

Deliverable achievement report

D1.1.1 – Data Management Plan (DMP)

Description of completion

Under WP1, Task 1.1 a detailed Data Management Plan (DMP) has been developed (this report) for making data findable, accessible, interoperable, and reusable (FAIR). These principles include (<https://www.go-fair.org/fair-principles/>):

- F1. (Meta)data are assigned a globally unique and persistent identifier
- F2. Data are described with rich metadata (defined by R1 below)
- F3. Metadata clearly and explicitly include the identifier of the data they describe
- F4. (Meta)data are registered or indexed in a searchable resource

- A1. (Meta)data are retrievable by their identifier using a standardised communications protocol
 - A1.1 The protocol is open, free, and universally implementable
 - A1.2 The protocol allows for an authentication and authorisation procedure, where necessary
- A2. Metadata are accessible, even when the data are no longer available

- I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (Meta)data use vocabularies that follow FAIR principles
- I3. (Meta)data include qualified references to other (meta)data

- R1. (Meta)data are richly described with a plurality of accurate and relevant attributes
 - R1.1. (Meta)data are released with a clear and accessible data usage license
 - R1.2. (Meta)data are associated with detailed provenance
 - R1.3. (Meta)data meet domain-relevant community standards

Project data type

MountainHER will generate three major type of data that can be grouped based on their possibility to be made public:

- *Project outputs, images, videos, articles, and reports*: the Project will produce scientific articles, press articles, reports, videos, pictures, and training models. All these outputs represent valuable information that the Project will strive to make available as rapidly as possible without restrictions. In case of images containing human subjects, written and consicnet approval will be sought before publication.
- *Datasets from field evaluations*: these are complex datasets that need to undergo statistical analysis and evaluations. They do not include any sensitive information, but the raw data might need to be embargoed until after a publication is achieved. The Project will strive to make analyzed data widely available as rapidly as possible.
- *Sensitive opinions, economic analysis, and policies*: the Project also includes several sensitive datasets that can not be shared publicly in their original formats. These include several surveys of opinions of stakeholders, socio-economic analysis of different community enterprises, human and animal subjects, and the study of various policies. These datasets will never be released in their original form. A sharable format will be generated by blacking out sensitive information and this will be released once formally approved by the Ethics team of the project. The following rules will be used for public release:

- Only data for which express and informed consent has been obtained can be made available.
- Privacy protection must be ensured by anonymizing the data in such a way that nobody can attribute the data point to the given person or institution that had provided it.
- Policy documents will only be verbatim or cited in the measure that is useful to generate knowledge of public interest, without affecting in any way the reputation of the legal entity that had created it.

Data accountability

The work package and task leaders are responsible for generating FAIR templates to record data in a consistent and precise manner. These templates require validation by WP4 leadership to meet gender responsive principles and by the project PI and Co-PI.

Each institution that generates data is accountable to safely store them until these are transferred to the defined project portals. An appointed staff within each partner institution has been identified as per the table below. This person is responsible to ensuring that all collected data by the institution are formatted into the requested template, and that these are correctly stored until they are fully transferred to the defined project portals. Standard procedures for data curation are defined as part of the template development process and must be followed by the data curator.

Institution	Name
INRA	Nourredine Bahri
ICARDA	Nada Oussoulous ¹ and Innocent Bikara ²
LARI	Rola Amil
Oxfam	Zakaria Nidkazza ¹
PINS	Natalija Mamula
Open Fields	Ilaria Mazzoli
INRAT	Hajer Ben Ghanem
Birrifacio Emiliano	Veronica Vincenzi
ITGC	Nawel Boukhobza

¹ Project data curator; ² MEL curator

Data storage

The CGIAR Monitoring Evaluation and Learning (MEL) platform has been selected as the project management portal for MountainHER. This is a flexible storage system that allows the upload of a vast array of datatypes and that ensures the recording of all minimum metadata criteria to adhere to the FAIR principles. It also allows the automatic creation of a public web interface where any user can retrieve the open access dataset, or can submit a formal request for release of embargoed datasets. MEL also assigns a DOI to all reported data allowing its direct citation and retrieval as needed. MEL data storage is ensured via a triple standard backup system that engages various storage locations and it also engages the CGIAR system with various fail-safe to prevent any data loss. MEL will be used substantially for all data type of the project

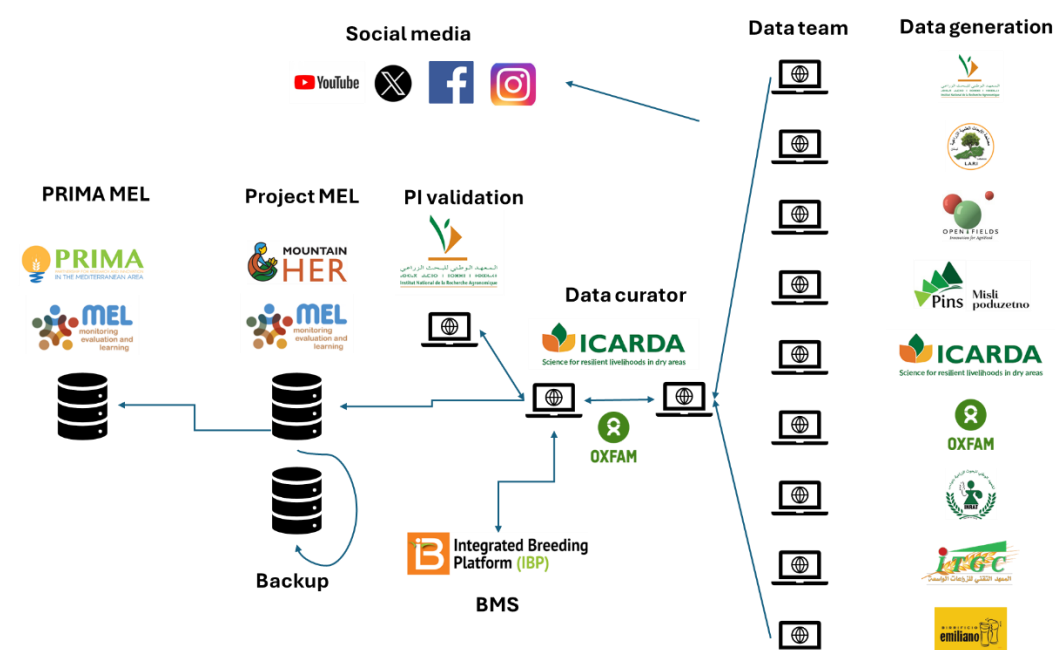
The Breeding Management System (BMS) is instead more suitable to record and store raw field data from trials. Only this data type will be stored here because BMS does not allow easy retrieval of data by external users. Instead, the analysed data will be transferred in the form of a dataset to MEL and stored there to follow FAIR principles.

Data flow

Data are generated and recorded by each institution. Each *institutional data manager* is responsible to reformat the data collected into the templates developed by the WP or Task lead. These data are then transferred to Oxfam and ICARDA *project data curators* that confirm the format and completeness of the data, eventually in an iterative process involving the partners. In most instances, the *project data curators* are also responsible of ensuring the merging of data from multiple institutions. After initial validation, the *project data curators* are responsible to upload data to the MEL project portal or the BMS project portal, depending on the data type. In case of BMS, the raw data are then retrieved by the institutions, WP or task leads for analysis. The analysed data are then provided back to the data curators for reformatting and uploading to MEL.

Only in the case of deliverable or milestone reports, the data curators provide these to the PI and Co-PI of the project for further validation. This might become an iterative process that involves the data curators, but also the original developer of the report. After full validation, the reports are uploaded to the project MEL portal by the data curators. Every six months, the charged MEL curator retrieves the deliverable reports from the MEL project space and transfers these to the PRIMA MEL space.

Social media posting in the form of video or pictures is done directly by the individual institutions communications teams, or through the WP7 social-media curators when using the project accounts. However, images and videos deemed of significant scientific scope are also uploaded to the MEL space following the guidelines above.



Data availability

Each output type will follow specific rules based on various IP considerations:

- The scientific articles will be published as Gold Open Access and made publicly available under MEL as repository.

- The press articles will be provided on MEL as a digital copy and/or as a direct link to the publisher and made publicly available.
- Reports, pictures, videos, analyzed datasets from BMS, and training modules will be maintained on MEL, assigned a DOI, and recognize the name(s) of the creator and contributor, and whenever adequate a geo tag will be applied.
 - Based on the considerations defined previously these will be made publicly available, available on requests, embargoed, or never shared.
- Pictures or videos that include human subjects will be shared only if informed express consent is given by the subject.

Justification of delay

This was due in M6 but it has been reported in M24. We did not feel the need to this DMP because the description in the project plan was sufficiently detailed for operationalization.

Supporting documents

- D1.1.1 18 months report: <https://repo.mel.cgiar.org/handle/20.500.11766/69071>
- D1.1.1 12 months report: <https://repo.mel.cgiar.org/handle/20.500.11766/68441>
- FAIR principles: <https://www.go-fair.org/fair-principles/>

