood	Leukocyte	Blood	Male	Molecular,Cytogenetics	Yes
urn:rdconnect.tngb:aaaacwyqsmwen6qwhz3mx2aaam	Ring chromosome 14	Immortalized Cell Lines	Lymphoblast	Blood	Male	Molecular,Cytogenetics	Yes
urn:rdconnect.tngb:aaaacwyqsmwen6qwhz3mx2aaaq	Ring chromosome 14	Immortalized Cell Lines	Lymphoblast	Blood	Female	Molecular,Cytogenetics	Yes
urn:rdconnect.tngb:aaaacwyqsmwet6qwhz3mx2aaae	Ring chromosome 14	Other	Fibroblast	Skin	Female	Molecular,Cytogenetics	Yes
urn:rdconnect.tngb:aaaacwyqsmwet6qwhz3mx2aaae	Ring chromosome 14	DNA	DNA	Blood	Female	Molecular,Cytogenetics	Yes
urn:rdconnect.tngb:aaaacwyqsmwet6qwhz3mx2aaai	Ring chromosome 14	DNA	DNA	Blood	Female	Molecular,Cytogenetics	Yes
urn:rdconnect.tngb:aaaacwyqsmwet6qwhz3mx2aaam	Ring chromosome 14	Immortalized Cell Lines	Lymphoblast	Blood	Female	Molecular,Cytogenetics	Yes

Rows per page: 20

Previous 1 2 3 4 5 6 ... 368 Next

7352 items found



Et voila: your own catalogue 😊



15

Biobank Samples Demo Demo may 2017

Biobank Samples Demo Delete

Search data values

Data item filters

Wizard

Data item selection

Select all Deselect all

- ☒ Participant ID
- ☒ Sample ID
- ☒ Sex
- ☒ Age at sampling
- ☒ Material type
- ☒ Orphanet code
- ☒ Diagnosis type
- ☒ Affected
- ☒ Family members available
- ☒ Genotype
- ☒ Date of last update

Data Aggregates Charts Annotators

	Participant ID	Sample ID	Sex	Age at sampling	Material type	Orphanet code	Diagnosis type	Affected	Family members available	Genotype	Date of last update
<input checked="" type="checkbox"/>	patient1	sample-1	M	42	serum	232	mol	1	1	1	2017-04-11T00:00:00+0200
<input checked="" type="checkbox"/>	patient1	sample-2	M	43	DNA	232	mol	1	1	1	2017-04-11T00:00:00+0200
<input checked="" type="checkbox"/>	patient1	sample-3	M	43	RNA	232	mol	1	1	1	2017-04-11T00:00:00+0200
<input checked="" type="checkbox"/>	patient2	sample-1	M	30	serum	170	clin	1	1	1	2017-04-11T00:00:00+0200
<input checked="" type="checkbox"/>	patient2	sample-2	M	30	plasma	170	clin	1	1	1	2017-04-11T00:00:00+0200
<input checked="" type="checkbox"/>	patient2	sample-3	M	31	urine	170	clin	1	1	1	2017-04-11T00:00:00+0200
<input checked="" type="checkbox"/>	patient2	sample-4	M	32	blood	170	clin	1	1	1	2017-04-11T00:00:00+0200
<input checked="" type="checkbox"/>	patient3	sample-1	F	45	blood	310	clin	1	1	1	2017-04-11T00:00:00+0200
<input checked="" type="checkbox"/>	patient3	sample-2	F	46	plasma	310	clin	1	1	1	2017-04-11T00:00:00+0200
<input checked="" type="checkbox"/>	patient3	sample-3	F	46	serum	310	clin	1	1	1	2017-04-11T00:00:00+0200
<input checked="" type="checkbox"/>	patient3	sample-4	F	47	urine	310	clin	1	1	1	2017-04-11T00:00:00+0200
<input checked="" type="checkbox"/>	patient4	sample-1	M	42	blood	662	clin	1	0	0	2017-04-11T00:00:00+0200
<input checked="" type="checkbox"/>	patient4	sample-2	M	45	blood	662	clin	1	0	0	2017-04-11T00:00:00+0200
<input checked="" type="checkbox"/>	patient4	sample-3	M	46	blood	662	clin	1	0	0	2017-04-11T00:00:00+0200
<input checked="" type="checkbox"/>	patient4	sample-4	M	47	blood	662	clin	1	0	0	2017-04-11T00:00:00+0200
<input checked="" type="checkbox"/>	patient5	sample-1	F	34	serum	461	mol	0	1	1	2017-04-11T00:00:00+0200
<input checked="" type="checkbox"/>	patient5	sample-2	F	35	serum	461	mol	0	1	1	2017-04-11T00:00:00+0200
<input checked="" type="checkbox"/>	patient5	sample-3	F	35	plasma	461	mol	0	1	1	2017-04-11T00:00:00+0200
<input checked="" type="checkbox"/>	patient5	sample-4	F	37	serum	461	mol	0	1	1	2017-04-11T00:00:00+0200
<input checked="" type="checkbox"/>	patient6	sample-1	F	23	serum	586	clin	1	1	1	2017-04-11T00:00:00+0200

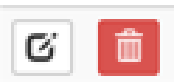
Rows per page: 20

Previous 1 2 Next

34 items found



Bonus: edit online



16

RD Connect Sample Catalogue EuroBioBank Data Integration Admin Feedback Account Sign out

Biobank Samples Demo Demo may 2017

Search data values

Data item filters

Data item selection

Select all Deselect all

	Participant ID	Sample ID	Sex	Age at sampling	Material type	Orphanet code	Diagnosis type	Affected	Family members available	Genotype	Date of last update
<input checked="" type="checkbox"/>	patient1	sample-1	M	42	serum	232	mol	1	1	1	2017-04-11T00:00:00+0200
<input checked="" type="checkbox"/>	patient1	sample-2	M	43	DNA	232	mol	1	1	1	2017-04-11T00:00:00+0200
<input checked="" type="checkbox"/>	patient1	sample-3	M	43	RNA	232	mol	1	1	1	2017-04-11T00:00:00+0200
<input checked="" type="checkbox"/>	patient2	sample-1	M	30	serum	170	clin	1	1	1	2017-04-11T00:00:00+0200
<input checked="" type="checkbox"/>	patient2	sample-2	M	30	plasma	170	clin	1	1	1	2017-04-11T00:00:00+0200
<input checked="" type="checkbox"/>	patient2	sample-3	M	31	urine	170	clin	1	1	1	2017-04-11T00:00:00+0200
<input checked="" type="checkbox"/>	patient2	sample-4	M	32	blood	170	clin	1	1	1	2017-04-11T00:00:00+0200
<input checked="" type="checkbox"/>	patient3	sample-1	F	45	blood	310	clin	1	1	1	2017-04-11T00:00:00+0200
<input checked="" type="checkbox"/>	patient3	sample-2	F	46	plasma	310	clin	1	1	1	2017-04-11T00:00:00+0200
<input checked="" type="checkbox"/>	patient3	sample-3	F	46	serum	310	clin	1	1	1	2017-04-11T00:00:00+0200
<input checked="" type="checkbox"/>	patient3	sample-4	F	47	urine	310	clin	1	1	1	2017-04-11T00:00:00+0200
<input checked="" type="checkbox"/>	patient4	sample-1	M	42	blood	662	clin	1	0	0	2017-04-11T00:00:00+0200
<input checked="" type="checkbox"/>	patient4	sample-2	M	45	blood	662	clin	1	0	0	2017-04-11T00:00:00+0200
<input checked="" type="checkbox"/>	patient4	sample-3	M	46	blood	662	clin	1	0	0	2017-04-11T00:00:00+0200
<input checked="" type="checkbox"/>	patient4	sample-4	M	47	blood	662	clin	1	0	0	2017-04-11T00:00:00+0200
<input checked="" type="checkbox"/>	patient5	sample-1	F	34	serum	461	mol	0	1	1	2017-04-11T00:00:00+0200
<input checked="" type="checkbox"/>	patient5	sample-2	F	35	serum	461	mol	0	1	1	2017-04-11T00:00:00+0200
<input checked="" type="checkbox"/>	patient5	sample-3	F	35	plasma	461	mol	0	1	1	2017-04-11T00:00:00+0200
<input checked="" type="checkbox"/>	patient5	sample-4	F	37	serum	461	mol	0	1	1	2017-04-11T00:00:00+0200
<input checked="" type="checkbox"/>	patient6	sample-1	F	23	serum	586	clin	1	1	1	2017-04-11T00:00:00+0200

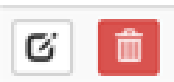
Rows per page: 20

Previous 1 2 Next

34 items found



Bonus: edit online



17

RD Connect Sample Catalogue EuroBioBank - Data Integration - Admin - Feedback Account

Biobank Samples Demo Demo may 2017

Search data values

Data item filters

Data item selection

- ☒ Participant ID
- ☒ Sample ID
- ☒ Sex
- ☒ Age at sampling
- ☒ Material type
- ☒ Orphanet code
- ☒ Diagnosis type
- ☒ Affected
- ☒ Family members available
- ☒ Genotype
- ☒ Date of last update

Select all Deselect all

Wizard

Edit Biobank Samples Demo

ID * patient1:sample-1

Participant ID * patient1

Sample ID * sample-1

Sex M

Age at sampling 42

Material type * serum

Orphanet code * 232

Diagnosis type * mol

Affected * 1

Family members available * 1

Genotype * 1

Date of last update 2017-04-11T00:00:00+0200

Cancel Save changes

Family members available	Genotype	Date of last update
1	1	2017-04-11T00:00:00+0200
1	1	2017-04-11T00:00:00+0200
1	1	2017-04-11T00:00:00+0200
1	1	2017-04-11T00:00:00+0200
1	1	2017-04-11T00:00:00+0200
1	1	2017-04-11T00:00:00+0200
1	1	2017-04-11T00:00:00+0200
1	1	2017-04-11T00:00:00+0200
1	1	2017-04-11T00:00:00+0200
1	1	2017-04-11T00:00:00+0200
0	0	2017-04-11T00:00:00+0200
0	0	2017-04-11T00:00:00+0200
0	0	2017-04-11T00:00:00+0200
0	0	2017-04-11T00:00:00+0200
0	0	2017-04-11T00:00:00+0200
1	1	2017-04-11T00:00:00+0200
1	1	2017-04-11T00:00:00+0200
1	1	2017-04-11T00:00:00+0200
1	1	2017-04-11T00:00:00+0200

patient5 sample-3 F 35 plasma 461 mol 0 1 1 2017-04-11T00:00:00+0200

patient5 sample-4 F 37 serum 461 mol 0 1 1 2017-04-11T00:00:00+0200

patient6 sample-1 F 23 serum 586 clin 1 1 1 2017-04-11T00:00:00+0200

Rows per page: 20

Previous 1 2 Next

34 items found



Converting data to standard

How we transform the data columns from the source into data columns desired in the catalogue (so all biobanks become comparable)



Remember, we are looking for:



19

Pseudonomized Sample ID *	Text
Material Type *	Value list
Anatomical Site *	Value list
Sex *	Male / Female
Disease *	Orphanet code
Diagnosis Type	Value list
Age at Sampling	Number
Age at Death	Number
Age at Diagnosis	Number
Age at Remission	Number
Affected	Yes / No / Empty

Family members available	Yes / No / Empty
Related samples available	Yes / No / Empty
Genotype data available	Yes / No / Empty
Registry	Yes / No / Empty
Sample Collection	hyperlink
Hosting Biobank *	Registry & Biobank Finder ID
Registry	Registry & Biobank Finder ID
Pseudonomized Patient ID *	Text

* Required

Template downloadable from:

<http://rd-connect.eu/biosamples-data/sample-catalogue/>

<https://samples.rd-connect.eu/>



Acceptable Values



20

Material Type: Blood, DNA, Faeces, Immortalized Cell Lines, Isolated Pathogen, Other, Plasma, RNA, Saliva, Serum, Tissue (Frozen), Tissue (FFPE), Urine

Diagnosis Type: Not Specified, Autoptic, Biochemical, Clinical, Echographic, Electrophysiological, Enzymatic, Hystological, Molecular, Neuroradiological, Radiological, Cytogenetics

Anatomical site: Muscle tissue, Blood, etc (ontology term)

Disease: e.g. Duchenne muscular dystrophy, Ring chromosome 14, (ORDO terms)

Affected / * Available: Yes, No, Unknown, Not asked, Not available

Age at Sampling, Diagnosis, etc.: Age of the patient in years, e.g. 5



Obviously, if you upload exactly this...



21

- Then no conversion is needed.
 - Template: <http://rd-connect.eu/biosamples-data/sample-catalogue/>
- But otherwise:
 - Use the “Mapping and re-coding service”



“Mapping and re-coding service”

Semi-automatically convert from biobank to catalogue



22

RD Connect Sample Catalogue EuroBioBank Data Integration Admin Feedback Account Sign out

Biobank Samples Demo Demo May 2017

Search data values

Data Explorer Mapping Service SORTA Tag Wizard Model registry

Data item filters

Data item selection

	Participant ID	Sample ID	Sex	Age at sampling	Material type	Orphanet code	Diagnosis type	Affected	Family members available	Genotype	Date of last update
<input checked="" type="checkbox"/>	patient1	sample-1	M	42	serum	232	mol	1	1	1	2017-04-11T00:00:00+0200
<input checked="" type="checkbox"/>	patient1	sample-2	M	43	DNA	232	mol	1	1	1	2017-04-11T00:00:00+0200
<input checked="" type="checkbox"/>	patient1	sample-3	M	43	RNA	232	mol	1	1	1	2017-04-11T00:00:00+0200
<input checked="" type="checkbox"/>	patient2	sample-1	M	30	serum	170	clin	1	1	1	2017-04-11T00:00:00+0200
<input checked="" type="checkbox"/>	patient2	sample-2	M	30	plasma	170	clin	1	1	1	2017-04-11T00:00:00+0200
<input checked="" type="checkbox"/>	patient2	sample-3	M	31	urine	170	clin	1	1	1	2017-04-11T00:00:00+0200
<input checked="" type="checkbox"/>	patient2	sample-4	M	32	blood	170	clin	1	1	1	2017-04-11T00:00:00+0200
<input checked="" type="checkbox"/>	patient3	sample-1	F	45	blood	310	clin	1	1	1	2017-04-11T00:00:00+0200
<input checked="" type="checkbox"/>	patient3	sample-2	F	46	plasma	310	clin	1	1	1	2017-04-11T00:00:00+0200
<input checked="" type="checkbox"/>	patient3	sample-3	F	46	serum	310	clin	1	1	1	2017-04-11T00:00:00+0200
<input checked="" type="checkbox"/>	patient3	sample-4	F	47	urine	310	clin	1	1	1	2017-04-11T00:00:00+0200
<input checked="" type="checkbox"/>	patient4	sample-1	M	42	blood	662	clin	1	0	0	2017-04-11T00:00:00+0200
<input checked="" type="checkbox"/>	patient4	sample-2	M	45	blood	662	clin	1	0	0	2017-04-11T00:00:00+0200
<input checked="" type="checkbox"/>	patient4	sample-3	M	46	blood	662	clin	1	0	0	2017-04-11T00:00:00+0200
<input checked="" type="checkbox"/>	patient4	sample-4	M	47	blood	662	clin	1	0	0	2017-04-11T00:00:00+0200
<input checked="" type="checkbox"/>	patient5	sample-1	F	34	serum	461	mol	0	1	1	2017-04-11T00:00:00+0200
<input checked="" type="checkbox"/>	patient5	sample-2	F	35	serum	461	mol	0	1	1	2017-04-11T00:00:00+0200
<input checked="" type="checkbox"/>	patient5	sample-3	F	35	plasma	461	mol	0	1	1	2017-04-11T00:00:00+0200
<input checked="" type="checkbox"/>	patient5	sample-4	F	37	serum	461	mol	0	1	1	2017-04-11T00:00:00+0200
<input checked="" type="checkbox"/>	patient6	sample-1	F	23	serum	586	clin	1	1	1	2017-04-11T00:00:00+0200

Rows per page: 20

Previous 1 2 Next

34 items found

Download Export to Galaxy



Create a new mapping project



23

RDConnect

Sample Catalogue

EuroBioBank

Data Integration

Admin

Feedback

Account

Mapping projects overview

Create and view mapping projects. Upload additional target entities and mapped sources [here](#).

+ Add Mapping Project

	Mapping name	Owner	Target entities	Mapped sources
	Demo May 2017	admin	rdconnect_Sample	biobank_samples
	Mappings for the EuroBiobank - Met UCL en nieuwe NCL	admin	rdconnect_Sample	ring14_samples, S NCL_samples

Create a new mapping project

Mapping project name

Demo mapping

Select the Target entity

rdconnect_Sample

Create project

Cancel



+ Add source



24

RD Connect

[Sample Catalogue](#)

[EuroBioBank](#)

[Data Integration](#)

[Admin](#)

[Feedback](#)

[Account](#)

[Sign out](#)

[← Back to mapping project overview](#)

Mappings for the Demo mapping project

Create and view mappings.

Curated algorithms

Generated algorithms with high quality

Generated algorithms with low quality

Algorithms to discuss

[+ Add source](#)

Target model: rdconnect_Sample

Sample ID (STRING) required unique
Sample ID as provided by the Biobank

Disease (MREF) required
Disease from an established ontology (e.g. ICD-10, OMIM, Orphanet)

MIABIS Material Type (XREF) required
Type of material

Material Type (XREF) required
Type of material

Anatomical Site (XREF) required
Code for the anatomic source of the material from an established ontology (<http://www.ontobee.org/ontology/UBERON>)

Sex (CATEGORICAL) required
Gender of the participant

Diagnosis Type (CATEGORICAL_MREF) required
Classification of how the participant was diagnosed



+ Add source



25

RD Connect

Sample Catalogue

EuroBioBank

Data Integration

Admin

Feedback

Account

Sign out

Back to mapping project overview

Mappings for the Demo mapping project

Create and view mappings.

Target model: rdconnect_Sample

Sample ID (STRING) required unique

Sample ID as provided by the Biobank

Disease (MREF) required

Disease from an established ontology (e.g. ICD-10, OMIM, Orphanet)

MIABIS Material Type (XREF) required

Type of material

Material Type (XREF) required

Type of material

Anatomical Site (XREF) required

Code for the anatomic source of the material from an established ontology (<http://www.ontobee.org/ontology/UBERON>)

Sex (CATEGORICAL) required

Gender of the participant

Diagnosis Type (CATEGORICAL_MREF) required

Classification of how the participant was diagnosed

Curated algorithms

Generated algorithms with high quality

Generated algorithms with low quality

Algorithms to discuss

+ Add source

Add new source

Select a new source to map against the target attribute

biobank_samples

Add source

Cancel



Software 'guesses' best mappings



26

RD Connect Sample Catalogue EuroBioBank - Data Integration - Admin - Feedback Account

Sign out

Info! The next attribute to map is Disease. "Target" → rdconnect_Sample, "Source" → biobank_samples

Back to mapping project overview

Mappings for the Demo mapping project

Create and view mappings.

Curated algorithms

Generated algorithms with high quality

Generated algorithms with low quality

Algorithms to discuss

+ Add source

Target model: rdconnect_Sample	Source: biobank_samples
Sample ID (STRING) required unique Sample ID as provided by the Biobank	Sample ID
Disease (MREF) required Disease from an established ontology (e.g. ICD-10, OMIM, Orphanet)	Orphanet code
MIABIS Material Type (XREF) required Type of material	Material type
Material Type (XREF) required Type of material	Material type
Anatomical Site (XREF) required Code for the anatomic source of the material from an established ontology (http://www.ontobee.org/ontology/UBERON)	Orphanet code
Sex (CATEGORICAL) required Gender of the participant	Sex
Diagnosis Type (CATEGORICAL_MREF) required Classification of how the participant was diagnosed	Diagnosis type
Genotype data available (CATEGORICAL) required Is genotype data available?	Genotype
Age at Sampling (INT) Age of the participant when the sample was taken	Age at sampling
Age at Death (INT) Age of the participant at death	Age at sampling
Age at Diagnosis (INT)	Diagnosis type



Per attribute 'mapping assistant'



27

Cancel and go back Save Save to discuss

Mapping to *rdconnect_Sample.AnatomicalSite* from *biobank_samples*

Algorithm state
Name AnatomicalSite (XREF)
Label Anatomical Site
Description Code for the anatomic source of the material from an established ontology (<http://www.ontobee.org/ontology/UBERON>)
Ontology Terms N/A
Categories
http://purl.obolibrary.org/obo/UBERON_0000178 = Blood = A fluid that is composed = 0000178
http://purl.obolibrary.org/obo/UBERON_0001088 = Urine = A fluid that is composed of blood plasma and erythrocytes. = 0001088
http://purl.obolibrary.org/obo/UBERON_0001969 = Plasma = Excretion that is the output of a kidney = 0001969
http://purl.obolibrary.org/obo/UBERON_0001977 = Serum = Liquid derived from blood plasma that has clotting factors removed. = 0001977
http://purl.obolibrary.org/obo/UBERON_0002097 = Skin = The organ covering the body that consists of the dermis and epidermis. = 0002097
http://purl.obolibrary.org/obo/UBERON_0002385 = Muscle tissue = Muscle tissue is a contractile tissue made up of actin and myosin fibers(GO). Vertebrate muscle is categorized into three major muscle types defined by their structural and functional properties: skeletal, cardiac and smooth. In Dmel the counterparts are somatic, heart/cardiac and visceral. Here we take a cell type based approach. = 0002385
http://purl.obolibrary.org/obo/UBERON_0004277 = Eye Muscle = A muscle that is part of the eye region. = 0004277

Attributes

Search all 12 attributes from biobank_samples

2 attributes have been found out of 12

Select	Attribute label	Name
<input type="checkbox"/>	Orphanet code (STRING)	Orpha
<input checked="" type="checkbox"/>	Material type (STRING)	MaterialType

Mapping

Script Map

Algorithm

Use the script editor to determine how the values of selected attributes are processed. See the list of available functions for more information.

Validate algorithm

```
# 1 -> "http://purl.obolibrary.org/obo/UBERON."+${'MaterialType'}.map({
2   "serum": "0000178",
3   "plasma": "0000178",
4   "blood": "0000178",
5   "urine": "0001088",
6   "DNA": "0000178",
7   "RNA": "0000178",
8 }).value();
```

Result

Validation

Algorithm validation starts automatically when the algorithm is updated. In case of errors, click the error label for more details

Total: 34 Success: 34 Errors: 0

Preview

	Source: MaterialType	Target: AnatomicalSite
<input type="checkbox"/>	serum	Blood
<input type="checkbox"/>	DNA	Blood
<input type="checkbox"/>	RNA	Blood
<input type="checkbox"/>	serum	Blood
<input type="checkbox"/>	plasma	Blood
<input type="checkbox"/>	urine	Urine
<input type="checkbox"/>	blood	Blood
<input type="checkbox"/>	blood	Blood
<input type="checkbox"/>	plasma	Blood
<input type="checkbox"/>	serum	Blood



Choose attribute(s)



28

Attributes ?

Search all 12 attributes from biobank_samples



2 attributes have been found out of 12

Select	Attribute label	Name
<input type="checkbox"/>	Orphanet code (STRING)	Orpha
<input checked="" type="checkbox"/>	Material type (STRING)	MaterialType



Edit mapping



29

Mapping ?

Script

Map

Algorithm

Validate algorithm

Use the script editor to determine how the values of selected attributes are processed. See the list of [available functions](#) for more information.

```
1 $('Orpha').value();|
```




Edit mapping



30

Mapping ?

Script

Map

Algorithm

Validate algorithm

Use the script editor to determine how the values of selected attributes are processed. See the [list of available functions](#) for more information.

```
1 - "http://purl.obolibrary.org/obo/UBERON_" + $('MaterialType').map({
2     "serum": "0000178",
3     "plasma": "0000178",
4     "blood": "0000178",
5     "urine": "0001088",
6     "DNA": "0000178",
7     "RNA": "0000178"
8 }).value();
```



Preview the result



31

Result ?

Validation

Algorithm validation starts automatically when the algorithm is updated. In case of errors, click the error label for more details

Total: 34 Success: 34 Errors: 0

Preview

	Source: MaterialType	Target: AnatomicalSite
<input type="button" value="Q"/>	serum	Blood
<input type="button" value="Q"/>	DNA	Blood
<input type="button" value="Q"/>	RNA	Blood
<input type="button" value="Q"/>	serum	Blood
<input type="button" value="Q"/>	plasma	Blood
<input type="button" value="Q"/>	urine	Urine
<input type="button" value="Q"/>	blood	Blood
<input type="button" value="Q"/>	blood	Blood
<input type="button" value="Q"/>	plasma	Blood
<input type="button" value="Q"/>	serum	Blood



Repeat for each attribute x biobank

When done, all is green



32

Target model: rdconnect_Sample	Source: ring14_samples	Source: SpainRDR_BioNER	Source: UCL_samples	Source: NCL_samples
Sample ID (STRING) required unique Sample ID as provided by the Biobank	Sample ID <input checked="" type="checkbox"/> <input checked="" type="checkbox"/>	Sample ID <input checked="" type="checkbox"/> <input checked="" type="checkbox"/>	Donor, Cell ID <input checked="" type="checkbox"/> <input checked="" type="checkbox"/>	Sample ID, Participant ID <input checked="" type="checkbox"/> <input checked="" type="checkbox"/>
Disease (MREF) required Disease from an established ontology (e.g. ICD-10, OMIM, Orphanet)	<input checked="" type="checkbox"/> <input checked="" type="checkbox"/>	DISEASE ORPHA NUMBER, DISEASE OMIM NUMBER, DISEASE ICD-10 <input checked="" type="checkbox"/> <input checked="" type="checkbox"/>	ORPHA CODE, OMIM <input checked="" type="checkbox"/> <input checked="" type="checkbox"/>	Disease <input checked="" type="checkbox"/> <input checked="" type="checkbox"/>
MIABIS Material Type (XREF) required Type of material	Sample Type <input checked="" type="checkbox"/> <input checked="" type="checkbox"/>	Material Type <input checked="" type="checkbox"/> <input checked="" type="checkbox"/>	Cell Code <input checked="" type="checkbox"/> <input checked="" type="checkbox"/>	Material type <input checked="" type="checkbox"/> <input checked="" type="checkbox"/>
Material Type (XREF) required Type of material	Sample Type <input checked="" type="checkbox"/> <input checked="" type="checkbox"/>	Material Type <input checked="" type="checkbox"/> <input checked="" type="checkbox"/>	Cell Code <input checked="" type="checkbox"/> <input checked="" type="checkbox"/>	Material type <input checked="" type="checkbox"/> <input checked="" type="checkbox"/>
Anatomical Site (XREF) required Code for the anatomic source of the material from an established ontology (http://www.ontobee.org/ontology/UBERON)	Sample Type <input checked="" type="checkbox"/> <input checked="" type="checkbox"/>	Anatomical Site <input checked="" type="checkbox"/> <input checked="" type="checkbox"/>	Cell Code <input checked="" type="checkbox"/> <input checked="" type="checkbox"/>	Anatomical site, Material type <input checked="" type="checkbox"/> <input checked="" type="checkbox"/>
Sex (CATEGORICAL) required Gender of the participant	Gender <input checked="" type="checkbox"/> <input checked="" type="checkbox"/>	SEX <input checked="" type="checkbox"/> <input checked="" type="checkbox"/>	Sex <input checked="" type="checkbox"/> <input checked="" type="checkbox"/>	Sex <input checked="" type="checkbox"/> <input checked="" type="checkbox"/>
Diagnosis Type (CATEGORICAL_MREF) required Classification of how the participant was diagnosed	Diagnosis type <input checked="" type="checkbox"/> <input checked="" type="checkbox"/>	<input checked="" type="checkbox"/> <input checked="" type="checkbox"/>	<input checked="" type="checkbox"/> <input checked="" type="checkbox"/>	Diagnosis type <input checked="" type="checkbox"/> <input checked="" type="checkbox"/>
Genotype data available (CATEGORICAL) required Is genotype data available?	Molecular analysis (FISH; array CGH) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/>	Mutation <input checked="" type="checkbox"/> <input checked="" type="checkbox"/>	Genotype <input checked="" type="checkbox"/> <input checked="" type="checkbox"/>	Genotype <input checked="" type="checkbox"/> <input checked="" type="checkbox"/>
Age at Sampling (INT) Age of the participant when the sample was taken	<input checked="" type="checkbox"/>	Year of Sampling, Year of Birth <input checked="" type="checkbox"/> <input checked="" type="checkbox"/>	Age at biopsy <input checked="" type="checkbox"/> <input checked="" type="checkbox"/>	Age at sampling <input checked="" type="checkbox"/> <input checked="" type="checkbox"/>
Age at Death (INT) Age of the participant at death	<input checked="" type="checkbox"/>	Year of Death, Year of Birth <input checked="" type="checkbox"/> <input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>



Time to push the button



33

▶ Create integrated dataset



Aim: enable 'sample search' use cases



34

E.g. As researcher I would like to find DNA with Ring chromosome 14 syndrome

RD Connect Sample Catalogue dataexplorer Account Sign out

Sample Information

Search data values

Data item filters

Material Type: ('DNA') ✕

Disease: ('Ring chromosome 14') ✕

Wizard

Data item selection

Select all Deselect all

- ☐ Sample ID
- ☒ MIABIS Material Type
- ☒ Material Type
- ☒ Anatomical Site
- ☐ Year of Sampling
- ☐ Sample Collection
- ☒ Hosting Biobank
- ☒ Disease
- ☒ Year of Birth
- ☒ Year of Death
- ☒ Year of Diagnosis

Data **Aggregates** **Charts**

MIABIS Material Type	Material Type	Anatomical Site	Hosting Biobank	Disease
DNA	DNA	Blood	Galliera Genetic Bank	Ring chromosome 1
DNA	DNA	Blood	Galliera Genetic Bank	Ring chromosome 1
DNA	DNA	Blood	Galliera Genetic Bank	Ring chromosome 1
DNA	DNA	Blood	Galliera Genetic Bank	Ring chromosome 1
DNA	DNA	Blood	Galliera Genetic Bank	Ring chromosome 1
DNA	DNA	Blood	Galliera Genetic Bank	Ring chromosome 1
DNA	DNA	Blood	Galliera Genetic Bank	Ring chromosome 1
DNA	DNA	Blood	Galliera Genetic Bank	Ring chromosome 1
DNA	DNA	Blood	Galliera Genetic Bank	Ring chromosome 1
DNA	DNA	Blood	Galliera Genetic Bank	Ring chromosome 1
DNA	DNA	Blood	Galliera Genetic Bank	Ring chromosome 1
DNA	DNA	Blood	Galliera Genetic Bank	Ring chromosome 1
DNA	DNA	Blood	Galliera Genetic Bank	Ring chromosome 1
DNA	DNA	Blood	Galliera Genetic Bank	Ring chromosome 1
DNA	DNA	Blood	Galliera Genetic Bank	Ring chromosome 1



What we need:

- Your **pseudonymized** data
- Description of how your data maps
- Contact details of the person within your organisation who is responsible for delivering your sample data for upload

What we do:

- Setup the catalogue to accept your data:
 - Validate and upload your data
 - Write the mappings of your data to the standard model
 - Provide you with login details



Webinar summary (hands-on)



36

- Learn how (catalogue team) uploads your sample parameters
 - 'as-is' from your biobank, i.e. easiest sample sheet you can generate
 - bonus: we actually create a little catalogue for each biobank
 - bonus: you can also enter data using online forms
- Learn how (catalogue team) converts to RD-Connect standard
 - We create mappings from your data columns to the standard
 - We use a tool called 'BiobankConnect' to aid conversion
 - bonus: automation
- Q&A
- ❖ We aim for the Catalogue to be self-service in longer term



Learn more



37

Reading

- Sample Catalogue: <http://rd-connect.eu/biosamples-data/sample-catalogue/>
- MOLGENIS <https://www.molgenis.org/>
- MOLGENIS docs @ <https://goo.gl/lg7J5n>
- BiobankConnect paper - <http://pubmed.org/25361575>
- MOLGENIS papers - <http://pubmed.org?term=MOLGENIS>

Videos

- Upload - <https://youtu.be/VSZNXdaGII4>
- BiobankConnect - <https://youtu.be/Gc1VKRCmTWU>
 - BiobankConnect is the underlying technology used to harmonize the data