

Preliminary Proof-of-Concept Catalogue Tool



Technical workshop on real-world metadata for regulatory purposes
Virtual meeting, April 12, 2021

Presented by Prof Morris Swertz
University Medical Center Groningen



Outline

- 1 The context: position in EMA MINERVA and its aims
- 2 Approach & description of open-source software we plan to reuse
- 3 A short demo of the proof of concept for the metadata access tool
- 4 Any questions

MINERVA



Metadata for data discoverability and study replicability
in observational studies

- To define a list of criteria to identify relevant real-world data sources from which the data sources to be included in this study will be selected
- To identify a minimum of 10 data sources to use in the study
- To define a set of metadata that should be collected from real-world data sources
- To conduct an in-depth stakeholders' consultation on the metadata identified
- To define a process to collect the set of metadata for the data sources included in the study
- To collect the defined set of metadata for data sources included in the study
- **To develop or provide a tool enabling access to the metadata collected (e.g., through a dynamic dashboard)**
- To draft a good practice guide with the description of the metadata defined and recommendations on the use of metadata for the purpose of identifying real-world data sources for a specific study purpose

Catalogue tool

From the proposal:

We will deliver a **proof-of-concept interactive catalogue and tools to test its functionality to access, search, and visualise metadata**, including the metadata that can be collected as part of this project from the selected data sources

The catalogue will be a metadata inventory with **searchability, dynamic dashboards, visualisation capabilities, and integration of different levels of access** (public vs. confidential information)

Approach

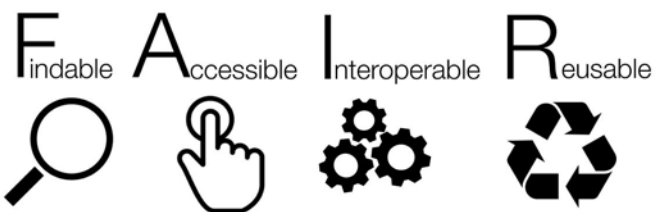


On top of MOLGENIS open source software

Data tools for scientific data

Application communities in

- cohorts & biobanks
- Rare disease
- Omics and imaging
- Large multi-center studies



Bulk data
Import your own data and metadata model using the EMX format, VCF, OBO

Data request
Find and request (biobank) data sets and items

Data explorer
Filter and download for further analysis

Genome browser
Data sharing and integration DAS protocol

Model registry
Metadata registry of models for biobanks and molecular data

Annotators
Data integration for diagnostics and personalized medicine

FAIR integration
Using ontologies to derive harmonization rule for data pooling

R statistics
Use R data api to up/d data and integrate graphics

Pipelines
Large scale computation on computational clusters, grids and clouds

Reports
Tailormade report functionality

App store
Add your own user interfaces and advanced tools.

Data capture
Questionnaires, forms, importer to easily add data to a running instance

Mostly funded via national and EU consortia since 2002; currently 27 active actions from:



Funded by the Horizon 2020 Framework Programme of the European Union



innovative medicines initiative



1. Configure the meta data items into a catalogue

Can change data model rapidly without coding

E.g. using the integrated EMX data model format & online metadata editor

Your Data

sex	birthyear	disease	consent	since
M	1976	NA	Y	1-1-201
F	1977	CV	N	1-1
M	1985	COPD	N	1-1
NA	1985	EB	Y	1-1
F	1940	NA	Y	1-1
F	2001	NA	Y	1-1

Your Meta Data

entity	attribute	dataType	refEntity	required
patient	id	autoid		true
patient	sex	xref	sexValues	
patient	birthyear	int		
patient	disease	xref	diseases	
patient	consent	boolean		true
patient	since	date		

2. Test against use cases

E.g. using the integrated data entry forms & query /data explorer tools

DATABANKS

Data collections such as registries or biobanks

filters columns download layout

1 - 19 of 19 Rows per page: 20

#	+	NAME	TYPE	PROVIDER	STARTYEAR	RECORDPROMPT
		Tuscany Registry of Congenital Defects	Congenital Anomaly Registry	CNR-IFC	1979	Congenital anomaly
		The Norwegian Prescription Database	Pharmacy Dispensation Records	UiO	2004	Prescription
		The Medical Birth Registry of Norway	Birth Registry	UiO	1967	Live birth Stillbirth Spontaneous abortion Induced termination
		The Norwegian Patient Registry	Primary care medical records	UiO	2008	Admission
		Inhabitant registry	Inhabitant Registry	ARS	2003	Registration with a healthcare system

TYPE

- Hospital Administrative Records
- Primary care medical records
- Pharmacy Dispensation Records
- Birth Registry
- Induced Terminations Registry
- Congenital Anomaly Registry
- Inhabitant Registry
- Registration with Healthcare System

[more](#) [clear](#)

POPULATION

- Belgium
- South Africa
- Japan

3. Tune user experience

Reuse practices from existing catalogues

E.g. **Databank overview**

(example courtesy of BBMRI-ERIC directory of biobanks, MIABIS working group)

▶ COVID-19

▶ COVID-19 Services

▼ Diagnosis available

Type to search


▼ Materials

Satisfy all


- Buffy Coat
- cDNA / mRNA
- Cell lines
- DNA
- Feces
- microRNA
- Nasal swab
- Not available
- Other
- Pathogen
- Peripheral blood cells

4 organisations with 770 collections matching the search criteria

REQUEST SAMPLES

 **Biobank Graz**

ISO-9001:2015 ✓



Collection types: Cohort, Disease specific, Hospital, Image collection, Longitudinal, Population-based, Sample collection, Prospective study, Case-control, Rare disease collection, Cross-sectional, Birth cohort

Juridical person: Medical University of Graz

Covid-19: Member COVID-19 Network, Proteomics studies including protein engineering and protein interactions, Screening tools for searching virus proteases inhibitors, Animal Testing Facility, Ability to set up clinical trials, Virus Sequencing Facility, BSL-2 laboratories available, BSL-3 laboratories available, Laboratories doing PCR-based diagnosis

Collection	Type	Materials	Standards	#Samples
Ability to collect COVID-19 cases	Cohort, Disease specific, Hospital, Prospective study, Sample collection	Serum, Throat swab, Buffy Coat, Feces, Plasma		<10
Bio-UV 2017	Cohort, Disease specific, Hospital, Longitudinal, Sample collection	Serum, Plasma, Buffy Coat, Other		2838
BioPersMed Cohort	Cohort, Disease specific, Hospital, Image collection, Longitudinal, Population-based	Serum, Plasma, Buffy Coat, Urine	CEN/TS 16945:2016 ✓	260680

3. Tune user experience

Reuse practices from existing catalogues

E.g. **Data items available**

(example courtesy LifeCycle project)

- Catalogue

- + Meta variables
- Maternal Characteristics
 - Socio-demographic characteristics
 - Maternal cohabitation
 - Maternal occupational status**
 - Maternal occupational status (>-1 year and <1 year)
 - Maternal education (years)
 - Maternal country of birth
 - Maternal ethnicity
 - Maternal age at birth
 - Maternal death
 - + Health-related characteristics
 - + Lifestyle Characteristics
 - + Obstetric Characteristics
- + Paternal Characteristics
- + Child
- + Household Characteristics
- + Urban environment

LifeCycle variables Harmonization

Maternal occupational status

Maternal occupational status (>-1 year and <1 year)

variable	occup_m_0
datatype	Categorical
values	<div style="border: 1px solid #ccc; padding: 5px; margin-top: 5px;"> 1 = Employed 2 = Self-employed 3 = Unemployed 4 = Student, apprentice 5 = Domestic tasks (housewife etc.) 6 = Inactive/other (receiving benefits) </div>
unit	-
match	-
comments	<div style="border: 1px solid #ccc; padding: 5px; margin-top: 5px;"> Occupational status of the mother who was employed (≥-1 year and <1 year) </div>

3. Tune user experience

Reuse practices from existing catalogues

E.g. **Mapping to standards**

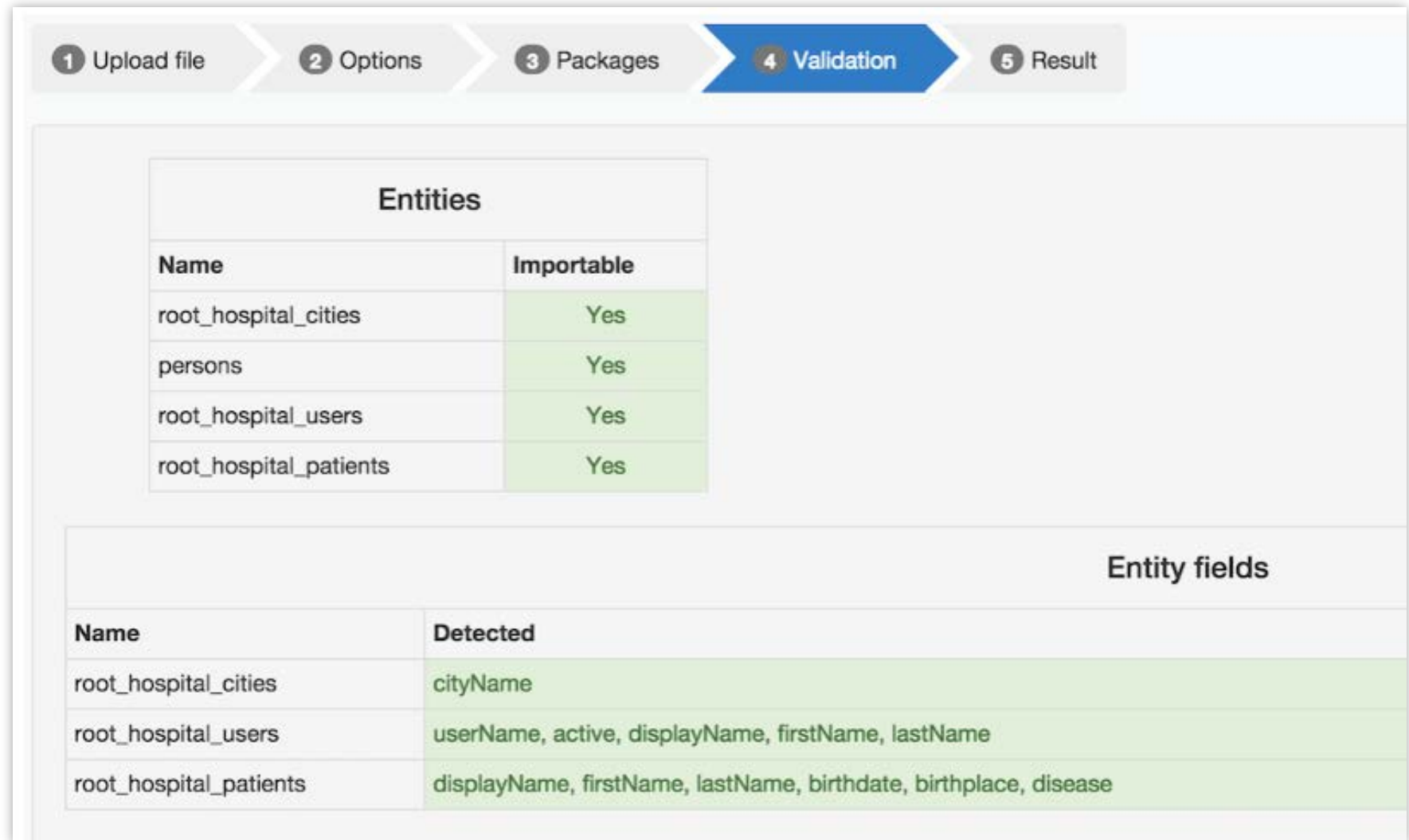
(example courtesy LifeCycle project)

Cohort	occup_m_0	occup_m_1	occup_m_2	occup_m_3	occup_m_4	occup_m_5
ALSPAC 12 / 18	✓	✓	✓	✓	✗	✓
BIB 18 / 18	✓	✓	✓	✓	✓	✓
CHOP 18 / 18	✓	✗	✗	✗	✗	✗
DNBC 18 / 18	✓	✗	✗	✗	✗	✗
EDEN 18 / 18	✓	✓	✓	✓	✓	✓
ELFE 18 / 18	✓	✓	✓	✗	✗	✗
GECKO 18 / 18	✓	✗	✗	✗	✗	✓
GenR 18 / 18	✓	✓	✓	✓	✓	✓
INMA 18 / 18	✓	✓	✓	✓	✓	✓

4. Automate data ingest

E.g. using the integrated batch file formats, data validation and error reporting tools

In collaboration with task on 'process'



The screenshot shows a web-based interface for data ingestion. At the top, there is a progress bar with five steps: 1 Upload file, 2 Options, 3 Packages, 4 Validation (highlighted in blue), and 5 Result. Below the progress bar, there are two tables.

Entities Table:

Name	Importable
root_hospital_cities	Yes
persons	Yes
root_hospital_users	Yes
root_hospital_patients	Yes

Entity fields Table:

Name	Detected
root_hospital_cities	cityName
root_hospital_users	userName, active, displayName, firstName, lastName
root_hospital_patients	displayName, firstName, lastName, birthdate, birthplace, disease

5. FAIR interfaces

Enables automated up- and downloads

E.g. using the integrated interoperability interfaces into linked data, FAIR data point, REST, JavaScript, Python, R, CSV, etc.

(courtesy of GO-FAIR / VODAN)

```

<http://localhost/fdpTest/repository> <http://localhost/fdpTest/repository> <http://localhost/fdpTest/repository/fdp1> .

<http://localhost/fdpTest/repository/fdp1> a <http://www.re3data.org/schema/3-0#Repository>,
  <http://www.w3.org/ns/dcat#Resource>;
  <http://purl.org/dc/terms/description> "test ground";
  <http://purl.org/dc/terms/hasVersion> "1.0";
  <http://purl.org/dc/terms/publisher> "http://molgenis.org";
  <http://purl.org/dc/terms/title> "test";
  <http://rdf.biosemantics.org/ontologies/fdp-o#metadataIdentifier> "fdp1";
  <http://rdf.biosemantics.org/ontologies/fdp-o#metadataIssued> "2020-09-22T22:22:00";
  <http://rdf.biosemantics.org/ontologies/fdp-o#metadataModified> "2020-09-22T22:22:00";
  <http://www.re3data.org/schema/3-0#dataCatalog> "http://localhost/fdpTest/catalog/catalog1",
    "http://localhost/fdpTest/catalog/catalog2" .

<http://localhost/fdpTest/cde_Patient> <http://localhost/fdpTest/cde_Patient> <http://localhost/fdpTest/cde_Patient/Patient001>,
  <http://localhost/fdpTest/cde_Patient/Patient002>, <http://localhost/fdpTest/cde_Patient/Patient003>,
  <http://localhost/fdpTest/cde_Patient/Patient004>, <http://localhost/fdpTest/cde_Patient/Patient005>,
  <http://localhost/fdpTest/cde_Patient/Patient006>, <http://localhost/fdpTest/cde_Patient/Patient007>,
  <http://localhost/fdpTest/cde_Patient/Patient008>, <http://localhost/fdpTest/cde_Patient/Patient009>,
  <http://localhost/fdpTest/cde_Patient/Patient010> .

<http://localhost/fdpTest/hpo> <http://localhost/fdpTest/hpo> <http://purl.obolibrary.org/obo/HP_0000252>,
  <http://purl.obolibrary.org/obo/HP_0011451> .

<http://localhost/fdpTest/patient> <http://localhost/fdpTest/patient> <http://localhost/fdpTest/patient/no123> .

<http://localhost/fdpTest/patient/no123> a <http://purl.obolibrary.org/obo/NCIT_C16960>;
  <http://purl.obolibrary.org/obo/NCIT_C16977> "http://purl.obolibrary.org/obo/HP_0000252",
  "http://purl.obolibrary.org/obo/HP_0011451";
  <http://purl.obolibrary.org/obo/NCIT_C83083> "no123" .

<http://localhost/fdpTest/catalog> <http://localhost/fdpTest/catalog> <http://localhost/fdpTest/catalog/catalog1>,
  <http://localhost/fdpTest/catalog/catalog2> .

<http://localhost/fdpTest/catalog/catalog1> a <http://www.w3.org/ns/dcat#Catalog>,
  <http://www.w3.org/ns/dcat#Resource>;
  <http://purl.org/dc/terms/description> "my first catalogue";
  <http://purl.org/dc/terms/hasVersion> "1.0";
  <http://purl.org/dc/terms/isPartOf> "fdp1";
  <http://purl.org/dc/terms/publisher> "http://molgenis.org";
  <http://purl.org/dc/terms/title> "cat1";
  <http://rdf.biosemantics.org/ontologies/fdp-o#metadataIdentifier> "catalog1";
  <http://rdf.biosemantics.org/ontologies/fdp-o#metadataIssued> "2020-09-22T22:22:00";
  <http://rdf.biosemantics.org/ontologies/fdp-o#metadataModified> "2020-09-22T22:22:00" .

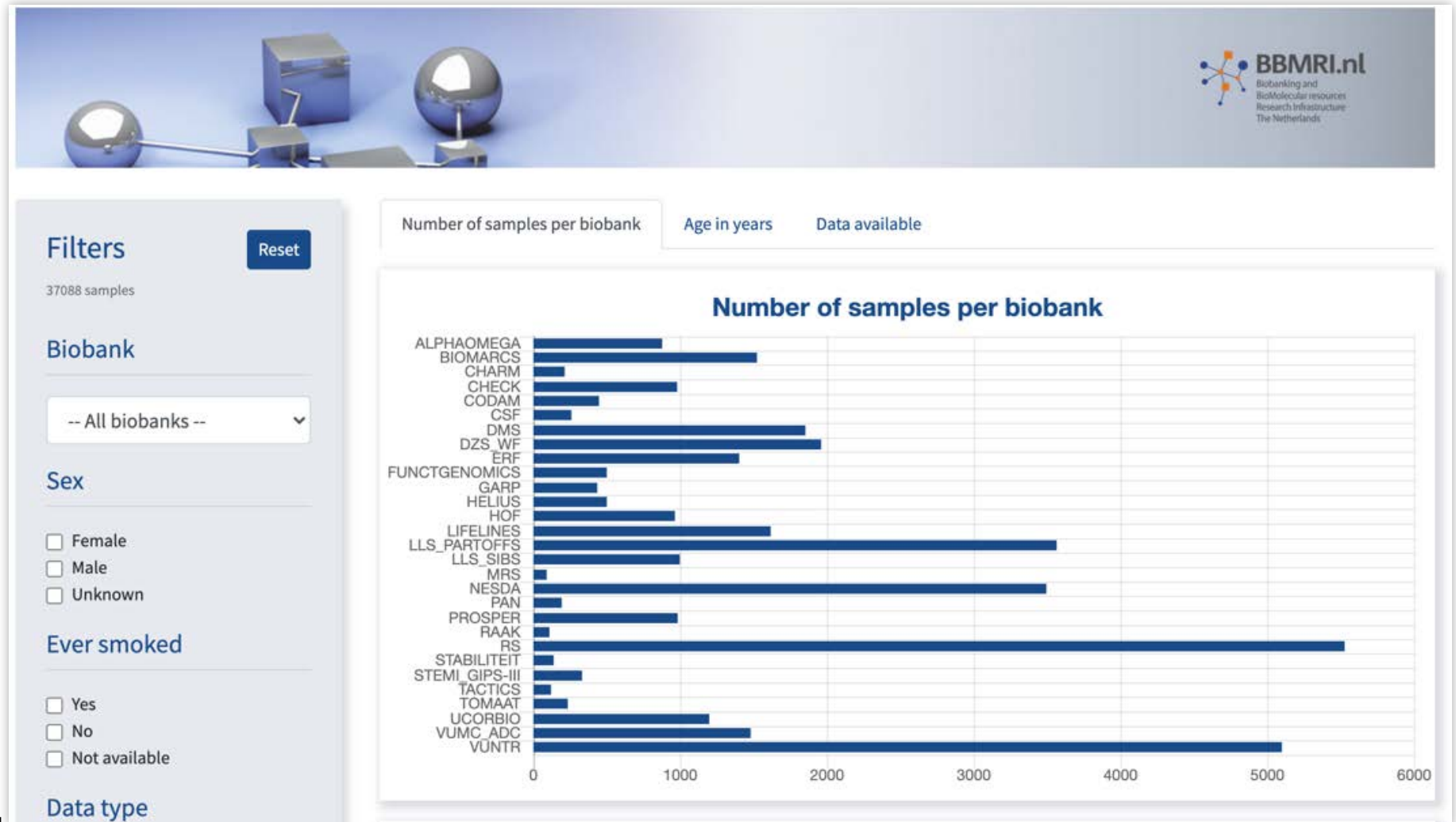
<http://localhost/fdpTest/catalog/catalog2> a <http://www.w3.org/ns/dcat#Catalog>,
  <http://www.w3.org/ns/dcat#Resource>;
  <http://purl.org/dc/terms/description> "my second catalogue";
  <http://purl.org/dc/terms/hasVersion> "1.0";
  <http://purl.org/dc/terms/isPartOf> "fdp1";
  <http://purl.org/dc/terms/publisher> "http://molgenis.org";
  <http://purl.org/dc/terms/title> "cat2";
  <http://rdf.biosemantics.org/ontologies/fdp-o#metadataIdentifier> "catalog2";
  <http://rdf.biosemantics.org/ontologies/fdp-o#metadataIssued> "2020-09-22T22:22:00";
  <http://rdf.biosemantics.org/ontologies/fdp-o#metadataModified> "2020-09-22T22:22:00" .

```

6. Add dashboards

To enable catalogue users to compare contents of the databanks or timepoints

(example courtesy of BBMRI-NL)



DEMO of PROOF OF CONCEPT



Thank you!

For any question on this presentation, please contact: Malgorzata.Durka-Grabowska@ema.europa.eu

Official address Domenico Scarlattilaan 6 • 1083 HS Amsterdam • The Netherlands

Address for visits and deliveries Refer to www.ema.europa.eu/how-to-find-us

Send us a question Go to www.ema.europa.eu/contact **Telephone** +31 (0)88 781 6000