

presentation of IMAGE data portal

Paolo Cozzi (paolo.cozzi@ibba.cnr.it)

GenResBridge WP4 (2021/04/28)

Gene Banks
using
Cryoweb

Gene
Banks with
other
databases

Other
collections
(templates)

Validation a
Metadata Standard

They describe Animals and Samples. The AIM
of IMAGE was to track information from
experiments to collections and to get in touch
with people which collect biological samples!

The screenshot displays the IMAGE database interface. The top window shows a PostgreSQL database dump with SQL commands like `SET statement_timeout = 0;` and `SET lock_timeout = 0;`. Below this, a table of data is visible, containing columns for sample IDs, species, sex, date, and location. The table includes entries for *Asturiana de los Valles* and *Asturiana de los Valles* with various sample IDs and dates.

Sample ID	Species	Sex	Date	Location
16M12099	Asturiana de los Valles	Bovine	male	17/02/1997 Asturias
16M12900	Asturiana de los Valles	Bovine	male	05/01/1995 Asturias
16M12901	Asturiana de los Valles	Bovine	male	15/02/2004 Asturias
16M12902	Asturiana de los Valles	Bovine	male	26/01/2003 Asturias
16M12903	Asturiana de los Valles	Bovine	male	08/02/2001 Asturias
16M12904	Asturiana de los Valles	Bovine	male	10/02/2003 Asturias
16M12905	Asturiana de los Valles	Bovine	male	19/01/1999 Asturias
16M12906	Asturiana de los Valles	Bovine	male	11/05/1987 Asturias
16M12907	Asturiana de los Valles	Bovine	male	14/12/1997 Asturias
16M12908	Asturiana de los Valles	Bovine	male	10/02/1989 Asturias
16M12909	Asturiana de los Valles	Bovine	male	30/10/2003 Asturias
16M12910	Asturiana de los Valles	Bovine	male	20/01/2002 Asturias
16M12911	Asturiana de los Valles	Bovine	male	24/02/2005 Asturias
16M12912	Asturiana de los Valles	Bovine	male	10/04/1988 Asturias
16M12913	Asturiana de los Valles	Bovine	male	06/01/2006 Asturias
16M12914	Asturiana de los Valles	Bovine	male	14/02/2001 Asturias
16M12915	Asturiana de los Valles	Bovine	male	16/02/1990 Asturias
16M12916	Asturiana de los Valles	Bovine	male	13/01/1991 Asturias
16M12917	Asturiana de los Valles	Bovine	male	28/03/2008 Asturias
16M12918	Asturiana de los Valles	Bovine	male	21/01/1994 Asturias
16M12919	Asturiana de los Valles	Bovine	male	02/07/1998 Asturias

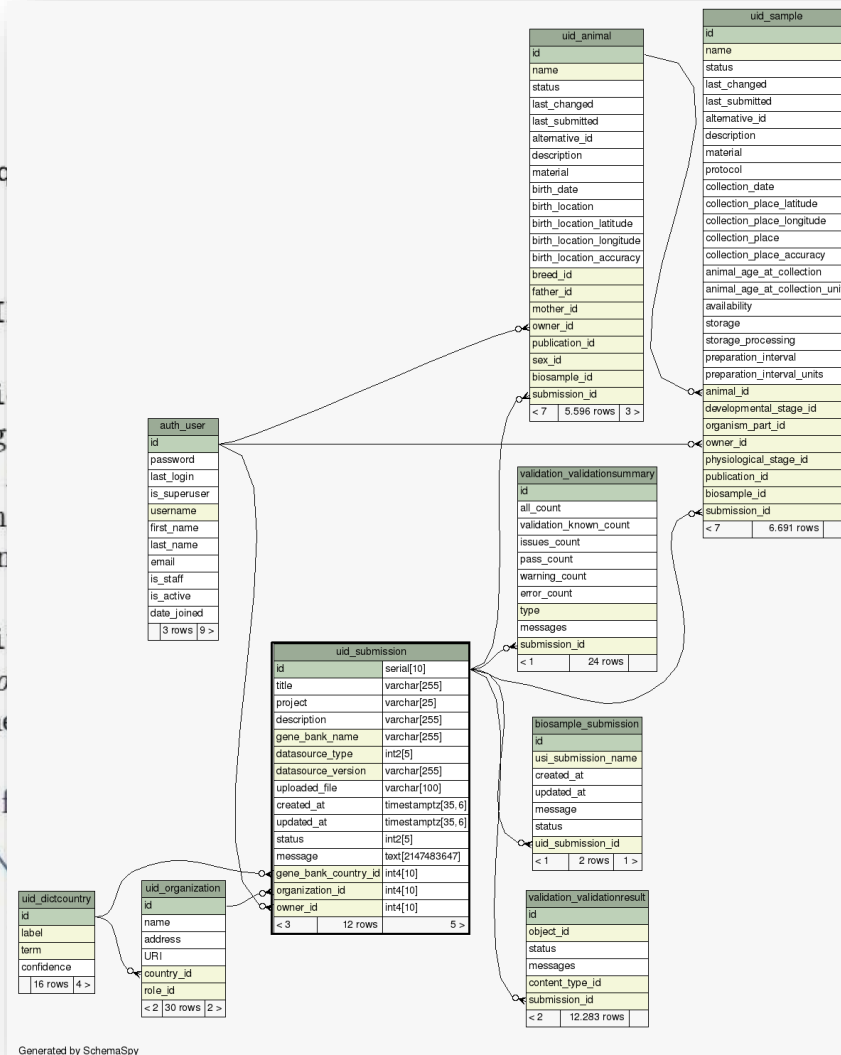


From metadata to Unified Internal Database

Mandatory:

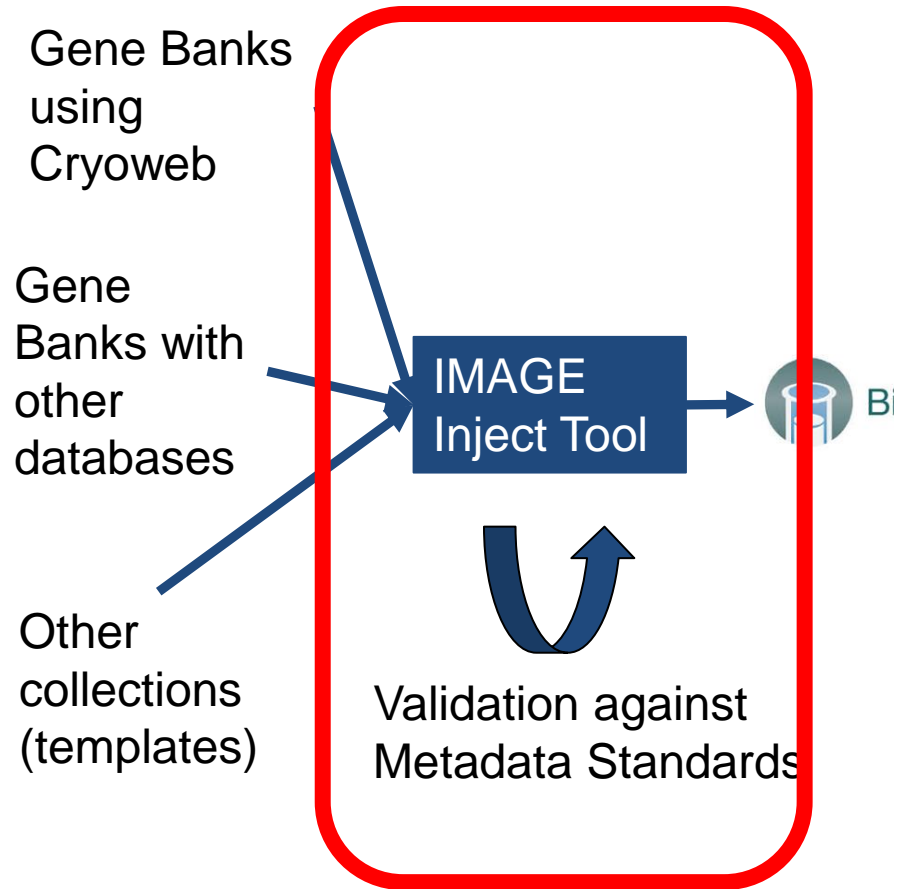
- Data source ID (the record)
- Alternative ID (data source, e.g. Project (text) while still belong to identify them in – IMAGE)
- Submission time
- Material (ontology sample (specimen) – Organism – Specimen type)

Field name



- Metadata standard: define which data to track and how
- Model data into a Unified Internal Database (UID)
 - store data from different sources
- Define all the other operations on top of UID
- Manage different data sources in the same way

IMAGE Data model



- User registration
- Data ingest
- Data validation
- Enrichment
- Data submission
- Data update
- Notification

FOOTNOTES



The IMAGE-InjectTool

- Manage submission to biosamples - <https://inject.image2020genebank.eu/>

The collage displays various components of the IMAGE-InjectTool web application:

- Dashboard:** Shows the 'IMAGE-InjectTool' logo, navigation links (Home, About), and a sidebar with options like 'Submit', 'Create', and 'Upload'.
- Submission Form:** A form for submitting data, including fields for 'Organization', 'Data Source Type', 'Data Source version', 'Data File', 'Created at', 'Status', 'N. of Animals', and 'N. of Samples'. It also features 'Edit data' and 'Submissions' buttons.
- Error Message:** A red box indicating 'Error: No value provided for field Birth date. ANIMAL::ID::CT01_1999'.
- Detail for Animal:** A section for viewing details of a specific animal, including 'Data source ID', 'Alternative id', 'Description', 'Project', 'Submission title', 'Submission description', 'Person last name', 'Person initial', and 'Person first name'.
- Edit Submission Data for Test:** A table showing submission data for a specific animal, with columns for Name, Action, Biosample ID, Validation, Last Changed, and Last Submitted.
- Successful submission into biosample:** A green notification box indicating a successful submission.
- Buttons:** 'Back to Submission Detail', 'Back to Submissions List', and 'Return to Dashboard'.

Name	Action	Biosample ID	Validation	Last Changed	Last Submitted
ANIMAL::ID::WF60/B0811	Edit Delete	SAMEA5649098	Pass	March 28, 2019, 11:35 a.m.	March 28, 2019, 11:37 a.m.
ANIMAL::ID::WF60/6532	Edit Delete	SAMEA5649097	Pass	March 28, 2019, 11:35 a.m.	March 28, 2019, 11:37 a.m.
ANIMAL::ID::WF001265	Edit Delete	SAMEA5649100	Pass	March 28, 2019, 11:36 a.m.	March 28, 2019, 11:37 a.m.
MA_052_KR_S_Snh_S03_Mariens...	Edit Delete	SAMEA5649099	Pass	March 28, 2019, 11:37 a.m.	March 28, 2019, 11:37 a.m.

IMAGE Data model

The screenshot shows the EMBL BioSamples website interface. At the top, there's a navigation bar with links to Home, Search, Graph Search, Submit, Documentation, and About. A search bar is present with the text 'leukemia, viridiplantae, chorio fruct'. Below the navigation bar, the 'Search results' section is active, showing 'showing 10 of 16674591 samples'. On the left, there's a table of filters with counts: organism (16480042), external reference (13514525), and sex (5723695). The main content area displays sample details for GSM2995402 and Arcobacter skirrowii 17-1201-4, including external IDs, INSDC center names, and various metadata fields like collection date, geo loc name, and isolation source.

collections
(templates)

Validation against
Metadata Standards

External cross
referencing

API
access

- Stores and supplies metadata about biological samples
- Search with website or API
- Required for crossreference or for submitting data to EBI
- Search for a certain record could be difficult (there are a lot of data!)

<https://www.ebi.ac.uk/biosamples/samples?filter=attr:project:IMAGE>



IMAGE Data model

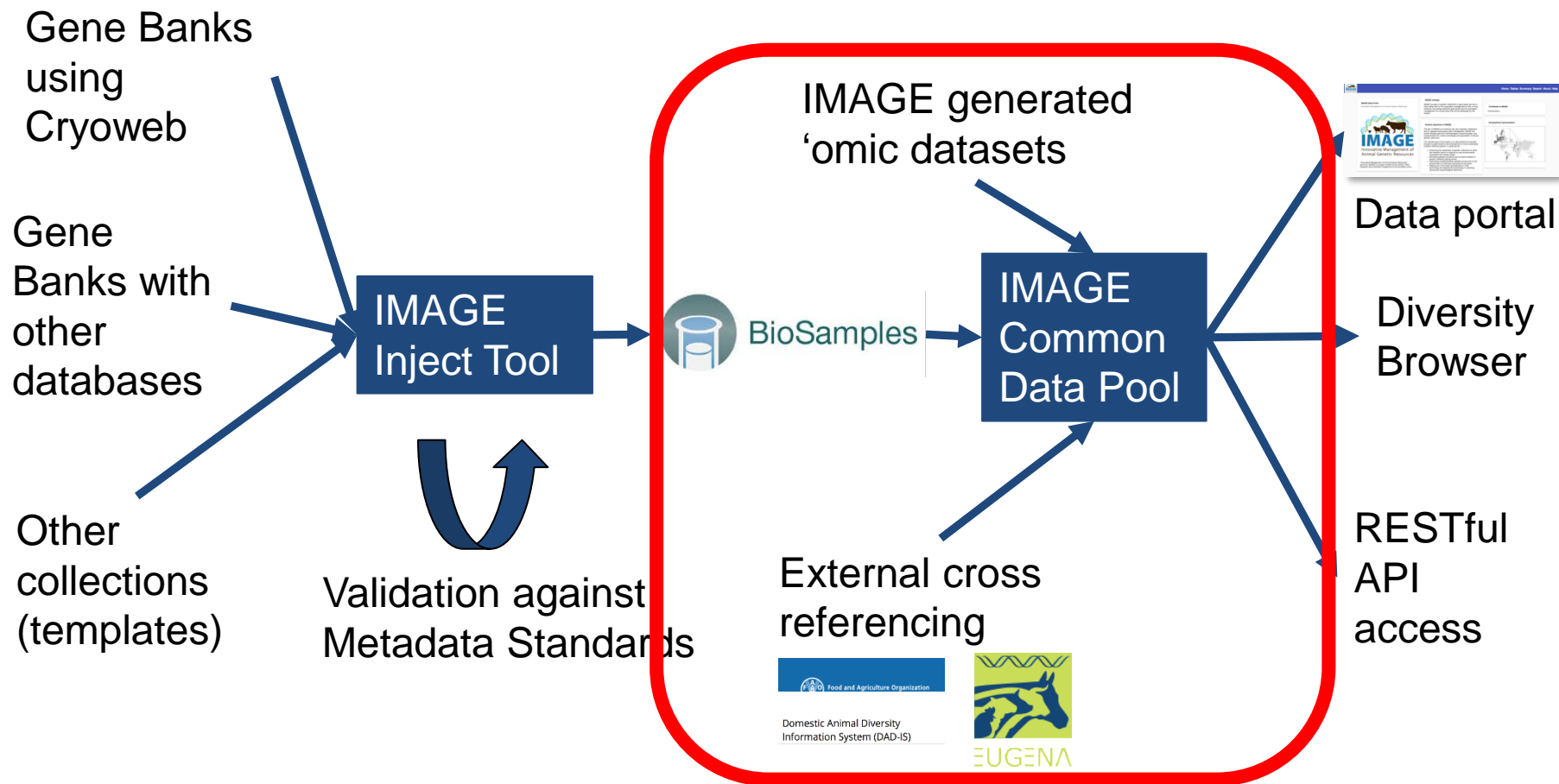


IMAGE Common Data Pool

- Retrieve all IMAGE data (each day)
 - <https://api.image2020genebank.eu/backend/organism/>
 - <https://api.image2020genebank.eu/backend/specimen/>
- Optimize Searches
- Provide additional informations
 - https://api.image2020genebank.eu/backend/dadis_link/
 - <https://api.image2020genebank.eu/backend/organism.geojson/>
 - <https://api.image2020genebank.eu/backend/specimen.geojson/>
- Application Programming Interface (API)

The IMAGE Portal

- Builded on top of Common Data Pool
- Simplifies common operations (search, browse, ...)

<https://www.image2020genebank.eu/>

Issues with data

- We are still waiting for partners data (genotypes, sequences)
 - There's a issue with a submission in particular (ex. [SAMEA7044752](#))
 - How a correct biosample record should be:
 - <https://www.ebi.ac.uk/biosamples/samples/SAMEA2591108>
 - <https://www.ebi.ac.uk/ena/browser/view/SAMEA2591108>
- Once issues are solved we plan to support searching data files within IMAGE Portal

References

- IMAGE Metadata
 - <https://github.com/cnr-ibba/IMAGE-metadata>
- IMAGE InjectTool
 - <https://inject.image2020genebank.eu/>
 - [About IMAGE-InjectTool](#)
 - [Submitting data into BioSamples](#) (video included!)
 - Source code: <https://github.com/cnr-ibba/IMAGE-InjectTool>
 - [Technical docs](#) (developers only!)
- IMAGE Common Data Pool (CDP)
 - <https://api.image2020genebank.eu/>
 - Source code: <https://github.com/cnr-ibba/IMAGE-CommonDataPool>
- IMAGE Portal
 - <https://www.image2020genebank.eu/>
 - Source code: <https://github.com/cnr-ibba/IMAGE-Portal>