

StatEO

5-7 May 2026 | ESA-ESRIN | Frascati (Rome), Italy



Standardised Reference Data Framework for Global Crop Mapping

Arun Pratihast, Hendrik Boogaard, Jeroen Degerickx, Belen Franch Gras, Steffen Fritz and Sven Gilliams

ESA UNCLASSIFIED - For ESA Official Use Only

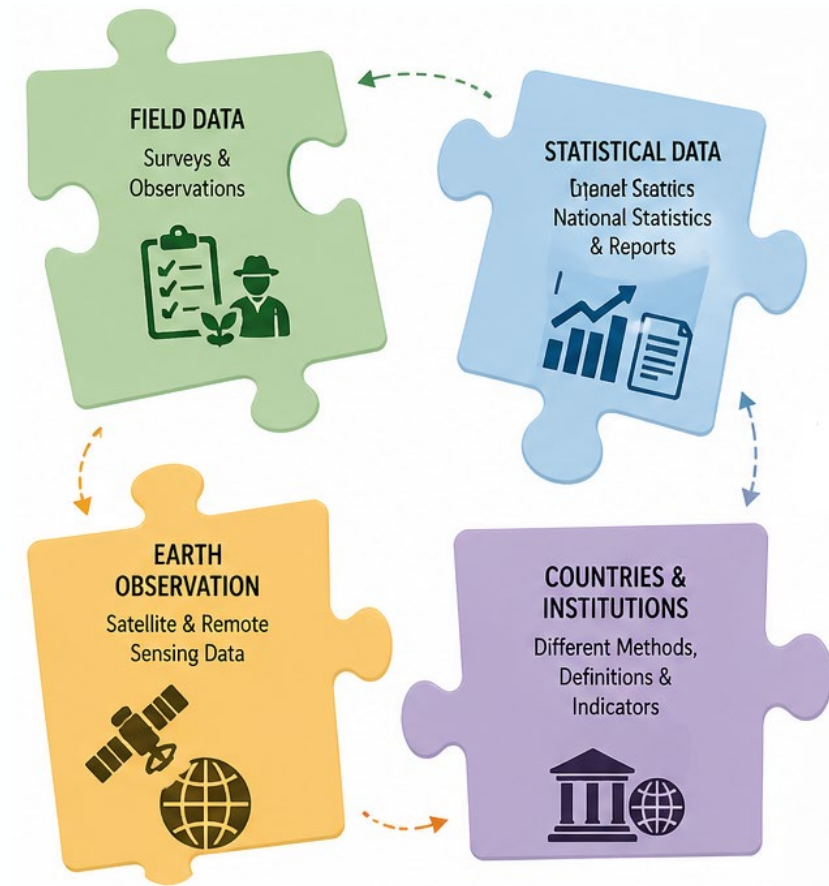
- Growing demand for reliable agricultural data to support food security and policy
- Fragmented data sources: field surveys, statistics, and Earth Observation
- Inconsistent methods across countries and institutions

Key Actors in Agricultural Data

- Ministries of Agriculture → better monitoring and policy support
- Statistics Departments → consistent, comparable national statistics

Core Problem

- No standardized data collection protocols
- Different methodologies and indicators
- Limited comparability and interoperability
- Weak integration with Earth Observation data

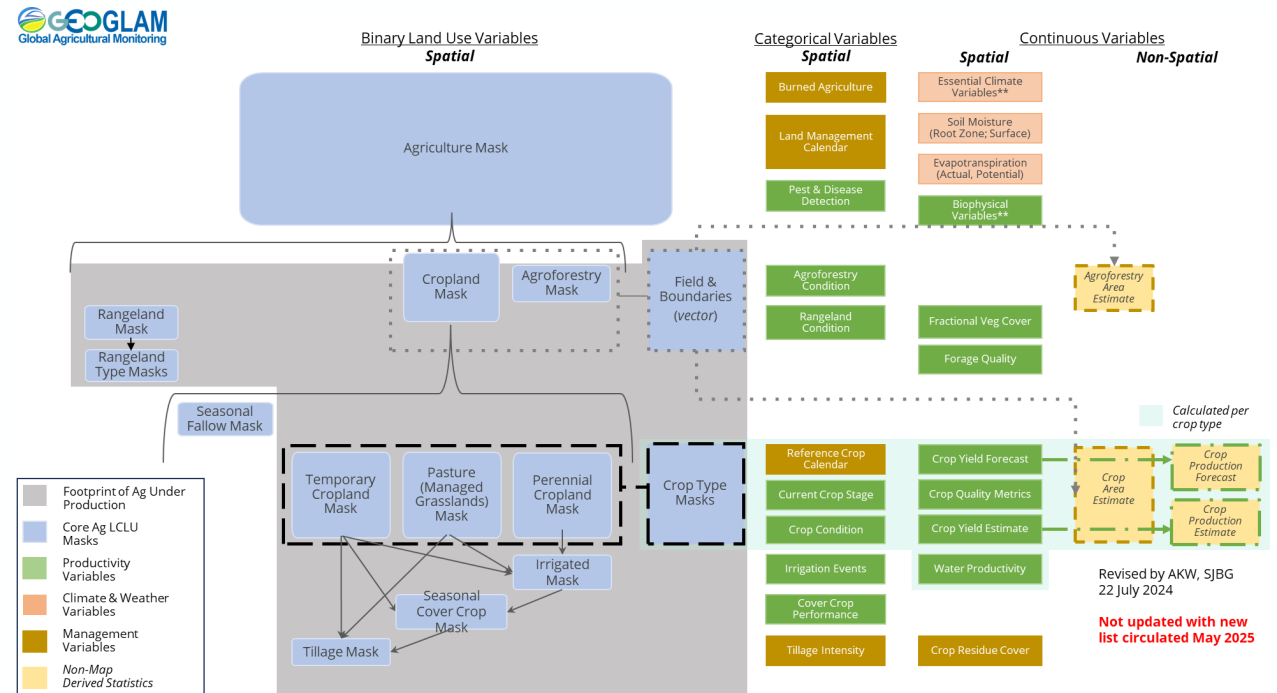


Essential Agricultural Variables (EAVs)

- A standardised global framework for agricultural monitoring
- Designed to connect field data, statistics, and Earth Observation

What EAVs Do

- Define common variables (crop type, area, condition, yield)
- Standardise how agricultural indicators are produced
- Enable consistent monitoring across countries and systems



Urgent Need → A shared language for agricultural data

Data Collection Protocols

JECAM

Sites

10+ countries
±5m accuracy

FAO AGROSTAT

Agricultural Statistics

190+ countries
Admin boundaries

WFP mVAM

Mobile Surveys

60+ countries
Daily snapshots

CIMMYT

Field Breeding Trials

Point-in-time
cultivar-specific

Standardisation Gaps

Taxonomy

Crop class
inconsistency

Metadata

No unified schema

Geolocation

±5m to ±1km
accuracy gap

Temporal

seasonal vs Yearly
Time Lag

Integration

No cross-
institutional link

Consequence

Unable to leverage

20M+ reference observations

for global ML development

Data harmonisation requires

60-70% manual effort

to integrate datasets



WorldCereal

1. Data Cleaning & Standardization

- Remove errors, missing labels, outdated records
- Standardize filenames & attributes
- Reproject to **EPSG:4326**

2. Attribute Completion & Harmonization

- Apply **standard crop legend** (common language)
- AI-based label suggestion (OpenAI model)
- Add **validity time** (derive if missing from year/season)
- Generate **unique, readable IDs**
- Add sampling labels for large datasets



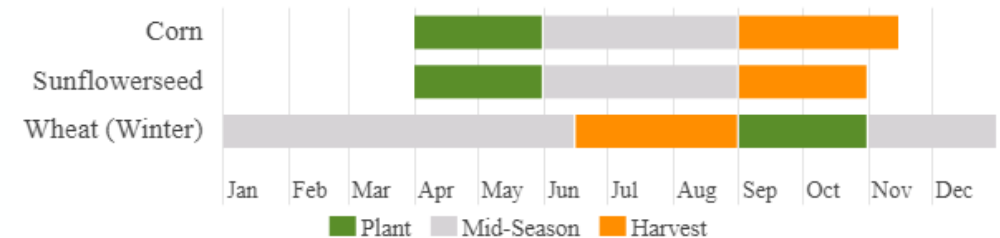
Land cover: Temporary crops

↳ Land cover subgroup: Cereals

↳ Crop group: Wheat

↳ Crop type: Durum wheat

↳ Variety/management: Winter durum wheat

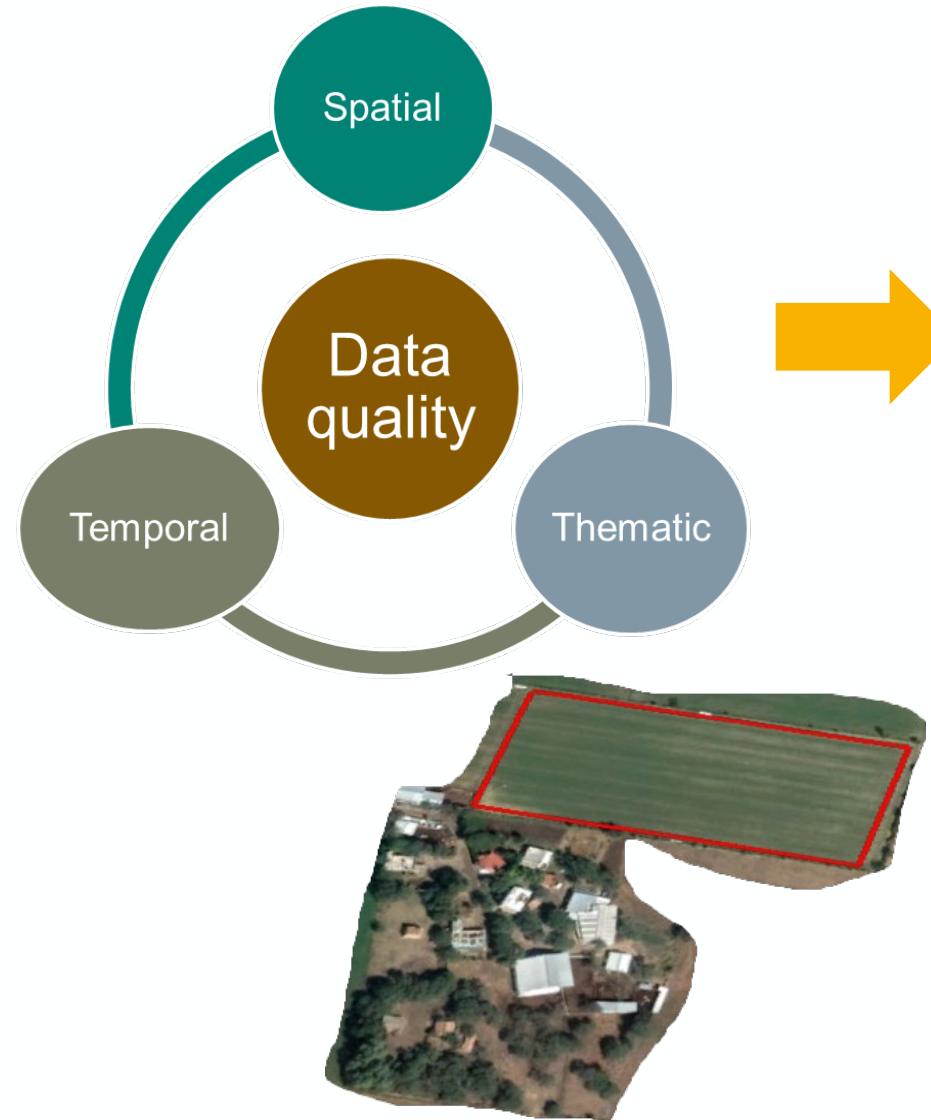


3. Quality & Confidence Assessment

- Spatial accuracy
- Temporal accuracy
- Thematic accuracy

4. Metadata Completion

- License, citation
- Data provider/holder
- Provenance / background



Confidence score



Reference Data Collections

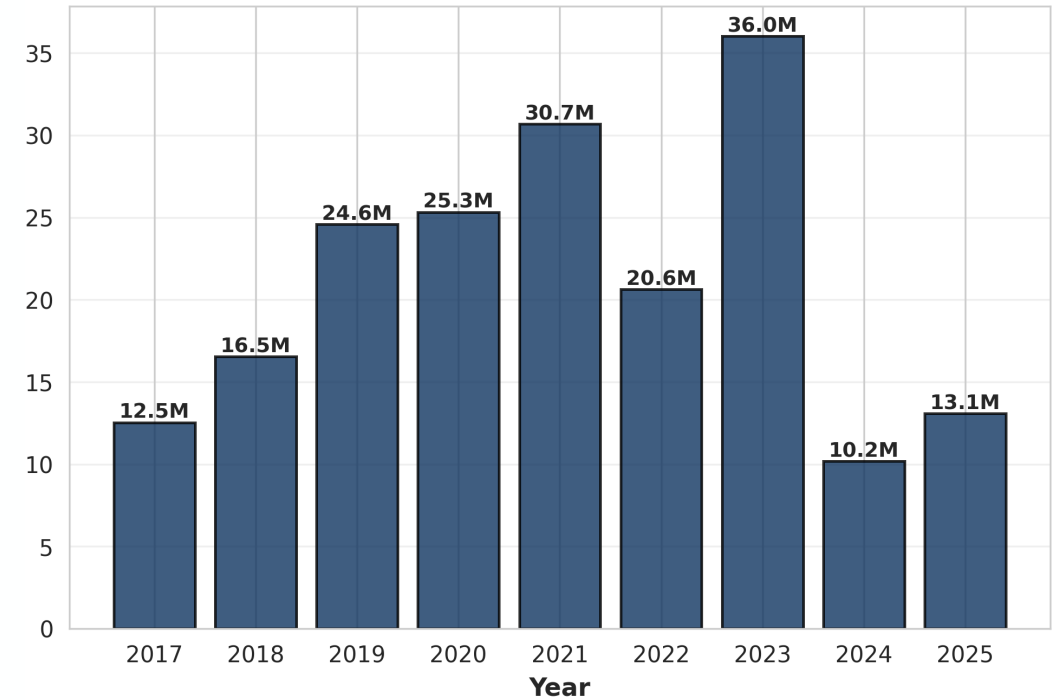
Public Collections available as input for processing

Collections 240	Features 189759707	Year All	Crop barley
---------------------------	------------------------------	-------------	----------------



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

Data Volume Timeline (2017-2025)



Completing metadata (following [FAIR principles](#))

- Title
- Unique ID
- Creator (name, URL, e-mail)
- License
- Citation
- Description (DataSet Name)
- Related publication (ReferenceDataSet)
- Coverage (spatial, temporal)

Title A crop type dataset on Central Asia, 2018 (Remelgado et al, 2020)	Collection ID 2018asremelgadopoly111	Feature... 2639	Dataset... 1	Sample... 0	Metada... 2
---	--	---------------------------	------------------------	-----------------------	-----------------------

Region AS	Geometry type Polygon
Observation Time Real Date	Date Range of Observations 1/3/2018 to 1/9/2018

Worldcereals Reference Documents

- [Crop type legend](#)
- [Irrigation Status legend](#)
- [About observation date](#)
- [Dataset confidence score calculation](#)

Downloads

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

Citation

Remelgado, R., Zaitov, S., Kenjabaev, S., Stulina, G., Sultanov, M., Ibrakhimov, 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.

Dataset Provider Details			
Code CAWa project (Remelgado et al, 2020)	Description Central Asia Waters (CAWa) (Remelgado et al, 2020)	Url www.cawa-project.net	Contact ruben.remelgado@idiv.de
DataSet Name A crop type dataset for consistent land cover classification in Central Asia	ReferenceDataSet https://doi.org/10.1038/s41597-020-00591-2	Type Of License CC_BY	ReferenceToLicense

Conclusion & Next Steps

- WorldCereal is the first coordinated initiative to standardise global crop mapping reference data
- Addresses five core dimensions:
 - Flexible crop and land-cover classification system
 - Minimum shared metadata (how, where, when data are collected)
 - Standardised spatial and temporal sampling rules and observation unit definitions
 - Repeatable field-based and remote sensing observation protocols
 - Robust data quality control, uncertainty reporting, and versioning
- Data harmonisation is an ongoing process; current efforts remain project-based and lack long-term sustainability
- Within GEOGLAM, JECAM data collection guidelines are being updated
 - Framework will be extended to additional EAVS: phenology, yield, and irrigation
 - Mixed cropping systems remain a key unresolved challenge for standardization
- Supports transition from FAIR data principles to AI-ready datasets as the emerging standard

Thank you for your attention!



Arun Pratihast

arun.pratihast@wur.nl

Wageningen University & Research,
Wageningen, Netherlands

[WorldCereal RDM documentation](https://esa-worldcereal.org/en)

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

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

<https://worldcereal.github.io/worldcereal-documentation/>

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

