Overview of procedures and standards in use at ISRIC WDC-Soils

N.H. Batjes
Overview of procedures and standards in use at ISRIC WDC-Soils

Niels H. Batjes

ISRIC Report 2016/02

Wageningen, December 2016
Contents

Preface 5
Summary 6
1 Introduction 7
2 Data acquisition 9
   2.1 Field description and soil classification 9
   2.2 Soil sampling and laboratory methods 9
   2.3 Soil monolith preparation 10
3 Data management and processing 13
   3.1 Data archiving 13
   3.2 ISRIC soil information system (point) 14
   3.3 SOTER world soil and terrain database (point, polygon) 14
   3.4 WISE world soil database (point, grid) 15
   3.5 Africa soil profile database (point) 15
   3.6 World soil information service (point, grid, polygon) 16
   3.7 SoilGrids products (grid) 17
4 Serving data 19
   4.1 Data and software policy 19
   4.2 Metadata service 19
   4.3 Spatial data infrastructure 20
   4.4 Towards data interoperability 23
5 Measures of uncertainty 24
6 Concluding remarks 24
7 Acknowledgements 25
Appendix Acronyms and web links 27
References 29
List of Figures

Figure 1. Guided tour through the ISRIC World Soil Museum. ............................................... 11
Figure 2. Referencing of soil samples using barcodes. .......................................................... 12
Figure 3. Testing new water-borne impregnants in the ISRIC workshop. .............................. 12
Figure 4. ISRIC adheres to international standards for research data repositories. ................. 13
Figure 5. Screenshot of SoilGrids250m (https://www.soilgrids.org). ...................................... 18
Figure 6. Example of a Geonetwork search (http://meta.isric.org/geonetwork). ................. 20
Figure 7. Example of a Geonode search (http://geonode.isric.org). ...................................... 21
Figure 8. Geonode layer description and download (http://geonode.isric.org). .................... 21
Figure 9. REST website (http://rest.soilgrids.org). ............................................................. 22
Figure 10. Example of data served through SoilInfo app (http://rest.soilgrids.org). ............... 23
**Preface**

ISRIC - World Soil Information has the mission to create and increase the awareness and understanding of the role of soils in major global issues. As an international institution we inform a wide audience about the multiple roles of soils in our daily lives; this requires scientific analysis of sound soil information. To be most useful, all data should be findable, accessible, interoperable, and reusable in compliance with so-called FAIR principles.

The purpose of this technical report is to give an overview of main procedures and standards in use at ISRIC – World Soil Information, a regular member of the ICSU World Data System. As a dynamic and innovative organisation we will issue updated versions of this document when appropriate.

ISRIC is seeking collaboration with national institutes with a mission for soil resource inventories in order to further develop its world soil information services for the benefit of the international community.

ir. Rik van den Bosch

Director, ISRIC — World Soil Information
Summary

This report serves to give an overview of main procedures and standards in use at ISRIC – World Soil Information, regular member of the ICSU World Data System. These cover the whole data life cycle from field sampling to serving quality-assessed soil data to the world community. Consistent procedures and de facto standards are used to screen (QA/QC) and standardise respectively harmonise the wide range of soil-related data that have been shared with us for consideration in our world-covering databases and web services. Ultimately, these processes are aimed at facilitating global data interoperability and citeability in compliance with FAIR principles: the data should be ‘findable, accessible, interoperable, and reusable’.
1 Introduction

ISRIC – World Soil Information is a regular member of the International Council for Science (ICSU) World Data System (WDS). It has a mission to ‘serve the international community with information about the world’s soil resources to help addressing major global issues.’ This process involves all stages of the data life cycle from data sampling in the field and laboratory analyses to collation into quality-assessed databases followed by data standardisation/analysis/modelling and publication of the data. This requires adoption and development of internationally recognised standards, such as the World Reference Base for Soil Resources (IUSS Working Group WRB 2015) for international soil correlation.

As indicated by RDA-CODATA (2016), the ‘ability of the research community to share, access, and reuse data, as well as to integrate data from diverse sources for research, education, and other purposes requires effective technical, syntactic, semantic, and legal interoperability rules and practices’. An important aspect here is that data should be ‘Findable, Accessible, Interoperable, and Reusable’ that is complying with FAIR principles (Wilkinson et al. 2016).

The purpose of this document is to provide an overview of (main) procedures and standards currently in use at ISRIC. As an international institute dealing with world soil information, with partners, we are working towards the maintenance and development of new, internationally, recognised standards. For example, we are proactive in working groups of the International Union of Soil Science (IUSS), the Open Geospatial Consortium (OGC, Ritchie 2016), GODAN (Global Open Data for Agriculture and Nutrition), GlobalSoilMap (Arrouays et al. 2014) and the Global Soil Partnership (Baritz et al. 2014; FAO-GSP 2014a).

The present document consists of the following, broadly defined, sections that cover successive stages of the data life cycle:

- Data acquisition (Chapter 2)
- Data management and processing (Chapter 3)
- Serving data (Chapter 4)
- Measures of uncertainty (Chapter 5)

A list of abbreviations, respectively acronyms, is provided in the Appendix. This list also provides the locator (URL) to the corresponding on-line resources.
2 Data acquisition

2.1 Field description and soil classification

From its inception, ISRIC (formerly International Soil Museum) has been supporting the initiative of the FAO concerning international soil characterisation in relation to the development of the FAO-UNESCO Soil Map of the World (FAO-Unesco 1974; FAO 1988) and its successors (FAO/IIASA/ISRIC/ISSCAS/JRC 2012).

Main procedures currently in use at ISRIC are described in the following documents:


2.2 Soil sampling and laboratory methods

Until closure of the ISRIC reference laboratory in 2001, the following standards were used for analysing samples of the ISRIC soil reference collection:


Quality control was performed according to:


Since 2010, soil samples collated in conjunction with the ISRIC soil exploration and sampling project (SOLEX) are sent to the National Soil Survey Center (NRCS), Lincoln NE. These samples are analysed according to:

According to earlier inter-laboratory comparisons, results of the ISRIC and USDA laboratories are considered to be similar for application at a broad scale (Pleijisier 1989; Vogel 1994).


### 2.3 Soil monolith preparation

ISRIC maintains the World Soil Museum, presently located on the Wageningen Campus. On display are some 80 monoliths (Figure 1), representing a selection of our collection of over 1000 monoliths covering all major soil types of the World (FAO-Unesco 1974; FAO 1988; IUSS Working Group WRB 2015). For each soil monolith we have sample data, a full profile description, soil chemical and physical data, and information on the landscape and land-use. Since 2016, this information is managed in WoSIS our central server database (see Section 3.2).

Procedures for taking and preserving soil monoliths are described in:

All samples in the collection have unique barcodes for ease of reference and traceability (Figure 2). New procedures for the preservation of soil monoliths that use environmentally safe lacquers are being tested in the ISRIC workshop (Figure 3). Details will be provided in the upcoming ‘Collection Management Plan for the World Soil Museum’ and associated documentation (Mantel 2016, pers comm.).
Figure 2. Referencing of soil samples using barcodes.

Figure 3. Testing new water-borne impregnants in the ISRIC workshop.
3 Data management and processing

3.1 Data archiving

As a World Data Centre, we ensure the long-term preservation of our collection of country documentation (reports, maps and slides), soil specimens (monoliths, samples, thin sections, hand specimens), as well as geo-referenced databases (point, polygon, and grid). Many of these holdings may be consulted both on site and on-line.

A substantial section of our soil map collection was digitised in collaboration with JRC-EU in the framework of the EuDASM project (e.g. Selvarajdou et al. 2005). Subsequently, a wider range of scanned reports and HR maps has been made available through the ISRIC on-line library (see Batjes 2007). Digitisation of the collections is ongoing and new materials1 are welcome!

Reports and maps are assigned unique identifiers (ISN) and registered in compliance with the Wageningen University & Research library management system (http://www.isric.org/content/search-library-and-map-collection). Since 2016, ISRIC can issue unique identifiers (DOI) to reports and databases that result from ISRIC projects; for this we have an agreement with DataCite, Netherlands.

We follow the certificates and standards of the ICSU World Data System. These are in accord with FAIR principles (Susanna-Assunta and Philippe 2016; Wilkinson et al. 2016). Accordingly, ISRIC World Data Centre for Soils (WDC-Soils) is listed as a trusted repository in the re3data.org registry (Figure 4).

![re3data.org](http://www.isric.org/content/search-library-and-map-collection)

http://doi.org/10.17866/R3X01J

World Data Centre for Soils

Figure 4. ISRIC adheres to international standards for research data repositories.

1 [http://www.isric.org/content/requests-soil-maps-and-reports](http://www.isric.org/content/requests-soil-maps-and-reports)
Preferred and acceptable formats for generating and submitting data are in line with those recommended by WDS, DANS (2015), and OGC. As a general guideline, according to DANS, file formats best suited for long-term sustainability and accessibility should be: ‘frequently used, have open specifications, and be independent of specific software, developers or vendors’. Different levels of support are provided for the various file formats depending on whether they are proprietary or open source; see list prepared by datacentrum.3tu.nl2.

3.2 ISRIC soil information system (point)

In the 1990’s the ISIS stand-alone database was specially developed to store data from the ISRIC Soil Reference collection, originally in dBase format. This collection characterises a selection of monoliths with morphological, analytical data that represent the main soil reference groups of the World (ISIS 2016; Kauffman et al. 1996; van de Ven and Tempel 1994).

As of 2015 the ISIS holdings are handled in WoSIS (see Section 3. 6), ISRIC’s central server database (Ribeiro et al. 2015). Standardised datasets served from WoSIS, however, do not (yet) consider the whole suite of attributes considered in ISIS. The full complement of ISIS holdings, as served from WoSIS, may be consulted on-line at http://isis.isric.org. Further, the ISIS holdings can be consulted through the ISRIC virtual soil museum at: http://wsm.isric.org

3.3 SOTER world soil and terrain database (point, polygon)

SOTER aims to establish a World Soils and Terrain Database, at scale 1:5 M, containing digitised map units and their attribute data in standardised format. The programme was implemented in 1986 by FAO, UNEP and ISRIC, under the aegis of the IUSS, to be carried out in collaboration with a wide range of national soil institutes. ISRIC played a lead role in methodology development and programme implementation, and de facto standard setting (Oldeman and van Engelen 1993; Van Engelen 2011; van Engelen and Dijkshoorn 2013). SOTER databases provided the soil-geographical basis for large sections of the Harmonised World Soil Database (FAO/IIASA/ISRIC/ISSCAS/JRC 2012). As of 2015, all SOTER holdings have been incorporated into WoSIS (Ribeiro et al. 2015).

SOTER procedures are described in the following manual:


2 http://datacentrum.3tu.nl/fileadmin/editor_upload/File_formats/Digital_Preservation_Support_levels.pdf
3.4 WISE world soil database (point, grid)

The WISE project, generally known as the ‘World Inventory of Soil property Estimates’ project, has developed homogenized sets of soil data, linked to a spatial data set, relevant for a wide range of environmental studies at global scale, such as agro-ecological zoning, assessments of crop production, soil vulnerability to pollution, soil carbon stocks and change, and soil gaseous emission potentials (Batjes 2016; Batjes and Bridges 1994; Batjes et al. 1997; Bouwman et al. 2002; Maire et al. 2015; Stoorvogel et al. 2016).

An important element of the WISE project has been to develop consistent procedures for filling gaps in measured soil (profile) data using taxotransfer rules (Batjes 2003; Batjes et al. 1997). These procedures have also been used to fill soil analytical gaps in SOTER, thus making such a database suited for modelling applications (Batjes et al. 2007). The resulting 'SOTWIS' databases also provided the soil attribute data for large sections of the Harmonised World Soil Database (FAO/IIASA/ISRIC/ISSCAS/JRC 2012). As of 2015, the point holdings from WISE have been incorporated into WoSIS (Ribeiro et al. 2015).

References to procedures and standards used for WISE may be found in:


3.5 Africa soil profile database (point)

The Africa Soil Profiles Database (AfSP) contains soil profile records compiled from a wide range of sources (Leenaars 2013). The database was compiled in support of the Africa Soil Information System (AFSIS 2016) in conjunction with the GlobalSoilMap project. Soil attribute values in AfSP are standardised according to SOTER conventions and validated according to routine rules. As of 2015, the point holdings from AfSP have been incorporated into WoSIS (Ribeiro et al. 2015)

References to procedures and standards used for AfSP are provided in:
3.6 World soil information service (point, grid, polygon)

ISRIC has developed a centralised and user-focused server database, known as ISRIC World Soil Information Service (WoSIS) that draws on contributions from many data providers. The aims are to:

1. Safeguard world soil data ‘as is’ (especially for soil legacy data).
2. Share soil data (point, polygon and grid) upon their standardisation and harmonisation.
3. Provide quality-assessed input for a growing range of environmental applications.

WoSIS, a PostgreSQL server database, includes the holdings from all our former stand-alone databases (see Sections 3.2, 3.3, 3.4 and 3.5). Any duplicate profiles have been removed in the process (Ribeiro et al. 2015).

Ultimately, the quality assessed and standardised resp. harmonised ‘shared’ data are made available to the international community through several web services and formats (Batjes et al. 2016; Ribeiro et al. 2015). An important aspect of WoSIS is the development of consistent procedures for describing disparate soil analytical method descriptions under a common (de facto) standard. This work is carried out in support of Pillar 5 (Harmonisation) of the Global Soil Partnership (FAO-GSP 2014b). The initial list of soil properties in WoSIS is based on the GlobalSoilMap (GlobalSoilMap 2013) specifications.

References to procedures and standards used for WoSIS are provided in:

Related procedures and standards:


3.7 SoilGrids products (grid)

SoilGrids™ predictions are generated at ISRIC — World Soil Information as a result of international collaboration. SoilGrids™, a system for automated soil mapping based on global soil profile and covariate data, produces a collection of updatable soil property and class maps of the world. It uses machine learning and statistics (Hengl et al. 2014; Hengl et al. 2016). Being automated, output can be generated at various resolutions (e.g. SoilGrids10km, SoilGrids1km, SoilGrids250m, and SoilGrids100m).

SoilGrids layers are available through a web portal with querying and download functionality3 (Figure 5). SoilGrids predictions can also be accessed and queried using the Soil Info4 app. In the future, this web application will also provide additional functionality such as an assessment of regional soil fertilizer requirements.

SoilGrids is based on Free and Open Source Software (Linux, PHP, Latex, R, GDAL, GRASS, SAGA GIS, PostgreSQL, Postages, Python, Google Earth, and similar).

SoilGrids™ predictions may be used to support the mapping activities of the Global Soil Partnership, mainly in data scarce regions (FAO-GSP 2014a; Montanarella and Vargas 2012; Omuto et al. 2012).

Detailed information on the procedures and standards used for SoilGrids™ are presented in:


3 https://www.soilgrids.org
4 http://soilinfo.isric.org
Figure 5. Screenshot of SoilGrids250m (https://www.soilgrids.org).
4 Serving data

4.1 Data and software policy

ISRIC’s Data and Software Policy governs the use by any user of the ISRIC website and web services. Recognising the benefits and importance of contributing to the growing international efforts of open data sharing, ISRIC has adopted the same general data sharing principles as ICSU-WDS, GEO/GEOSS, GODAN and the INSPIRE directive.

- ICSU-WDS data sharing principles: https://www.icsu-wds.org/services/data-sharing-principles

4.2 Metadata service

Our metadata are managed by a GeoNetwork 3.0 instance that implements the ISO 19115 standard, defining metadata sections, entities and elements. The GeoNetwork catalogue is divided into 2 sub sections: ISRIC Catalogue and ISRIC Library records. This is due to the fact that the WUR-Library is used to store maps and scanned content from ISRIC, while the ISRIC Catalogue contains the metadata of datasets and web services developed by ISRIC. Figure 6 gives an example of a GeoNetwork search for ‘SOTER’.

GeoNetwork provides a Catalogue Web Service (CWS) which is an important part of a Spatial Data Infrastructure (SDI) and a requirement for the Global Earth Observations System of Systems (GEOSS) and interoperability with ICSU-WDS.

Resources:

- ISRIC Geonetwork: http://meta.isric.org/geonetwork
- ISRIC CWS: http://meta.wurnet.nl/geonetwork/srv/eng/csw

---

5 http://www.geoportal.org/
6 https://www.icsu-wds.org/services/data-portal
4.3 Spatial data infrastructure

ISRIC’s spatial data infrastructure (SDI) is based on open and free OGC standards. Since late 2016, it includes a GeoNode\(^7\) providing a better integration between different components of a SDI (search, description, display and download of data).

With Geonode, a user can search for data content and easily obtain information about a specific dataset (Figure 7 and 8). The download of data is done directly from a Geoserver instance that allows users to choose between multiple file formats (e.g. GeoTiff and shapefiles).

\(^7\) [http://geonode.org/](http://geonode.org/)
Figure 7. Example of a Geonode search (http://geonode.isric.org).

Figure 8. Geonode layer description and download (http://geonode.isric.org).
As part of our SDI (especially for SoilGrids™), point data are served using a REST (Representational State Transfer) interface (Figure 9). This interface is mainly used for web mobile development where a simple query using an URL will return formatted data that is easily understood by programming languages.

![REST website](http://rest.soilgrids.org)

Figure 9. REST website (http://rest.soilgrids.org).

Point data provided by the REST can be visualized on the soilinfo app as illustrated in Figure 10.

Resources:

- [http://geonode.isric.org](http://geonode.isric.org)
- [http://rest.soilgrids.org](http://rest.soilgrids.org)
4.4 Towards data interoperability

ISRIC is actively contributing to various international activities aimed at ultimately achieving global interoperability of soil data. Such activities are/were undertaken in conjunction with our eSOTER, AfSIS, GlobalSoiMap, and WoSIS-related work and much progress has been made. However, concerted work is still needed to bring the various approaches to a common markup language for soils ('soilML') under a common standard (Wilson 2016). At present, there are several complementary efforts to the ISO and OGC work (e.g. GODAN 2016; GS Soil 2008; INSPIRE 2015):


Figure 10. Example of data served through SoilInfo app (http://rest.soilgrids.org).
5 Measures of uncertainty

All ISRIC databases are submitted to rigorous, routine in-house checks prior to their distribution. Typically, measures for uncertainty or accuracy assessments are described in technical reports and/or manuscripts for external peer-review, in compliance with common scientific practice (e.g. Batjes 2016; Hengl et al. 2014; Hengl et al. 2015; Hengl et al. 2016; Heuvelink 2014). Nonetheless, our products are unlikely to be free of errors. Users may report potential ‘bugs’ in the data or software using various automated procedures, for example through a web-based, Git-repository hosting service8.

6 Concluding remarks

As indicated, the purpose of this technical report is to give an overview of main procedures and standards in use at ISRIC. As such, it should be seen as ‘work in progress’. Newly emerging standards of relevance to ISRIC’s mission and activities will be gradually incorporated in subsequent releases of this document. Often, these will be new de facto standards that evolve from long-term international collaboration.

8 https://github.com/ISRICWorldSoil/SoilGrids250m/issues
7 Acknowledgements

Thanks are due to my colleagues at ISRIC for their editorial comments, in particular Jorge Jesus de Mendes who rewrote Sections 4.2 and 4.3.
Appendix  Acronyms and web links

AfSP: Africa Soil Profile Database, http://www.isric.org/content/africa-soil-profiles-database-afsp
CC: Creative Commons (licences), https://creativecommons.org/
DataCite: A leading global non-profit organisation that provides persistent identifiers (DOIs) for research data, https://www.datacite.org/
GeoNetwork: Open-source catalogue application to manage spatially referenced resources, http://geonetwork-opensource.org/
GitHub: web-based Git repository hosting service, https://en.wikipedia.org/wiki/GitHub
GSM: GlobalSoilMap, http://www.globalsoilmap.net/
INSPIRE: http://inspire.ec.europa.eu/
INSPIRE, code list register: http://inspire.ec.europa.eu/codelist
JRC: Joint Research Centre – European Commission, http://driver-project.eu/content/jrc-joint-research-centre-european-commission
NRCS: Natural Resources Conservation Service (USA),

OdBL: Open Database License, http://opendatacommons.org/licenses/odbl/


OGC Standards and Supporting Documents: http://www.opengeospatial.org/standards


Re3data.org: Registry of Research Data Repositories, http://www.re3data.org/

RDA: Research Data Alliance, https://www.rd-alliance.org/


* All URLs last accessed on 5 December 2016.
References

DANS 2015. Preferred formats (ver. 3.0), Data Archiving and Networked Services (DANS), The Hague (NL), 18 p.

GODAN 2016. Soil Data Working Group http://www.godan.info/working-groups/soil-data


ISRIC – World Soil Information has a mission to serve the international community as custodian of global soil information and to increase awareness and understanding of soil in major global issues.

More information: www.isric.org

ISRIC – World Soil Information has a strategic association with Wageningen UR (University & Research centre)