Metadata fields description with examples for ICRISAT open data platform

**\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\***

**(\* fields are mandatory)**

# Title\*

A simple title that is meaningful and fits the dataset.

Standard practice is to start the title with xxx Data on followed by the title, where xxx can be Phenotype/Genotype/Spatial/Social science/Soil/Weather etc.

***(Exampl****e:* *GBS Data for concurrent improvement of sorghum fodder yield and quality under water-stressed condition****)***

# Description\*

A summary describing the data, its purpose, nature, scope of the Dataset, and its utilization.

***(Example****: Data on sequential observations of agronomically important traits like Leaf biomass, Stem biomass, Total biomass, Leaf area, and Grain yield during cropping season for the year 2016 in Kharif for Sorghum at ICRISAT Patancheru campus research fields****)***

# Author/authors\*

The person(s), corporate body(ies), or agency(ies) responsible for creating the work.

It should be of the format First name, Last name (b) affiliation (Organization name), (d) ORCID id. All the fields including ORCID id are mandatory. Include all the authors corresponding to the dataset

(***Example****:*

*K N S Usha Kiranmayee (ICRISAT) - ORCID: https://orcid.org/0000-0002-7413-965X*

*C. T. Hash (ICRISAT) - ORCID: https://orcid.org/0000-0003-3138-9234*

*S Sivasubramani (ICRISAT) - ORCID: https://orcid.org/0000-0001-9696-3562*

*P B Kavi Kishor (Osmania University)*

*Rajeev Gupta (ICRISAT) - ORCID: https://orcid.org/0000-0002-1451-215X****)***

# Keywords\*

Keywords describe the important aspects of the dataset. They are used for building keyword indexes, for classification, and search a dataset. Providing as many keywords with controlled vocabulary helps generate visibility to the dataset.

|  |  |
| --- | --- |
| Purpose | Ontology |
| <http://www.cropontology.org/> | Trait and other crop-related terms |
| <http://agrovoc.uniroma2.it/agrovoc/agrovoc/en/> | General terms related to socio-economic, phenotypic, or genotype |

***(Example:***

*Shoot fly incidence (CROP ONTOLOGY)* [*http://www.cropontology.org/terms/CO\_324:0000067/*](http://www.cropontology.org/terms/CO_324:0000067/)

*Stay-green (CROP ONTOLOGY)* [*http://www.cropontology.org/terms/CO\_320:0001215/*](http://www.cropontology.org/terms/CO_320:0001215/)

*Livelihoods (AGROVOC)*[*http://agrovoc.uniroma2.it/agrovoc/agrovoc/en/page/c\_1374498089962*](http://agrovoc.uniroma2.it/agrovoc/agrovoc/en/page/c_1374498089962)

*Surveys (AGROVOC)*[*http://agrovoc.uniroma2.it/agrovoc/agrovoc/en/page/c\_7537*](http://agrovoc.uniroma2.it/agrovoc/agrovoc/en/page/c_7537)***)***

## Ontology procedure

# Publication

The publication that uses the data from this dataset. This should include a full bibliographic citation for the related publication. Type of digital identifier for this publication and its identifier(DOI number or URL of the publications.

*(****Example****:* Varshney, R.K., Thudi, M., Nayak, S.N. et al. Theor Appl Genet (2014) 127: 445. https://doi.org/10.1007/s00122-013-2230-6 doi: 10.1007/s00122-013-2230-6 <https://link.springer.com/article/10.1007%2Fs00122-013-2230-6>***)***

# Notes

Additional important information about the dataset.

# Producer\*

The producer has financial or administrative responsibility for this Dataset. It can be an organization name or person name who produced/generated the data.

For every dataset, at least two producers should be specified:

1. Information about the project. This includes the project name, and link to the website(if any)
2. Information about institution

* If the data is generated only by ICRISAT then the producer is ICRISAT
* If the data is produced by the partner institute then the producer is the partner institute.
* If the data is generated by both the ICRISAT and partner institute, then both are producers.

***(Example****:*

1. ICRISAT Project: Improving Postrainy Sorghum Varieties to Meet the Growing Grain and Fodder Demand in India – Phase 2 <http://gems.icrisat.org/gems-projects/aciar-improving-postrainy-sorghum-varieties-to-meet-the-growing-grain-and-fodder-demand-in-india-phase-2/>
2. International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) <http://www.icrisat.org/> Logo URL***)***

# Contributor

The organization or person, responsible for either collecting, managing, or otherwise contributing in some form to the development of the resource.

Various available contributor types are: Data collector, Data curator, Data Manager, Editor, Hosting Institution, Project Leader/Manager, Research Group, Supervisor, other

***(Example****:*

*Research Group: International Rice Research Institute(IRRI)*

*Research Group: National Centre for Agricultural Economics and Policy Research (NCAP)****)***

# Grant information\*

* Funding agency name.
* Grant number

(***Example****: Bill & Melinda Gates Foundation: 12345*)

# Distributor\*

The organization designated by the author or producer to generate copies of the particular work including any necessary editions or revisions.

Name of organization, Affiliation, Abbreviation, URL

(***Example****: International Crops Research Institute for the Semi-Arid Tropics, ICRISAT, http://www.icrisat.org/*)

# Depositor

The person (Family Name, Given Name) or the name of the organization that deposited this Dataset to the repository.

# Time Period Covered

The period to which the data refer. It reflects the time covered by the data, not the dates of coding or making documents machine-readable or the dates the data were collected.

Experiment duration (start date and end date)

# Date of Collection

It contains the date(s) when the data were collected.

Data collection dates (start date and End date)

***(Example****: start: 2012-08-14; End: 2013-11-25****)***

Production place\*

Place where the survey/ experiment was conducted.

Name of the place, Province, City, Country. If available please provide the geographical location using latitude and longitude.

***(Example****: ICRISAT, Patancheru, Hyderabad, India, RP-14A*

*Babrol, Gujarat, India, longitude - 73.753818; latitude – 22.643788***)**

# Kind of data\*

Type of data included in the file.

**(*Example***: *Trial data, phenotypic data, socio-economic data, survey data, census/enumeration data, aggregate data, administrative records data, experimental data, coded textual, coded documents, observation data/ratings, process-produced data, or other)*

# Software

Informationabout the software used to generate the Dataset. Name of the software used along with version information

(***Example***: *TASSEL-12)*

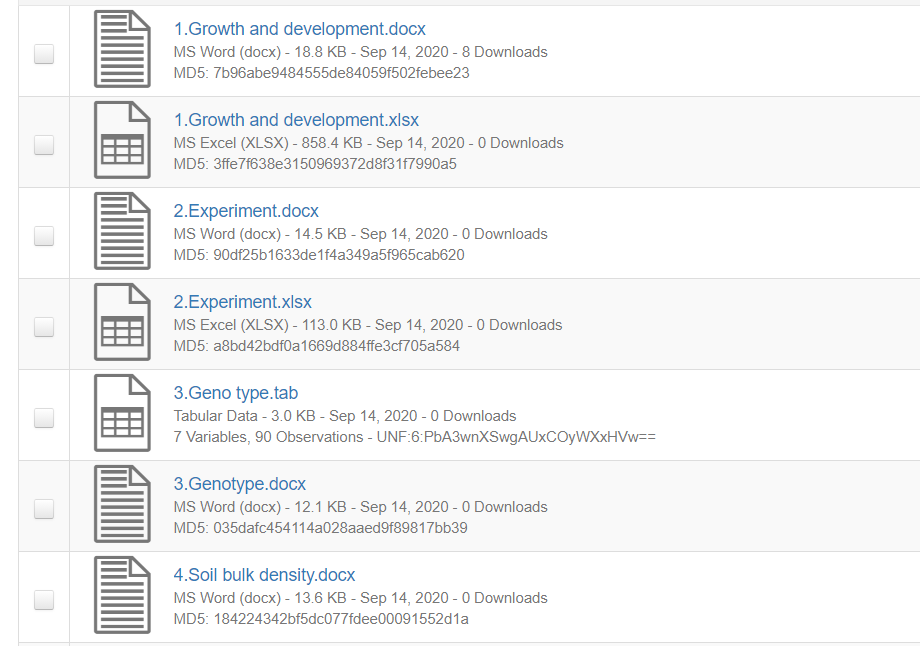
# Data source

List of books, articles, serials, or machine-readable data files that served as the sources of the data collection

***(Examples***: *If the project used ‘Survey of India’ for primary data, then the data source is Survey of India.)*

## File Name

Filename should contain only characters and should follow sentence case. All phenotype, survey data should be in XLSX format. If there are multiple files they should be numbered.



## Reference Datasets

Please use recently published datasets for reference while preparing metadata and data:

|  |  |
| --- | --- |
| **Data Type** | **Reference Link** |
| Genotypic | <http://dataverse.icrisat.org/dataverse/genotypic> |
| Phenotypic | <http://dataverse.icrisat.org/dataverse/phenotypic> |
| Social Science | <http://dataverse.icrisat.org/dataverse/socialscience> |
| Soil & Weather | <http://dataverse.icrisat.org/dataverse/soilandweather> |
| Spatial | <http://dataverse.icrisat.org/dataverse/spatial> |