

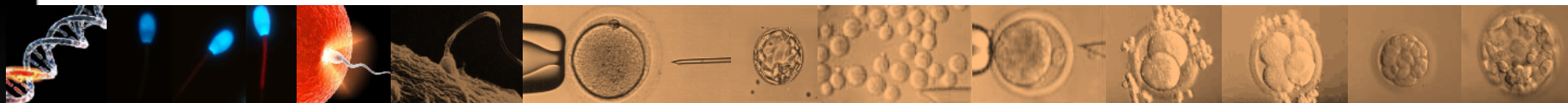


European Project n° 677353

# Future of IMAGE: information system and portal

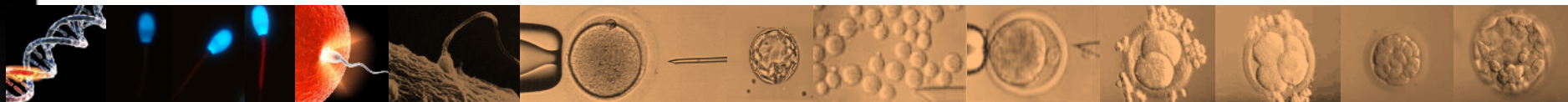
Ghent

August 25th, 2019



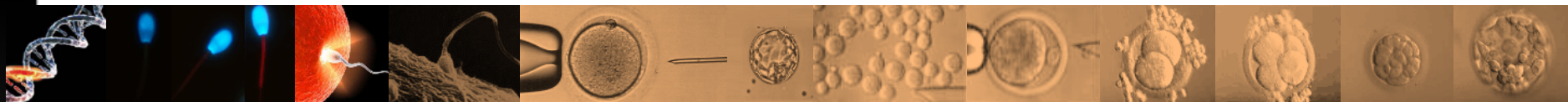
# WP5 IMAGE

One of the main goals of IMAGE is the creation of a European web portal which integrates data from gene banks and collections, all over Europe, with genomics data, geographical information systems data, and other information generated by IMAGE



# Objectives and development

- Specific objective towards the development of such infrastructure were:
  - to define and refine rules and implement the metadata standardization; use of the ontologies to improve data quality and comparability
  - to construct (coding and testing) a custom Inject Tool that applies the ruleset to import data from gene banks, unifies units, terms and languages, and submits the enhanced data to public EMBL-EBI BioSamples archive

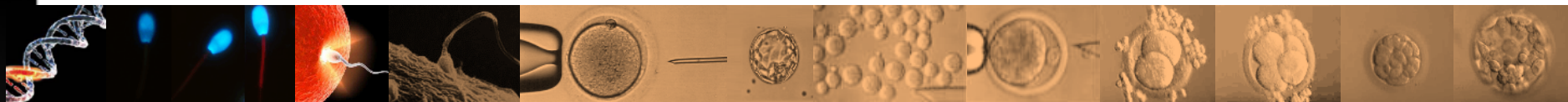


- to create the Common Data Pool to integrate:
  - Data from participating gene banks (archived in BioSamples)
  - IMAGE 'omics datasets
  - External data (DAD-IS, EuGeNa)
- to guide partners in the input of gene bank /collection data
- to start populating the common data pool and feed a diversity browser and the IMAGE portal

# IMAGE gene bank metadata standardisation



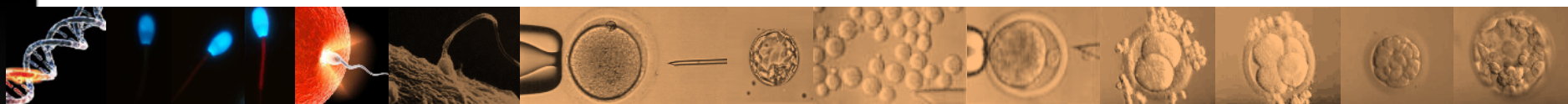
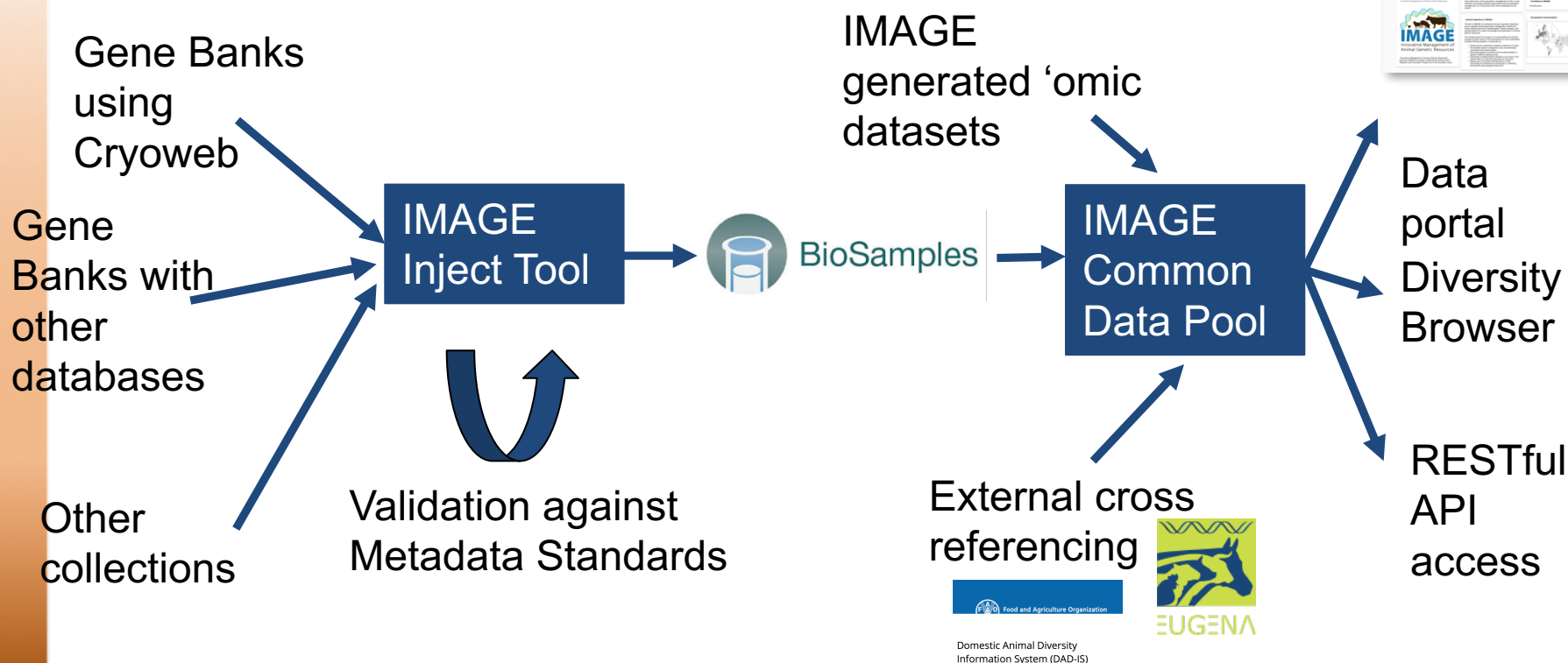
- IMAGE metadata standards and data model
- IMAGE ruleset standardises metadata from diverse gene bank records
- Utilises ontologies (controlled dictionaries for biology) to ensure consistency in supplied terms and synonyms, and provide structure
- Metadata standards are not fixed and evolve with requirements of consortium (now version 1.6 June 2019), always open to consider further changes from the consortium.
- <https://github.com/cnr-ibba/IMAGE-metadata/tree/master/docs>



# IMAGE Data model



Simplified schematic of IMAGE data model.



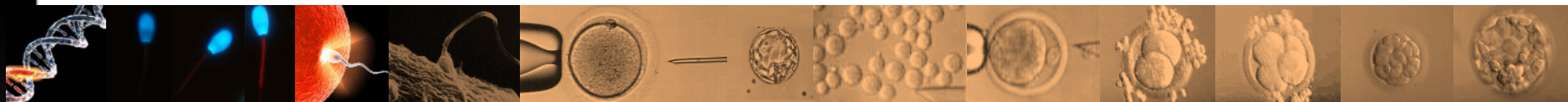
# Validation



- Rule-based contextual validation
- Standardises language, terminology, units and ontologies
- Ensures minimum mandatory fields have been completed correctly
- Prevents errors and duplications
- Feeds back errors clearly for correction through Inject tool interface
- Test driven development in Python and Django framework
- Validation code built into Inject Tool, code viewable at <https://github.com/cnr-ibba/IMAGE-ValidationTool>



Jun  
Fan





# EMBL-EBI BioSamples archive



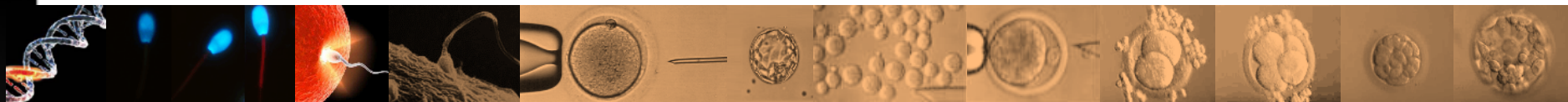
BioSamples



- Inject tool manages submission of IMAGE data to BioSamples archive.
- BioSamples assigns universally recognized unique identifier to each organism and specimen e.g. [SAMEA5159537](#)
- Provides sustainable and secure long term storage of information
- IMAGE utilising new EMBL-EBI Unified Submissions Interface as part of the Inject Tool for a more stable and responsive submission process
- Easier to broker and track submissions on users behalf
- Fully supports IMAGE metadata standards
- Ready and awaiting IMAGE data from Inject Tool



Photo: Peter Harrison





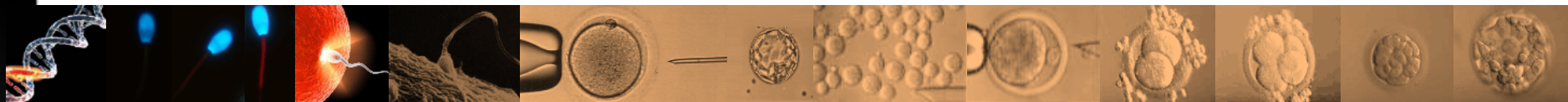
# IMAGE Common Data Pool



- The IMAGE common database for all IMAGE data
- Daily sync of gene bank metadata tagged as being from the 'IMAGE' project from EMBL-EBI BioSamples archive
- Collates this metadata with geographic information, IMAGE produced 'omics sequencing data and cross references to external breed and gene bank resources
- Built in PostgreSQL to be compatible with GIS tools from EPFL
- Acts as backend data source for IMAGE Data Portal, Diversity Browser and programmatic access (<https://wp5image.eu/help/api>)

```
curl https://wp5image.eu/data_portal/backend/organism/?format=json
```

- Production release 2018, accepting IMAGE data



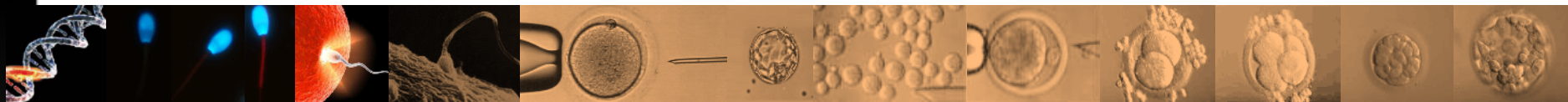
# The IMAGE data portal



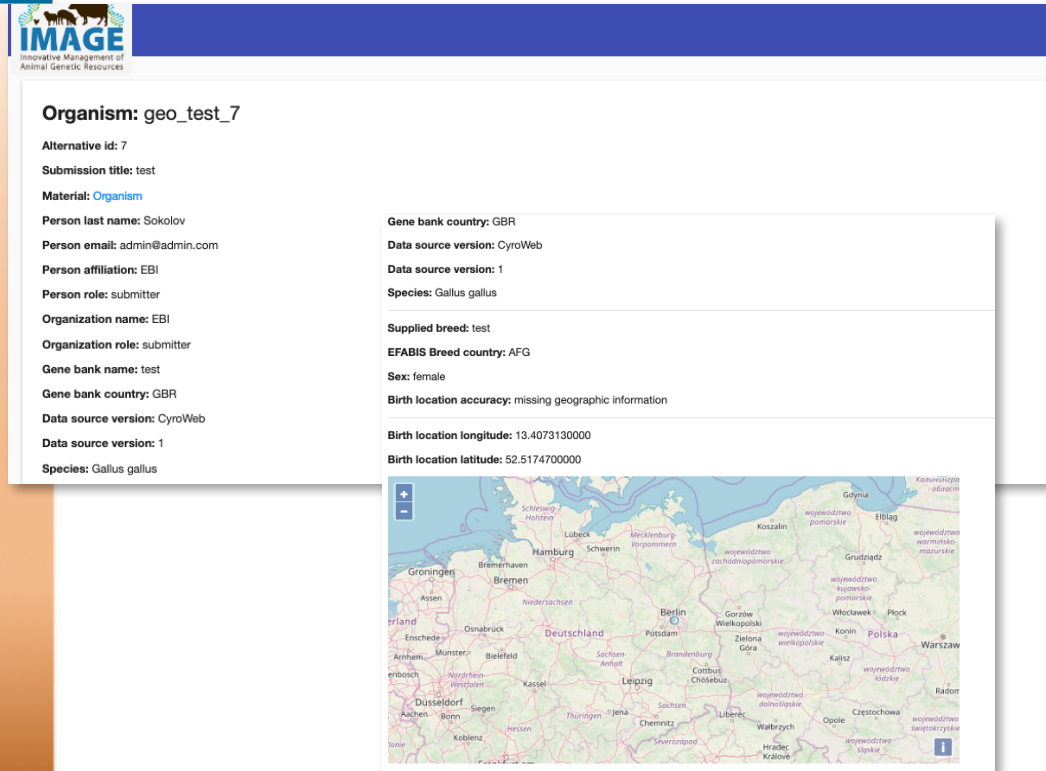
- Single access point to all IMAGE data
- Powerful metadata filters, e.g. by species
- Download filtered results
- Access additional information
- Modern Angular design

The screenshot displays the IMAGE data portal interface. At the top, there is a blue navigation bar with the IMAGE logo on the left and links for Home, Tables, Summary, Search, About, and Help on the right. Below the navigation bar, there are two tabs: Organisms and Specimens. The Specimens tab is currently selected. Under the Specimens tab, there is a 'Filters' section with a dropdown arrow. The filters include Species (set to 'Sus scrofa: 2'), Supplied breed, and Sex. Below the filters, there is an 'Active filters' section with a 'Remove all filters' button and a 'Download data' button. The main content area shows a table with the following columns: Data source ID, Species, Supplied breed, and Sex. The table contains two rows of data, both for 'Sus scrofa' from 'Cinta Senese', with 'male' sex. The first row has Data source ID 'SAMEA5159538' and the second row has 'SAMEA5159535'. At the bottom right of the table, there is a pagination control showing 'Items per page: 5' and '1 - 2 of 2'.

Data source ID	Species	Supplied breed	Sex
<a href="#">SAMEA5159538</a>	Sus scrofa	Cinta Senese	male
<a href="#">SAMEA5159535</a>	Sus scrofa	Cinta Senese	male



# The IMAGE data portal – detailed record view



**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

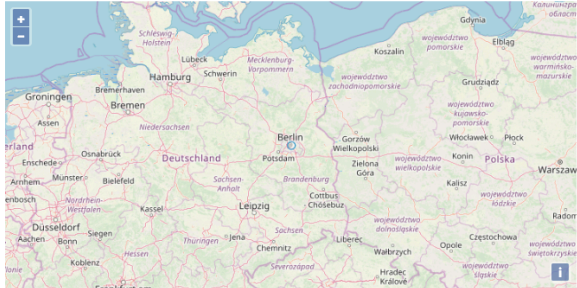
**EFABIS Breed country:** AFG

**Sex:** female

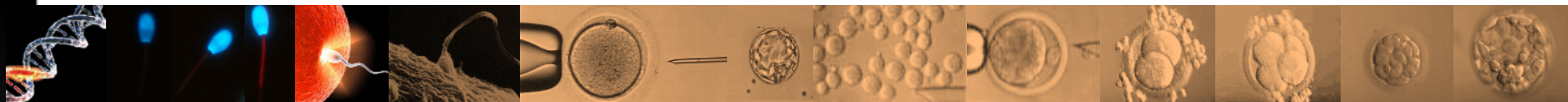
**Birth location accuracy:** missing geographic information

**Birth location longitude:** 13.4073130000

**Birth location latitude:** 52.5174700000



- Detailed metadata descriptions for each gene bank record.
- Gene bank contact information for access.
- Cross references to EUGENA and DAD-IS (coming soon).
- Coordinates automatically generate interactive geographic views, utilizing code developed by EPFL.

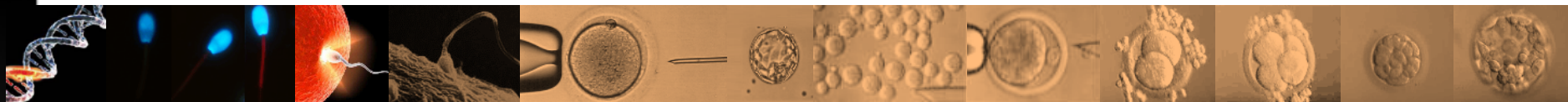


# The IMAGE data portal – Advanced Search

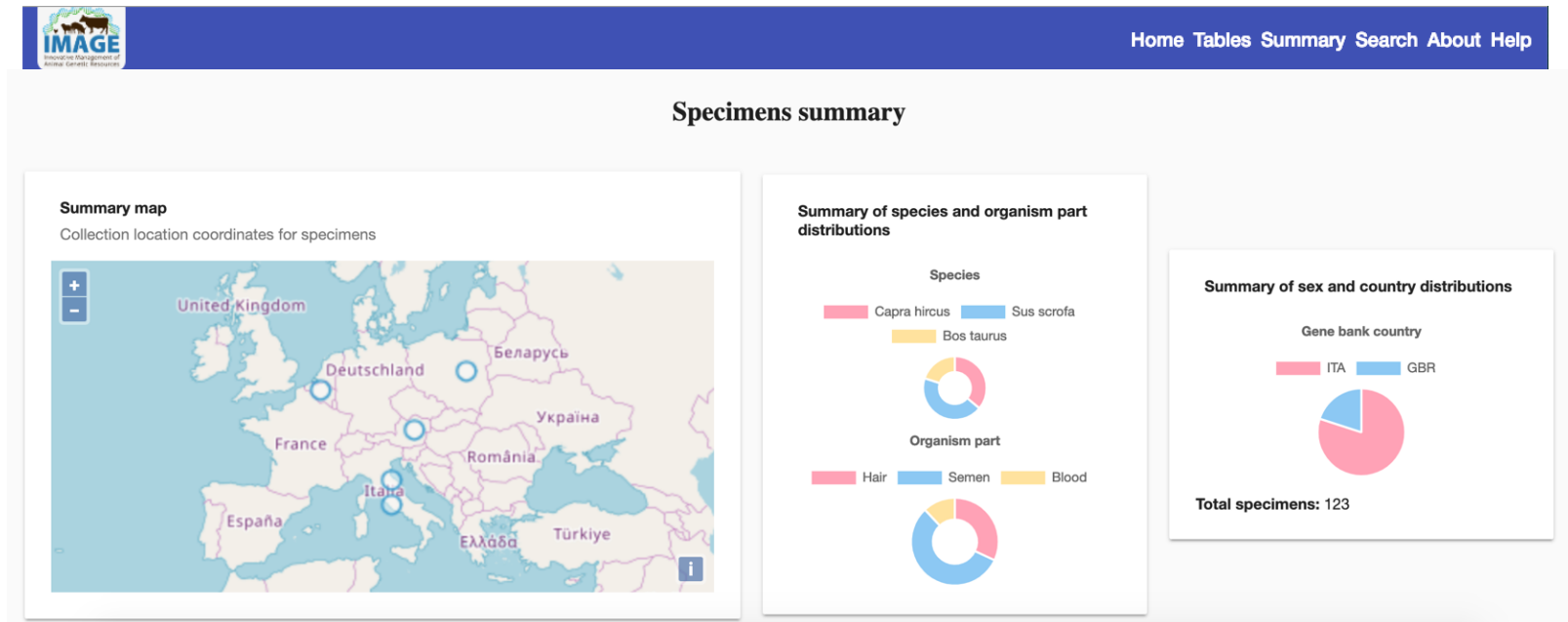


 <span>Home Tables Summary Search About Help</span>			
Search cap			
Organisms:			
Data source ID	Species	Supplied breed	Sex
<a href="#">SAMEA5159526</a>	Capra hircus	Verzaschese	male
<a href="#">SAMEA5159534</a>	Capra hircus	Verzaschese	male
<a href="#">SAMEA5159537</a>	Capra hircus	Verzaschese	male

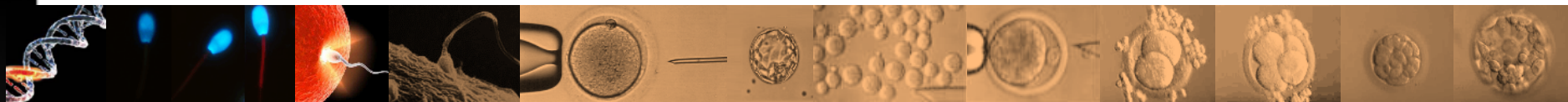
- Predictive text based search across all IMAGE metadata fields
- Returns organisms, specimens and datasets (when available)
- Search results link directly to individual detail pages



# The IMAGE data portal – Summary statistics



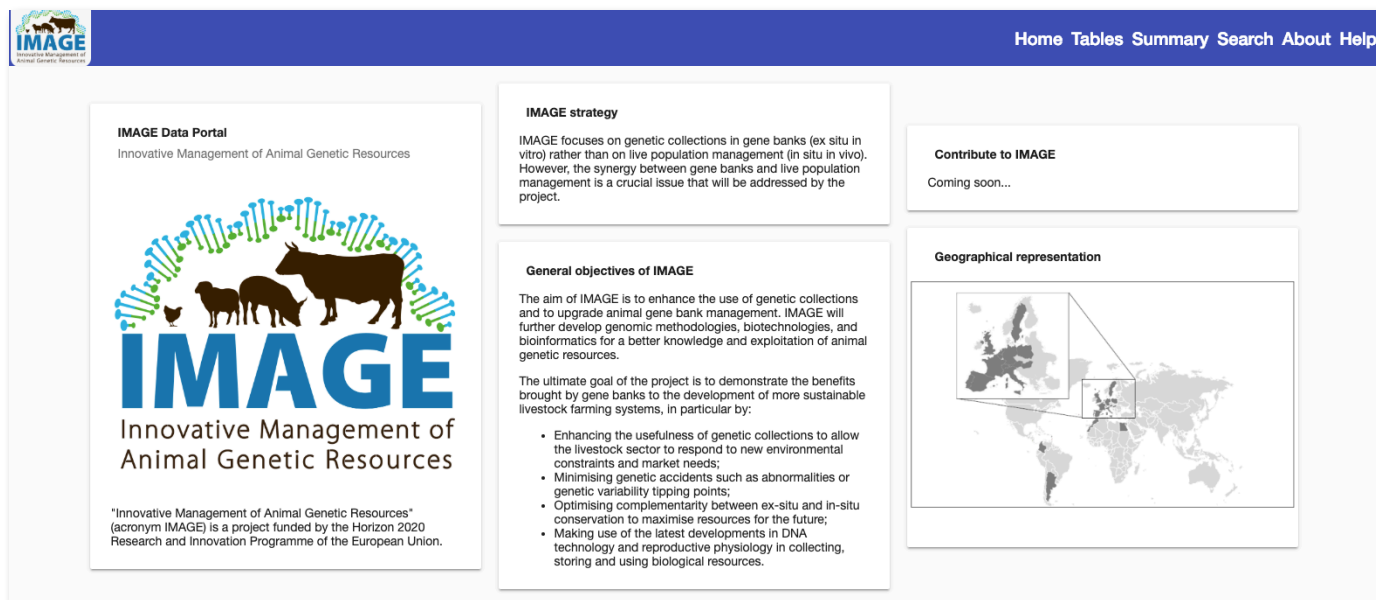
- Automatically generated, self updating and interactive summary statistics.



# The IMAGE data portal – feedback welcome

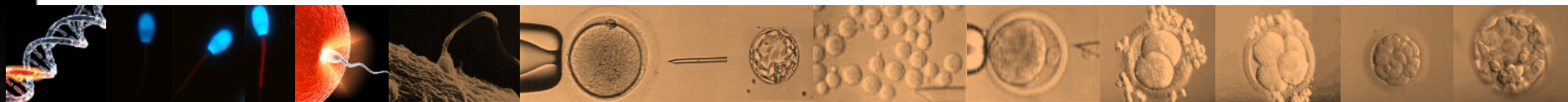


- Currently contains a small test dataset.
- Portal service fully ready to accept IMAGE data.
- Awaiting permanent URL from CNR
- Feedback and feature requests very welcome



<https://wp5image.eu>

Alexey  
Sokolov





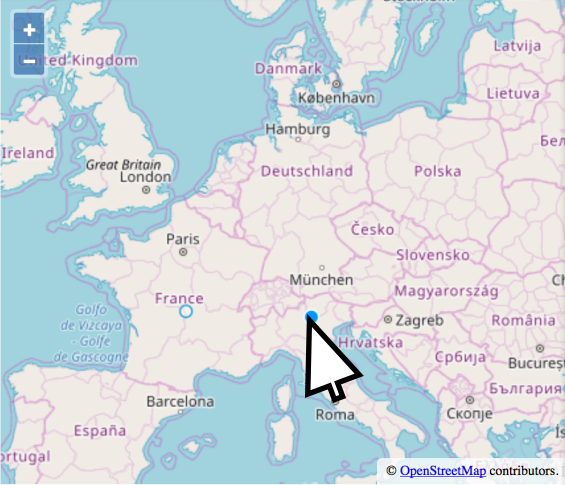
# Geographic referencing tool

- For new submissions (via excel template) and curation
- Not mandatory
- Facilitate georeferencing of samples → '*click-on-map*' interface
- Allow to set coordinates accuracy/confidence level

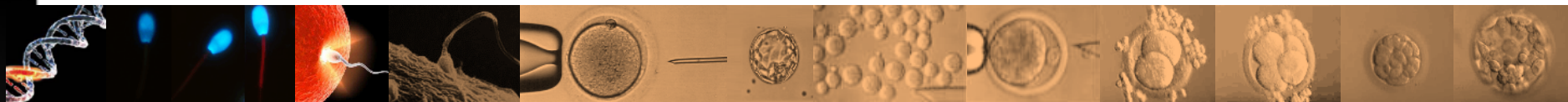
id	species	breed	longitude	latitude	accuracy	sample	
a1	Cattle	hollstein				semen	<input checked="" type="checkbox"/>
a2	Cattle	hollstein				semen	<input checked="" type="checkbox"/>
a3	Sheep	moroccan				semen	<input checked="" type="checkbox"/>
a4	Sheep	moroccan				semen	<input type="checkbox"/>
a5	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   
Cancel   
Send Data to DB

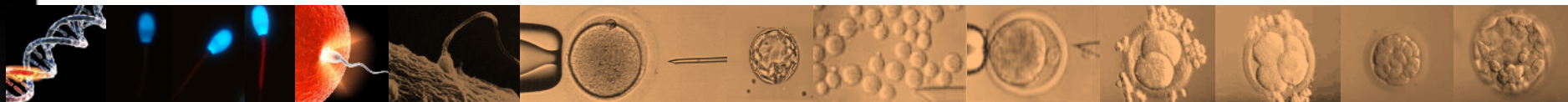


© OpenStreetMap contributors.



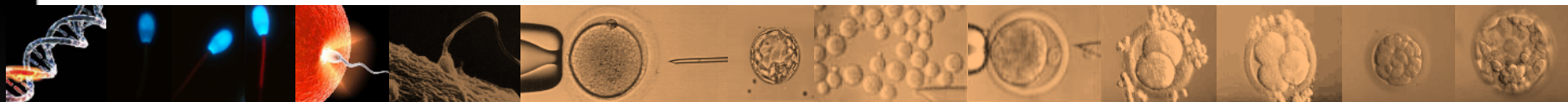
# The Inject Tool

- Video will drive you through the registration and submission (Paolo Cozzi)
- [https://youtu.be/ etbl5wPwxY](https://youtu.be/etbl5wPwxY)



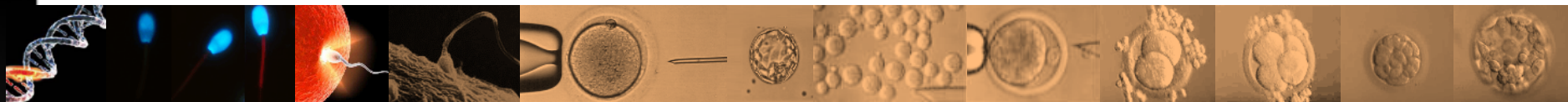
# The Inject Tool

- INSTRUCTIONS ON CREATING A CRYOWEB DUMP
  - Please see  
[https://www.wp5image.eu/image/about\\_uploading/#cryoweb\\_dump](https://www.wp5image.eu/image/about_uploading/#cryoweb_dump)
- INSTRUCTIONS FOR COMPLETING THE EXCEL TEMPLATE
  - Please see  
[https://www.wp5image.eu/image/about\\_uploading/#template\\_file](https://www.wp5image.eu/image/about_uploading/#template_file)
- INSTRUCTIONS ON CREATING A CRB-anim DUMP
  - Please see  
[https://www.wp5image.eu/image/about\\_uploading/#crbanim\\_dump](https://www.wp5image.eu/image/about_uploading/#crbanim_dump)



# Current activities and issues

- Selection of data to be submitted
  - Importing subsets of the whole DB
- Re-submission and editing of submitted data
- Revisions of ruleset
  - Feedback received from submitters
- GDPR compliance, Privacy Policy



## Next steps

