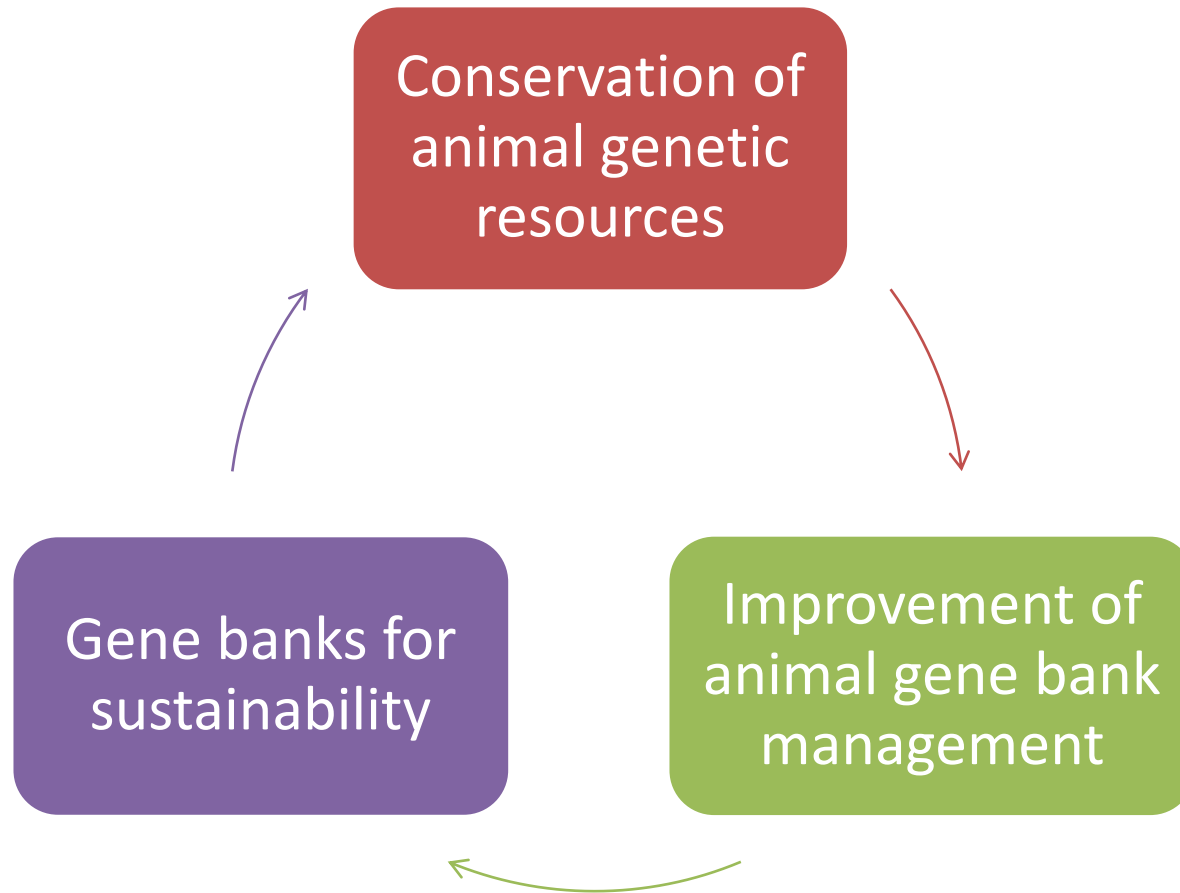


# The IMAGE unified data portal to integrate and represent European gene bank data

Paolo Cozzi, Jun Fan, Alexey Sokolov, Oliver Selmoni, Elia Vajana, Stéphane Joost, Eildert Groeneveld, Paul Flicek, Peter Harrison, Alessandra Stella & the IMAGE Consortium



# The Innovative Management of Animal Genetic Resources (IMAGE)



# Exploiting data as a Community

**01**

## Poorly used collections

Gene banks collections are a backup of animal genetic diversity that are poorly used. Data are complex by nature

**02**

## Information Integrations

Challenging task due to heterogeneity and large amount of data generated

**03**

## Standard definition

We need to unify access to gene bank data and create standards and protocols to exchange data



# Data Management Component Objectives

1

European web portal which integrates data from gene banks and collections

Unify data access to gene bank data  
(60+ databases, 20+ countries)

2

3

Data management conceptual model definition



# IMAGE Conceptual Model

Ruleset to ensure  
comparable data

Collect,  
standardize and  
submit data

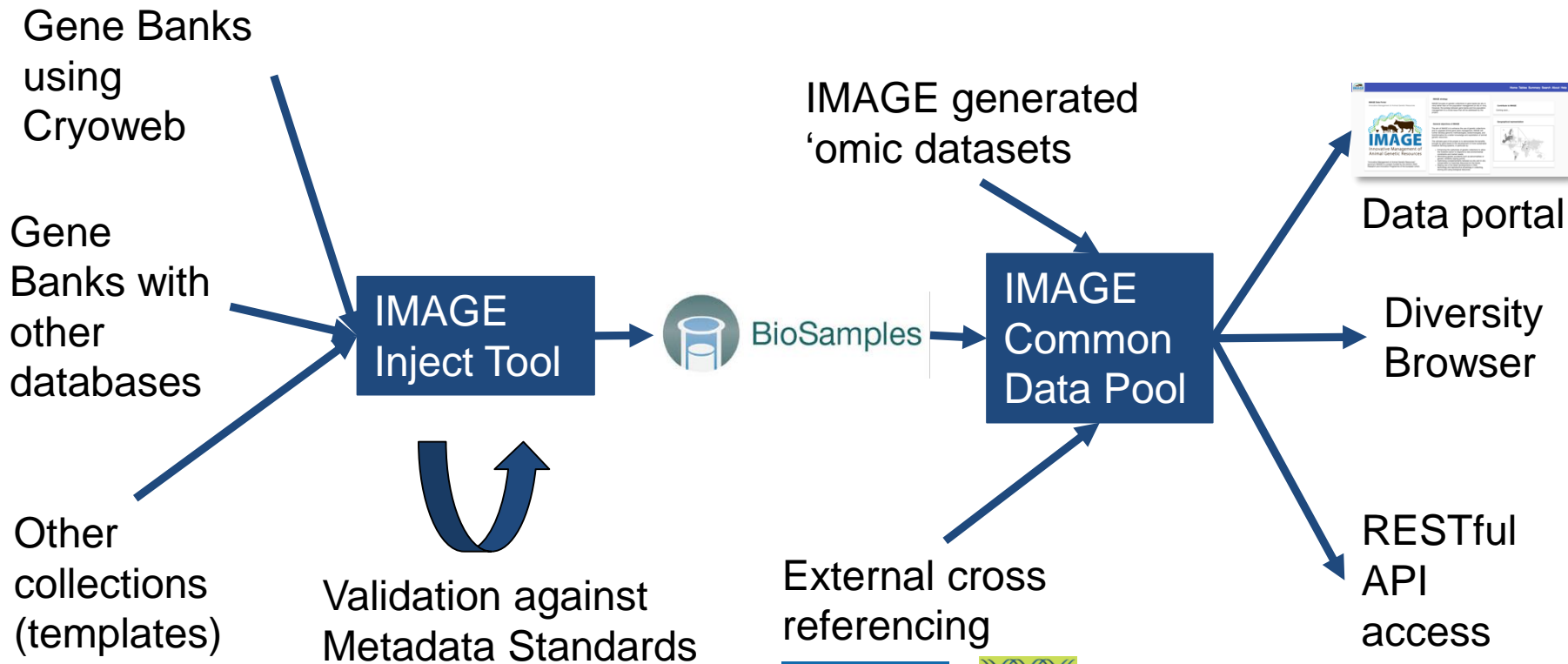
Data sustainability  
and persistence

Data integration

- Support for GIS
- Querying accross data



# IMAGE Data model



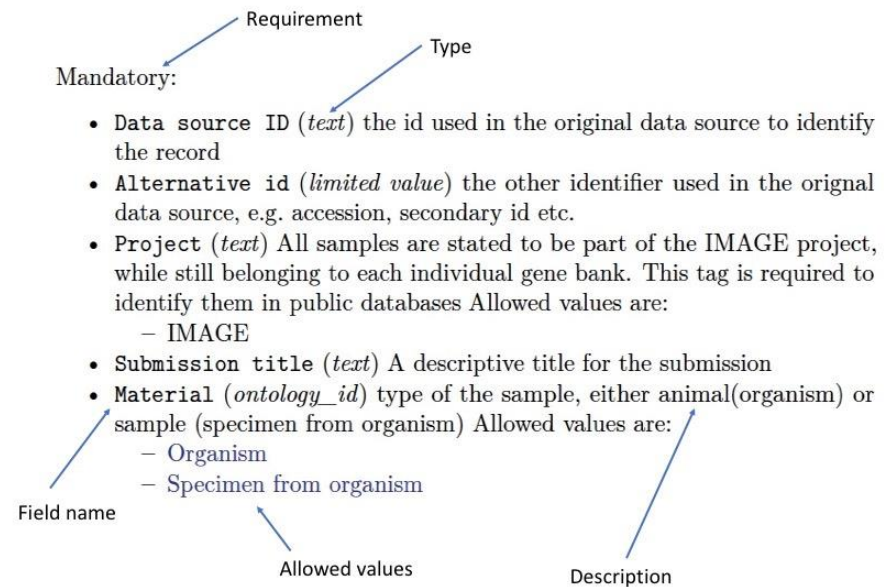
Food and Agriculture Organization  
Domestic Animal Diversity  
Information System (DAD-IS)



# Make data comparable

<https://github.com/cnr-ibba/IMAGE-metadata>

**Metadata ruleset:** Provides information about the data which is a list of fields with different names, **data types**, **descriptions**, **requirements**, **cardinalities**, **allowed values**.



# IMAGE-ValidationTool

<https://github.com/cnr-ibba/IMAGE-ValidationTool>

Preliminary  
check

- USI structure
- Duplicates
- Ruleset

Ruleset-based  
validation

- Individual attribute
- Cardinality
- Valid values/terms/units

Context  
validation

- Cross more than one attribute
- Birth location ⇔ Birth location accuracy
- Species ⇔ Breed
- Child ⇔ Parent(s)



# Data Validation

## Validation summary for samples

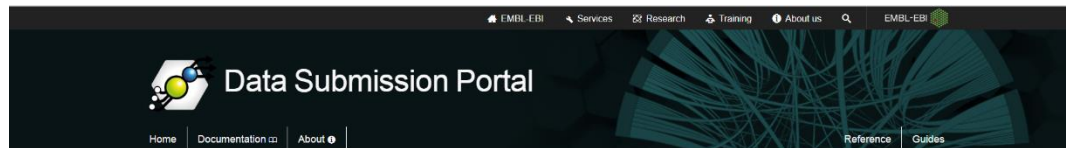
Pass: 4 Warnings: 0 Errors: 180 JSON Errors: 0

176	Error: Value Milano provided for field Collection place but value in field Collection place accuracy is missing geographic information	<a href="#">Fix this</a>
4	Error: No url found for the field Organism part which has the type of ontology_id (specimen from organism section)	<a href="#">Fix this</a>
2	Warning: Ontology provided for field EFABIS Breed country however there is no requirement in the ruleset (organism section)	<a href="#">Fix this</a>
5	Error: No value provided for field Birth location but value in field Birth location accuracy is not missing geographic information	<a href="#">Fix this</a>
10	Error: decimal degree for field Collection place latitude is not in the valid units list (decimal degrees) (specimen from organism section)	<a href="#">Fix this</a>
15	Warning: Fail to retrieve record SAMEA0000001 from BioSamples as required in the relationship	<a href="#">Fix this</a>
3	Error: Mandatory field Material in standard section could not be found	<a href="#">Fix this</a>

A report to understand where errors are



# Biosamples and Unified Submission Interface



- Sustainability
- Cross-references

## Overview

### Contents

Accessing the service  
HTTP Verbs  
HTTP status codes  
Authentication  
Errors  
Hypermedia

This describes the resources that make up the EMBL-EBI submissions API. If requests, please contact our helpdesk.

### Accessing the service

The submission API is available from <https://submission.ebi.ac.uk/api>.

- Programmatic Access

The screenshot displays the 'Create a submission' page. It includes a 'Request' section with a curl command and a 'Response' section showing an HTTP 201 Created status and a JSON response. A sidebar on the right lists navigation links: Overview, API root, Submissions, Teams, Projects, Samples, Studies, Assays, Assay data, File upload, Validation results, UI Support, Spreadsheets and templates, Submission plans, and Data types.

**Create a submission**

**Request**

```
$ curl 'https://submission-test.ebi.ac.uk/api/teams/subs.dev-team-1/submissions' -i -X POST -H 'Content-Type: application/json;charset=UTF-8' -H 'Accept: application/hal-json' -H 'Authorization: Bearer TOKEN' -d '{}'
```

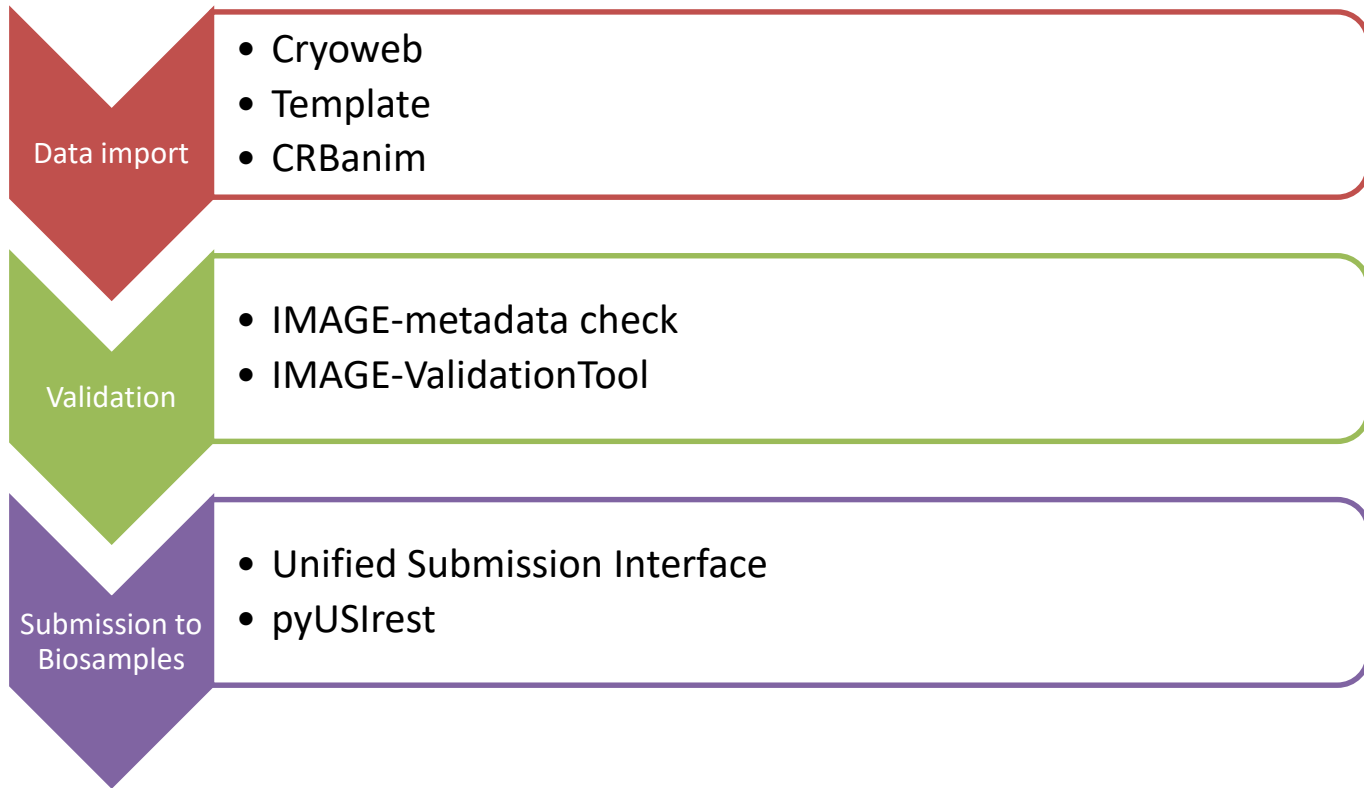
**Response**

```
HTTP/1.1 201 Created
ETag: "0"
Last-Modified: Tue, 09 Jul 2019 10:56:39 GMT
Location: https://submission-test.ebi.ac.uk/api/submissions/998780f9-d5cd-434b-92b7-9d27a9585c48
Content-Type: application/hal-json;charset=UTF-8
Content-Length: 3064

{
  "id": "998780f9-d5cd-434b-92b7-9d27a9585c48",
  "submitter": {
    "email": "alice@example.com",
    "name": "no name"
  },
  "team": {
    "name": "subs.dev-team-1",
    "profile": {
      "centre name": "An Institute"
    }
  },
  "_links": {
    "submissions": {
      "href": "https://submission-test.ebi.ac.uk/api/submissions/search/by-team?teamName=subs.dev-team-1"
    },
    "submissions:create": {
      "href": "https://submission-test.ebi.ac.uk/api/teams/subs.dev-team-1/submissions"
    },
    "items": {
      "href": "https://submission-test.ebi.ac.uk/api/teams/subs.dev-team-1/items"
    }
  }
}
```



<https://www.wp5image.eu/image/>



# Create a submission and upload data

powerpoint smart... slides and videos... Validation tool Ma... PowerPoint FAQ: H... 4 Text Boxes for Po... 10 ways to avoid de... 6 Alternatives to Bu... Upload a Datasour...

https://www.wp5image.eu/image/submissions/create/

App Bio Linux - admin Programmazione Tempo Libero Bookmarks

IMAGE-InjectTool Home About Admin Dashboard paolo

example: CryoWeb

Gene bank country:  
Italy (NCIT\_C16761)

Organization:  
IBBA-CNR (Italy)

Who owns the data

Data source type:  
-----

CryoWeb  
Template  
CRB-Anim

examples: "2018-04-27", "version 1.5"

Uploaded file:  
Scegli file Nessun file selezionato

Need to be in UTF-8 or ASCII format

Create Submission Return to dashboard

Choose your data type



# Submission detail

The screenshot displays the 'Submission Detail for Italian Cryoweb Subset' page in the IMAGE-InjectTool web application. The page features a table with submission details and a set of action buttons at the bottom. A red rectangle highlights the 'Edit data', 'Validate', and 'Submit' buttons.

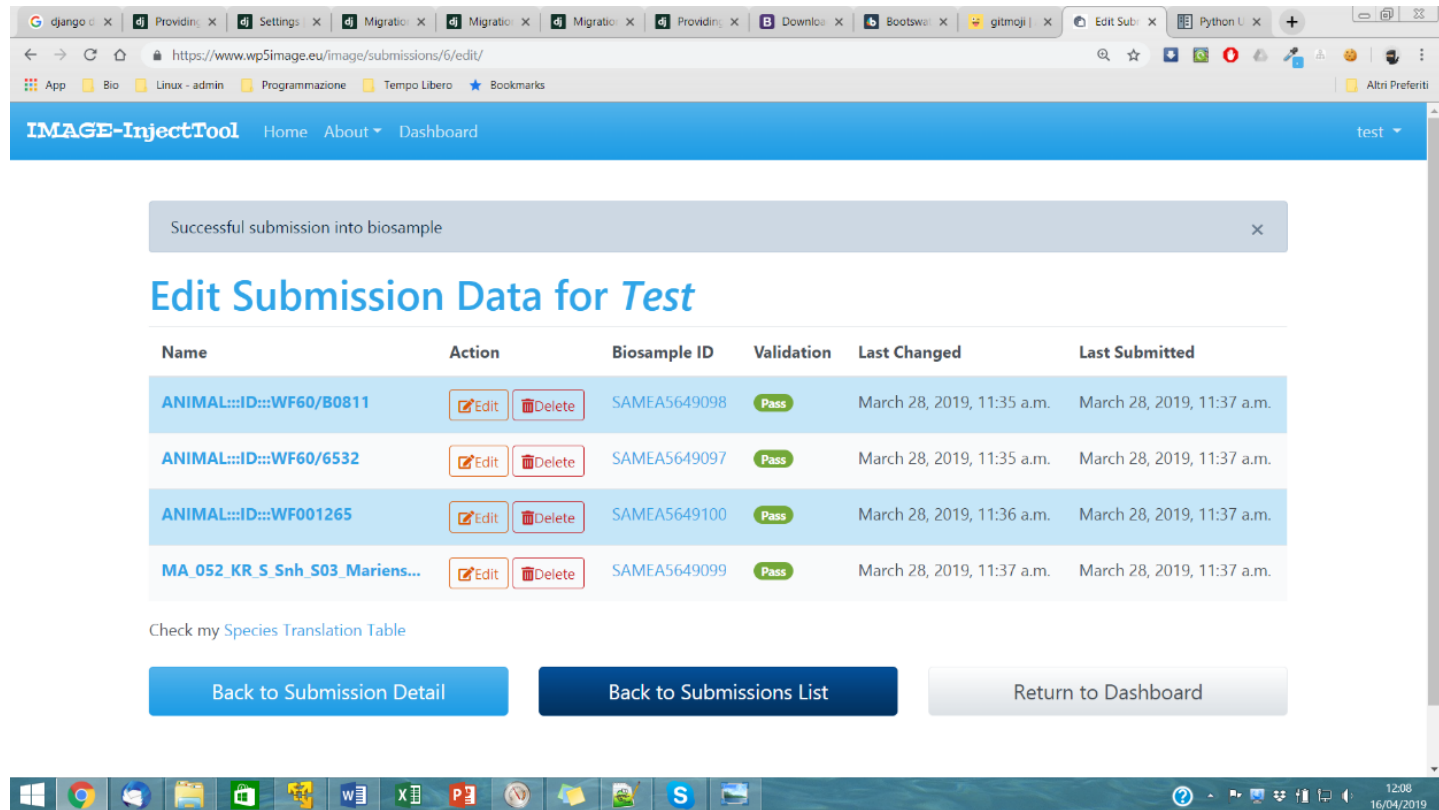
Field	Value
Submission Title:	Italian Cryoweb Subset
Description:	Italian Cryoweb Subset
Gene Bank Name:	Cryoweb IT
Gene Bank Country:	Italy (NCIT_C16761)
Organization:	IBBA-CNR (Italy)
Data Source Type:	CryoWeb
Data Source version:	Subset
Data File:	cryoweb_alessandra_data_only.sql
Created at:	Feb. 5, 2019, 4:41 p.m.
Status:	Need Revision
N. of Animals	26 (8 with issues in validation)
N. of Samples	184 (184 with issues in validation)

Buttons: Edit data, Validate, Submit, Submissions List, Return to Dashboard, Danger Zone

Edit, Validate, Submit data by clicking on buttons



# A complete submission



The screenshot shows the IMAGE-InjectTool web interface. At the top, a blue navigation bar contains the logo and links for Home, About, and Dashboard. A success message at the top reads "Successful submission into biosample". Below this, the page title is "Edit Submission Data for *Test*". A table lists four submissions with columns for Name, Action, Biosample ID, Validation, Last Changed, and Last Submitted. Each row includes "Edit" and "Delete" buttons. At the bottom, there are three 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	<a href="#">Edit</a> <a href="#">Delete</a>	SAMEA5649098	Pass	March 28, 2019, 11:35 a.m.	March 28, 2019, 11:37 a.m.
ANIMAL::ID::WF60/6532	<a href="#">Edit</a> <a href="#">Delete</a>	SAMEA5649097	Pass	March 28, 2019, 11:35 a.m.	March 28, 2019, 11:37 a.m.
ANIMAL::ID::WF001265	<a href="#">Edit</a> <a href="#">Delete</a>	SAMEA5649100	Pass	March 28, 2019, 11:36 a.m.	March 28, 2019, 11:37 a.m.
MA_052_KR_S_Snh_S03_Mariens...	<a href="#">Edit</a> <a href="#">Delete</a>	SAMEA5649099	Pass	March 28, 2019, 11:37 a.m.	March 28, 2019, 11:37 a.m.

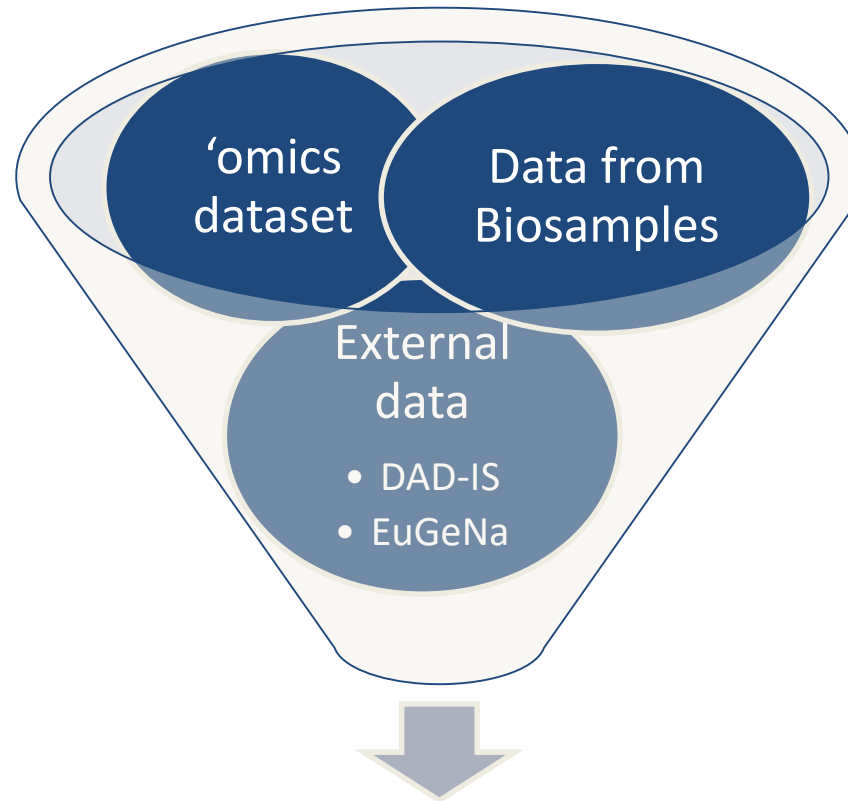
Check my [Species Translation Table](#)

[Back to Submission Detail](#) [Back to Submissions List](#) [Return to Dashboard](#)

Update Biosamples records through InjectTool



# Common Data Pool



Common Data Pool

- Daily sync, collate data, API



-

# Geographic Referencing Tool

- Click-on-map interface
- Set coordinates and their accuracies

id	species	breed	longitude	latitude	accuracy	sample
a1	Cattle	hollstein				semen <input checked="" type="checkbox"/>
a2	Cattle	hollstein				semen <input checked="" type="checkbox"/>
a3	Sheep	moroccan				semen <input checked="" type="checkbox"/>
a4	Sheep	moroccan				semen <input type="checkbox"/>
a5	Cattle	brown	2.56465	46.21111	regional	semen <input type="checkbox"/>
a6	Goat	wild				semen <input type="checkbox"/>
a7	Goat	wild				semen <input type="checkbox"/>
a8	Goat	wild				semen <input type="checkbox"/>
a9	Goat	wild				semen <input type="checkbox"/>

Choose coordinate accuracy level:

Exact Coordinate



# Data portal Detail



**Organism:** geo\_test\_7

**Alternative id:** 7

**Submission title:** test

**Material:** [Organism](#)

**Person last name:** Sokolov

**Person email:** admin@admin.com

**Person affiliation:** EBI

**Person role:** submitter

**Organization name:** EBI

**Organization role:** submitter

**Gene bank name:** test

**Gene bank country:** GBR

**Data source version:** CyroWeb

**Data source version:** 1

**Species:** Gallus gallus

**Gene bank country:** GBR

**Data source version:** CyroWeb

**Data source version:** 1

**Species:** Gallus gallus

**Supplied breed:** test

**EFABIS Breed country:** AFG

**Sex:** female

**Birth location accuracy:** missing geographic information

**Birth location longitude:** 13.4073130000

**Birth location latitude:** 52.5174700000



[Home](#) [Tables](#) [Summary](#) [Search](#) [About](#) [Help](#)

## Organisms summary

### Summary map

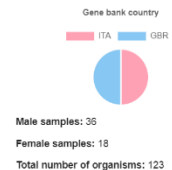
Birth location coordinates for organisms



### Summary of species and breed distributions



### Summary of sex and country distributions



## Specimens summary

### Summary of species and organism part distributions

Geographical coordinates on maps



# Conclusions

- Effective data sharing from gene banks across Europe requires defining clear standards and providing tools to validate against these standards.
- Long term sustainability ensured through submission to public data archives at EMBL-EBI
- Our data portal provides the community with a single point to access to gene bank data, added value 'omics datasets and cross references to external resources e.g. EugENA
- Our tool simplifies data submission process for genebanks, handling validation and brokering submission to public archives and IMAGE.
- Currently in final stages of testing, production release for full datasets in coming months



# Acknowledgement

Financial support from the European Union's Horizon 2020 research and innovation programme under grant agreement No 677353

Jun Fan  
Alexey Sokolov  
Paul Flicek  
Peter Harrison  
Eildert Groeneveld  
Alessandra Stella  
Oliver Slemoni  
Elia Vajana  
Stéphane Joost  
Sylvain Marthey  
Michèle Boichard  
Sipke-Joost Hiemstra

Richard Crooijmans  
Dirkjan Schokker  
Licia Colli

& The IMAGE Consortium

