



WorldCereal



# WorldCereal MOOC I



## Introduction to the WorldCereal Reference Data Module (RDM)

Santosh Karanam (IIASA)

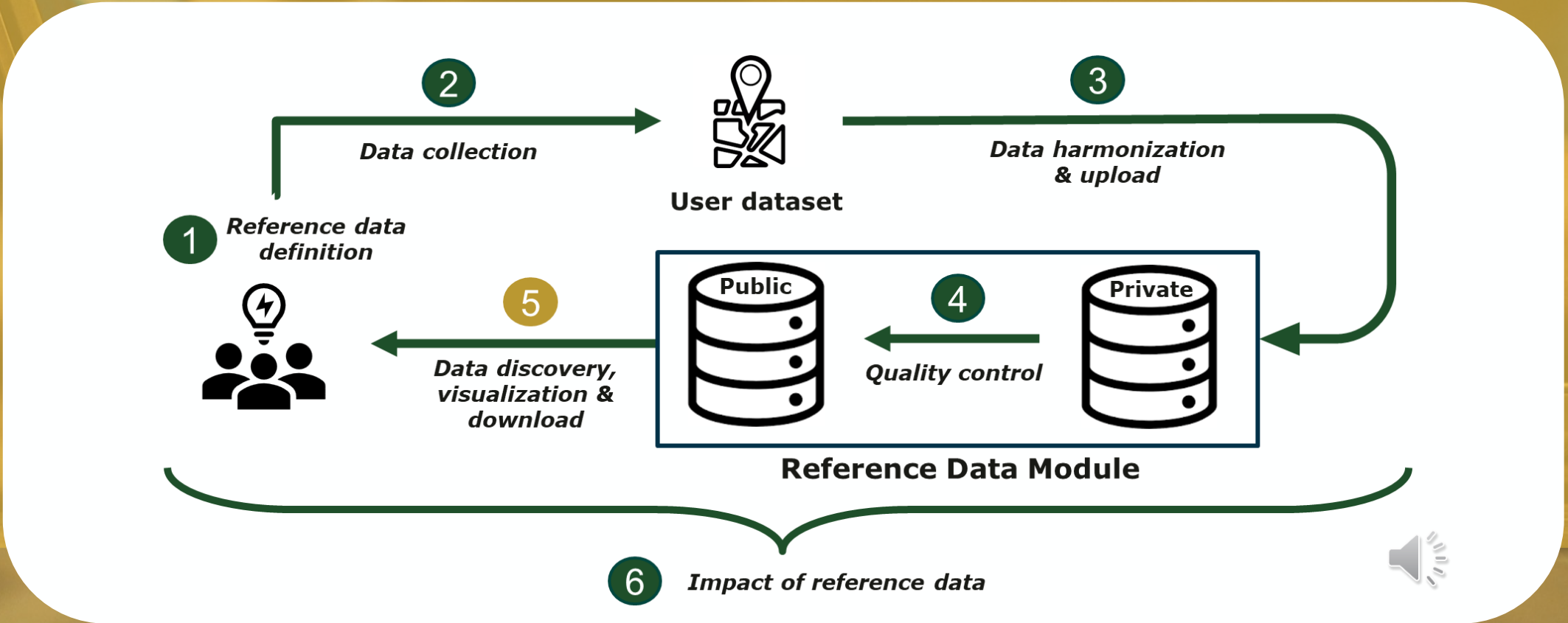


→ THE EUROPEAN SPACE AGENCY



WorldCereal

# MOOC I: Outline

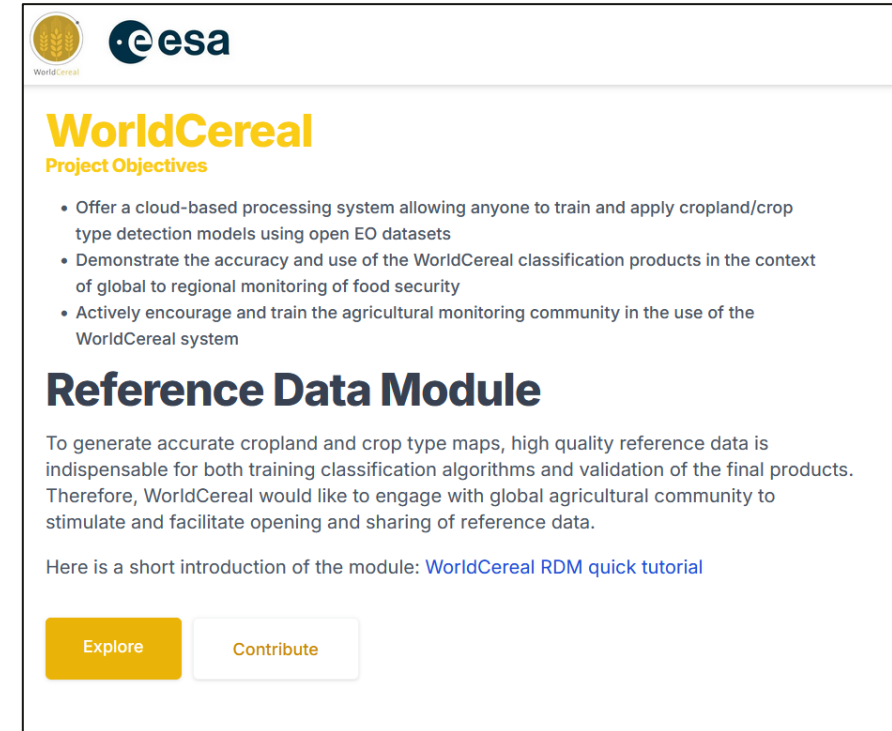


- The WorldCereal RDM
- Architecture & technologies
- Dataset exploration – RDM Website
- Dataset exploration – RDM APIs & Notebook
- Uploading new datasets



WorldCereal

- 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



The screenshot shows the top of the WorldCereal website. It features the WorldCereal logo (a golden wheat stalk) and the ESA logo. Below the logos, the heading "WorldCereal" is followed by "Project Objectives". Three bullet points describe the project's goals: offering a cloud-based processing system, demonstrating accuracy, and encouraging the agricultural community. Below this is the heading "Reference Data Module" and a paragraph explaining the need for high-quality reference data. A link to a "WorldCereal RDM quick tutorial" is provided. At the bottom of the screenshot are two buttons: "Explore" (yellow) and "Contribute" (white).

<https://rdm.esa-worldcereal.org>

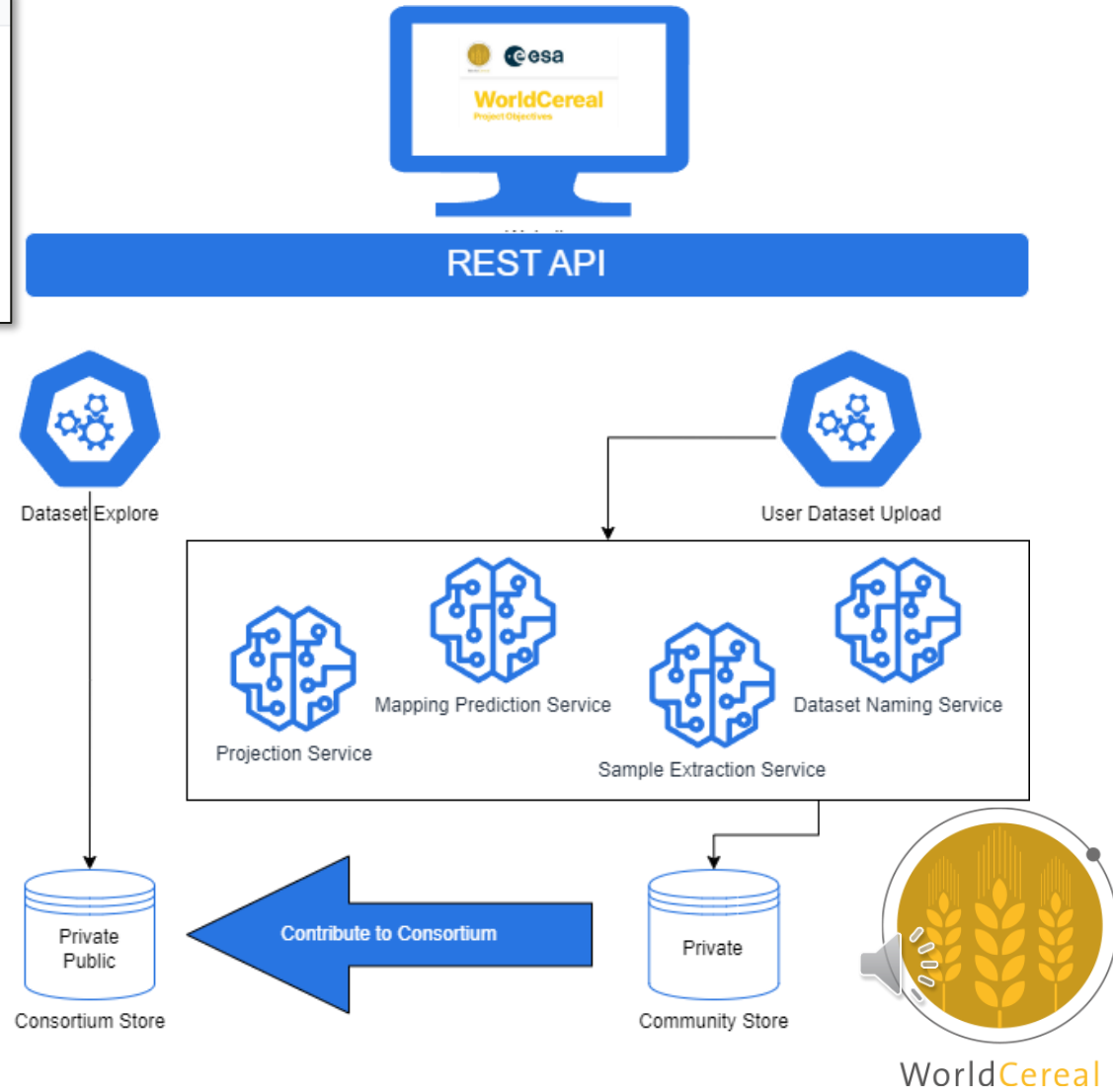


WorldCereal

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

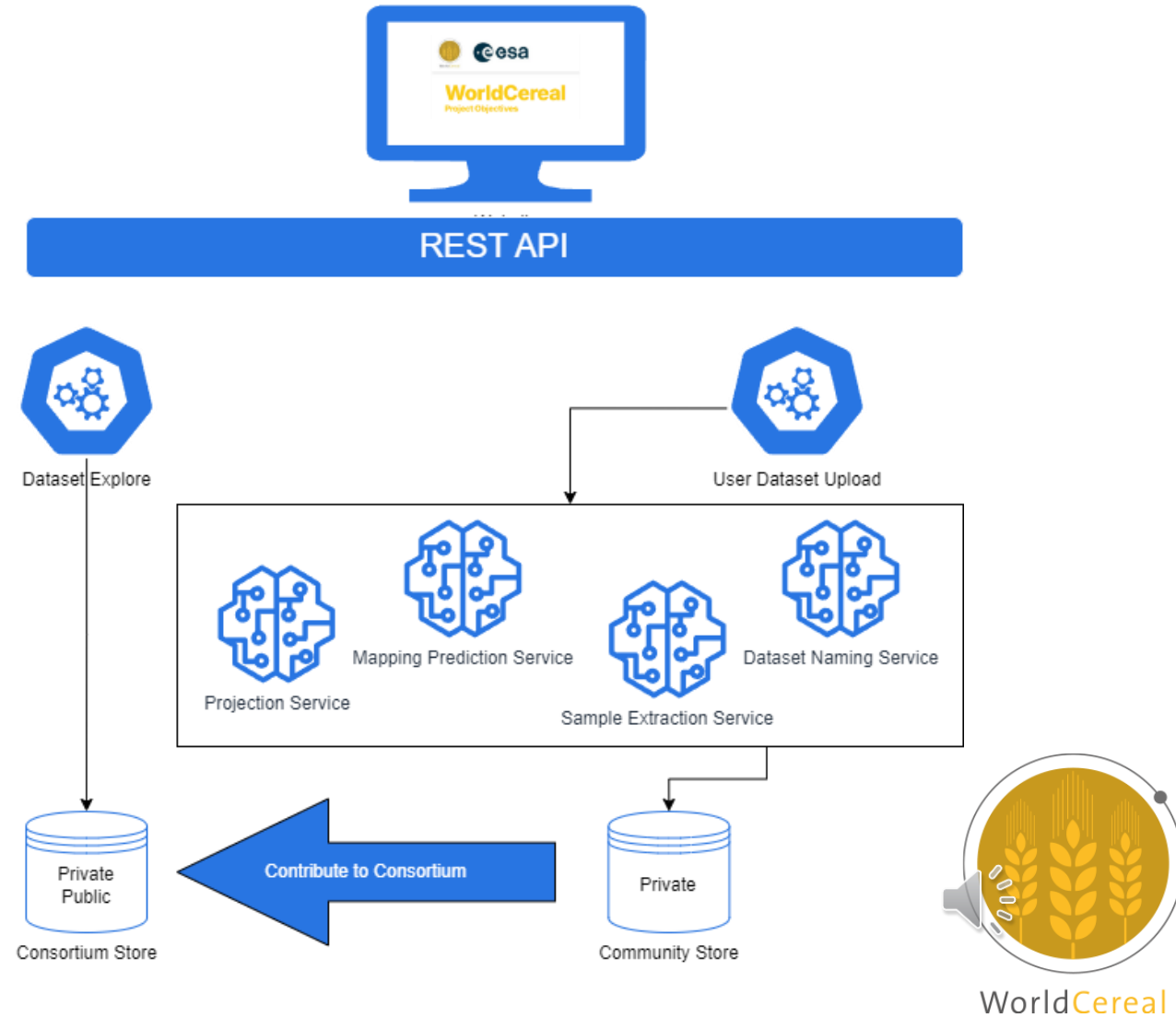
\* See Creative Commons licenses

- 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



# 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.



# Data exploration – RDM website



WorldCereal [Dashboard](#) [Explore](#) [Collections](#) [User Collections](#) [API](#) [About](#) [Login](#)

## Reference Data Collections

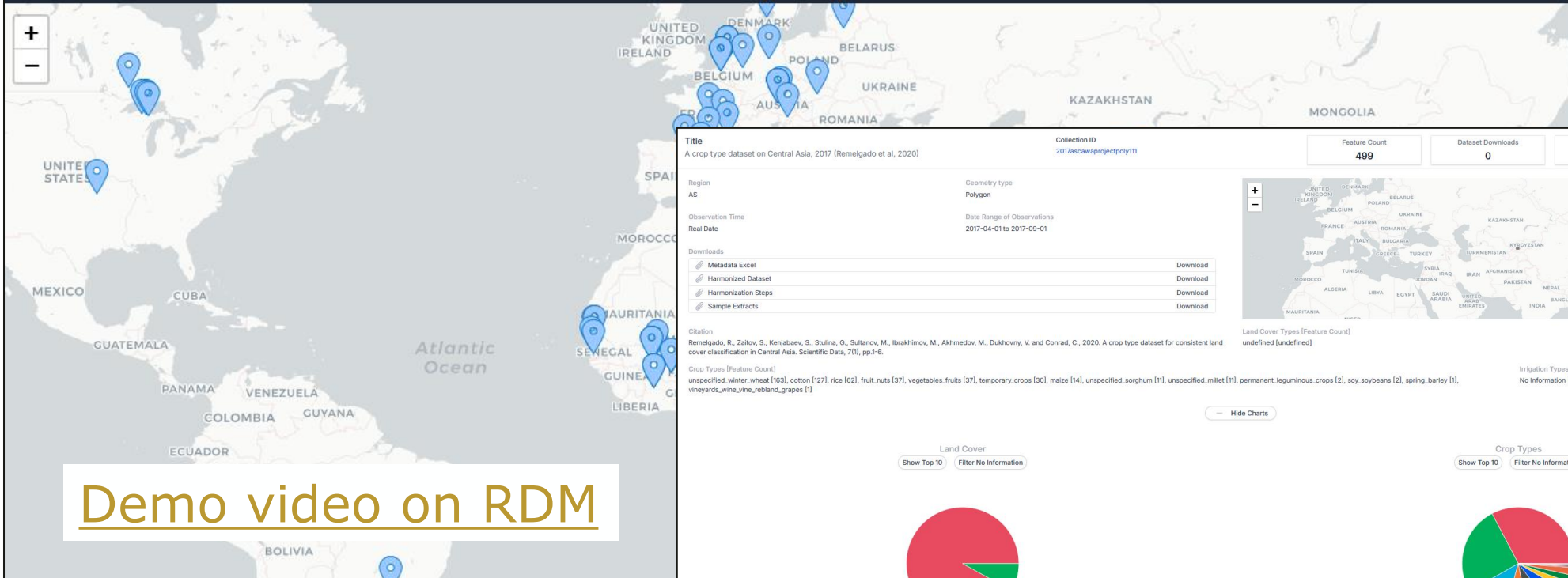
Public Collections available as input for processing

Collections  
**108**

Features  
**66659005**

Year  
**All**

Crop  
**bare\_spars**



[Demo video on RDM](#)

**Title**  
A crop type dataset on Central Asia, 2017 (Remelgado et al, 2020)

**Collection ID**  
2017ascawprojectpoly11

<b>Feature Count</b> 499	<b>Dataset Downloads</b> 0	<b>Sample Downloads</b> 0	<b>Metadata Downloads</b> 2
-----------------------------	-------------------------------	------------------------------	--------------------------------

**Region**  
AS

**Geometry type**  
Polygon

**Observation Time**  
Date Range of Observations  
Real Date  
2017-04-01 to 2017-09-01

**Downloads**

Metadata Excel	Download
Harmonized Dataset	Download
Harmonization Steps	Download
Sample Extracts	Download

**Citation**  
Remelgado, R., Zaitov, S., Kenjabaev, S., Stulina, G., Sultanov, M., Ibrahimov, M., Akhmedov, M., Dukhovny, V. and Conrad, C., 2020. A crop type dataset for consistent land cover classification in Central Asia. Scientific Data, 7(1), pp.1-6.

**Crop Types [Feature Count]**  
unspecified\_winter\_wheat [163], cotton [127], rice [62], fruit\_nuts [37], vegetables\_fruits [37], temporary\_crops [30], maize [14], unspecified\_sorghum [11], unspecified\_millet [11], permanent\_leguminous\_crops [2], soy\_soybeans [2], spring\_barley [1], vineyards\_wine\_vine\_rebland\_grapes [1]

**Land Cover Types [Feature Count]**  
undefined [undefined]

**Irrigation Types [Feature Count]**  
No Information [498]

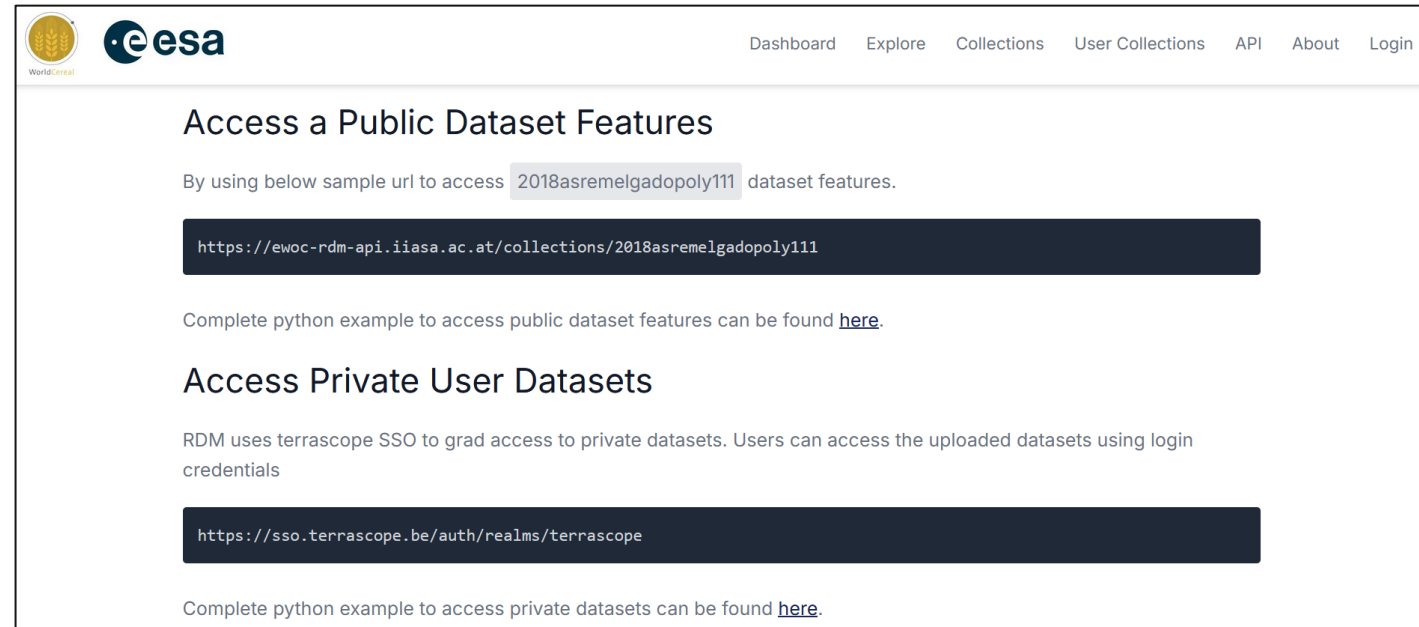
— Hide Charts

Land Cover: Show Top 10 | Filter No Information

Crop Types: Show Top 10 | Filter No Information



- 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



The screenshot shows the RDM API documentation page. At the top left, there are logos for WorldCereal and ESA. The top right navigation bar includes links for Dashboard, Explore, Collections, User Collections, API, About, and Login. The main content area is titled "Access a Public Dataset Features" and explains that users can access public dataset features using a sample URL: `https://ewoc-rdm-api.iiasa.ac.at/collections/2018asremelgadopoly111`. It also mentions that a complete python example can be found [here](#). Below this, there is a section titled "Access Private User Datasets" which states that RDM uses terrascope SSO for access to private datasets and that users can access uploaded datasets using login credentials. A sample URL for the SSO endpoint is provided: `https://sso.terrascope.be/auth/realms/terrascope`. A final note indicates that a complete python example for private datasets can be found [here](#).

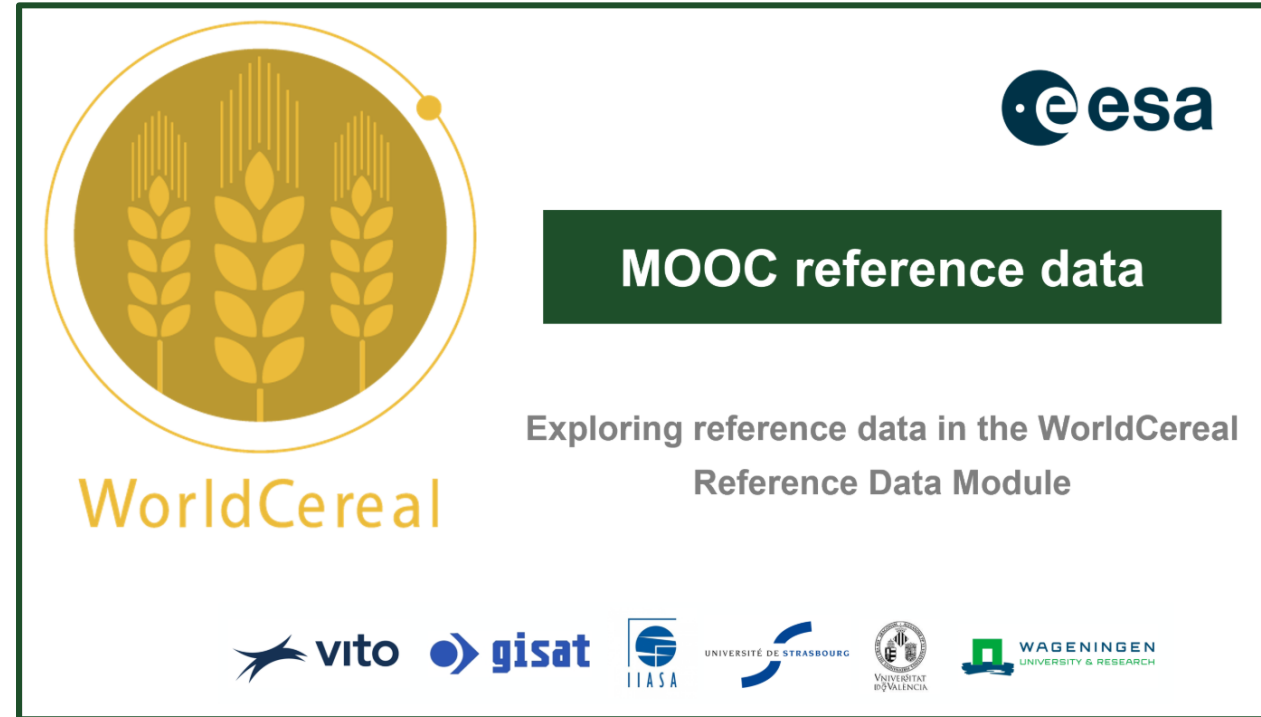
<https://rdm.esa-worldcereal.org/api>



WorldCereal



- 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

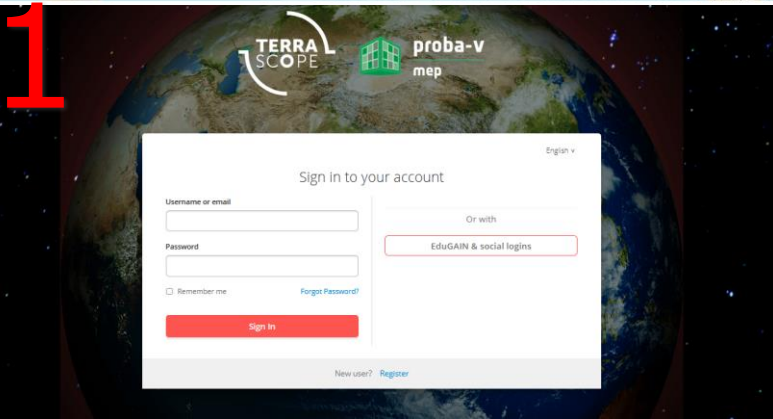


The slide features the WorldCereal logo (a golden circle with wheat stalks) on the left. On the right, the ESA logo is at the top, followed by a dark green box with the text "MOOC reference data". Below this, the text "Exploring reference data in the WorldCereal Reference Data Module" is centered. At the bottom, a row of logos includes vito, gisat, IIASA, UNIVERSITÉ DE STRASBOURG, UNIVERSITAT DE VALÈNCIA, and WAGENINGEN UNIVERSITY & RESEARCH.

**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...



# Uploading new datasets to the RDM



## 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?  
*Land cover and crop type information should be combined into a single dataset attribute. The format of this attribute should be a string.*  Yes  No

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

Is the validity time of your samples later than January 1st 2017?  
*Due to reduced availability of satellite data prior to 2017, we currently do not support datasets before 2017.*  Yes  No

## 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).

**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.

## 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>).



# Uploading data: AI-assisted crop harmonization

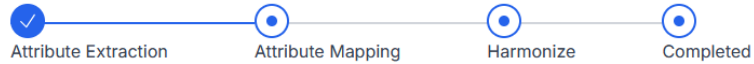
Attribute Selection

AI Based Mapping

Harmonization

Private Dataset

Public



## Dataset Harmonization

### Step 1: Automated harmonization of land cover/crop type labels

All land cover and crop type labels in your dataset will be automatically converted to be compatible with our [hierarchical WorldCereal legend](#).

Select the dataset attribute that contains land cover or crop type labels (supported types: String and Int):

### Step 2: Selecting validity time

Option 1: Select the dataset attribute containing information on validity time for each individual observation (supported types: date and string). In case of string format, please specify the used date format:

Option 2: Define a single validity time for all observations in the dataset. [Link to instructions](#).

Define a date for all observations in case observation date is missing.

### Step 3 (optional): Irrigation information

In case your dataset contains information on irrigation practices, we harmonize this information to our [WorldCereal irrigation legend](#).

Select the dataset attribute containing irrigation labels (supported types: String and Int):

5

Submit

## Automated Mappings for Harmonization.

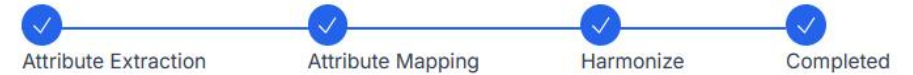
### Crop Types

Definition of [Worldcereal Crop Types](#)

Land_Cover	Worldcereal Code
Bare soil	bare_soil
Beans	beans
Cabbages	brassica_oleracea_cabb
Carrots	carrots_daucus
Clover	clover
Eggplant	aubergine_eggplant

6

Accept and Submit



7

## Share with us!

- By accepting this option, you permit the WorldCereal consortium to use your dataset to improve services such as global land cover and crop type maps.
- The dataset will only be shared with the WorldCereal consortium and will not be made public. It will be handled in accordance with our data privacy policy.

We are v...  
WorldCereal...  
Please check t...  
data to society (<https://esa-worldcereal.org/en/situ-data-global-crop-mapping>).

Share

Later



WorldCereal



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

