WorldCereal

WorldCereal MOOC I

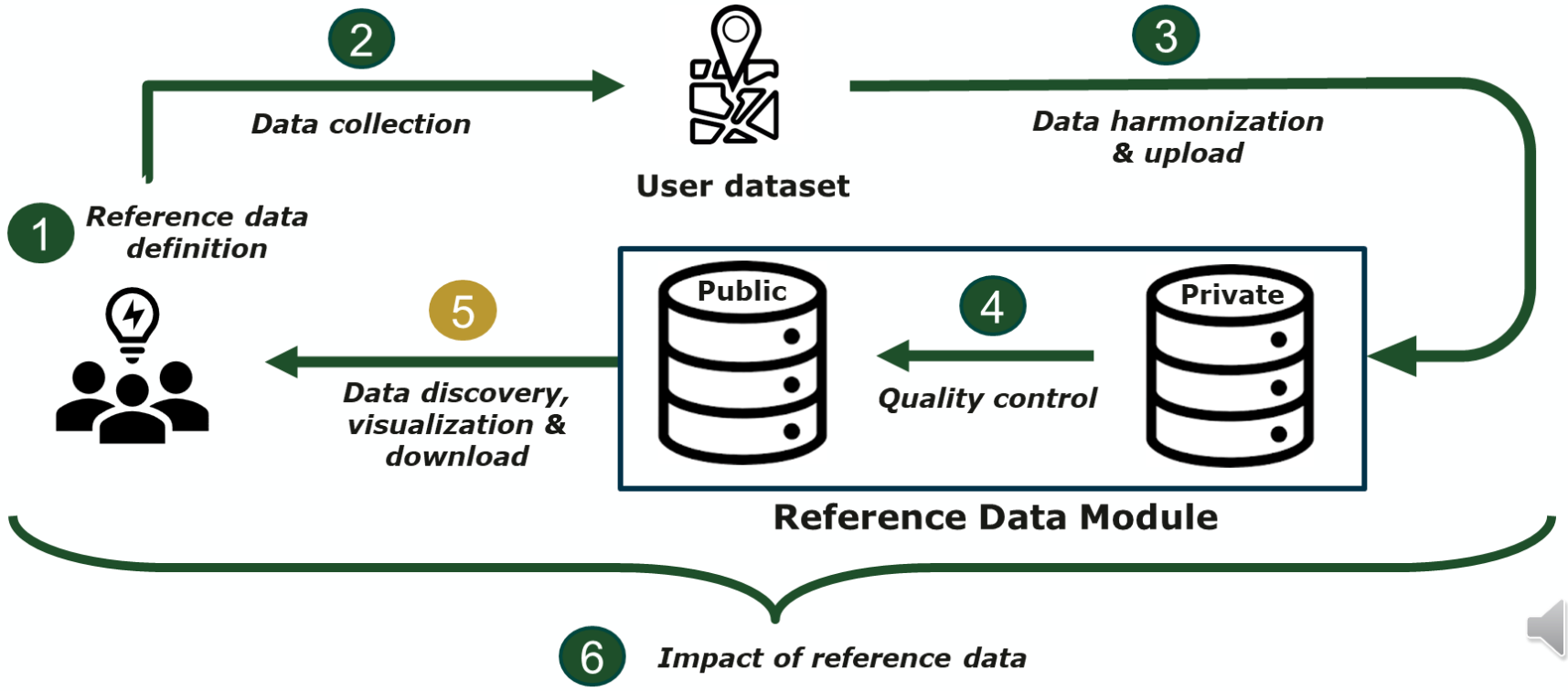
Introduction to the WorldCereal Reference Data Module (RDM)

Santosh Karanam (IIASA)

vito · WAGENINGEN UNIVERSITY & RESEARCH · IIASA · UNIVERSITÉ DE STRASBOURG · Universitat de València · gisat

→ THE EUROPEAN SPACE AGENCY

MOOC I: Outline



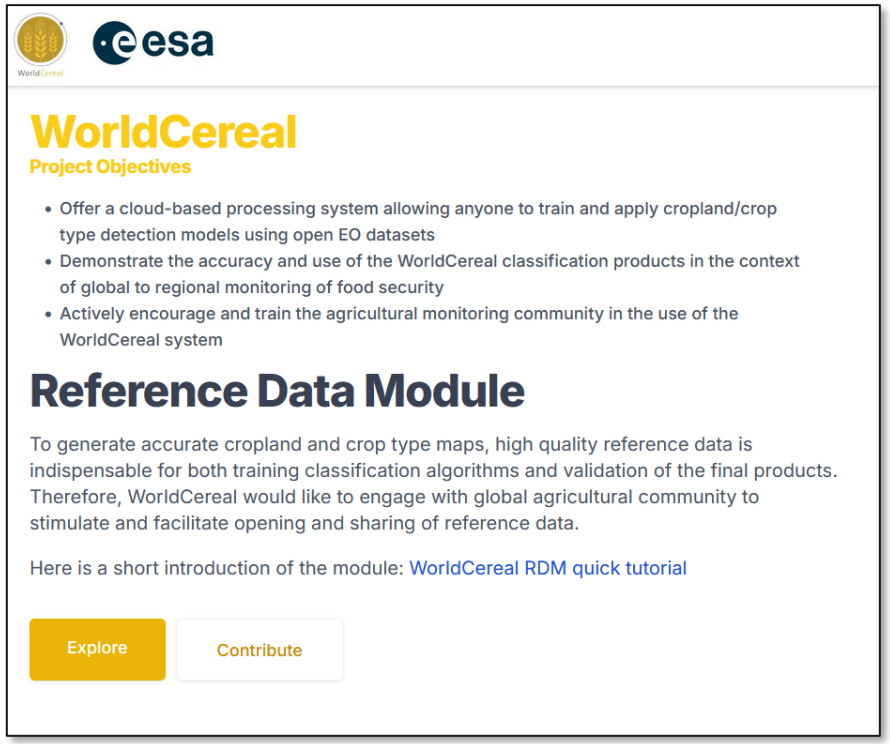
Contents

- The WorldCereal RDM
- Architecture & technologies
- Dataset exploration – RDM Website
- Dataset exploration – RDM APIs & Notebook
- Uploading new datasets

WorldCereal

The WorldCereal Reference Data Module

- Harmonized data storage
- Data + metadata (web + API)
  - License
  - Citation
  - Original dataset information, etc.
- New datasets can be uploaded by users
- Public datasets are harmonized + QC by WorldCereal moderators



https://rdm.esa-worldcereal.org

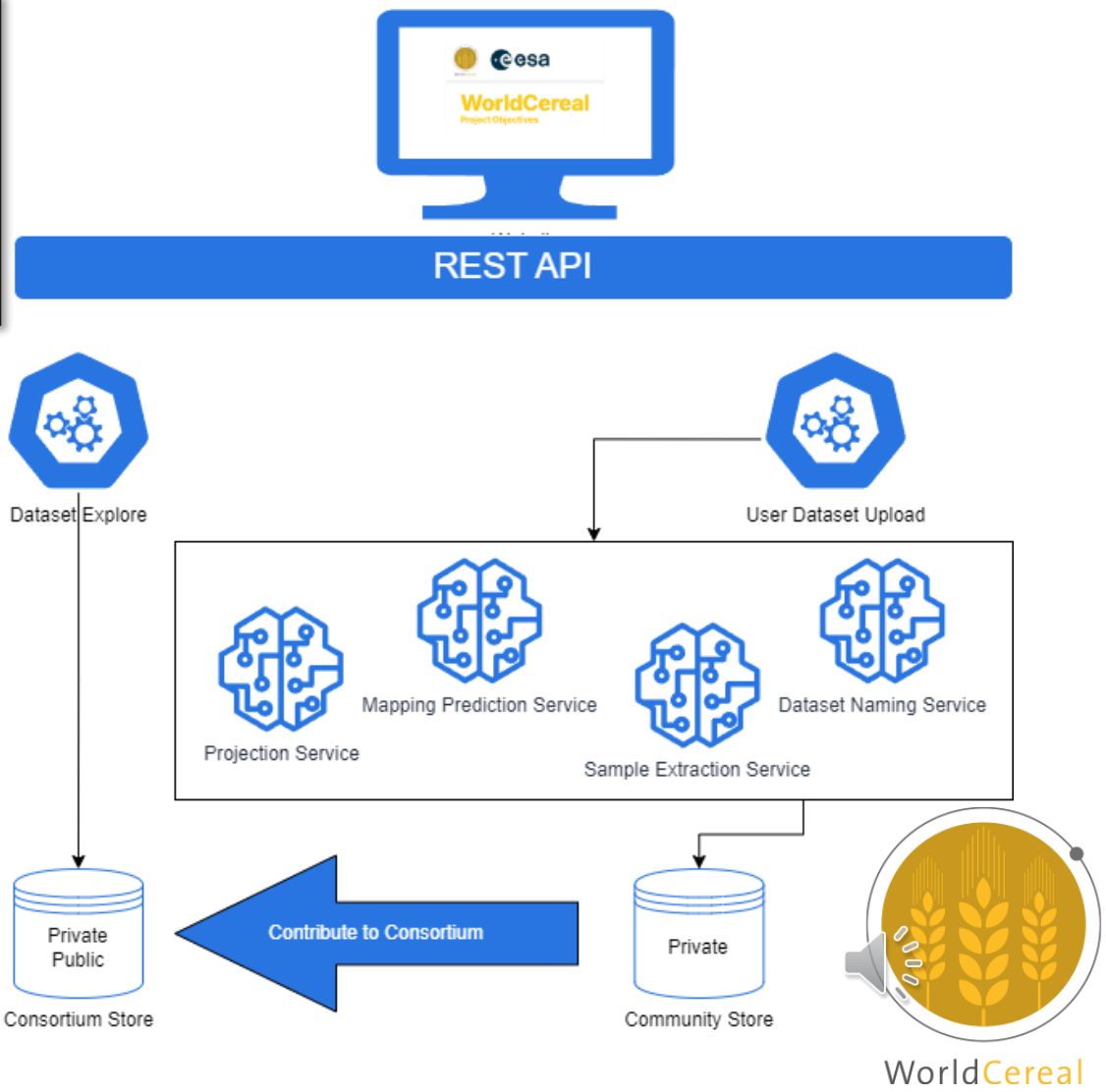
WorldCereal

Architecture

- RDM Datasets
  - Consortium Store
  - Community Store
- Consortium Dataset Access
  - Public and private access.
- User datasets are private when uploaded.
- Users can choose to share their datasets privately to consortium. Users can make datasets public after QC review by Worldcereal moderators.
- User dataset upload workflow consists of
  - Projection service
  - Mapping service
  - Dataset Naming service
  - Sampling service
- Share datasets with License

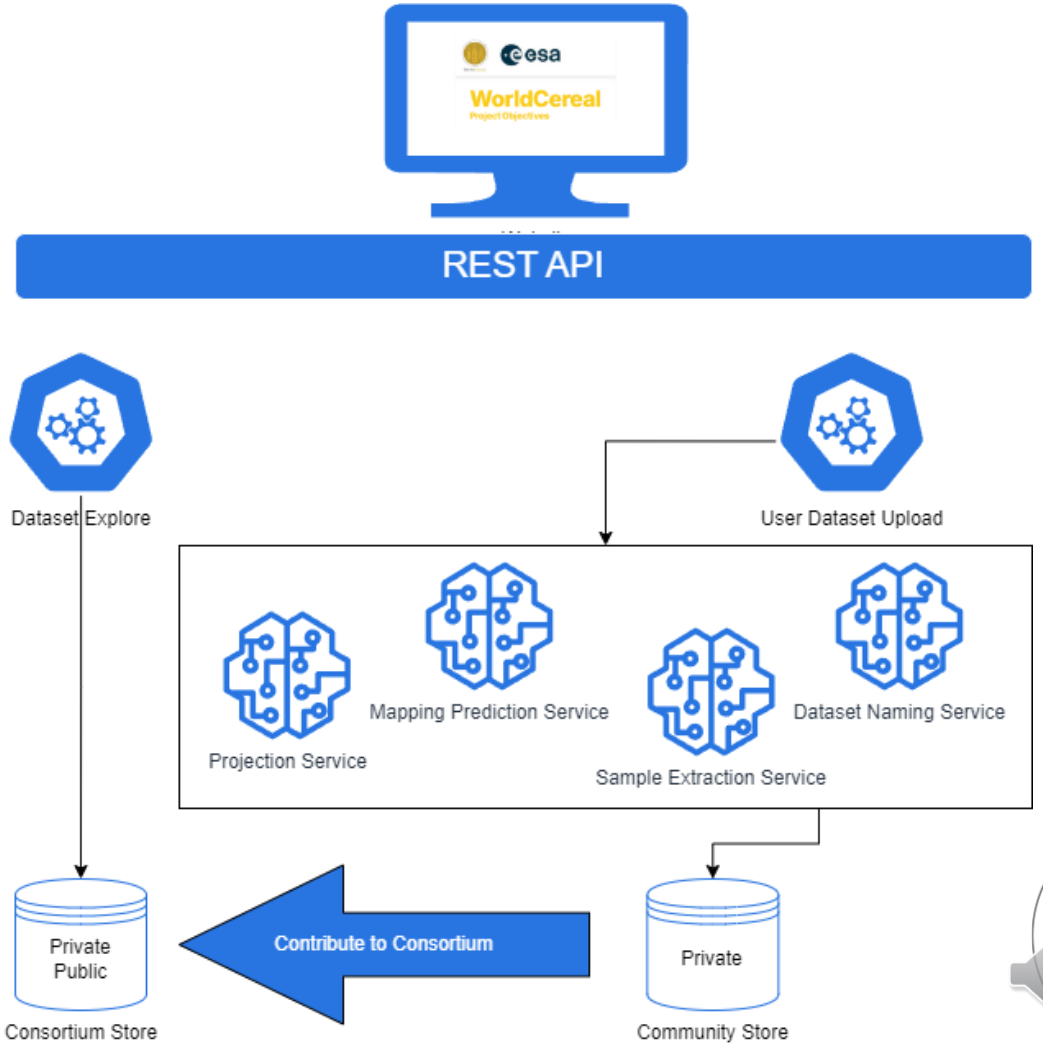
LICENSE TYPES*	REMARKS
CC0	No Rights Reserved
CC BY	Attribution
CC BY-SA	Attribution-ShareAlike
CC BY-NC	Attribution-NonCommercial
CC BY-NC-SA	Attribution-NonCommercial-ShareAlike
Private	Only accessible for the owner
Other	To be defined by the owner

* See Creative Commons licenses



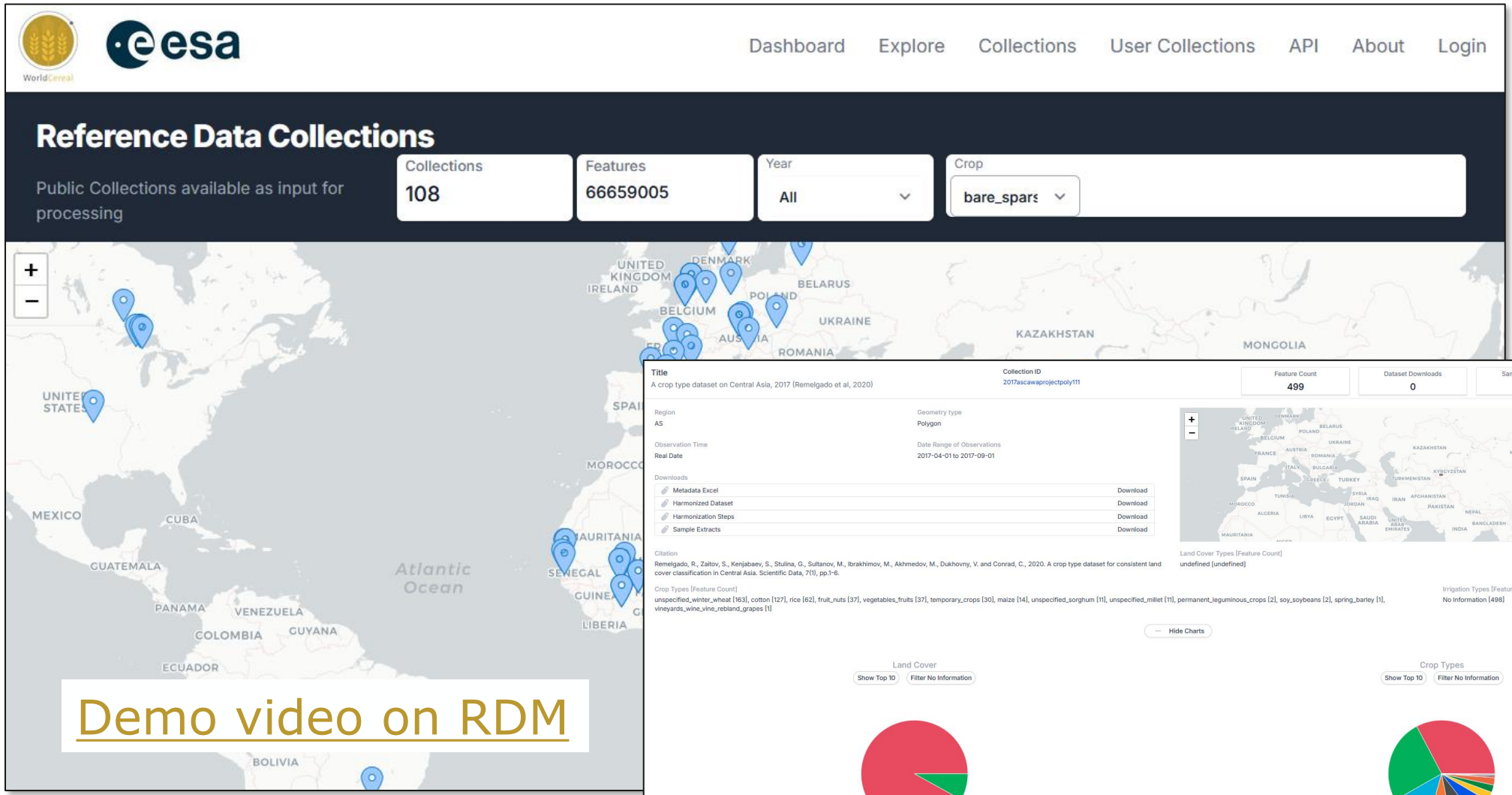
Technologies Used

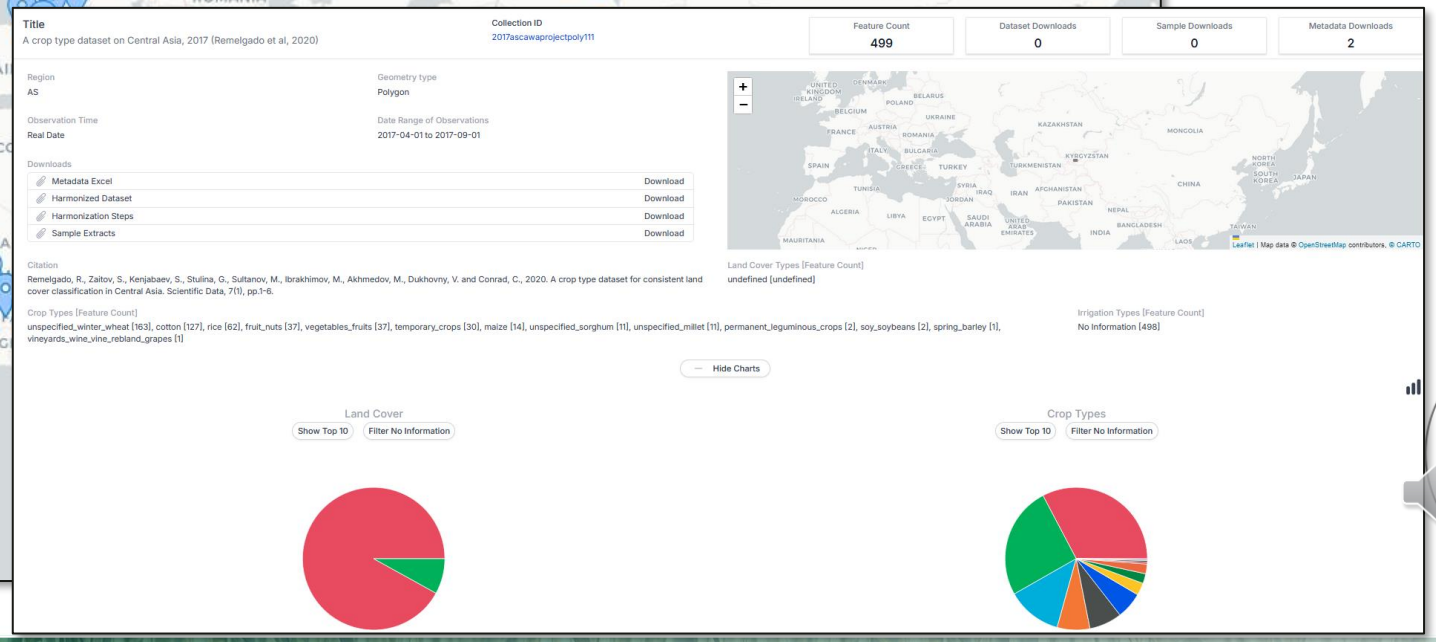
- IIASA Kubernetes cluster for hosting
- Docker based containers/pods
- React with Next JS for the website development.
- Postgres with PostGIS for database and .Net Web API for REST API.
- Python (ver 3) based services for reprojecting, harmonizing and sampling the user uploaded datasets.
- OpenAI LLMs for legend mapping predictions.



WorldCereal

Data exploration – RDM website

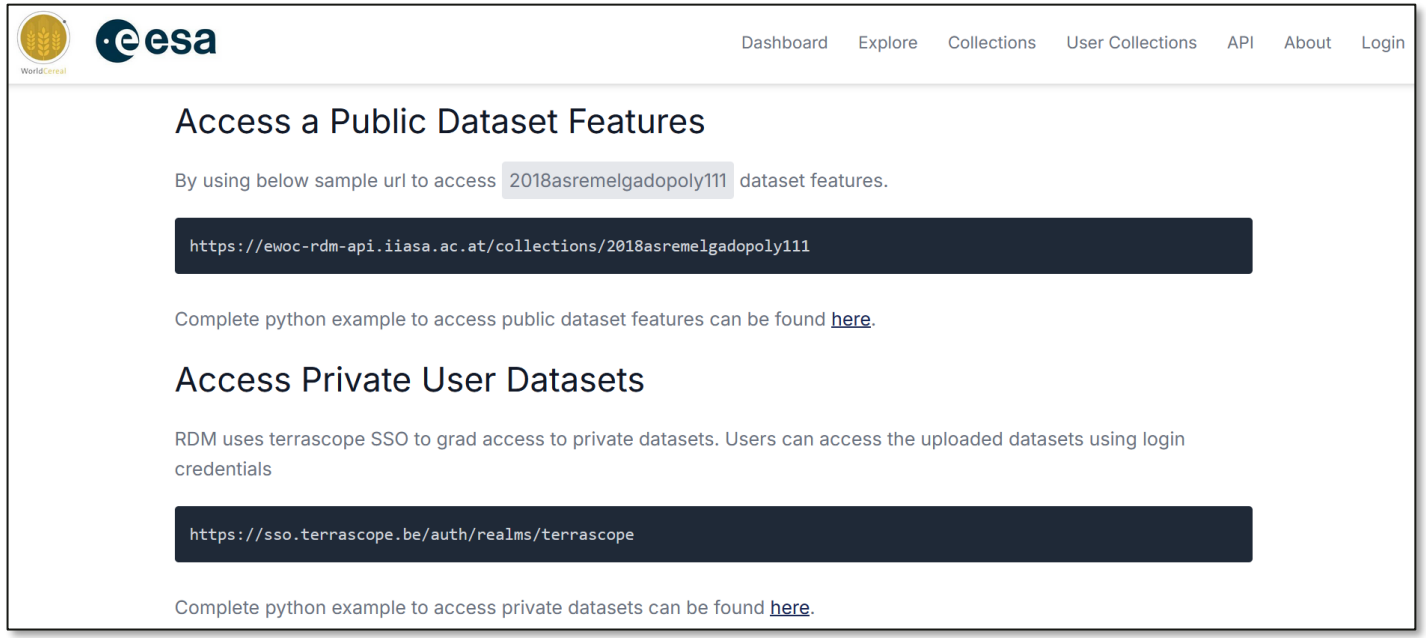




WorldCereal

Data exploration – RDM API

- Access public datasets without credentials
- Download datasets and samples
- Download metadata
- Interaction with private datasets requires a (free!) Terrascope account
- Full API documentation on Swagger

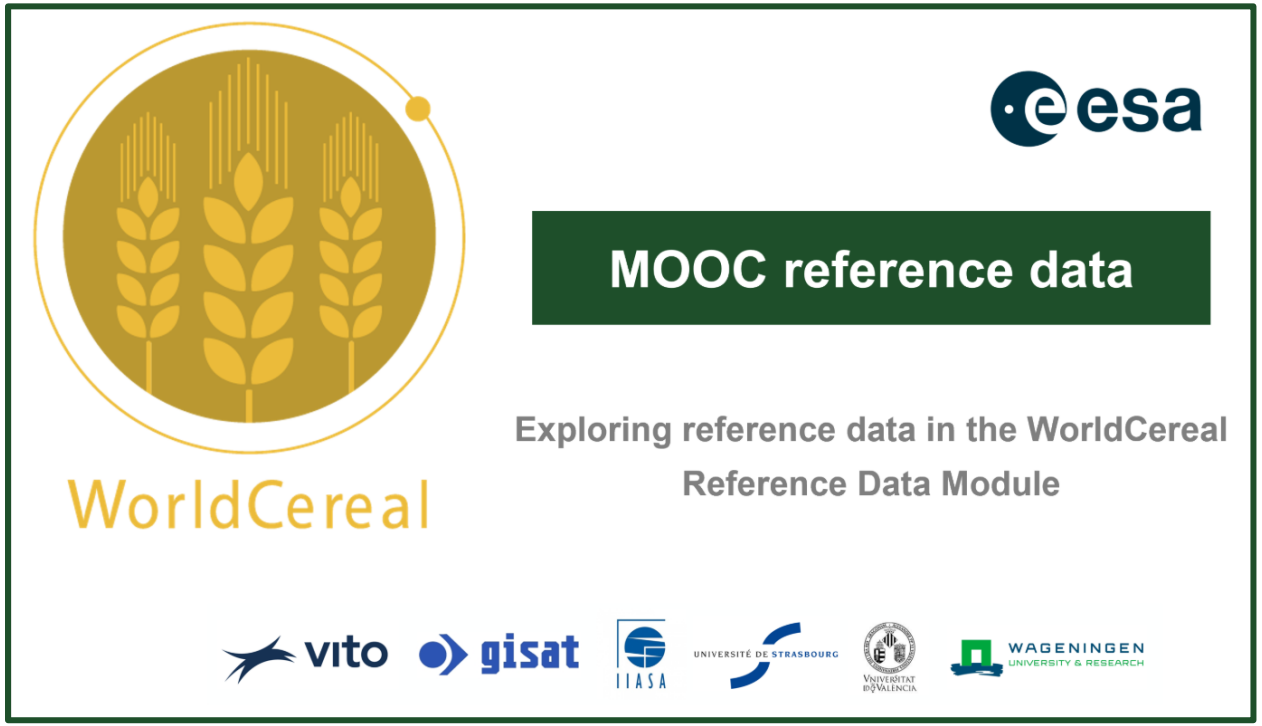


https://rdm.esa-worldcereal.org/api

WorldCereal

Data exploration – Jupyter Notebook

- Access the notebook HERE
- RDM API interaction made easy!
- Filtering and downloading full collections and individual samples
- Checking collection metadata
- Quickly assessing availability of samples for a particular crop type
- Both for public and private datasets

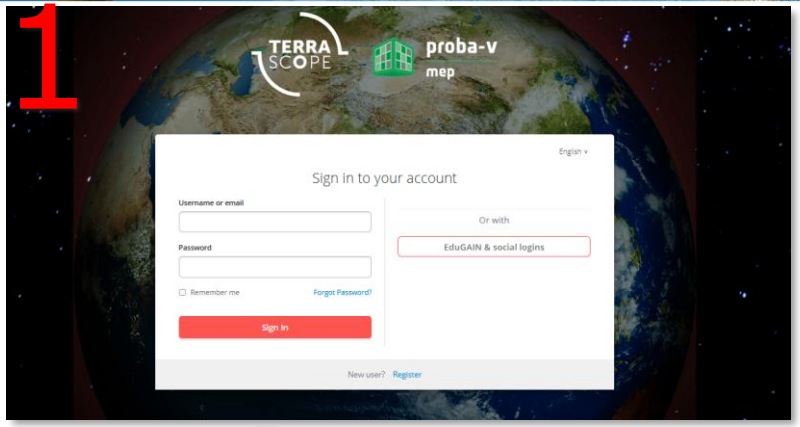


Exercise: Walk through the different steps in the notebook and try to answer the question on how many maize samples can be found around Kenya currently…

WorldCereal

Uploading new datasets to the RDM

1



2

Dataset Qualification Check ✓

The dataset must meet certain criteria to be eligible for upload. Please use the following checks to verify its compliance.

Does your dataset contain spatial geometry (e.g., points or polygons)? Yes No

Does your dataset include land cover and/or crop type labels for each observation? Yes No

Land cover and crop type information should be combined into a single dataset attribute. The format of this attribute should be a string.

Do you have information about the validity time of the observations in your dataset? Yes No

Validity time represents the date when the declared land cover or crop type was present at the specified location. It can be specified in one of two ways: (1) separately for each individual observation (in which case it should be included as a dataset attribute in string or date format), or (2) as a single date for the entire dataset (which can be selected during harmonization).

Is the validity time of your samples later than January 1st 2017? Yes No

Due to reduced availability of satellite data prior to 2017, we currently do not support datasets before 2017.

Next Skip to Upload

3

Dataset Upload

Note on data privacy:

After a successful upload, your dataset will remain fully private by default and will not be shared with anyone. You will have the option to share your data with the WorldCereal consortium or make it fully public later.

Upload File

Supported file formats: (geopackage[.gpkg] /zipped shapefiles[.zip] / geoparquet / parquet).

ZIP

Upload a file
or drag and drop file
file up to 100Mb

Identifier

Min 3, max 20 alphanumeric identifier. Refer dataset naming convention for more information.

abcde

Back Submit

4

Attribute Extraction — Attribute Mapping — Harmonize — Completed

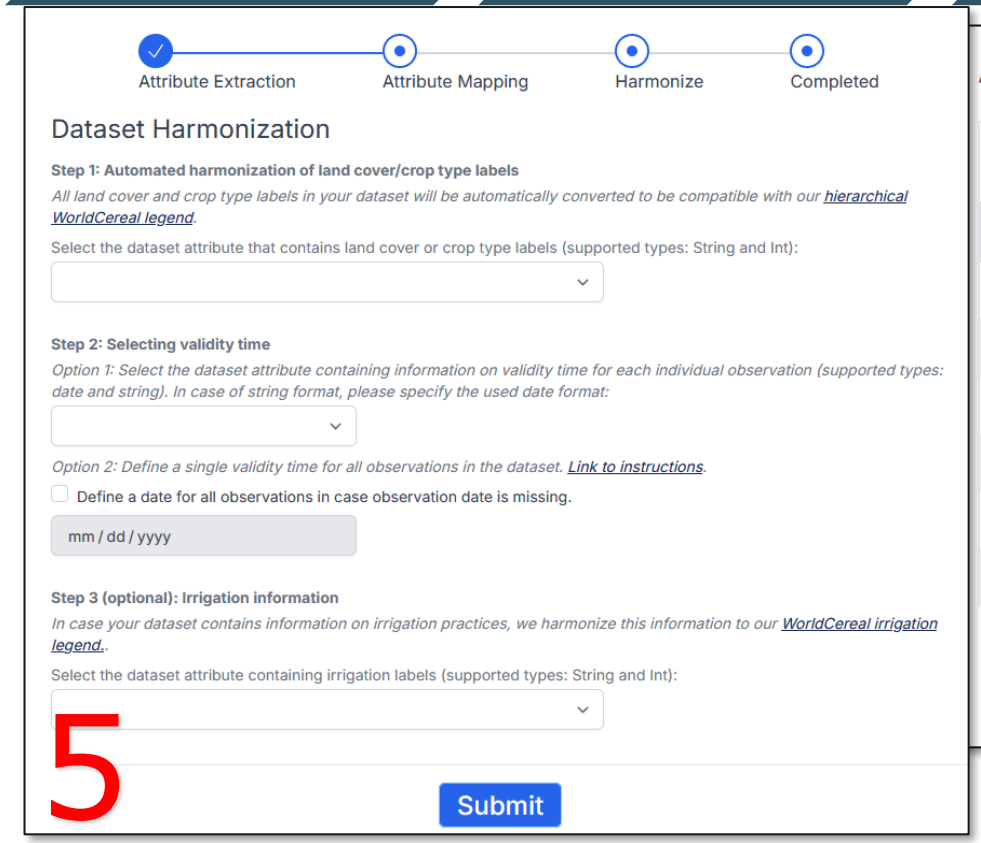
Uploaded Attribute Extraction In Progress

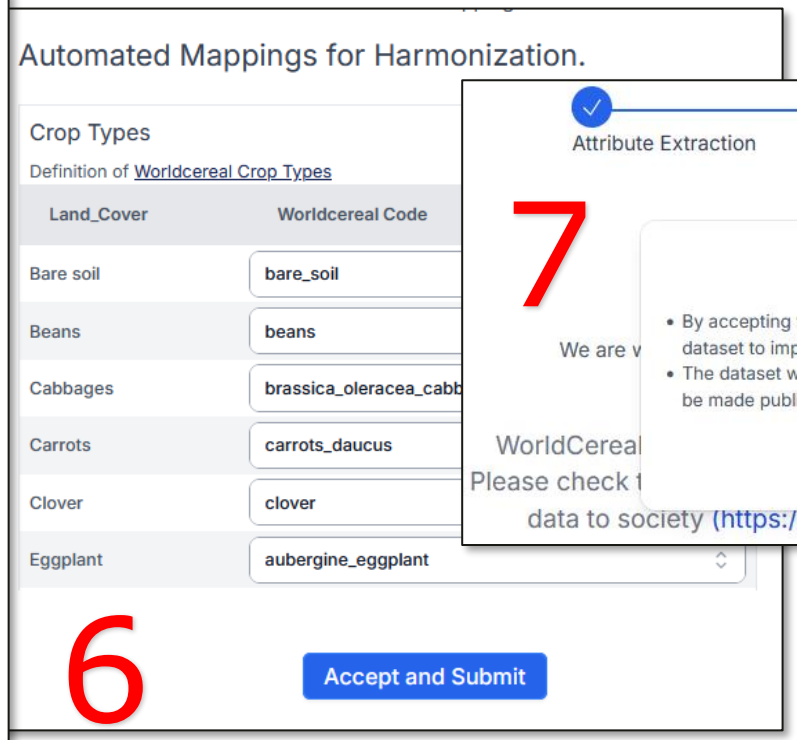
WorldCereal supports the general movement towards data sharing and open science. Please check the below link to learn more about WorldCereal's view on opening reference data to society (https://esa-worldcereal.org/en/situ-data-global-crop-mapping).

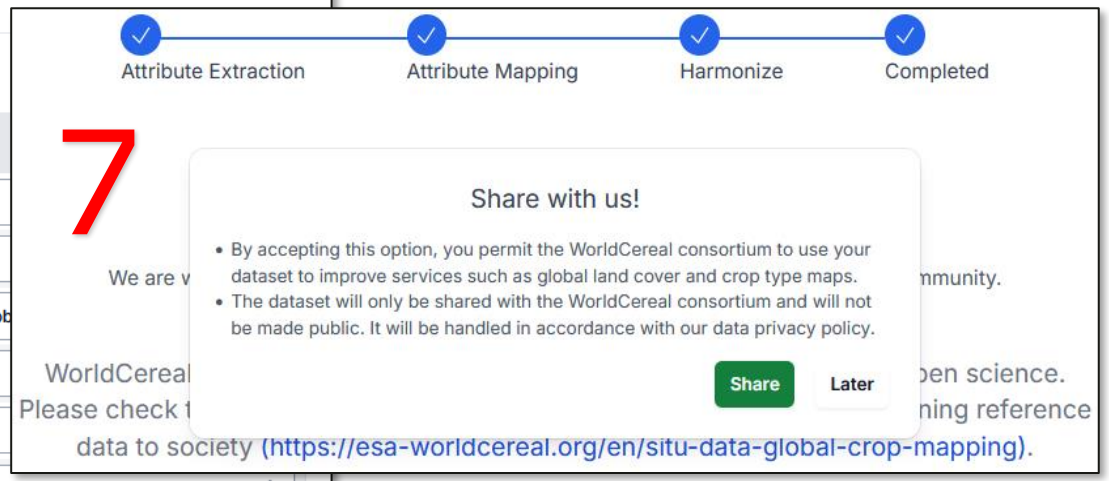
WorldCereal

Uploading data: AI-assisted crop harmonization

Attribute Selection → AI Based Mapping → Harmonization → Private Dataset → Public







WorldCereal

Uploading data: full demo

Watch the following step-by-step guide for more information on dataset upload:

How to Upload Datasets to RDM (video)

Exercise: Try out the dataset upload workflow with the points you collected through Crop-Capture Geo-Quest!
https://crop-capture.main.geo-wiki.org/

WorldCereal

WorldCereal

THANK YOU

Interesting links:

About ref data → https://esa-worldcereal.org/en/reference-data

RMD UI → https://rdm.esa-worldcereal.org/

Documentation → https://worldcereal.github.io/worldcereal-documentation/rdm/overview.html

Questions? → WorldCereal Forum MOOC I

Subscribe to our mailing list