ISRIC Data Management Protocol (Version 1.0)

Summary: This Data Management Protocol describes the procedures and processes of ISRIC – World Soil Information (hereafter referred to as ISRIC) with respect to the management and long-term preservation of data produced by the center. It consists of nine sections:

1) Introduction (purpose),

2) Roles and responsibility,

3) Data ownership, sharing and accessibility,

4) Safe and shareable storage during research,

5) Data archiving and registration after research,

6) Data documentation and metadata,

7) Data management and research projects,

8) Relationship to other ISRIC WDC-Soils policies and documents,

9) Document control.

Last updated: 29 March 2022 [https://doi.org/10.17027/isric-wdcsoils.202201]

_____

1. Introduction

The purpose of this Data Management Protocol is to set general data management practices at ISRIC – World Soil Information. More specifically, these data management practices aim to prevent data loss and increase reusability of data by complying with the ISRIC Data Policy which is guided by, amongst others, the FAIR principles (Findable, Accessible, Interoperable, and Reusable). ISRIC is a regular member of the International Science Council (ISC, formerly ICSU) World Data System (WDS); in that

capacity ISRIC is the 'WDC Soils'. As such, the document aligns with the longer-term ISRIC strategy as trustworthy provider of soil information.

ISRIC – World Soil Information, legally registered as International Soil Reference and Information Centre, is an independent science-based foundation with a service level agreement with Wageningen University formalising operational support (see SLA's).

The data management practices outlined in this Data Management Protocol[1] apply to all data produced within projects performed at ISRIC – World Soil Information. This includes data derived within internal projects, PhD thesis projects, guest researcher projects, as well as externally funded. External projects (e.g. EU H2020, DeSIRA), however, typically will have their own requirements with respect to project-specific Data Management Plans (DMP).

2. Roles and responsibilities

The ISRIC director identifies the following roles and responsibilities within the context of the management of research and project data (Table 1).

Table 1. Management of research and project data, roles and responsibilities

Who	Roles and responsibilities
Institutional level	• The final responsibility for proper storage of the data lies with the Director, respectively ISRIC Management Board.
Data steward	• Updates the Data Management Protocol when necessary. • Functions as a primary contact point for questions about data management plans (DMP) and refers colleagues to (data management) support when necessary. • Informs new colleagues, PhD candidates and guest researchers about the Data Management Protocol, and related need for project-specific Data Management Plans (see also ISRIC Operational Handbook). • Develops and maintains a template for internal data management plans. • Keeps a log file monitoring the status of Data Management Plans (for projects started after 2019; updated yearly or more often when needed).
Project leader	• Supervises the writing of a Data Management Plan for his/her project(s).

---

[1] This protocol has been adapted from the template of Wageningen UR with which ISRIC has a Service Level Agreement for operational support. It has been tailored to meet ISRIC's own requirements as an independent foundation and host of the WDC-Soils.

- Responsible for good data management during the project using storage solutions and software (e.g. provided by WUR and/or on the whitelist) that match the classification of the data.
- Gives read/write permissions to the project's data (folders and/or files) and periodically checks these authorisations.
- When applicable, ensures careful handling of any sensitive (e.g. personal) data during and after the research project. This must be registered in MyProjects by the Project Leader when a project starts[2]. Subsequently, the Project Leader will be asked to fill in a *SmartPia* form for their project for central evaluation/rating and possible further action.
- At the end of a project, Project Leaders ensure all data is archived on the W-drive (ISRIC_archive server) (see ISRIC Handbook and File Management folder)[3].
- Archives data and code underlying any publication after the research project ends for at least 10 years (Note: for datasets this is permanent, see ISRIC Data Policy).
- Ensures all project-related articles/reports are registered in Pure (via ISRIC secretariat) respectively the ISRIC Soil Data Hub for datasets.

3. Data ownership, sharing and accessibility

ISRIC has ownership over research data created by its staff, PhD candidates and guest researchers. Sharing of and accessibility to the data created by ISRIC will be arranged on a case-to-case basis, depending on whether restrictions apply. When needed, an agreement will be made with the third party, in which ownership, sharing and accessibility to the data are arranged (see ISRIC Data Policy). When data is created by a party other than ISRIC, the data ownership policy of that party should be consulted and respected.

4. Safe and shareable storage during research

The Data Management Protocol requires secure and shared storage of research and project data. The nature of the data will determine how secure the storage must be. For high risk data (e.g. personal data), increased security measures are needed. Several secure and shared storage solutions are provided to ISRIC by Wageningen UR, within the framework of its service level agreement (SLA). Access to these safe storage solutions ensures compliance with the ISRIC data policy for the data collected. As such, data are stored in locations that are backed up automatically / synchronized (so NOT on (external) hard disks, USBs and laptops). This procedure will prevent data loss and ensure that data can be accessed by authorised parties in case of emergency or staff departure from ISRIC.

Available storage solutions, based on our SLA with FB-IT Wageningen UR, are listed in Table 1; the corresponding workflows are continuously being updated by our Spatial Data Infrastructure (SDI)

---

[2] For this, select 'Ja' in the 'Privacy impact assessment (PIA)' box in tab 'Hoofd Kenmerken' of the project budget (begroting) (Note: Only accessible to ISRIC Project leaders).
[3] Note: Like the Operational Handbook, these files are only accessible to ISRIC staff.

team. Additional details about the various options are provided at https://tinyurl.com/yeytm5za. For ISRIC, specific recommendations and guidelines are provided in the Procedure Handbook, version 2021-11-24 (*not for external use*).

All Servicedesk IT manuals available to Wageningen University & Research employees, including ISRIC staff, and students can be found here.

Table 2: Data storage options available to ISRIC

Options	W: drive	M: drive	Teams	OneDrive for Business	External hard drive, USB, Laptop C:
1	x	x	x	x	x
2	x		x	x	
3			x	x	
4	x	x	x	x	
5			x	x	x
6	x	x	x	x	
7	x		x		
8	x				x

Options: 1) relatively cheap; 2) sharing with Wageningen UR; 3) sharing with external parties; 4) no risk of breakage or theft, files can be restored; 5) work offline (synchronise to local disk); 6) access on any device, anywhere, anytime; 7) durable, long life span; and 8) suitable for bulk storage (TB). Derived from: https://tinyurl.com/y4c85yvx, Wageningen UR.

Upon termination of an ISRIC's staff employment, all his/her files on the M-drive and/or OneDrive for Business will be automatically and permanently deleted. Therefore, the overarching ISRIC policy on data storage stipulates that research data in a personal folder/personal cloud storage should also be stored in a departmental or project folder on the W-drive or MS Teams. All data needs to be saved to these media on a regular basis (e.g. monthly).

5. Data archiving and registration after research

Table 3 summarises main aspects to be considered during archiving. These are: a) what data needs to be archived?; b) where should the data be archived?; c) in which format should the data be stored?; d) how should the data set be registered?; and e) who has access to the data archived at ISRIC?

Table 3. Main steps to be considered during archiving

What data needs to be archived?	In compliance with the ISRIC data policy, all data underlying publications (articles/reports) are archived to enable verification of the research and reusability of the data. Data underlying publications include: • Raw or source data

	- Processed or cleaned data - Analysed data - Source code - Software developed during the research project - Data documentation including metadata (e.g. a README.txt file).
Where should the data be archived?	- Data/code/software underlying publications that <u>can</u> be made publicly available for reuse is deposited in an online repository: ISRIC Soil Data Hub (data) or git.wur.nl (code/software). - Data/code/software underlying publications that <u>cannot</u> be made publicly available for reuse (e.g. personal data, company interest, otherwise sensitive) is archived on the W-drive (data) or git.wur.nl (code, software). The metadata, however, is deposited in a repository. - Data/code/software that does not underly publications but <u>can</u> be made publicly available for reuse, should be deposited in an online repository: ISRIC Soil Data Hub (data) or git.wur.nl (code/software). - Data/code/software that does not underly publications and that <u>cannot</u> be made publicly available for reuse can be archived on git.wur.nl (code/software) or W-drive (data), but this is not mandatory and decided upon per project (for specifics, are listed in the ISRIC Procedures Handbook)
In which format should the data be stored?	Data can be stored in any format suitable for analysing the data. However, insofar possible data files should also be stored in open file formats (e.g. txt, csv, pdf, GeoTiff, <u>COG</u> etc.) to facilitate accessibility and reusability. Open file formats will continue to be readable (for software) in the future. When this is not possible the software, the version and the provider of the software should be included in the data documentation. For preferred formats see list prepared by <u>DANS-EASY</u>. ISRIC is in the process of updating this list, taking into consideration recent OGC and INSPIRE standards.
How should the data set be registered?	Data sets published in a data repository and archived on the W-drive need to be registered in <u>Pure</u>, by sending the persistent identifier or URL and accompanying publication(s) to the ISRIC secretariat. Data sets that do not underly publications yet should also be documented and registered to allow reuse.
Who has access to the data archived at ISRIC?	Data on the W-drive (ISRIC_archive) can only be changed or removed by the secretary of ISRIC and a few dedicated staff members.

	Code on private git repositories can be changed and removed by repository members.

6. Data documentation and metadata

Data/code/software should be accompanied by documentation in the form of a README.txt file. Documentation should cover items such as authors, methodology, explanation variables, analytical and procedural information. Metadata should at least have the following components [e.g. title data set, author(s) and affiliation(s), keywords, license, etc.]. ISRIC follows the ISO 19115 standard through its GeoNetwork metadata catalogue and uses the DataCite Metadata Schema (4.4) for assigning a DOI.

7. Data management for research projects

In the initial stages of a project proposal, data management practices should be outlined in a Data Management Plan. In general, the Data Management Plan will be an evolving document. It should be updated and maintained throughout the research project when changes in data management practices are made. Upon completion of a project, the final version of the DMP should be stored in the corresponding, project archive folder (on W-drive).

The Data Management Plan for a research project can be written by using:

- The funder's DMP template, e.g. for EU H2020 projects, or
- The ISRIC DMP template which is available via this page.

Further information on operational guidance is provided in the ISRIC Procedures Handbook (access restricted to ISRIC staff members only). A log file of DMP's, for projects started after 2019, is maintained on the ISRIC Teams site to monitor adherence to this Protocol.

8. Relationship to other policies and documents

This protocol does not exist in isolation. Other corporate policies and contextual links, which cover our daily functioning as ISRIC WDC-Soils, include:

o   ISRIC Data and Software policy

o   ISRIC Collection Management Policy

o   ISRIC Digital Data Preservation Policy

o   <u>ISRIC Service Level Agreements</u>

o   <u>ISRIC Privacy Statement</u>

o   <u>ISRIC Procedures and Standards</u> for data deposit and ingest (*under revision*)

In addition, the stakeholder requirements (SLAs, MoUs, formal contracts etc.) and the role of ISRIC within national and international soil-related communities will also be considered in the preservation framework.

9. Document control

This policy will be reviewed on a regular basis at the instigation of the ISRIC Data Steward.

Version: 1.0
Last updated: 29 March 2022

Written by: Niels H. Batjes, Coordinator WDC-Soils / ISRIC Data Steward
Approved by: Rik van den Bosch, Director ISRIC
Review date: 29 March 2026 (or earlier when needed)