



Excellence in Agronomy

TRANSFORM Work Package

Useful Resources: Data Management, Analytics & Modeling

Tutorials	2
AKILIMO	2
R	2
Data science for agriculture training	2
Aquacrop	2
Regional Agronomy	2
Responsible data management guidelines	2
Guideline for agronomy and soil fertility data collection in Ethiopia	2
Generation of standards-compliant (FAIR) data	2
FAIRscribe	2
VMapper+	3
DataScribe	3
AgroFIMS	3
RHoMIS	3
Data processing, transformation, analysis	3
Carob	3
Collaborative GARDIAN Labs.....	3
Data access	3
GARDIAN.....	3
Data Catalog	3
WorldClim.....	3
Database of Global Administrative areas	4
Planet's Basemap imagery	4
Geoglam.....	4
Seasonal forecasts.....	4
Agroecological Zones of Africa.....	4
Other data management resources	4
CG Core Metadata Schema	4
Ontologies	4
Controlled vocabularies	4

| Tutorials

AKILIMO

[AKILIMO](#) is an agronomic advisory service developed by the International Institute for Tropical Agriculture with and for smallholder farmers. It employs state-of-the-art analytics to provide site-specific recommendations that optimize productivity and profits.

R

R is an integrated suite of free software facilities for data manipulation, calculation and graphical display.

[Introduction to R](#): tutorial with examples to get familiar with R developed by Robert Hijmans. University of California - Davis.

[R tutorial with exercises](#): developed by Amit Srivastava, International Rice Research Institute.

Data science for agriculture training

Training on data science applied to agriculture in [Spanish](#) and [English](#) developed by the Alliance of Bioversity and CIAT.

Aquacrop

[Aquacrop](#) is a crop growth model developed by FAO to assess the effect of environment and management on crop production. It simulates the yield response of crops to water and is well suited to conditions in which water is a key limiting factor in crop production.

Regional Agronomy

Regional agronomy refers to the use of agronomy and related natural and social sciences “*at scale*”. [This resource](#) presents concepts, methods, tools, and workflows to study spatial and temporal variation in agricultural production across large geographic regions.

Responsible data management guidelines

Researchers need to handle privacy and Personally-Identifiable Information (PII) throughout the research project data lifecycle. CGIAR’s Platform for Big Data in Agriculture has developed [Responsible Data Guidelines](#) to help navigate these challenges.

Guideline for agronomy and soil fertility data collection in Ethiopia

The purpose of [this guideline](#) is to provide guidance on standardizing soil and agronomy data collection and thereby enhance temporal and spatial data interoperability.

| Generation of standards-compliant (FAIR) data

FAIR data meet principles of findability, accessibility, interoperability, and reusability. Data produced by CGIAR should be annotated with metadata at the repository and the variable level with terms from ontologies or controlled vocabularies to facilitate interoperability and reuse.

FAIRscribe

[FAIRscribe](#) allows users to easily conform to metadata schemas, ontologies and controlled vocabularies to make research outputs FAIR. It auto-checks for the presence of PII, and offers a wizard to help select an appropriate license.

VMapper+

[VMapper+](#) is a service for annotation of tabular data that allows the association of each data column with entities defined in controlled vocabularies (e.g.: ICASA), [ontologies](#), or custom entities defined by the user.

DataScribe

[DataScribe](#) helps users to create ODK compliant forms with questions or variables linked to ontologies or controlled vocabularies terms.

AgroFIMS

[AgroFIMS](#) allows users to create agronomic field books for digital data collection, using ontology-based units, variables and protocols.

RHoMIS

[Rural Household Multi-Indicator Survey](#) is a tool to efficiently collect a series of harmonized and standardized performance indicators at farm household level, especially targeting smallholder farmers in low-income countries.

| Data processing, transformation, analysis

Carob

[Carob](#) is an open-source R framework that reshapes primary agricultural research data from experiments and surveys into a standard format, aggregating individual data sets into larger collections that can be used in further research.

Collaborative GARDIAN Labs

[CG Labs](#) is an open collaborative data science platform that allows researchers to work together on the same data science project using datasets securely transferred from GARDIAN and other trusted sources, and scripts co-developed and/or seamlessly pushed/pulled from GitHub.

| Data access

GARDIAN

[GARDIAN](#) enables access to over 219,000 publications and 26,000 datasets across CGIAR Centers and organizations working in the agricultural sector. Assets from these sources are presented in standard format, with the possibility of advanced search and filtering.

Data Catalog

EiA data catalog provides a compiled list of data resources available through [CG Labs](#) with a description of the data source, procedures implemented, units, access procedures, etc. The catalog is continuously growing and suggestions to add new resources are welcomed. <https://github.com/reagro/carob>

WorldClim

[WorldClim](#) is a database of high spatial resolution global weather and climate data. These data can be used for mapping and spatial modeling.

Database of Global Administrative areas

[GADM](#) provides maps of administrative areas for all countries, at all levels of sub-division at high spatial resolution, and for an extensive set of attributes.

Planet's Basemap imagery

[A useful source](#) of high-resolution imagery.

Geoglam

[Geoglam](#)'s Crop Monitor sub-national crop calendars are based on a combination of the best available data from different international agencies, national ministries, and expert knowledge from country partners.

Seasonal forecasts

ECMWF has made [all forecast data graphics](#) available for free!

Agroecological Zones of Africa

Developed by Harvard and available [here](#).

| Other data management resources

CG Core Metadata Schema

Metadata addresses the “who, what, where, when, how” of data, and is essentially “data describing data”. The [CG Core metadata schema](#) is a minimum set of metadata elements to describe the products published by CGIAR.

Ontologies

An ontology is a set of concepts about a domain organized to show their properties and interrelationships. Ontology terms have unique URIs, making them machine-readable. They should be used to describe data variables such that variables with the same meaning are uniformly annotated, contributing to data interoperability.

[Crop ontology](#) can be used to describe crop-related and plant breeding data.

[Agronomy Ontology](#) (AgrO) can be used to describe agricultural data.

[Ontology Lookup Service](#) (OLS) is a general query service to search ontology terms among multiple ontologies.

Controlled vocabularies

[AGROVOC](#) is a controlled vocabulary for agricultural sciences. It is a standardized arrangement of words that provides a consistent way to describe data, but terms may or may not have unique URIs (with implications for machine-readability), and typically do not specify relationships.