World Soil Information Service (WoSIS) – Towards the standardization and harmonization of world soil data

Procedures Manual 2018

ISRIC Report 2018/01

Eloi Ribeiro
Niels H. Batjes
Ad van Oostrum

World Soil Information Service (WoSIS) - Towards the standardization and harmonization of world soil data

Procedures Manual 2018

Eloi Ribeiro, Niels H. Batjes and Ad van Oostrum

ISRIC Report 2018/01

Wageningen, February 2018

Director, ISRIC-World Soil Information
Droevendaalsesteeg 3 (building 101)
6700 PB Wageningen
The Netherlands

Additional information on ISRIC can be accessed through `http://www.isric.org`

Citation:
Ribeiro E, Batjes NH and van Oostrum AJM, 2018. World Soil Information Service (WoSIS)-Towards the standardization and harmonization of world soil data. Procedures manual 2018, Report 2018/01, ISRIC - World Soil Information, Wageningen.

doi:10.17027/isric-wdcsoils.20180001.

This report is only available as an online PDF version. Being based on a LaTeX template, the lay out differs from that used for printed reports in the ISRIC series.

ISRIC Report 2018/01

Contents

List of Tables

List of Figures

Acronyms and abbreviations

AfSP	Africa Soil Profiles database (a compilation of soil legacy data for Africa)
DOI	Digital Object Identifier
eSOTER	Regional pilot platform as EU contribution to a Global Soil Observing System
FAO	Food and Agriculture Organization of the United Nations
FOSS	Free and Open Source Software
GEOSS	Global Earth Observation System of Systems
GML	Geography Markup Language
GLOSOLAN	Global Soil Laboratory Network (established by the Global Soil Partnership)
GODAN	Global Open Data for Agriculture and Nutrition
GSP	Global Soil Partnership
ICSU	International Council for Science
INSPIRE	Infrastructure for Spatial Information in the European Community
ISIS	ISRIC Soil Information System (holds the ISRIC World Soil Reference Collection)
ISN	Unique record number; as used by ISRIC & Wageningen University Library
ISO	International Organization for Standardization
ISRIC	ISRIC - World Soil Information (legally registered as: Inter. Soil Refer. and Info. Centre)
IUSS	International Union of Soil Sciences
JSON	JavaScript Object Notation
OGC	Open Geospatial Consortium
REST	Representational State Transfer
SDI	Spatial Data Infrastructure
SDF	Soil Data Facility; a facility hosted by ISRIC for the GSP
SOAP	Simple Object Access Protocol
SoilML	Soil Markup Language
SOTER	Soil and Terrain database programme
UNEP	United Nations Environmental Programme
USDA	United States Department of Agriculture
UUID	Universally unique identifier (for soil profiles)
WDC-Soils	World Data Center for Soils, regular member of the ICSU World Data System (ICSU-WDS)
WISE	World Inventory of Soil Emission potentials (harmonized soil profile data for the world)
WoSIS	World Soil Information Service (server database)

WSDL	Web Services Description Language
XML	Extensible Markup Language
XSL	Extensible Stylesheet Language

Preface

The ISRIC - World Soil Information foundation, legally registered as the International Soil Reference and Information Centre, has a mission to serve the international community as custodian of global soil information. We are striving to increase awareness and understanding of soils in major global issues.

We have updated the procedures for WoSIS (World Soil Information Service), our centralized enterprise database to safeguard and share soil data (point, polygons and grids) upon their standardization and harmonization. Everybody may contribute data for inclusion in WoSIS. However, data providers must indicate how their data may be distributed through the system. This may be ‘please safeguard a copy of our dataset in your data repository’, ‘you may distribute any derived soil data but not the actual profile data’ or ‘please check and help us to standardize our data in WoSIS, there are no restrictions on use (open access)’.

Conditions for use are managed in WoSIS together with the full data lineage to ensure that data providers are properly acknowledged. In accord with these conditions, the submitted data are quality-assessed, standardized and harmonized, ultimately to make them ‘comparable as if assessed by a single given (reference) method’. The most recent set of quality-assessed data served from WoSIS, commonly referred to as ‘WoSIS latest’, may be accessed freely through our Soil Data Hub (`http://data.isric.org`).

ISRIC, in its capacity as World Data Center (WDC) for Soils, also serves its data products to the global user community through auxiliary portals, in particular those of the ICSU World Data System and GEOSS (Global Earth Observing System of Systems).

WoSIS is the result of collaboration with a steadily growing number of partners and data providers, whose contributions we gratefully acknowledge. New releases of WoSIS-derived products, that consider a broader range of quality-assessed soil data, will gradually be released by us for the shared benefit of the international community and national stakeholders.

Ir H. van den Bosch
Director, ISRIC - World Soil Information

Summary

To better address the growing demand for soil information ISRIC - World Soil Information has developed a centralized database for the shared benefit of the international community. This database, hereafter referred to as WoSIS (World Soil Information Service), has been designed in such a way that, in principle, any type of soil data (point, polygon, and grid) may be accommodated. However, WoSIS will only provide quality-assessed data in a consistent format, with detailed information on data lineage and conditions for use. Data derived from WoSIS may be used to address pressing challenges of our time including food security, land degradation, water resources, and climate change.

At present, the focus in WoSIS is on developing consistent procedures for standardizing and harmonizing soil analytical data as submitted by a wide range data providers. The general procedure for processing profile data in WoSIS is as follows. First, new source data are imported ‘as is’ into a PostgreSQL database, with the original naming and coding conventions, abbreviations, domains, lineage and data licence; thereby copies of the source materials are safeguarded at ISRIC. Second, the source databases are imported into WoSIS proper, forming the first major step of data standardization (into a single data model). The next step of data standardization, applied to the values for the various soil properties as well as to the naming conventions themselves, is needed to make the data queryable and useable.

Special attention has been paid to the standardization of analytical method descriptions, focusing on the list of soil attributes considered in the GlobalSoilMap (GSM, 2013) specifications (e.g. organic carbon, soil pH, soil texture (sand, silt, and clay), coarse fragments, cation exchange capacity, bulk density, and water holding capacity), to which we have added electrical conductivity. Further, we checked and added the soil classification (FAO, WRB and USDA Soil Taxonomy) and horizon designations as provided in the source databases.

During the standardization of the analytical method descriptions, major characteristics of commonly used methods for determining a given soil property are identified first. For soil pH, for example, these are the sample pretreatment, extractant solution (water or salt solution), and in case of salt solutions the salt concentration (molarity), as well as the soil/solution ratio; a further descriptive element is the type of instrument used for the actual laboratory measurement. Similar schemes were developed for the other soil properties under consideration here, with accompanying flowcharts.

A third step in the standardization / harmonization process will require data harmonization to make the analytical data comparable that is as ’if assessed by a single given (reference) method’. Such work will require further international collaboration and data sharing to the benefit of the international user community as foreseen in the framework of Pillar 5 of the Global Soil Partnership.

Inherently, the present standardization procedures are only applied to soil profiles flagged as having adequate permissions (i.e. ’shared’ profiles with at least a Creatice Commons Licence type CC BY or CC BY-NC). The resulting standardized data can be accessed through our GeoNetwork instance (`http://data.isric.org/`). The latest, dynamic dataset is available through a web feature service (WFS); the corresponding data layers are referred to as ‘WoSIS latest’. For consistent citation purposes, we also produce ‘static’ snapshots of the standarized data in comma delimited format (CSV), most recently ‘WoSIS snapshot - July 2016’ (Batjes et al., 2017).

WoSIS forms an important building block of ISRIC‘s Spatial Data Infrastructure (SDI). Further developments will allow for the fulfilment of future demands for global soil information, and enable further incorporation of soil data shared by third parties in an inter-operable way, within a federated system.

Chapter 1

Introduction

ISRIC‘s mission, as custodian of global soil information, is to ‘serve the international community with quality-assessed information about the world’s soil resources to help addressing major global issues’. Since the 1980‘s, ISRIC has developed and managed a number of stand-alone soil databases that are freely available to the scientific community and other non-commercial groups. However, dissemination opportunities have changed drastically in the past two decades permitting faster and more efficient forms of information delivery. Strategies adjusted to these opportunities led to the development of a prototype for ‘A centralized and user-focussed database containing only validated and authorized data with a known and registered accuracy and quality’ in 2010 (Tempel et al., 2013). Subject to initial testing, a revised system was released in 2015: WoSIS or ISRIC World Soil Information Service (Ribeiro et al., 2015).

WoSIS is a server database for handling and managing multiple soil datasets in an integrated manner, subsequent to proper data screening, standardization and ultimately harmonization. A key element is that the system allows for inclusion of soil data shared by third parties, while keeping track of the data lineage (provenance) and possible restrictions for use (licences). Ultimately, the terms of these licences will determine which set of quality-assessed and standardized data can be served freely to the international community.

WoSIS forms an important component of the ISRIC Spatial Data Infrastructure (SDI)[1], ISRIC’s overarching approach for generating open soil data (Figure 1.1). The approach aims to collate/safeguard soil data (both legacy/historic data and new soil data) and upon their standardisation use the standardized soil data in combination with spatial co-variate layers for the automated production of grid maps using digital soil mapping at various spatial resolutions (SoilGrids) (Hengl et al., 2017).

WoSIS and SoilGrids are key components of ISRIC‘s evolving Spatial Data Infrastructure (SDI), through which quality-assessed data about soils can be made accessible and shared across disciplines to address global challenges such as climate change, food security, and the degradation of land and water resources (Batjes et al., 2013).

A vast amount of soil data have been collated throughout the world yet only a fraction thereof is as yet freely available for use by the international community (Omuto et al., 2012; Arrouays et al., 2017). Realistically, however, only part of the shared site, profile (e.g. classification) and horizon (e.g. morphological, chemical and physical) data can be standardized in programmes such as WoSIS (Batjes et al., 2017). The present focus in WoSIS is on serving quality-assessed soil data for those properties considered in the GlobalSoilMap specifications (GSM, 2013).

This report presents the latest changes to the WoSIS procedures; as such it supercedes the preceding version (Ribeiro et al., 2015). It consists of six Chapters and seven Appendices. Following up on the introduction, and prior to describing the database design (Chapter 3), basic principles for flagging repeated (e.g. duplicate) soil profiles originating from different international databases, basic measures for defining data quality (i.e. level of trust and accuracy), and the main steps towards standardization and harmonization of numerical soil data are discussed in Chapter 2. Aspects of data interoperability and web services (Chapter 4), and possible approaches to federated databases (Chapter 5), set in the context of ISRIC’s

---

[1] `http://isric.org/explore`

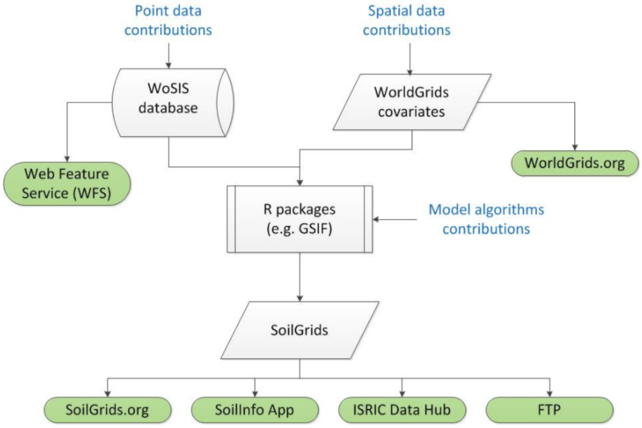


Figure 1.1: Simplified representation of ISRIC's workflow for data processing.

evolving Spatial Data Infrastucture (SDI), are discussed next. Subsequently, an outlook concerning future developments is presented in Chapter 6.

Appendix A explains how the (steadily growing number of) quality-assessed and standardized data managed in WoSIS can be accessed freely by users. Appendix B describes basic principles for compiling soil profile data to facilitate entry into WoSIS, and provides standardized templates for this. Appendix C focuses on quality aspects related to soil laboratory data. Subsequently, Appendix D provides the rationale and criteria for standardizing soil analytical procedures descriptions in WoSIS. Flowcharts for this are presented in Appendix E, while the corresponding option tables (i.e. look up tables) are described in Appendix F. Descriptions for each data table in WoSIS are provided in Appendix G.

Chapter 2

Basic principles for processing data

As indicated, the present focus in WoSIS is on uniformly characterizing point soil data for the world, using a normalized and structurally sound data model Chapter 3. For this, soil profile data are considered as the result of observations and measurements (O&M), see Appendix 2. Such a systematic grouping of the available information is a prerequisite for soil-data-interoperability (OGC, 2015; Wilson, 2016), as further discussed in Chapter 4 and 5.

2.1 Flagging repeated profiles

One of the first tasks in the process of importing data into WoSIS is the search for repeated profiles. This is necessary as the same profile may have been described in multiple source databases, albeit using different procedures and profile identifiers. Such a situation is likely to arise with stand-alone databases that are data compilations, such as those developed for projects such as SoTER (van Engelen and Dijkshoorn, 2013), WISE (Batjes, 2009) or the Africa Soil Profile Database (Leenaars et al., 2014b). This screening process will yield a unique set of soil profiles and thus produce a truthful profile count (Figure 2.1).

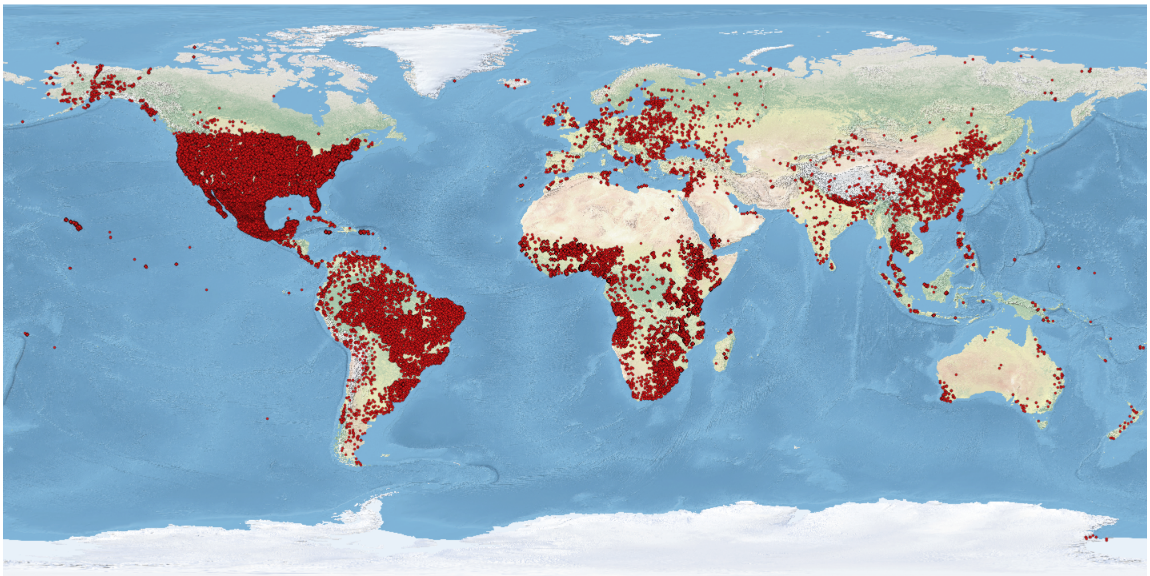


Figure 2.1: Location of shared, unique, geo-referenced profiles held in WoSIS (February 2018).

Repeated profiles may be identified using various approaches. Two main approaches or checks are applied in WoSIS, on lineage and on geographical proximity. The lineage check considers the data source identifiers, uses this information to trace the original data source, and from there look for duplicates.

Alternatively, the proximity check is based on the geographic coordinates. The procedure first identifies profiles that are suspiciously close to another (e.g. <10 m). Subsequently, the information for these profiles is compared and the database manager assesses the likelihood of such profiles being identical. In case of duplicates, only the standardized soil profile derived from the most detailed source database will be served to the international community in accord with the corresponding data licence (see Figure 2.1). The above screening process is a rather time consuming task as it cannot be automated fully; additional techniques for identifying possible duplicates are under investigation.

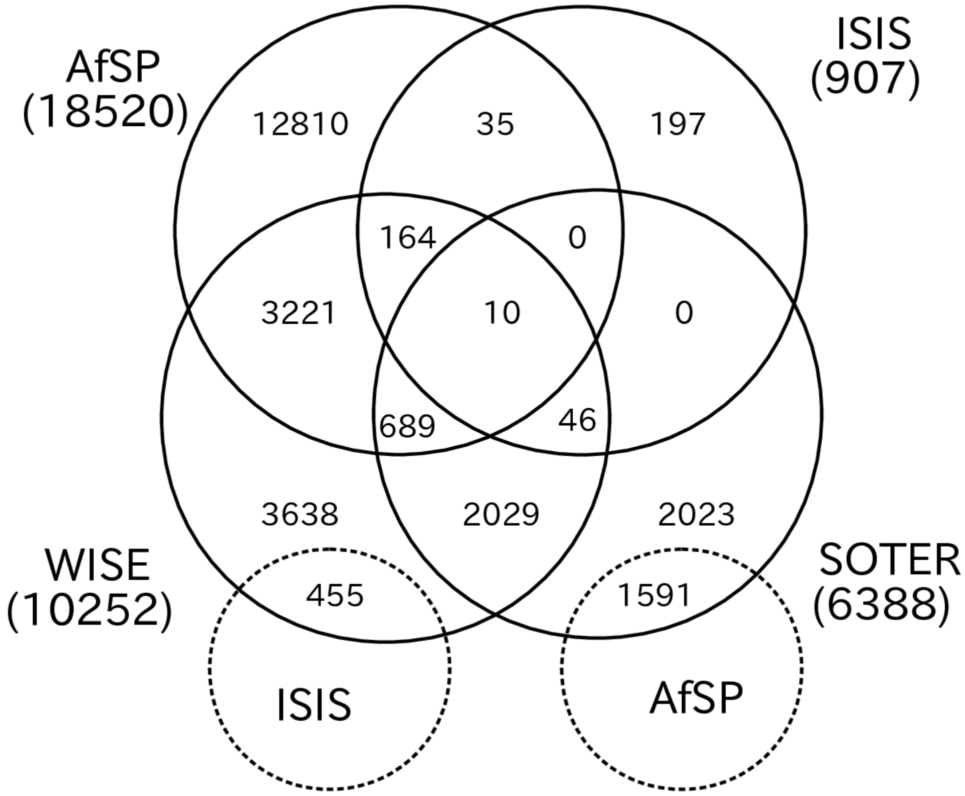


Figure 2.2: Intersection between ISRIC stand-alone profile databases showing the number of overlapping profiles (AfSP, Africa Soil Profile Database; ISIS, ISRIC Soil Information Service, SOTER, Soil and Terrain Database; WISE, World Inventory of Soil Emissions potentials database).

Figure 2.2 illustrates the intersection between four profile databases compiled in the framework of collaborative ISRIC projects: AfSP (Africa Soil Profile Database) (Leenaars et al., 2014a), ISIS (ISRIC Soil Information System)[1], WISE (World Inventory of Soil Emission potentials; (Batjes, 2009)) and various national and continental scale SOTER databases (Soil and Terrain Databases)[2]. Except for ISIS, which holds profile data for the ISRIC World Soil Reference Collection, the other datasets are project-specific compilations from various (possibly overlapping) data sources. As shown in Figure 2.2, 12,810 profiles are exclusively present in AfSP; 35 are shared among AfSP and ISIS; 164 are shared between AfSP, ISIS and WISE; 10 profiles are present in the 4 databases, and so on. In case of duplicate profiles, data for the most complete source data set will be prioritized when serving the standardized-data (see Appendix A).

---

[1] `http://isis.isric.org/`

[2] `http://www.isric.org/projects/soil-and-terrain-soter-database-programme`

2.2 Measures for data quality

2.2.1 Data lineage

As indicated by Chapman (2005), ‘Too often, data are used uncritically without consideration of the error contained within, and this can lead to erroneous results, misleading information, unwise environmental decisions and increased costs’. WoSIS is being populated using datasets produced for different types of studies ranging from routine soil surveys to more specific assessments, each of these having their own specific quality requirements (Landon, 1991; SSDS, 1993). The corresponding samples were analysed in a range of laboratories or in the field according to a range of methods (e.g. wet chemistry or spectroscopy), each with their own uncertainty. As indicated by Kroll (2008), issues of soil data quality are not restricted to uncertainty issues, they also include aspects like completeness and accessibility of data.

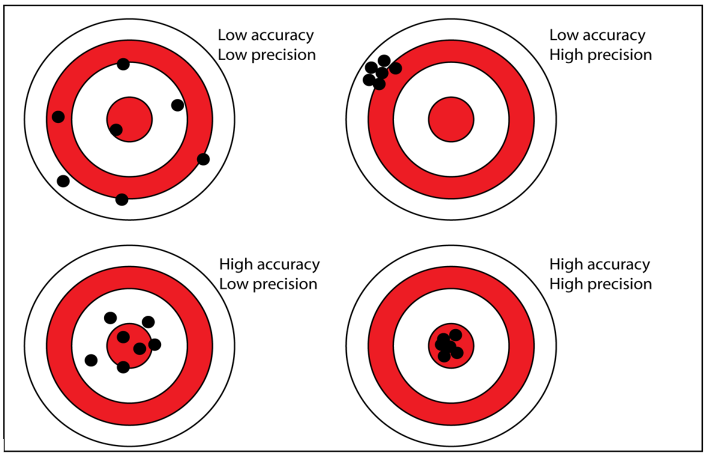


Figure 2.3: Depiction of the accuracy and precision of measurements.

Quality of data can be evaluated against a range of properties, for example positional accuracy, attribute accuracy, logical consistency, completeness and lineage. Underlying these properties are always the two central themes in data quality assessment, the concepts of accuracy and precision. In the case of environmental data, accuracy can be defined as ‘the degree of correctness with which a measurement reflects the true value of the property being assessed’, and precision as ‘the degree of variation in repeated measurements of the same quantity of a property’ (EH&E, 2001). A high degree of precision and accuracy need not occur simultaneously in a process (Figure 2.3), thereby determining attribute uncertainty. When results are both precise and accurate, confidence in data quality is maximized. The desired accuracy and precision, however, will vary with user requirements and scale of application (Leenaars et al., 2014a; Finke, 2006).

Similarly, differences between accuracy and precision in a positional context can be visualized (Figure 2.4) (Chapman, 2005): a red spot shows the true location, a black spot, represents the locations as reported by a collector. For point data, the aspect of positional accuracy, in the context of digital soil mapping, has been discussed in detail with respect to legacy soil profiles collated for the Africa Soil Profile Database (Leenaars et al., 2014a).

To address and document the above issues, three quality indicators are applied throughout the WoSIS database. These are:

- Observation date: date of observation or measurement (*sensu* data lineage),
- Level of Trust, a subjective measure based on soil expert knowledge (column: trust; see Section 2.2.2), and
- Accuracy and precision, the Laboratory/Field/Location related uncertainty (column: accuracy; see Section 2.2.3).

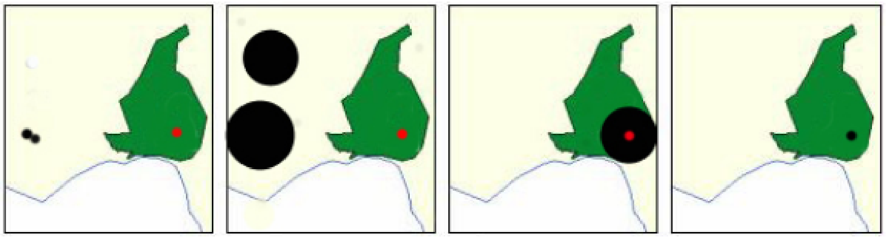


Figure 2.4: Differences between accuracy and precision in a spatial context (From left to right: High precision, low accuracy; Low precision, low accuracy showing random error; Low precision, high accuracy; High precision and high accuracy.

The above indicators were introduced to provide 'flags' that allow investigators to recognize factors that may compromise the quality of certain data, hence their suitability for use. Consideration of all three indicators ensures that objective methods are applied for evaluating data in the database, while at the same time it enables soil expert knowledge to override these assessments when needed. In practice, however, the information provided with the source materials may not allow for a full characterization of all three indicators (Appendix C).

2.2.2 Level of trust

Different attributes managed in WoSIS need to be characterized in terms of inferred trust. The lowest level 'A' is used for data entered 'as is'. Subsequently, such 'A' level data can be standardized to level 'B'; this step considers the soil property, analytical method and unit of measurement. Ultimately, B-level data can be harmonized ('C') to an agreed reference (target) method, subject to international agreement about the recommented target method (Baritz et al., 2014). Step B includes automated error-checking for possible inconsistencies with some visual checks. Level 'C' is the highest achievable degree of harmonization in WoSIS; such values have been approved by a (regional) expert who has performed an in-depth check, ideally considering the value in relation to the full soil profile, and found no apparent anomalies.

2.2.3 Accuracy and precision

The precision and accuracy of results from laboratory measurements can be derived from the random error and systematic error in repeated experiments on reference materials or with reference methods. This information generally is only available in the originating laboratories, as further discussed in Appendix C.

Any given measurement has a specific measurement error, which can be determined using a range of methods. The accuracy of values derived in a laboratory can be characterized using blind samples or based on repeated measurements on reference materials. Any laboratory should be able to provide these parameters according to good laboratory practice (OECD, 1998; van Reeuwijk and Houba, 1998), but in practice this need not be the case. For measurements that use other devices, such as GPS and soil moisture sensors, the accuracy can be obtained from manufacturers, literature and even expert knowledge.

2.3 Main steps towards data harmonization

2.3.1 Data lineage and access conditions

WoSIS aims to facilitate the exchange and use of soil data collated within the context of collaborative efforts at global, regional, national and local level. As indicated, such data have been collected and analyzed using numerous approaches and procedures; typically, these conform to the prevailing national standards. Subsequently, these data have been compiled in databases using specific templates with underlying data models and data conventions. These ‘raw’ data often meet specific goals and are not necessarily meant to contribute to international transboundary studies. Standardization of such data for wider use may imply a loss of appropriateness for originally intended purposes. However, once compiled under a global common standard they importantly gain in appropriateness for use for international purposes.

A priori standardization of data, for the purpose of being shared with the global community as in SoTER and WISE, implies a serious burden for data providers while not necessarily contributing to their direct goals. It often implies a loss of lineage and traceability (Leenaars et al., 2014a). Consequently, data standardization generally occurs a posteriori. Such is preferably done by the data provider who is best able to correctly interpret the data; this would yield a ‘double dataset’ holding both original data as well as standardized data (Leenaars et al., 2014b). Alternatively, data standardization would need to be done by a ‘central compiler’. Therefore, any soil dataset intended for being shared through WoSIS should be sufficiently documented, with adequate metadata, to make the data understandable and usable.

Data providers that submit data for possible inclusion in WoSIS must specify conditions for access to the data they deposit. This may be done using a Creative Commons[3] license or other existing licence. In practice, this information is provided as part of the data lineage (i.e. possible ‘inherited restrictions’). Access conditions for third parties to each dataset managed in WoSIS are enforced through ‘access registers’; overall conditions are in accord with the ISRIC Data and Software Policy[4]. Ultimately, only standardized data derived from ’shared’ sources are provided to the international community (Appendix A).

2.3.2 Data standardization and harmonization

As indicated, the WoSIS database has been designed in such a way that, in principle, any type of soil data can be accommodated irrespective of the data source (with associated data models and data conventions as originally compiled).

Basic principles, and standardized templates, for compiling soil profile datasets for use in WoSIS are given in Appendix B. Adoption thereof will permit to: a) keep track of data sources and identify uniqueness of profile records (through their lineage), and b) describe the full (source) data so that they may be correctly collated into WoSIS.

Main steps for processing data in WoSIS are schematized in Figure 2.5. First, all submitted datasets are imported as is in the ISRIC Data Repository keeping the original data model, naming and coding conventions, abbreviations, domains and so on. Subsequently, these source datasets are converted into PostgreSQL format. Thereafter, these PostgreSQL data sets are mapped to the WoSIS Data Model (Batjes et al., 2017). Basically, this is the first major step of data standardization in WoSIS. The second step of data standardization, applied to the values for the various soil properties as well as to the naming conventions, is applied to make the data queryable and usable. A desired third step, full data harmonization, would involve making similar data comparable, that is as if assessed by a commonly endorsed single reference method (for pH, CEC, organic carbon, etc.). At present, however, these reference (or target) methods still have to be agreed upon by the international soil community (Baritz et al., 2014). Once this has been done, regionally calibrated pedotransfer functions will need to be derived drawing on results from large scale laboratory method inter-comparisons, such as GLOSOLAN[5] (Global Soil Laboratory Network). As indicated by GSM (2013) and others, the necessary pedotransfer functions should be region and soil type specific.

Ultimately, the quality-assessed, standarized/harmonized data are served to the international community, in accord with the licence specified by each data provider (Appendix A).

---

[3] `http://creativecommons.org/licenses/`

[4] `http://www.isric.org/about/data-policy`

[5] `http://www.fao.org/global-soil-partnership/resources/events/detail/en/c/1037455/`

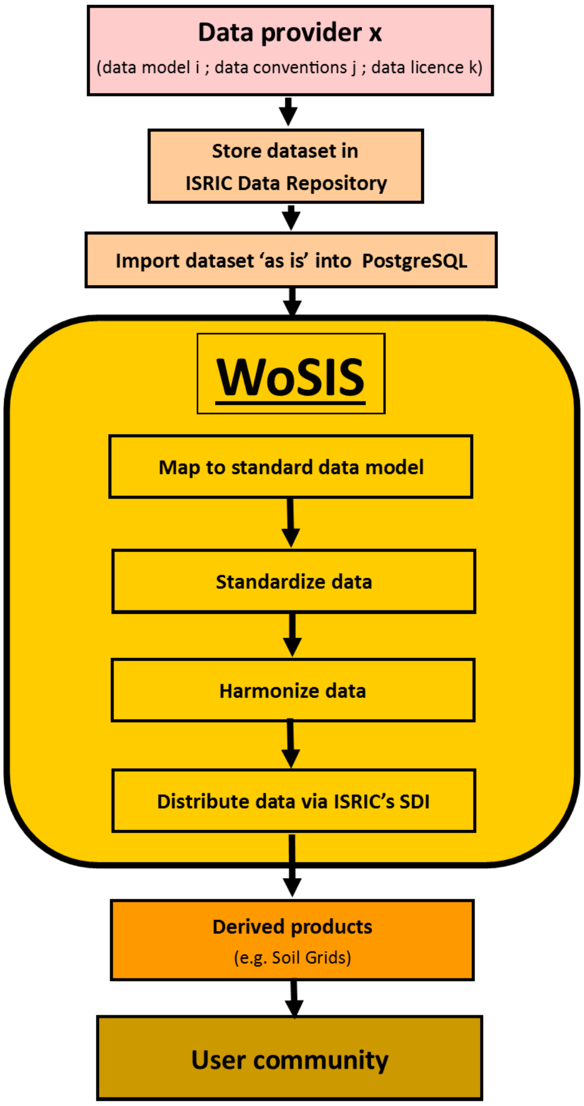


Figure 2.5: Main stages of data standardization and harmonization.

Chapter 3

Database design

3.1 General concept

The database design for WoSIS consists of 46 interrelated tables following a standard relational model implemented in PostgreSQL, a powerful, open source object-relational database system. Each table has an unique identifier (Primary key). Primary key fields are based on the natural key fields such as *dataset_id* or *country_id*, rather than artificial key fields. When this was not possible, artificial keys were used together with a Sequence to automatically generate the next unique value on new data inserts. Foreign keys were created to build the data model and enforce data referential integrity. In other words, Foreign keys establish links between tables and define the way they behave (e.g. ON DELETE CASCADE / RESTRICT / NO ACTION / SET NULL / SET DEFAULT). Other constraints, such as Check, Not-Null or Unique, were implemented when necessary in accordance with certain attributes properties. Functions and Triggers were created to facilitate management of the database, for instance to batch rename all the Primary keys according to a certain rule or to facilitate the import of data into the database. Further, Views and Materialized Views were generated to output results.

Objects in WoSIS are named using the following set of rules:

Common rules

- lower-case characters
- separate words and prefixes with underlines (snake_case)
- no numbers
- no symbols
- no diacritical marks
- short descriptive names (example: profile_layer)
- the name of the object should indicate what data it contains (example: reference_author)

Table names

- singular names
- avoid abbreviated, concatenated, or acronym-based names
- use same prefix for related tables

Column names

- singular names
- the primary key column is formed by the table name suffixed with '_id'

- foreign key columns have the same name as the primary key to which they refer
- in views, all column names derived directly from tables stay the same

For the rest of the objects, default PostgreSQL names are used:

- Primary key: <table_name>_pkey
- Sequence: <table_name> _<column_name>_seq
- Foreign key: <table_name> _<column_name>_fkey
- Index: <table_name> _<column_name>_idx
- Check: <table_name> _<column_name>_check
- Views: vw_<view_name>
- Function: fun_<descriptive_name>
- Trigger: trg_<table_name> _<column_name>

WoSIS uses one single database schema to logically group objects such as tables, views, triggers and so on. Different schemas are used for other purposes, such as Database Auditing and Web Applications, to enforce role grant access and use different tablespaces. This will allow other systems, such as GeoNetwork, to operate using the same database.

In order to enhance legibility, tables have been grouped according to their functions to better show relationships (see different colours in Figure 3.1):

- Metadata
- Soil classification
- Attribute definition
- Reference
- Profile data
- Map unit data
- Raster data

The individual 'components' are described in detail in Sections 3.2.1 to 3.2.7, while the structure of each table is documented in Appendix G.

By convention, in the text, table names appear in **bold** and column names in *italic*.

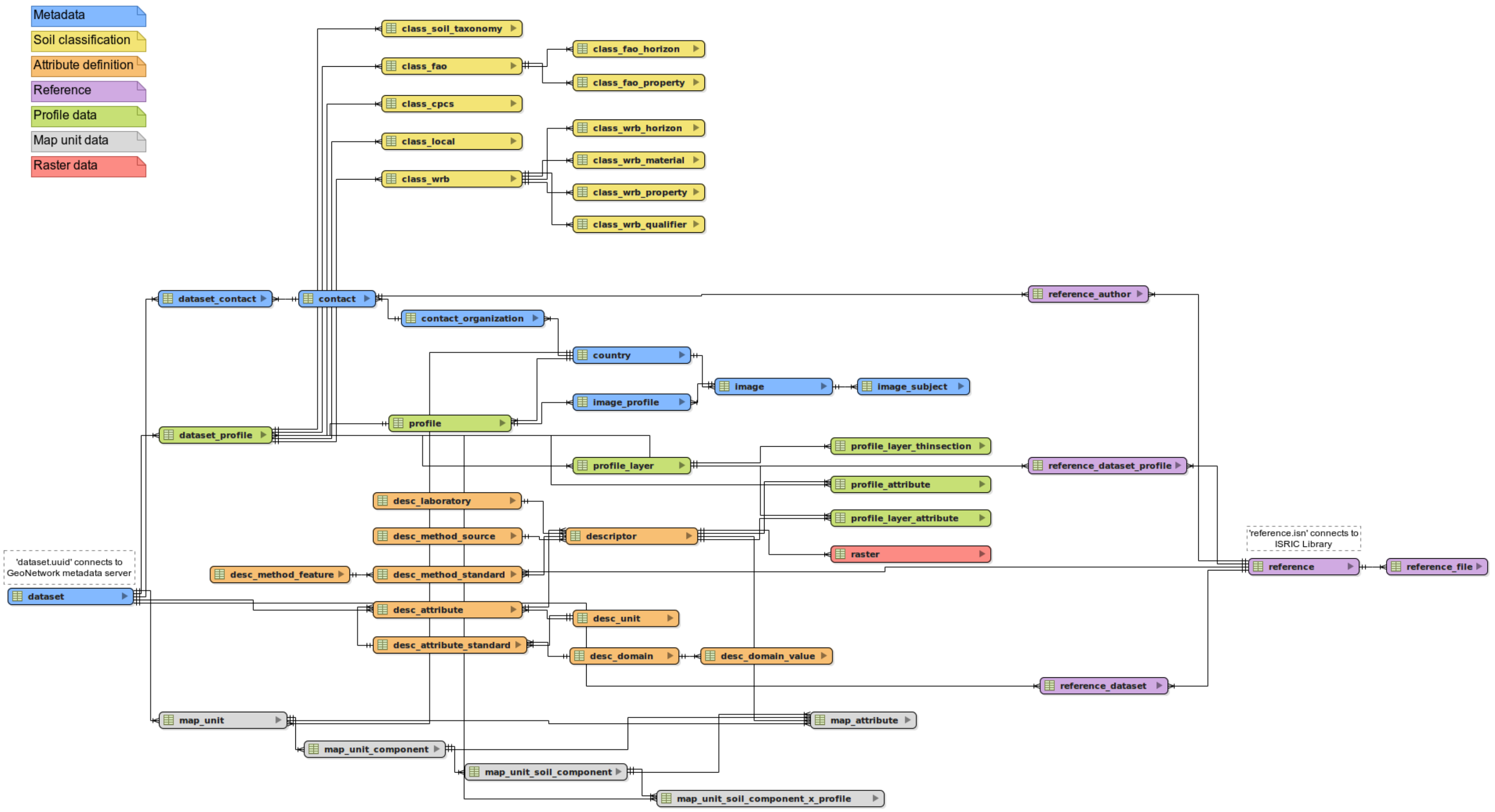


Figure 3.1: Main reference groups and components of the WoSIS database model.

3.2 Main components

3.2.1 Metadata

Metadata are data that define and describe other data. They also define the terms and conditions for use of the data and ensure that all data can be properly attributed and cited (Figure 3.2). GeoNetwork, a catalogue application for spatially referenced resources[1], provides powerful metadata editing and search functions as well as an embedded interactive web map viewer. It is based on the principles of Free and Open Source Software (FOSS) and International and Open Standards for services and protocols (a.o. from ISO/TC211 and OGC). GeoNetwork, is interoperable with standards used by the ICSU World Data System [2] and GEOSS (Global Earth Observation System of Systems) portal [3].

WoSIS links its **dataset** table with ISRIC's Geonetwork 3.0 database[4].

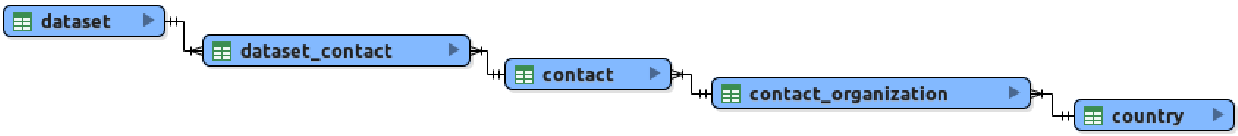


Figure 3.2: Metadata reference group tables.

The **dataset** table stays at the top of the hierarchy of the database defining where the data come from. Most importantly, the **dataset** table is used to manage and enforce the access rights, for example whether the associated data may be shared freely with the general public or not; conditions for this are specified by each data provider in accordance with the ISRIC Data Policy[5]. The **dataset** table also serves to make a bridge, as mentioned before, to the Geonetwork database where detailed metadata is stored.

Table **dataset_contact** is an intermediary table that enables a dataset have more than one contact assigned. In tables **contact** and **contact_organization** describe organizations and/or persons that have been instrumental in collating or providing the observation results (either descriptive or measured) that are stored in the database. It is the single entry point to authoritative names and contact information in the overall database. This is to prevent the use of different names or spellings for the same organization or individual in various parts of the database (e.g., KIT, Tropen Instituut, Royal Tropical Institute, Koninklijk Instituut voor de Tropen).

The **contact_organization** table may also store organization components like departments and regional centres. A contact field stores the contact information for a real person. A contact can be linked to one or more organization - in the sense of a 'collaborated with' or 'is employed by' relationship. The **contact_organization** table links to the **country** table which defines codes for the names of countries, dependent territories and special areas of geographical interest based on ISO 3166 and their geometry from the Global Administrative Units Layer (GAUL)[6], release 2015, a spatial database of the world administrative areas (or administrative boundaries). GAUL describes where these administrative areas are located (the 'spatial features'), and for each area it provides attributes such as the name and variant names.

3.2.2 Soil classification

Soil classification involves the systematic categorization of soils based on distinguishing characteristics as well as criteria that dictate choices in use. It is probably one of the most controversial, and debated, soil science subjects. Many countries have therefore developed their own classification systems (FAO, 2015); international correlation of the various systems is being addressed by the World Reference

---

[1] `http://geonetwork-opensource.org/`

[2] WDC-Soils, see `https://www.icsu-wds.org/services/data-portal`

[3] GEOSS portal, see `http://www.geoportal.org/`

[4] `http://data.isric.org/`

[5] `http://www.isric.org/data/data-policy`

[6] See `http://www.fao.org/geonetwork/srv/en/metadata.show?id=12691&currTab=simple`

Base for Soil Resources (IUSS WG-WRB, 2015) and earlier through the FAO-Unesco Soil Map of the World (FAO-Unesco, 1974; FAO, 1988).

The classification tables (Figure 3.3) in WoSIS support three widely used soil classification systems:

- FAO Soil Map of the World: This system was originally intended as legend for the Soil Map of the World, at a scale of 1:5M, but in the course of time it has been used increasingly as a classification system (FAO-Unesco, 1974; FAO, 1988); the FAO system has now been subsumed into the WRB (tables **class_fao**, **class_fao_horizon** and **class_fao_property**).
- World Reference Base for Soil Resources: the international, standard soil classification system endorsed by the International Union of Soil Sciences (IUSS WG-WRB, 2015), and earlier versions as indicated by the year of publication (tables **class_wrb**, **class_wrb_horizon**, **class_wrb_material**, **class_wrb_property** and **class_wrb_qualifier**).
- USDA Soil Taxonomy (SSDS, 2010), and earlier approximations as indicated by the year of publication (table **class_soil_taxonomy**).

In addition to the above, the national or local classification can be stored in table **class_local**. Also, having been used extensively in Western Africa, the French soil classification (CPCS) can be specified in a specific table (**class_cpcs**). In the future, once fully developed, a table for the Universal Soil Classification (Michéli et al., 2016) may be accommodated in WoSIS.

Sometimes, as discussed earlier, the same soil profile may have been considered/processed in different ISRIC and international datasets. As a result, the same profile may have been classified/correlated differently in each source dataset, based on the same soil classification system, depending on the classifier's perspectives (Kauffman, 1987). Therefore, WoSIS contains a link **dataset_profile** table to assign profiles (**profile** table) to specific source datasets (**dataset** table). All classifications refer to an entry in the **dataset_profile** link table (that is, a profile in a particular dataset), thus enabling one classification per profile and dataset.

In some cases, the USDA Soil Taxonomy coding is inconsistent between editions as different standard notations have been used in successive versions (SSDS, 1975, 1992, 1998, 2003, 2010); examples are given elsewhere (Spaargaren and Batjes, 1995). Alternatively, the original (FAO-Unesco, 1974) and revised Legend (FAO, 1988) to the FAO Soil Map of the World use a well-established coding scheme. Similarly, there is now an agreed coding scheme for the WRB Legend (IUSS WG-WRB, 2015). Nonetheless, to avoid any ambiguity in soil classification names, for any soil classification system full descriptive names are stored in the database together with the edition (year) of the classification system.

Soil classifications in WoSIS are given as they were in the source database; soil names were only checked for spelling errors. Similarly, horizon or layer designations are given 'as is', but cleaned. Harmonization, for example to the FAO (2016) nomenclature, is considered the reponsibility of the individual data providers. This in view of the large differences in conceptual approaches and coding systems used internationally, and their versioning (Bridges, 1993).

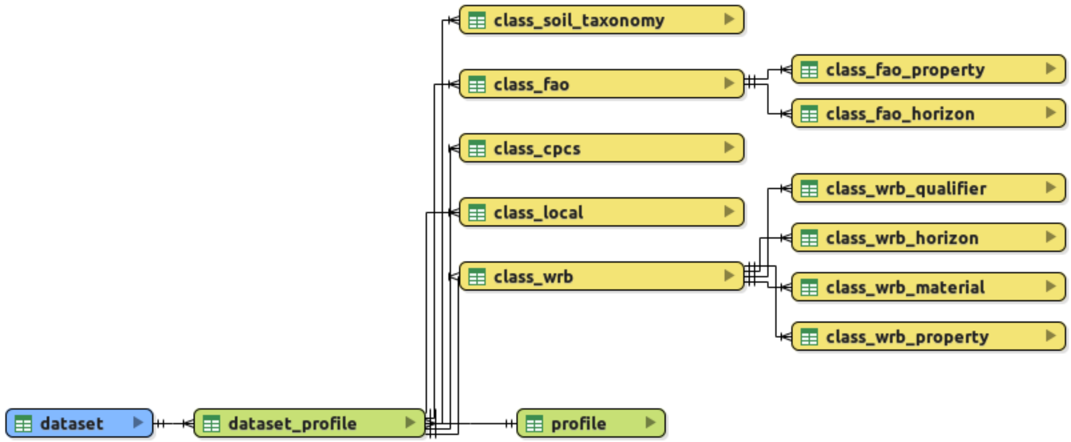


Figure 3.3: Classification tables.

3.2.3 Attribute definition

Each dataset (as described in **dataset** table) comes with a list of attributes (or parameters or properties) in order to express a description or measurement. These source attributes are described in the table **desc_attribute** (Figure 3.4). The naming or coding of the source attributes need to be standardized to permit querying for a certain attribute across the entire database with multiple (source) datasets. For example, the following terms are used to describe soil organic carbon content in various source databases: organic carbon, carbone organique, organischer Kohlenstoff, and carbono orgânico. Standard attributes are described in table **desc_attribute_standard** with basic information about their data type, unit and domain.

For each attribute (**desc_attribute**), the definition, analytical methods and source laboratory must be defined explicitely to allow for standardization and ultimately full harmonization. Soil analytical methods, and a description of their main characteristics, however, is a complex topic as many of these analyses are soil type specific (SSDS, 2011; van Reeuwijk, 2002). Analytical methods are often poorly defined in the source materials. Alternatively, the same analytical method may have been described in various ways. To preserve the lineage, the analytical method descriptions, as defined in the source materials, are preserved 'as is' in table **desc_method_source**, for example 'Exchangeable Potassium - Neutral Salt (meq/100g)'.

Each standard method, as encountered during data compilation, is described in table **desc_method_standard** using a defined number of standardized options, as documented in table **desc_method_feature**. Further, details about the laboratory where the measurement took place are stored in table **desc_laboratory**. The next table in the data model is **descriptor** in which the attribute, analytical method and laboratory id's are combined into a new, unique id (*descriptor_id*). The *descriptor_id* is later used in tables such as **profile_attribute**, **profile_layer_attribute**, **map_attribute** and **map_raster** in which the measured respectively description values are stored.

According to their nature, data are stored in a specific table:

- Profile (point 2D): **profile_attribute**
- Layer (point 3D): **profile_layer_attribute**
- Map-unit (polygon): **map_attribute**
- Matrices (pixel): **raster**

Recognizing the broad scope of the domain of knowledge that can be accommodated in the WoSIS database, every effort was made to be as accurate as possible in the definition of the entities of interest as well as their characteristics.

In data management and database analysis, a data domain refers to all unique values that a given data element may contain. The rule for determining the domain boundary can be as simple as a data type with an enumerated list of values. For example, a table about soil drainage may contain one record per spatial soil feature. The observed 'drainage class' may be declared as a string data type, and allowed to have one of seven known code values: V, P, I, M, W, S, E for very poorly drained, poorly drained and so on in compliance with (FAO, 2006a) conventions. The data domain for 'drainage class' then is: V, P, I, M, W, S, E. Alternatively, other datasets with information about soil drainage may employ other code values (e.g. '0' for very poorly drained, '1' for poorly drained, ...) for the same 'drainage' phenomenon. Since the database should allow users to enter data in their primary form - that is, in principle, users should not be burdened with conversion issues upon entering or submitting (their) data - a mechanism is needed to link a phenomenon to more than one data domain. This mechanism is in the **desc_domain** table which essentially links an attribute to a data domain in **desc_domain_value**. Our 'soil drainage' example would require one record to link 'soil drainage' to the corresponding data domains in
desc_domain_value. Conversely, a data domain may be used to describe more than one characteristic. For example, in the FAO Guidelines for Soil Description (FAO, 2006a), several surface characteristics are defined using the same surface coverage classes, ergo the same data domain.

Since a data domain may be referenced by more than one characteristic, the relationship between the **desc_domain_values** table and the **desc_attribute** table would be of a many-to-many nature. To circumvent such many-to-many relationships in the database, a **desc_domain** table was added between the table with **desc_attribute** and the **desc_domain_values** table (Figure 3.4).

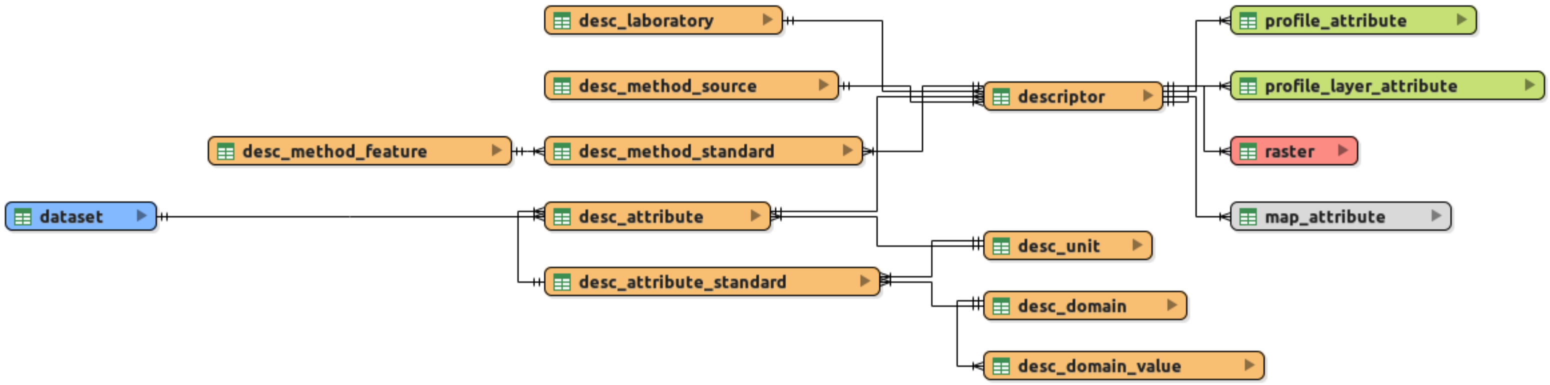


Figure 3.4: Attribute reference group tables.

3.2.4 Source materials

Data, definitions, and descriptions may be drawn from a variety of data and information sources. Potential sources include publications, grey literature, maps, web sites (URL's) and digital media. These sources vary widely in their nature and in the way they are described. Table **reference** provides a harmonized structure to refer to these heterogeneous sources. It also allows for the description of the following types of information sources:

- Publications and grey literature
- Web site (URL)
- Map
- Digital media (CD-ROM, DVD, etc.)

The Reference group consists of five tables. The main table, **reference**, describes the full reference of the source materials. When available, it is linked to the actual document in the ISRIC World Soil Library through the unique code in column *isn*.

In WoSIS, three entities may have a reference: a dataset (**dataset** table), a profile (**dataset_profile** table), and an analytical method (**desc_method_standard** table). The **desc_method_standard** table has a 1:1 relation with the **reference** table (Figure 3.5). It is mainly used for documenting the laboratory manual procedure for a specific method. Dataset and profile have an intermediate table so that a profile or a dataset may have more than one reference (e.g. dataset, journal article, reports and maps), which permits to reconstruct the full lineage to the original data source. Through table **reference_author**, an author is linked to a given reference. A reference can have multiple authors, therefore these names are stored in a separate table **reference_author**. The same applies for table **reference_file**.

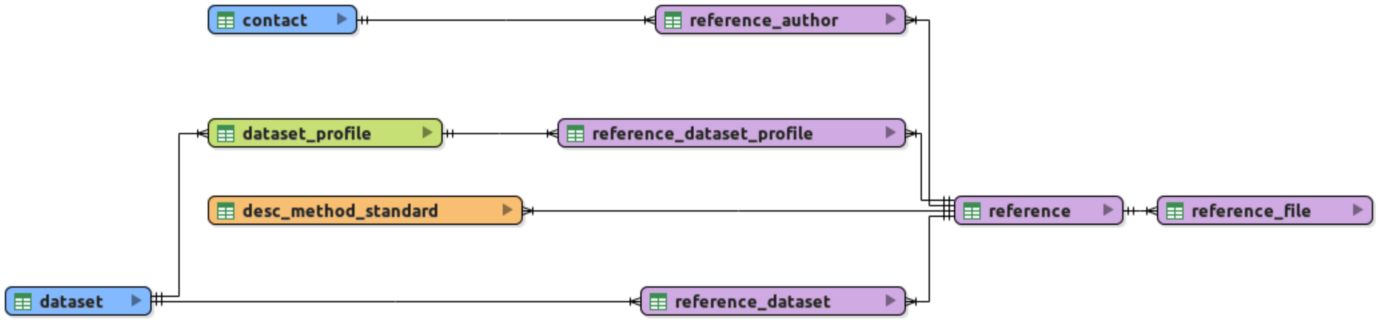


Figure 3.5: Reference data group tables.

3.2.5 Profile data

Tables in the profile data group (Figure 3.6) describe two basic entities from the domain of discourse underlying the database: a soil profile (pedon) and its properties or attributes (e.g. land use, position in the terrain, signs of erosion, and drainage), as well as its constituent layers with their respective attributes or properties (e.g. horizon designation, structure, colour, texture, and pH).

Table **profile** holds unique soil profiles along with their geometry (x,y). The coordinates are stored in column *geom*, in binary mode, using the PostGIS[7] spatial extension for PostgreSQL. The default coordinate system used in WoSIS is WGS84, EPSG code 4326. The accuracy of the profile coordinates is stored in column *geom_accuracy* in decimal degrees. Further, the country in which a profile is located is registered using the 2 character ISO code (e.g. BE for Belgium) in column *country_id*.

Each soil profile in WoSIS is given a specific integer ID as well as a UUID[8]: for example, profile id 50000 corresponds with UUID of b7b86368-b8f2-11e4- 90de-8851fb5b4e87'. The UUID is automatically generated when a record is inserted into WoSIS. UUID's allow for easy profile identification in diverse computer systems like harvesting environment, web services, broadcasting in social networks (e.g.Twitter and Facebook), or integration with GeoNetwork.

As indicated, some profiles are represented in more than one (source) dataset, together with their respective soil property values. In order to preserve the original soil properties and soil property values from the different source datasets, the tables (**profile_attribute** and **profile_layer_attribute**) containing the measured values link to table **dataset_profile**. Figure 3.6 shows that the **dataset_profile** table forms the node or the backbone of the database as it represents the inventory of soil profiles and soil profile source datasets. All tables that link to **dataset_profile** always have a foreign key formed by *dataset_id* and *profile_id*.

Table **profile_attribute** is used to manage the properties about the profile and profile‘s site, including drainage, terrain, vegetation, land use, and climate. In order to store the soil‘s properties for a given layer, this layer has to be defined first in table **profile_layer**. This table stores information about the upper and lower depths of the layer (and horizon), measured from the surface, including organic layers (O)[9] and mineral covers, downwards in accord with current conventions (FAO, 2006b; SSDS, 2012), together with the corresponding soil samples and dataset.

Table **profile_layer** links to table **profile_layer_attribute** in which the chemical, physical, morphological and biological soil properties of a layer are recorded, such as structure, colour, texture and pH. Soil properties are defined in table **descriptor**, as explained in Section 3.2.3.

---

[7] PostGIS is an open source software programme that adds support for geographic objects to PostgreSQL - `https://en.wikipedia.org/wiki/PostGIS`.

[8] Universally unique identifier, `https://en.wikipedia.org/wiki/Universally_unique_identifier`

[9] Prior to 1993, the begin (zero datum) of the profile was set at the top of the mineral surface (the solum proper), except for ‘thick' organic layers as defined for peat soils (FAO, 1977; FAO and ISRIC, 1986). Organic horizons were recorded as above and mineral horizons recorded as below, relative to the mineral surface (SSDS, 2012) p. 2-6).

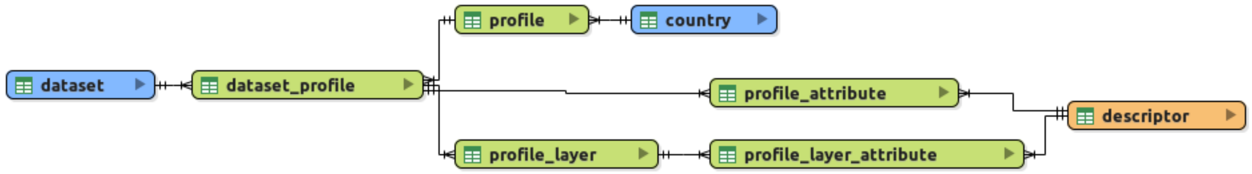


Figure 3.6: Profile data group tables.

3.2.6 Map unit (polygon) data

'Traditional soil surveys describe kinds of soils that occur in an area in terms of their location on the landscape, profile characteristics (classification), relationships to one another, suitability for various uses and needs for particular types of management' (SSDS, 1983). Soils are grouped into map units for display purposes. A soil map unit is a conceptual group of one-to-many delineations. It is defined by the same name in a soil survey that represents similar landscape areas in terms of their components soils plus inclusions or miscellaneous areas (SSDS, 1983). For example, in the SOTER methodology, which has been used by ISRIC, FAO and their partners to develop a range of continental and national scale databases (van Engelen, 2011; Omuto et al., 2012), a map unit identifies areas of land with a distinctive, often repetitive, pattern of landform, lithology, surface form, slope, parent material, and soil types (van Engelen and Dijkshoorn, 2013). Tracts of land demarcated in this manner are named SOTER (map) units; again, each map unit may consist of one or more individual areas or polygons on the map.

In WoSIS, each map unit is stored as a single or multi polygon geometry. All polygon maps, and therefore their mapping units, are stored in a single table called **map_unit**. The *map_unit_id* identifies each individual map unit within the table. Hence, every map unit must refer to a dataset as uniquely defined in the **dataset** table. As indicated, in WoSIS, the reference datum for any point or polygon on the Earth's surface is WGS84[10].

In WoSIS, information about a map unit, its component soils and their attributes and their values is stored in a separate table called **map_attribute** (Figure 3.7). The id's of the profiles associated with each component soil are listed in **map_unit_soil_component_x_profile**.

---

[10] World Geodetic System 1984 (WGS84), `http://en.wikipedia.org/wiki/World_Geodetic_System`.

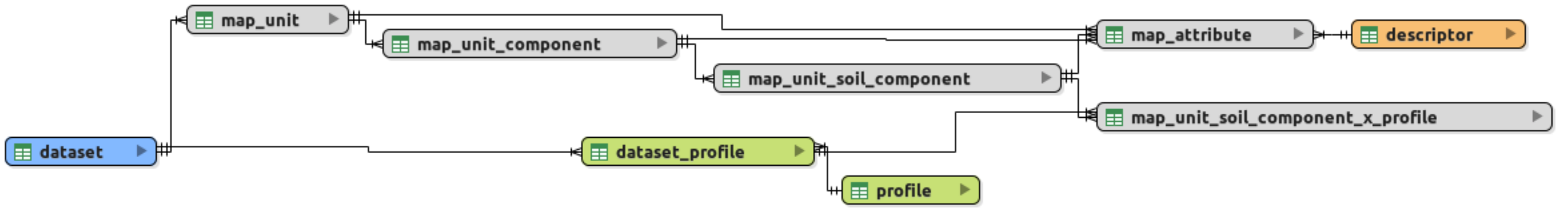


Figure 3.7: Map unit tables.

3.2.7 Raster data

With PostGIS 2.x raster data can be stored in PostgreSQL. In WoSIS this functionality is used to accommodate raster data such as those produced by SoilGrids (Hengl et al., 2014, 2015), the Africa Soil Information System (AfSIS), GlobalSoilMap (Arrouays et al., 2014), and other projects. The source of the product can be described in table **dataset**, while **desc_attribute** describes the raster image or the attribute they represent (Figure 3.8). **Descriptor** is an intermediate table which defines the attribute plus the method and source laboratory using one single id (*descriptor_id*). All raster databases are registered in table **raster**.

Unlike for polygon data, raster data are only registered in WoSIS. The actual (generally large) raster files are kept in the original file system for ease of handling (thus not imported). For consistency, the default coordinate reference system for raster data is also WGS 84 (EPSG 4326)[11] and the preferred format GeoTiff. Together with the Geospatial Data Abstraction Library (GDAL)[12] tools, a broad range of processing procedures can be applied, for example warping, tiling, and compression before registration in the database.

Figure 3.8: Raster tables.

---

[11] `http://spatialreference.org/ref/epsg/4326/`.

[12] Geospatial Data Abstraction Library (GDAL), `http://www.gdal.org/`.

Chapter 4

Interoperability and web services

An overview of procedures and standards in use at ISRIC WDC-Soils is presented in Batjes (2017). We use the term web services[1] to describe a standardized way of integrating web-based applications using an agreed-upon format for transmitting data between different devices. Various protocols can be used for this: XML (Extensible Markup Language), SOAP (Simple Object Access Protocol), WSDL (Web Services Description Language), and UDDI (Universal Description, Discovery, and Integration) open standards.

XML serves to tag the data, SOAP serves to transfer the data, WSDL permits to describe the services available and UDDI is used for listing what services are available. Web services are used mainly as a means for ISRIC to communicate with other organisations and with clients. The web services, which form an important part of ISRIC's Spatial Data Infrastructure (SDI), allow several organizations to exchange and communicate data without having detailed knowledge of each other's IT systems behind their respective firewalls.

Interoperability of the data exchanged or processed by the web services is achieved through a priori standardization of the data themselves (see Section 2.3); the latter is done according to agreed upon data conventions that express the (soil) data in a (machine) understandable 'soil-vocabulary'. Multiple soil data types and sources can be managed in WoSIS. For this, the original soil data have first to be modelled into the WoSIS database structure respecting its schema, tables and relationships as described elsewhere in the present Procedures Manual.

Standardized data (i.e. known modelled data) are of extreme importance here since, for SoilML, web services have to translate the database data model into a simplified data model that is more compatible with web communication. A Web Feature Service (WFS) is implemented using Geoserver that connects to the WoSIS PostgreSQL system, reading its views and tables. OGC's (Open Geospatial Consortium)[2] WFS standard provides an interface allowing requests for geographical features across the web, using platform-independent calls. The client's web services are totally independent from WoSIS, as these clients are located in a very broad range of platforms, from mobile phones to GIS software.

The approach of using OGC web services and model data in XML is necessary for fulfilment of INSPIRE requirements (GSSoil, 2008; INSPIRE, 2015). The output of the data can be customized between different XML standards using Extensible Stylesheet Language (XSL) templates or using server schema mapping. For example, converting generic GML (Geographic Markup Language) into soilML (Soil Markup Language) or to INSPIRE compliant XML describing soil profiles. As yet, there is no common standards for this; recent developments are discussed elsewhere (Mendes de Jesus et al., 2017; Wilson, 2016).

Data transfer between the providing web-service and client operates both ways. For example, the client first calls the web-service provider with a specific request after which the request is processed and the response provided to the client. The request objectives can be: a) Determine capabilities of the providing service, b) Get data based on query, and c) Submit data from the client into the provider (here, the client itself becomes a provider).

---

[1] `https://en.wikipedia.org/wiki/Web_service`.

[2] `http://www.opengeospatial.org/standards/common`.

Figure 4.1 illustrates how soil layers (point, polygon and grid), managed at ISRIC, can be provided to the client. Metadata for these layers can be accessed through the ISRIC Soil Data Hub[3] using a GeoNetwork instance; this facility provides a central location for searching and downloading soil data layers from around the world. As indicated, soil layers are also accessible via a Web Feature Service (WFS), implemented using Geoserver, which connects to WoSIS reading its views and tables. Further, a Representational State Transfer (REST) service is available that permits download/streaming for the web-service, querying based on coordinates (latitude (X) and longitude (Y)) as provided by the client, as well as linkage with mobile phone applications such as SoilInfoApp[4]. An ongoing development at ISRIC is to allow a clients web-service to become a data provider to WoSIS, for example in the context of anticipated crowd-sourced projects (Hobley et al., 2017).

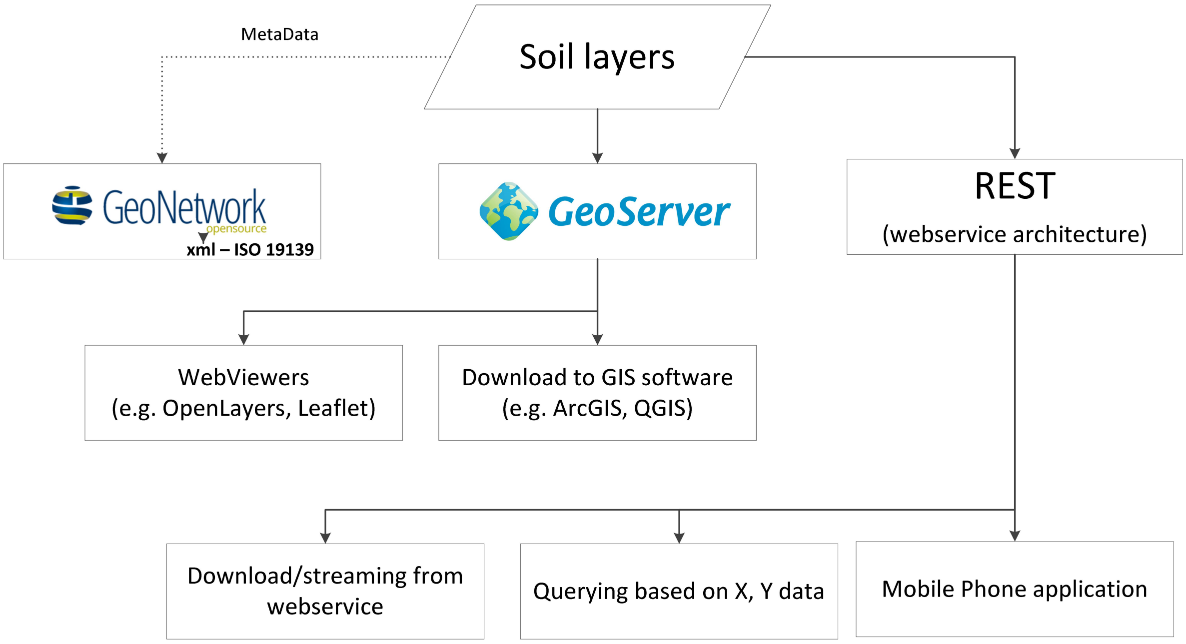


Figure 4.1: Serving soil layers from WoSIS to the user community.

In 2017, the Plenary Assembly of the Global Soil Partnership (GSP) elected ISRIC - World Soil Information as the institution to host the Soil Data Facility (SDF) that will be built in Pillar 4 of the partnership. This means that ISRIC, as member of the Pillar 4 Working Group ('Enhance the quantity and quality of soil data and information'), will: a) contribute to the design of the Global Soil Information System, b) participate in capacity building programmes, and c) provide a system that integrates the national facilities into a global soil information system.

The de-centralized information system envisioned by the GSP will rely on multiple network sources; this is unlike WoSIS which is set up as a centralized database to which clients may provide (part of) their data for further standardisation and harmonisation.

Various components of ISRIC's own SDI may be used to develop modules for a system of distributed inter-operable national systems, using a bottom-up-approach, as envisaged for the Soil Data Facility (SDF) of the Global Soil Partnership (GSP, 2016).

Once implemented at a satisfactory level of detail and authority, 'shared' data collated through GSP's SDF may also be considered for processing according to the WoSIS work stream.

---

[3] `http://www.isric.org/explore/isric-soil-data-hub`.

[4] `http://www.isric.org/explore/soilinfo`.

Chapter 5

Federated databases

Worldwide there are many organizations with valuable soil data (Omuto et al., 2012; Arrouays et al., 2017). Yet, these data are accomodated in different databases using a range of data models and conventions (or may be available in paper format only). Merging all these different sources into a common, inter-operable system (Figure 5.1) is a daunting task.

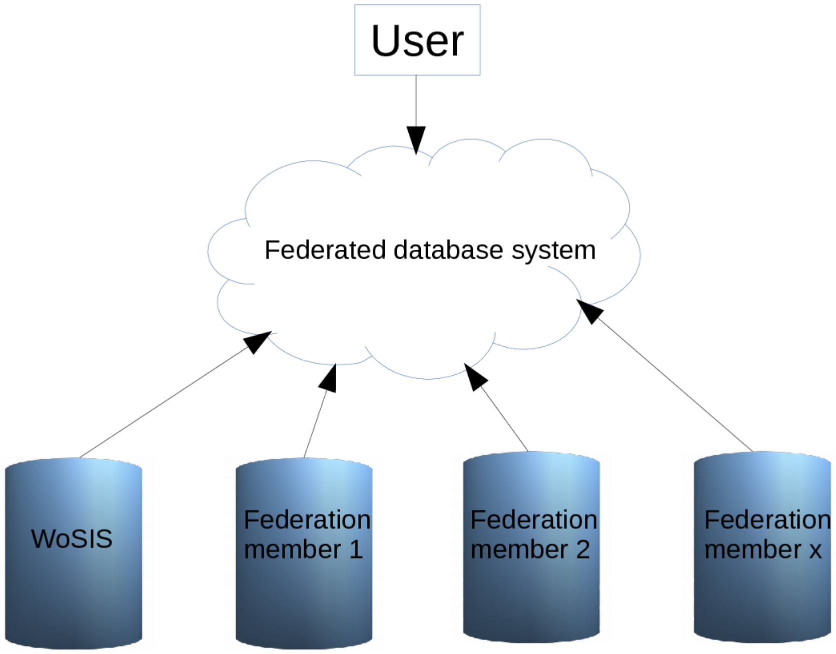


Figure 5.1: Federated databases.

A federated database, also called a virtual database, is a way to view and query several databases as if they formed a single entity. The constituent databases are interconnected and often geographically decentralized. As such, there is no actual data integration in the constituent databases; the respective servers are managed independently, yet cooperate to process requests on the database (Wikipedia, 2017).

Through data abstraction, federated database systems can provide a uniform user interface, enabling users and clients to store and retrieve data from multiple non-contiguous databases with a single query even if the constituent databases are heterogeneous. Because various database management systems employ different query languages, federated database systems can apply wrappers to the subqueries to translate them into the appropriate query languages.

PostgreSQL introduced the Foreign Data Wrapper (FDW) feature for accessing external data at the query level in 2003 (version 9.1). There are now a variety of FDW available which enable PostgreSQL

Server to interact with different remote data stores, ranging from other SQL databases through to flat files.

At ISRIC, we strongly encourage organizations with soil data, that want to maintain their databases autonomously, to use e.g. FDW technology to connect to the WoSIS database and to join efforts to build a federated system of soil databases in order to better serve the global soil scientific community. Developing a federated system, using a bottom-up-approach, is an important goal of the Global Soil Partnership and its emerging Soil Data Facility (GSP, 2016).

Chapter 6

Future developments

At the time of writing the ‘WoSIS latest’ dataset comprised standardized analytical data for some 95,000 globally distributed soil profiles. Inherently, there are various gaps (e.g. geographic, taxonomic, soil properties) in the derived dataset as not all soil properties were measured routinely in the underpinning ’shared’ databases. Further, the number of observations generally decreases with depth.

In view of its global scope, and ISRIC‘s role as regular member of the ICSU World Data System, WoSIS will always remain ‘work in progress’ as new source datasets become available and web technologies evolve.

For the future, the following activities will be considered (as realistic within the allocated project time):

- Expand the number of soil properties for which standardized soil analytical method descriptions are developed, in first instance working towards the list of soil properties considered in recent WISE-derived soil property databases[1].
- Development of procedures for handling soil profile data derived from proximal sensing.
- Processing of ‘new’ soil profile datasets into WoSIS when such are shared by new data providers (in principle in order of receipt of the various datasets/permissions, with priority for ‘fully shared’ datasets from so far under-represented regions).
- Add map unit based soil datasets to WoSIS, starting with those derived from ISRIC-related projects.
- Develop a facility to upload ‘raw’ soil data into WoSIS, and ideally standardize/harmonize on the fly, in accord with SoilInfoApp[2] developments.
- Regularly udate and expand the documentation for WoSIS and make it available online as PDF’s with clear time stamps (DOI’s).
- Besides the dynamic WFS version, publish static versions of WoSIS on a 1-2 yearly basis (after incorportation of say $>$100,000 new soil profiles).
- Further develop and implement procedures towards a federated system, in close collaboration with our international partners.

Capacity building and collaboration with (inter)national soil institutes around the world on data collection and sharing, data screening, standardization and harmonization as well as mapping and the subsequent dissemination of the derived information will be essential to create ownership of the newly standardized (or ultimately harmonized) soil data as well as to create the necessary expertise and capacity to further develop and test the system worldwide. The ultimate goal is to arrive at a federated system of inter-operable databases that can interact with WoSIS, where national data providers maintain and update their own data.

---

[1] `http://dx.doi.org/10.1016/j.geoderma.2016.01.034`

[2] `http://soilinfo.isric.org/`

Acknowledgements

The authors would like to thank the following experts that have contributed in various ways to the development of WoSIS since its conceptualization in 2010: Piet Tempel (2009-2012), Koos Dijkshoorn (2010-2012), Vincent van Engelen (2010-2012), Bob MacMillan (2010-2013), Hannes I. Reuter (2011-2013), Daniel van Kraalingen (2010-2012), Rob Lokers (2010-2012), and Jorge Jesus de Mendes (2010-2016); ISRIC Directors Prem Bindraban (2009-2013), Hein van Holsteijn (2013) and Rik van den Bosch (2014-present) were instrumental in these developments. Alessandro Samuel-Rosa, Bas Kempen, Ingrid Haas, Marcos Angelini, Thomas Caspari, Tom Hengl, Piet van Reeuwijk and Johan Leenaars provided advice on specific WoSIS-related topics for which they are gratefully thanked.

Importantly, our special thanks go to the steadily growing number of data providers that have contributed data for consideration in WoSIS. So far these include [3,4]:

- Africa Soil Profiles Database (`http://www.isric.org/projects/africa-soil-profiles-database-afsp`).
- Belgium, historic soil profiles Flanders (Aardewerk-Vlaanderen-2010, Vlaamse overheid, Departement LNE, dienst Land en Bodembescherming), see: `http://www.geopunt.be/catalogus/datasetfolder/78e15dd4-8070-4220-afac-258ea040fb30`
- Bhutan, National Soil Services Centre (NSSC), Ministry of Agriculture & Forests, Thimphu. `http://www.nssc.gov.bt/`
- Brazilian national soil profile data (compiled by Cooper et al. 2005, ESALQ-USP). `https://dl.sciencesocieties.org/publications/sssaj/pdfs/69/3/0649` and `http://www.esalq.usp.br/`
- Brazilian-Amazon: Soil properties database of the Amazon part from the RADAMBRASIL project (compiled by Bernoux et al. 2003, CENA-USP; data shared in the framework of the GEFSOC project). `http://www.cena.usp.br/`
- Canadian Soil Information Service (CanSIS), Agriculture and Agri-Food Canada, Government of Canada. // `http://wms1.agr.gc.ca/xgeng/pedon.zip` and `http://sis.agr.gc.ca/cansis/`
- Food and Agriculture Organization of the United Nations (FAO): various inputs through SOTER and WISE-related projects (see: `http://www.fao.org/3/a-i3161e.pdf`).
- Instituto Nacional de Estadistica y Geografi (INEGI), Mexico. `http://www.inegi.org.mx/geo/contenidos/recnat/edafologia/PerfilesSuelo.aspx`
- International Soil Carbon Network (ISCN, `http://iscn.fluxdata.org/Pages/default.aspx`)
- ISRIC Soil Information System (ISIS, `http://www.isric.org/projects/establishment-national-soil-reference-collections-nasrec`)
- ISRIC World Inventory of Soil Emission Potentials (WISE) soil profile database (`http://www.isric.org/index.php/explore/wise-databases/wise-cooperating-institutions-and-experts`)
- SOTER-AR, Soil and Terrain Database for Argentina (`http://isric.org/index.php/projects/soil-and-terrain-soter-database-programme`)

---

[3] The most up to date overview is available at: `http://www.isric.org/explore/wosis/wosis-cooperating-institutions-and-experts`

[4] Names of collaborating institutions and individual experts may be found in the Acknowledgement section of the Technical Reports that accompany the various databases.

- SOTER-CN, Soil and Terrain Database for China (`http://isric.org/index.php/projects/soil-and-terrain-soter-database-programme`)
- SOTER-CU, Soil and Terrain Database for Cuba (`http://isric.org/index.php/projects/soil-and-terrain-soter-database-programme`)
- SOTER-KE, Soil and terrain database for Kenya (`http://isric.org/index.php/projects/soil-and-terrain-soter-database-programme`)
- SOTER-KET, Soil and Terrain Database for Upper Tana River Catchment (`http://isric.org/index.php/projects/soil-and-terrain-soter-database-programme`)
- SOTER-MW, Soil and Terrain Database for Malawi (`http://isric.org/index.php/projects/soil-and-terrain-soter-database-programme`)
- SOTER-NP; Soil and Terrain Database for Nepal (`http://isric.org/index.php/projects/soil-and-terrain-soter-database-programme`)
- SOTER-TN, Soil and Terrain Database for Tunisia (`http://isric.org/index.php/projects/soil-and-terrain-soter-database-programme`)
- SOTER-SN&GM, Soil and Terrain Database for Senegal and the Gambia (`http://isric.org/index.php/projects/soil-and-terrain-soter-database-programme`)
- SOTER-ZA, Soil and Terrain Database for South Africa (`http://isric.org/index.php/projects/soil-and-terrain-soter-database-programme`)
- SOTER-CAF, Soil and Terrain Database for Central Africa (`http://isric.org/index.php/projects/soil-and-terrain-soter-database-programme`)
- SOTER-EUR, Soil and Terrain Database for Central and Eastern Europe (`http://isric.org/index.php/projects/soil-and-terrain-soter-database-programme`)
- SOTER-LAC, Soil and Terrain Database for Latin America and the Caribbean (`http://isric.org/index.php/projects/soil-and-terrain-soter-database-programme`)
- SOTER-SAF, Soil and Terrain Database for Southern Africa (`http://isric.org/index.php/projects/soil-and-terrain-soter-database-programme`)
- United Nations Environment Programme (UNEP, various inputs mainly through regional SOTER activities)
- USA: National Cooperative Soil Survey, National Cooperative Soil Characterization Database (NCSS, `http://ncsslabdatamart.sc.egov.usda.gov/`)

Figure credits:

Figure 2.4: Chapman (2005)

Figure 4.1: Bas Kempen

Bibliography

Arrouays, D., Grundy, M., Hartemink, A., Hempel, J., Heuvelink, G., Hong, S., Lagacherie, P., Lelyk, G., McBratney, A., McKenzie, N., Mendonca-Santos, M., Minasny, B., Montanarella, L., Odeh, I., Sanchez, P., Thompson, J., Zhang, G.-L., and Donald, L. (2014). Globalsoilmap: Toward a fine-resolution global grid of soil properties. *Advances in Agronomy*, 125:93–13.

Arrouays, D., Leenaars, J., Richer-de Forges, A., Adhikari, K., Ballabio, C., Greve, M., Grundy, M., Guerrero, E., Hempel, J., Hengl, T., Heuvelink, G., Batjes, N., Ribeiro, E., Hartemink, A., Hewitt, A., Hong, S.-Y., Krasilnikov, P., Lagacherie, P., Lelyk, G., Libohova, Z., Lilly, A., McBratney, A., McKenzie, N., Vasquez, G., Mulder, V., Minasny, B., Montanarella, L., Odeh, I., Padarian, J., Poggio, L., Roudier, P., Saby, N., Savin, I., Searle, R., Solbovoy, V., Thompson, J., Smith, S., Sulaeman, Y., Vintila, R., Rossel, R., Wilson, P., Zhang, G.-L., Swerts, M., Oorts, K., Karklins, A., Feng, L., Ibelles-Navarro, A., Levin, A., Laktionova, T., Dell'Acqua, M., Suvannang, N., Ruam, W., Prasad, J., Patil, N., Husnjak, S., Pásztor, L., Okx, J., Hallet, S., Keay, C., Farewell, T., Lilja, H., Juilleret, J., Marx, S., Takata, Y., Kazuyuki, Y., Mansuy, N., Panagos, P., van Liedekerke, M., Skalsky, R., Sobocka, J., Kobza, J., Eftekhari, K., Alavipanah, S., Moussadek, R., Badraoui, M., da Silva, M., Paterson, G., Gonalves, M., Theocharopoulos, S., Yemefack, M., Tedou, S., Vrscaj, B., Grob, U., Kozák, J., Boruvka, L., Dobos, E., Taboada, M., Moretti, L., and Rodriguez, D. (2017). Soil legacy data rescue via globalsoilmap and other international and national initiatives. *GeoResJ*, 14:1–19. `https://doi.org/10.1016/j.grj.2017.06.001`.

Baize, D. (1993). *Soil Science Analyses: A Guide to Current Use*. John Wiley & Sons, Chichester.

Baritz, R., Erdogan, H., Fujii, K., Takata, Y., Nocita, M., Bussian, B., Batjes, N., Hempel, J., Wilson, P., and Vargas, R. (2014). Harmonization of methods, measurements and indicators for the sustainable management and protection of soil resources (providing mechanisms for the collation, analysis and exchange of consistent and comparable global soil data and information). Technical report, Global Soil Partnership, Food and Agriculture Organization of the United Nations. `http://www.fao.org/fileadmin/user_upload/GSP/docs/ITPS_Pillars/annexV_pillar5.pdf`.

Batjes, N. (2009). Harmonized soil profile data for applications at global and continental scales: updates to the wise database. *Soil Use and Management*, 25:124–127.

Batjes, N. (2016). Harmonised soil property values for broad-scale modelling (WISE30sec) with estimates of global soil carbon stocks. *Geoderma*, 269:61–68. `http://dx.doi.org/10.1016/j.geoderma.2016.01.034`.

Batjes, N. (2017). Overview of procedures and standards in use at ISRIC WDC-Soils. Technical Report ISRIC Report 2017/01, ISRIC - World Soil Information, Wageningen. `http://www.isric.org/sites/default/files/isric_report_2017_01doi.pdf`.

Batjes, N., Bindraban, P., and Reuter, H. (2013). Towards improved soil information for quantification of environmental, societal and economic sustainability. Technical Report ISRIC Report 2013/05, ISRIC - World Soil Information, Wageningen. `http://dx.doi.org/10.13140/2.1.4831.0560`.

Batjes, N., Ribeiro, E., van Oostrum, A., Leenaars, J., Hengl, T., and Mendes de Jesus, J. (2017). Wosis: providing standardized soil profile data for the world. *Earth System Science Data*, 9:1–14. `http://dx.doi.org/10.5194/essd-9-1-2017`.

Bolt, G. and Bruggenwert, M. (1978). *Soil Chemistry*. Elsevier Scientific Publishing Company, Amsterdam.

Botula, Y., van Ranst, E., and Cornelis, W. (2014). Pedotransfer functions to predict water retention for soils of the humid tropics: A review. *Revista Brasileira de Ciencia do Solo*, 38:679–698.

Bridges, E. (1993). Soil horizon designations: past use and future prospects. *CATENA*, 20:363–373.

Buurman, P. and van Doesburg, J. (2007). *Laser-diffraction grain-size analyses of reference profiles*, chapter Soils of Volcanic Regions in Europe, pages 451–468. Springer-Verlag, Heidelberg.

Chapman, A. (2005). *Principles of Data Quality, version 1.0. Global Biodiversity Information Facility*. Copenhagen. `http://www.gbif.org/orc/?doc_id=1229`.

Ciampalini, R., Lagacherie, P., Gomez, C., Grünberger, O., Hamrouni, M., Mekki, I., and Richard, A. (2013). Detecting, correcting and interpreting the biases of measured soil profile data: A case study in the cap bon region (tunisia). *Geoderma*, 192:68–76.

De Gruijter, J., Brus, D., Bierkens, M., and Knottters, M., editors (2006). *Sampling for natural resource monitoring*. Springer, Heidelberg.

Dixon, J., Weed, S., Kittrick, J., Milford, M., and White, J. (1977). *Minerals in soil environments*. Soil Science Society of America, Madison (WI).

EH&E (2001). The precision and accuracy of environmental measurements for the building assessment survey and evaluation program. Technical Report EH&E Report 11995, Environmental Health and Engineering Inc. (EH&E) for US Environmental Protection Agency, Washington DC. `http://www.epa.gov/iaq/base/pdfs/precisionandaccuracy.pdf`.

Elrashidi, M. (2010). Selection of an appropriate phosphorus test for soils. Technical report, Soil Survey Laboratory, USDA, Lincoln (NE). `http://www.nrcs.usda.gov/Internet/FSE_DOCUMENTS/nrcs142p2_051918.pdf`.

EURACHEM (2015). A network of organizations in europe having the objective of establishing a system for the international traceability of chemical measurements and the promotion of good quality practices (EURACHEM). `https://www.eurachem.org/`.

FAO (1977). Guidelines for the description of soils (2nd edition). Technical report, Food and Agriculture Organization of the United Nations, Rome. `http://www.fao.org/docrep/019/a0541e/a0541e.pdf`.

FAO (1988). FAO-Unesco soil map of the world, revised legend, with corrections and updates; reprinted with updates as technical paper 20 by isric, wageningen, 1997, pp. 140. Technical Report World Soil Resources Reports 60, Food and Agriculture Organization of the United Nations, Rome. `http://www.fao.org/fileadmin/user_upload/soils/docs/isricu_i9264_001.pdf`.

FAO (2006a). Guidelines for soil description (4th edition). Technical report, Food and Agriculture Organization of the United Nations, Rome. `ftp://ftp.fao.org/agl/agll/docs/guidel_soil_descr.pdf`.

FAO (2006b). World reference base for soil resources - a framework for international classification, correlation and communication. Technical Report World Soil Resources Reports 103, International Union of Soil Sciences, ISRIC - World Soil Information and Food and Agriculture Organization of the United Nations, Rome. `http://www.fao.org/ag/agl/agll/wrb/doc/wrb2007_corr.pdf`.

FAO (2015). Soils portal, soil classification. Technical report, Food and Agriculture Organization of the United Nations. `http://www.fao.org/soils-portal/soil-survey/soil-classification/en/`.

FAO and ISRIC (1986). Guidelines for soil description (3rd edition, rev.). Technical report, Food and Agriculture Organization of the United Nations and ISRIC - World Soil Information, Rome.

FAO-Unesco (1974). Legend of soil map of the world, 1:5,000,000. Technical report, United Nations Educational, Scientific, and Cultural Organization, Paris.

Finke, P. (2006). *Quality assessment of digital soil maps: producers and users perspectives*, chapter Digital soil mapping: An introductory perspective, pages 523–541. Elsevier, Amsterdam.

Goodchild, M. (1994). *Sharing imperfect data*, chapter Proceedings UNEP and IUFPRO international workshop in cooperation with FAO in developing large environmental databases for sustainable development, pages 102–11. UNEP/GRID, Sioux Falls.

Goodchild, M. and Gopal, S. (1989). *The accuracy of spatial databases*. Taylor & Francis, London.

Grewal, K., Buchnan, G., and Sherlock, R. (1991). A comparison of three methods of organic carbon determination in some new zealand soils. *Journal of Soil Science*, 42:251–257.

Grossman, R. and Reinsch, T. (2002). *Bulk density and linear extensibility, Methods of Soil Analysis: Part 4*. Physical Methods. `https://dl.sciencesocieties.org/publications/books/tocs/sssabookseries/methodsofsoilan4`.

GSM (2013). Specifications Version 1 GlobalSoilMap.net products Release 2.1. Technical report, GlobalSoilMap.net. `http://www.globalsoilmap.net/system/files/GlobalSoilMap_net_specifications_v2_0_edited_drafttaylor_Sept_2011_RAM_V12.pdf`.

GSP (2016). Global soil partnership pillar 4: Implementation plan - towards a global soil information system. Technical report, Food and Agriculture Organization of the United Nations (FAO), Rome. `http://www.fao.org/3/a-bl102e.pdf`.

GSSoil (2008). Discover INSPIRE compliant harmonized soil data and services. Technical Report GS Soil 06/2009 -05-2012, GS Soil Consortium, Paris-Lodron University Salzburg. `http://www.bgr.bund.de/DE/Themen/Boden/Projekte/Informationsgrundlagen-abgeschlossen/GSSoil/Bod_GSSoil_Broschuere.pdf?__blob=publicationFile&v=2`.

Hengl, T., Heuvelink, G., Kempen, B., Leenaars, J., Walsh, M., Shepherd, K., Sila, A., MacMillan, R., Mendes de Jesus, J., Tamene, L., and Tondoh, J. (2015). Mapping soil properties of africa at 250 m resolution: Random forests significantly improve current predictions. *PLoS ONE*, 10:e0125814. `http://dx.doi.org/10.1371/journal.pone.0125814`.

Hengl, T., Mendes de Jesus, J., Heuvelink, G., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M., Geng, X., Bauer-Marschallinger, B., Guevara, M., Vargas, R., MacMillan, R., Batjes, N., Leenaars, J., Ribeiro, E., Wheeler, I., Mantel, S., and Kempen, B. (2017). Soilgrids250m: Global gridded soil information based on machine learning. *PLoS ONE*, 12:e0169748. `http://dx.doi.org/10.1371/journal.pone.0169748`.

Hengl, T., Mendes de Jesus, J., MacMillan, R., Batjes, N., Heuvelink, G., Ribeiro, E., Samuel-Rosa, A., Kempen, B., Leenaars, J., Walsh, M., and Ruiperez Gonzalez, M. (2014). Soilgrids1km - global soil information based on automated mapping. *PLoS ONE*, 9(8):e105992. `http://dx.doi.org/10.1371/journal.pone.0105992`.

Heuvelink, G. (1996). Identification of field attribute error under different models of spatial. *International Journal of Geographical Information Systems*, 10:921–935.

Heuvelink, G. (2014). *Uncertainty quantification of GlobalSoilMap products*, chapter GlobalSoilMap. Basis of the Global Spatial Soil Information System, pages 335–340. GlobalSoilMap.

Hilke, I. (2015). General principles for the quantification of total organic carbon (TOC) in environmental solid and liquid samples. Technical Report Technical Reports 30, Max-Planck-Institut fr Biogeochemie, Jena, Germany. `https://www.bgc-jena.mpg.de/www/uploads/Publications/TechnicalReports/30_2015_TechnicalReport_Hilke.pdf`.

Hobley, E., Schultze, D., Robinson, D., Jahanshiri, E., Aitkenhead, M., Batjes, N., and Hengl, T. (2017). Open soil science: technology is helping us discover the mysteries under our feet, the conversation. `https://theconversation.com/open-soil-science-technology-is-helping-us-discover-the-mysteries-under-our-feet-81727`.

INSPIRE (2015). Data specifications - infrastructure for spatial information in the european community (INSPIRE). Technical report, European Commission Joint Research Centre. `http://inspire.ec.europa.eu/index.cfm/pageid/2`.

ISO (1995). ISO 10694:1995 soil quality - determination of organic and total carbon after dry combustion (elementary analysis). Technical report, International Organization for Standardization (ISO). `https://www.iso.org/obp/ui/#iso:std:iso:10694:ed-1:v1:en`.

ISO (2005). ISO 10390:2005 soil quality - determination of ph. Technical report, International Organization for Standardization (ISO). `http://www.iso.org/iso/catalogue_detail.htm?csnumber=40879`.

ISO (2016). ISO/TC 190:2016 standards catalogue. Technical report, International Organization for Standardization (ISO). `https://www.iso.org/standards-catalogue/browse-by-ics.html`.

IUSS WG-WRB (2015). World reference base for soil resources 2015 - international soil classification system for naming soils and creating legends for soil map. Technical Report World Soil Resources Reports 106, Global Soil Partnership, International Union of Soil Sciences and Food and Agriculture Organization of the United Nations, Rome. http://www.fao.org/3/a-i3794e.pdf.

Jankauskas, B., Jankauskiene, G., Slepetiene, A., Fullen, M., and Booth, C. (2006). International comparison of analytical methods of determining the soil organic matter content of lithuanian eutric albeluvisols. *Communications in Soil Science and Plant Analysis*, 37:707–720. http://dx.doi.org/10.1080/00103620600563499.

Jury, A., Gardner, W., and Gardner, W. (1991). *Soil Physics (5th edition)*. Wiley, New York.

Katschinski, N. (1956). Die mechanische bodenanalyse und die klassifikation der böden nach ihrer mechanischen zusammensetzung. *Rapports au Sixième Congrès International de la Science du Sol*, pages 321–327. http://iuss.boku.ac.at/files/vi_e_congres_international_de_la_science_du_sol_volume_b__paris_1956_compressed.pdf#page=323.

Kauffman, J. (1987). Comparative classification of some deep, well-drained red clay soils of mozambique. Technical Report Technical Paper 16, ISRIC - World Soil Information, Wageningen. http://www.isric.org/sites/default/files/ISRIC_TechPap16.pdf.

Ketterings, Q., Czymmek, K., Reid, W., and Wildman, R. (2002). Conversion of modified morgan and mehlich iii soil tests to morgan soil test values. *Soil Science*, 167(12):830–837.

Koorevaar, P., Menelik, G., and Dirksen, C. (1983). Elements of soil physics. *Developments in Soil Science*, 13:228.

Kroll, B. (2008). *Towards a data management framework for digital soil mapping with limited data*, chapter Digital soil mapping with limited data, pages 137–149. Springer.

Kuhlbusch, T. A., Borowiak, A., Gelenscer, A., Genberg, J., Gladtke, D., Maenhaut, W., Pio, C., Popoviecheva, O., Putaud, J. P., Quincey, P., Sciare, J., ten Brink, H., Viana, M., and Espen Yttri, K. (2009). Measurement of Elemental and Organic Carbon in Europe. Technical report, European Commission, Joint Research Centre, Institute for Environment and Sustainability, Luxembourg. http://publications.jrc.ec.europa.eu/repository/bitstream/JRC54045/ecoc%20eur%20report%20240909.pdf.

Landon, J. (1991). *Booker Tropical Soil Manual*. Longman Scientific & Technical, New York.

Leenaars, J., Kempen, B., van Oostrum, A., and Batjes, N. (2014a). *Africa Soil Profiles Database: a compilation of georeferenced and standardized legacy soil profile data for Sub-Saharan Africa*, chapter GlobalSoilMap - Basis of the global spatial information system, pages 51–57. Taylor & Francis Group, Orléans. http://books.google.nl/books?hl=en&lr=&id=S5ClAgAAQBAJ&oi=fnd&pg=PA51&ots=ilrpyldHtj&sig=8KsouYF6P23GR5R1UQ2oaV6U_30&redir_esc=y#v=onepage&q&f=false.

Leenaars, J., van Oostrum, A., and Ruiperez Gonzalez, M. (2014b). Africa soil profiles database version 1.2. a compilation of georeferenced and standardized legacy soil profile data for sub-saharan africa (with dataset). Technical Report ISRIC Report 2014/01, ISRIC - World Soil Information, Wageningen. http://www.isric.org/sites/default/files/isric_report_2014_01.pdf.

Lettens, S., Vos, B., Quataert, P., van Wesemael, B., Muys, B., and van Orshoven, J. (2007). Variable carbon recovery of walkley-black analysis and implications for national soil organic carbon accounting. *European Journal of Soil Science*, 58:1244–1253. http://dx.doi.org/10.1111/j.1365-2389.2007.00916.x.

Louis, B., Saby, N., Orton, T., Lacarce, E., Boulonne, L., Jolivet, C., Ratié, C., and Arrouays, D. (2014). Statistical sampling design impact on predictive quality of harmonization functions between soil monitoring networks. *Geoderma*, 213:133–143.

Loveland, P. (2016). Particle size analysis: What does it mean? *The Auger*, 2016, December:25–25. http://soils.org.uk/auger.

Loveland, P. and Whalley, W. (2000). *Particle size analysis*, chapter Soil and environmental analysis - Second edition (Revised and Expanded), pages 281–314. Marcel Dekker Inc., New York & Basel.

Magnusson, B. and Örnemark, U., editors (2014). *Eurachem Guide: The fitness for purpose of anlytical methods - A laboratory guid to method validation and related topics (2nd edition)*. Eurachem. `https://www.eurachem.org/`.

Maire, V., Wright, I., Prentice, I., Batjes, N., Bhaskar, R., van Bodegom, P., Cornwell, W., Ellsworth, D., Niinemets, U., Ordonez, A., Reich, P., and Santiago, L. (2015). Global effects of soil and climate on leaf photosynthetic traits and rates. *Global Ecology and Biogeography*, 24:706–715. `http://dx.doi.org/10.1111/geb.12296`.

Mallarino, P. (1995). *Comparison of Mehlich-3, Olsen and Bray-P1 procedures for phosphorus in calcareous soils*, chapter Proceedings of the twenty-fifth North Central extension-industry soil fertility conference, pages 96–101. Potash & Phosphate Institute, Manhattan (KS). `http://www.agronext.iastate.edu/soilfertility/info/ComparisonofMehlich-3OlsenandBray-P1Procedures.pdf`.

Mendes de Jesus, J., Ribeiro, E., Batjes, N., and Kempen, B. (2017). Spatial data infrastructures for handling soil data. In *Abstract Book Pedometrics 2017*. `https://static1.squarespace.com/static/5653202ee4b037d305e7fd3e/t/594a53f7e4fcb553cd43b9c0/1498043388899/Abstract+Book+Pedometrics+2017.pdf#page=154`.

Michéli, E., Láng, V., Owens, P., McBratney, A., and Hempel, J. (2016). Testing the pedometric evaluation of taxonomic units on soil taxonomy - a step in advancing towards a universal soil classification system. *Geoderma*, 264:340–349. `http://dx.doi.org/10.1016/j.geoderma.2015.09.008`.

Motsara, M. and Roy, R. (2008). Guide to laboratory establishment for plant nutrient analysis. Technical Report Fertilizer and Plant Nutrition Bulletin 19, Food and Agriculture Organization of the United Nations, Rome.

NAPT (2015). North american proficiency testing (NAPT) program. `http://www.naptprogram.org/`.

NCSS (2010). National cooperative soil survey (NCSS). national cooperative soil characterization database. `http://ssldata.nrcs.usda.gov/`.

Nelson, D. and Sommers, L. (1982). *Total carbon, organic carbon, and organic matter*, chapter Methods of Soil Analysis, pages 539–579. American Society of Agronomy, Madison, WI.

Nemes, A., Wösten, J., Lilly, A., and Oude Voshaar, J. (1999). Evaluation of different procedures to interpolate particle-size distributions to achieve compatibility within soil databases. *Geoderma*, 90:187–202. `http://www.sciencedirect.com/science/article/pii/S0016706199000142`.

OECD (1998). *OECD Principles of Good Laboratory Practice*. OECD Environmental Health and Safety Publications. `http://dx.doi.org/10.1787/9789264078536-en`.

OGC (2013). OGC abstract specification geographic information - observations and measurements (ver. 2.0). Technical report, Open Geospatial Consortium (OGC). `https://portal.opengeospatial.org/files/?artifact_id=41579`.

OGC (2015). 97th OGC technical committee, soil data interoperability experiment, 3 december 2015. Technical report, Open Geospatial Consortium (OGC), Sydney, Australia. `http://www.youtube.com/watch?v=kNVfKVhtKJ8`.

Omuto, C., Nachtergaele, F., and Vargas Rojas, R. (2012). State of the art report on global and regional soil information: Where are we? where to go? Technical report, Food and Agriculture Organization of the United Nations, Rome. `http://typo3.fao.org/fileadmin/templates/nr/kagera/Documents/Soil_information_Report.pdf`.

Pansu, M. and Gautheyrou, J. (2006). *Handbook of Soil Analysis: Mineralogical, Organic and Inorganic Methods*. Springer.

Pribyl, D. (2010). A critical review of the conventional SOC to SOM conversion factor. *Geoderma*, 156:75–83. `http://www.sciencedirect.com/science/article/pii/S0016706110000388`.

Rayment, G., Lyons, D., and Hill, R. (2017). ASPAC soil proficiency testing program report 2011-12. Technical report, ASPAC, Melbourne, Victoria. `http://www.aspac-australasia.com/images/PDF/Soil-ASPAC-Annual-Report-2011-12_GER_August17f.pdf`.

Rhoades, J., Chanduvi, F., and Lesch, S. (1999). Soil salinity assessment: Methods and interpretation of electrical condunductivity measurements. Technical report, Food and Agriculture Organization of the United Nations, Rome. `http://www.fao.org/docrep/019/x2002e/x2002e.pdf`.

Ribeiro, E., Batjes, N., Leenaars, J., van Oostrum, A., and Mendes de Jesus, J. (2015). Towards the standardization and harmonization of world soil data: Procedures manual. Technical Report ISRIC Report 2015/03, ISRIC - World Soil Information, Wageningen. `http://www.isric.org/sites/default/files/isric_report_2015_03.pdf`.

Ross, D., Bailey, S., Briggs, R., Curry, J., Fernandez, I., Fredriksen, G., Goodale, C., Hazlett, P., Heine, P., Johnson, C., Larson, J., Lawrence, G., Kolka, R., Ouimet, R., Paré, D., Richter, D., Schirmer, C., and Warby, R. (2015). Inter-laboratory variation in the chemical analysis of acidic forest soil reference samples from eastern north america. *Ecosphere*, 6:73. `http://onlinelibrary.wiley.com/doi/10.1890/ES14-00209.1/epdf`.

Rousseva, S. (1997). Data transformations between soil texture schemes. *European Journal of Soil Science*, 48:749–758. `http://dx.doi.org/10.1111/j.1365-2389.1997.tb00574.x`.

Spaargaren, O. and Batjes, N. (1995). Report on the classification into FAO-Unesco soil units of profiles selected from the NRCS pedon database for IGBP-DIS. Technical Report ISRIC Report 1995/01, ISRIC - World Soil Information, Wageningen. `http://www.isric.org/sites/default/files/isric_report_1995_01.pdf`.

SSDS (1951). *Soil Survey Manual. United States Agriculture Handbook 18*. United States Department of Agriculture, Natural Resources Conservation Service, Washington. `https://www.nrcs.usda.gov/wps/portal/nrcs/detailfull/soils/ref/?cid=nrcs142p2_054262`.

SSDS (1975). *Soil Taxonomy: A basic system of soil classification for making and interpreting soil surveys. Handbook 436*. Soil Conservation Service, United States Department of Agriculture, Washington.

SSDS (1983). *Soil Survey Manual (rev. edition). Handbook 18*. United States Department of Agriculture, Natural Resources Conservation Service, Washington.

SSDS (1992). *Keys to Soil Taxonomy (5th edition), SMSS Technical Monograph No. 19*. Agency for International Development, United States Department of Agriculture, Soil Conservation Service, Soil Management Support Services, Pocahontas Press, Inc., Blacksburg (Virginia).

SSDS (1993). *Soil Survey Manual, Handbook 18*. Soil Conservation Service, United States Department of Agriculture, Washington. `https://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/ref/?cid=nrcs142p2_054262`.

SSDS (1998). *Keys to Soil Taxonomy (8th edition)*. United States Department of Agriculture, Natural Resources Conservation Service, Blacksburg (Virginia).

SSDS (2003). *Keys to Soil Taxonomy (9th edition)*. United States Department of Agriculture, Natural Resources Conservation Service, Blacksburg (Virginia).

SSDS (2010). *Keys to Soil Taxonomy (11th edition)*. United States Department of Agriculture, Natural Resources Conservation Service, Washington, DC. `ftp://ftp-fc.sc.egov.usda.gov/NSSC/Soil_Taxonomy/keys/2010_Keys_to_Soil_Taxonom.pdf`.

SSDS (2011). *Soil Survey Laboratory Information Manual (Ver. 2.0)*, volume Soil Survey Investigations Report No. 45. National Soil Survey Center, Soil Survey Laboratory, Lincoln (NE). `ftp://ftp-fc.sc.egov.usda.gov/NSSC/Lab_Info_Manual/SSIR_45.pdf`.

SSDS (2012). *Field book for describing and sampling soils (ver. 3.0)*. National Soil Survey Center Natural Resources Conservation Service, U.S. Department of Agriculture, Lincoln (NE). `ftp://ftp-fc.sc.egov.usda.gov/NSSC/Field_Book/FieldBookVer3.pdf`.

SSDS (2014). *Kellog Soil Survey Laboratory Methods Manual (ver. 5.0)*, volume Soil Survey Investigations Report No. 42. Kellogg Soil Survey Laboratory, National Soil Survey Center, Natural Resources Conservation Service, U.S. Department of Agriculture, Lincoln (NE). `http://www.nrcs.usda.gov/Internet/FSE_DOCUMENTS/stelprdb1253872.pdf`.

Tempel, P., van Kraalingen, D., Mendes de Jesus, J., and Reuter, H. (2013). Towards an isric world soil information service (WoSIS version 1.0). Technical Report ISRIC Report 2013/02, ISRIC - World Soil Information, Wageningen. `http://www.isric.org/sites/default/files/isric_report_2013_02.pdf`.

US-EPA (2015). Quality management tools - best practices for laboratory quality systems. Technical report, U.S. Envrionmental Protection Agency (US-EPA). `http://www.epa.gov/quality/bestlabs.html`.

van Bemmelen, D. (1890). Über die bestimmung des wassers, des humus, des schwefels, der in den colloïdalen silikaten gebundenen kieselsäure, des mangans u. s. w. im ackerboden. *Die Landwirthschaftlichen Versuchs-Stationen*, 37:279–290. `http://edepot.wur.nl/211282`.

van Engelen, V. (2011). Standardizing soil data (e-SOTER regional pilot platform as EU contribution to a Global Soil Information System). *International Innovation*, pages 48–49.

van Engelen, V. and Dijkshoorn, J. (2013). Global and national soils and terrain digital databases (SOTER) - procedures manual (ver. 2.0). Technical Report ISRIC Report 2013/04, The International Union of Soil Sciences (IUSS) and ISRIC - World Soil Information and Food and Agriculture Organization of the United Nations (FAO), Wageningen. `http://www.isric.org/sites/default/files/isric_report_2013_04.pdf`.

van Reeuwijk, L. (1998). Guidelines for quality management in soil and plant laboratories. Technical Report FAO Soil Bulletin 74, ISRIC - World Soil Information and Food and Agriculture Organization of the United Nations. `http://www.fao.org/docrep/W7295E/W7295E00.htm`.

van Reeuwijk, L. (2002). Procedures for soil analysis (6th edition). Technical Report Technical Paper 9, ISRIC - World Soil Information, Wageningen. `http://www.isric.org/sites/default/files/ISRIC_TechPap09.pdf`.

van Reeuwijk, L. and Houba, V. (1998). Guidelines for quality management in soil and plant laboratories. external quality control of data. Technical report, ISRIC - World Soil Information and Food and Agriculture Organization of the United Nations. `http://www.fao.org/docrep/W7295E/W7295E00.htm`.

Viscarra Rossel, R., Behrens, T., Ben-Dor, E., Brown, D., Demattê, J., Shepherd, K., Shi, Z., Stenberg, B., Stevens, A., Adamchuk, V., Aïchi, H., Barthès, B., Bartholomeus, H., Bayer, A., Bernoux, M., Böttcher, K., Brodský, L., Du, C., Chappell, A., Fouad, Y., Genot, V., Gomez, C., Grunwald, S., Gubler, A., Guerrero, C., Hedley, C., Knadel, M., Morrás, H., Nocita, M., Ramirez-Lopez, L., Roudier, P., Campos, E., Sanborn, P., Sellitto, V., Sudduth, K., Rawlins, B., Walter, C., Winowiecki, L., Hong, S., and Ji, W. (2016). A global spectral library to characterize the world's soil. *Earth-Science Reviews*, 155:198–230.

Walkley, A. and Black, I. (1934). An examination of the degtjareff method for determining soil organic matter, and a proposed modification of the chromic acid titration method. *Soil Science*, 34:29–38.

WEPAL (2015). Wageningen evaluating programmes for analytical laboratories. Technical report, ISRIC - World Soil Information, Wageningen. `http://www.wepal.nl/website/about_wepal/History.htm`.

WHO (2011). Laboratory quality management system: handbook. Technical report, World Health Organization, Lyon, France. `http://whqlibdoc.who.int/publications/2011/9789241548274_eng.pdf?ua=1`.

Wikipedia (2017). Wikipedia article on federated database system. `https://en.wikipedia.org/wiki/Federated_database_system`.

Wilson, P. (2016). Global harmonisation and exchange of soil data. In *Advancing the Frontiers of Data in Research*, Colorado, USA.

Appendices

Appendix A

Procedures for accessing WoSIS

This appendix describes procedures for accessing data served from WoSIS. To access soil profile (point) data you can either use: a) QGIS[1] or similar GIS software that supports Web Feature Service (WFS), and b) the statistical software R.

A.1 Accessing WoSIS from QGIS using WFS

To open and view WoSIS points in QGIS, you first need to create a WFS connection. To do this press the Add WFS layer button, then press 'New', give it a connection name, for example, 'WoSIS' in the URL field put **http://data.isric.org/geoserver/wosis_latest/wfs/**, then press the OK button. Press 'Connect' and all available layers will appear on screen (Figure A.1). No authentication is needed here, so leave blank the 'User name' and 'Password' boxes.

Next, select the layer of interest listed under 'Title' and press 'Add' button (Figure A.2). Note that it may take some time until you fetch all points for the global coverage. Be aware that you are working online and that some layers have over half a million records. So probably a more efficient approach would be to request only the points for a specific area. In this case, first zoom to the area of interest and then tick on 'Only request features overlapping the current view extent'. Another option, when adding the layer, is to click on the 'Build query' button, expand 'Fields and Values' and, for example, build the expression based on the country code, like this: country_id = 'BR' (Figure A.3).

After loading the layer you can save it as a shape file on your local machine for further offline use. Check for updates on the WFS every month, as this is a dynamic dataset. Note that the symbology can be adapted to better document the data as in Figure A.4, which shows pH-water observations across the world.

Detailed instructions for ArcGIS users may be found at: `http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#//00sp0000001m000000.htm`.

A.2 Accessing WoSIS from R using WFS

This assumes that you have already installed:

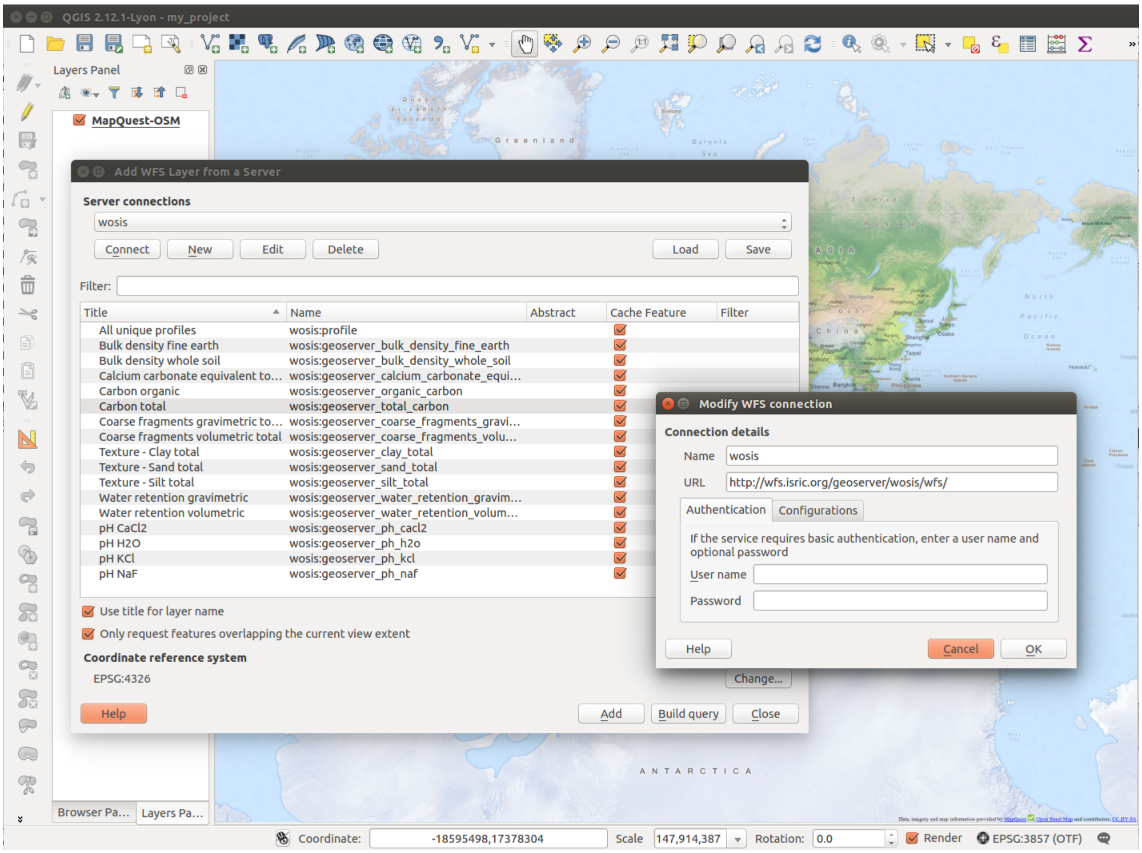
- R [2];
- R packages: rgdal, gdalUtils;
- GDAL [3];

---

[1] `http://www.qgis.org/en/site/`

[2] `http://cran.r-project.org/bin/`

[3] `http://www.gisinternals.com/release.php`



Server connections: wosis

Filter:

Title	Name	Abstract	Cache Feature	Filter
All unique profiles	wosis:profile		☑	
Bulk density fine earth	wosis:geoserver_bulk_density_fine_earth		☑	
Bulk density whole soil	wosis:geoserver_bulk_density_whole_soil		☑	
Calcium carbonate equivalent to...	wosis:geoserver_calcium_carbonate_equi...		☑	
Carbon organic	wosis:geoserver_organic_carbon		☑	
Carbon total	wosis:geoserver_total_carbon		☑	
Coarse fragments gravimetric to...	wosis:geoserver_coarse_fragments_gravi...		☑	
Coarse fragments volumetric total	wosis:geoserver_coarse_fragments_volu...		☑	
Texture - Clay total	wosis:geoserver_clay_total		☑	
Texture - Sand total	wosis:geoserver_sand_total		☑	
Texture - Silt total	wosis:geoserver_silt_total		☑	
Water retention gravimetric	wosis:geoserver_water_retention_gravim...		☑	
Water retention volumetric	wosis:geoserver_water_retention_volum...		☑	
pH CaCl2	wosis:geoserver_ph_cacl2		☑	
pH H2O	wosis:geoserver_ph_h2o		☑	
pH KCl	wosis:geoserver_ph_kcl		☑	
pH NaF	wosis:geoserver_ph_naf		☑	

☑ Use title for layer name

☑ Only request features overlapping the current view extent

Coordinate reference system: EPSG:4326

Connection details

Name: wosis

URL: http://wfs.isric.org/geoserver/wosis/wfs/

If the service requires basic authentication, enter a user name and optional password

User name:

Password:

Coordinate: -18595498,17378304 Scale 147,914,387 Rotation: 0.0 ☑ Render EPSG:3857 (OTF)

Figure A.1: Adding WoSIS WFS configuration in QGIS.

Start the R session and insert the following commands; this will install the necessary packages.

```
install.packages("rgdal", dependencies = TRUE)

install.packages("gdalUtils", dependencies = TRUE)
```

Load the packages.

```
library(gdalUtils)

library(rgdal)
```

Define the WFS url.

```
dsn <- "WFS:http://data.isric.org/geoserver/wosis_latest/wfs/"

ogrinfo(dsn, so=TRUE)
```

Outputs the WFS metadata and the list of available layers from the server.

```
ogr2ogr(dsn, "ph_h2o.shp", "wosis:ph_h2o")
```

At this point some warnings are shown about the renaming of long field names.

```
ph_h2o <- readOGR("ph_h2o.shp", "ph_h2o", stringsAsFactors=FALSE)
```

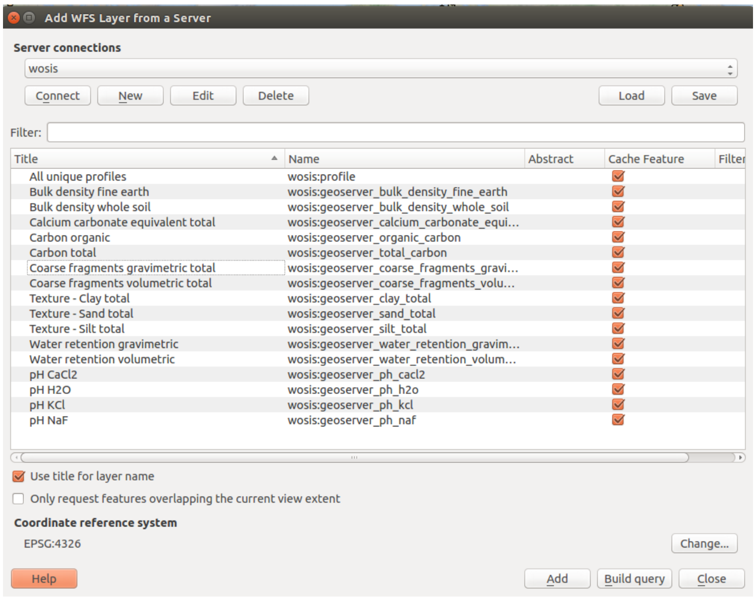
Outputs the number of records, in this case 446780, and the number of fields (16).

```
str(ph_h2o@data)
```

Outputs sample data from each of the dataset fields. Outputs a simple map with the point distribution and then histogram.

```
plot(ph_h2o)

hist(ph_h2o@data$value, col = "red", freq = FALSE, breaks = "Sturges")
```



Server connections

wosis

Title	Name	Abstract	Cache Feature	Filter
All unique profiles	wosis:profile		☑	
Bulk density fine earth	wosis:geoserver_bulk_density_fine_earth		☑	
Bulk density whole soil	wosis:geoserver_bulk_density_whole_soil		☑	
Calcium carbonate equivalent total	wosis:geoserver_calcium_carbonate_equi...		☑	
Carbon organic	wosis:geoserver_organic_carbon		☑	
Carbon total	wosis:geoserver_total_carbon		☑	
Coarse fragments gravimetric total	wosis:geoserver_coarse_fragments_gravi...		☑	
Coarse fragments volumetric total	wosis:geoserver_coarse_fragments_volu...		☑	
Texture - Clay total	wosis:geoserver_clay_total		☑	
Texture - Sand total	wosis:geoserver_sand_total		☑	
Texture - Silt total	wosis:geoserver_silt_total		☑	
Water retention gravimetric	wosis:geoserver_water_retention_gravim...		☑	
Water retention volumetric	wosis:geoserver_water_retention_volum...		☑	
pH CaCl2	wosis:geoserver_ph_cacl2		☑	
pH H2O	wosis:geoserver_ph_h2o		☑	
pH KCl	wosis:geoserver_ph_kcl		☑	
pH NaF	wosis:geoserver_ph_naf		☑	

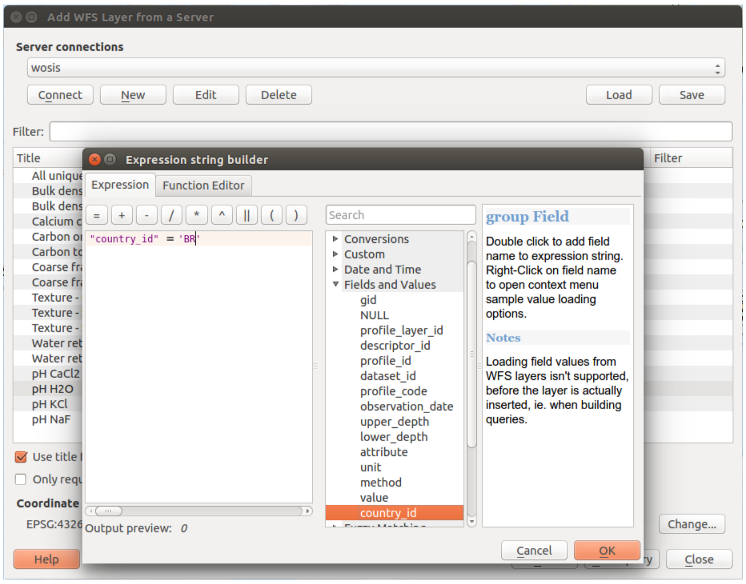
☑ Use title for layer name

☐ Only request features overlapping the current view extent

Coordinate reference system

EPSG:4326

Figure A.2: Selecting WFS available layers.



```
"country_id" = 'BR'
```

Output preview: *0*

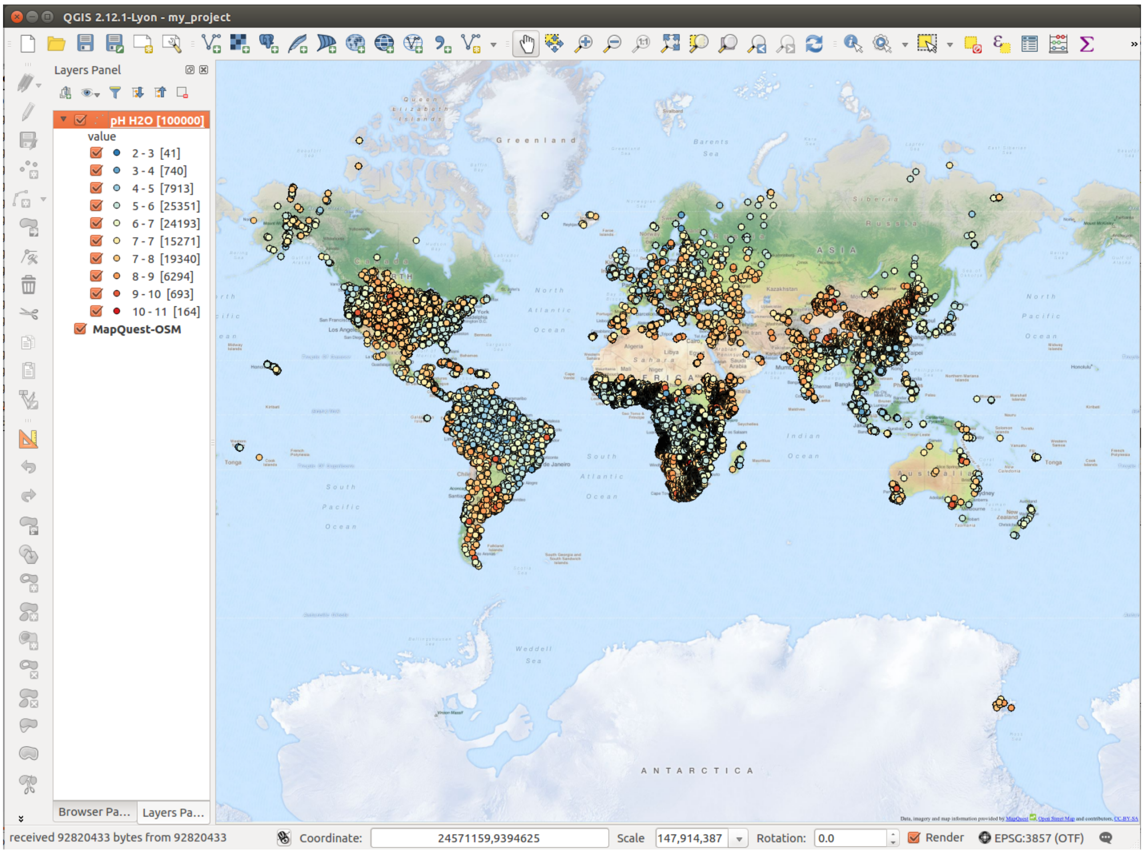
group Field

Double click to add field name to expression string. Right-Click on field name to open context menu sample value loading options.

Notes

Loading field values from WFS layers isn't supported, before the layer is actually inserted, ie. when building queries.

Figure A.3: WFS filter records.



☑ pH H2O [100000]

value

☑ ● 2 - 3 [41]

☑ ● 3 - 4 [740]

☑ ○ 4 - 5 [7913]

☑ ○ 5 - 6 [25351]

☑ ○ 6 - 7 [24193]

☑ ○ 7 - 7 [15271]

☑ ○ 7 - 8 [19340]

☑ ○ 8 - 9 [6294]

☑ ● 9 - 10 [693]

☑ ● 10 - 11 [164]

☑ **MapQuest-OSM**

received 92820433 bytes from 92820433

Coordinate: 24571159,9394625

Scale 147,914,387

Rotation: 0.0

☑ Render

EPSG:3857 (OTF)

Figure A.4: Overview of pH water data provided with the WFS.

Appendix B

Basic principles for compiling a soil profile dataset

To be considered in WoSIS, a soil dataset should include data for commonly required soil properties (FAO, 2006a; SSDS, 2012; van Engelen and Dijkshoorn, 2013; Batjes, 2016), but no minimum dataset is prescribed. However, sufficient information (metadata) should be provided to assess the source and quality of the data as well as the access category (licence). A dictionary table describing the meaning of all (often abbreviated) column headings used in the dataset tables should be provided with the metadata. Similarly, the use of dictionary tables is recommended for describing all coded data entries, like ID‘s as well as any abbreviated descriptive soil property value (e.g. ’W’ means ’well drained’ when defined according to the FAO Guidelines (FAO, 2006a)).

Soil profile data should be consistently given as the result of observations and measurements (O&M). In this manual, an observation (O) is the outcome of an ‘act of measuring or otherwise determining the value of a property’, while a measurement (M) is the outcome of a ‘set of operations having the object of determining the value of a quantity’ (OGC, 2013).

Soil records are considered complete and thus processable into WoSIS when: 1) the lineage[1] of the soil record is well described and 2) the soil data are consistently expressed as the result of observations and measurements (O&M). A single soil profile record can be considered as a collection of observations and measurements, with similar lineage, applied to the soil (to the soil profile site, to the profile itself, and to the profile depth intervals (e.g. horizons or layers)), by using (defined) field and laboratory methods to assess values for the soil properties or attributes under consideration. Those values, either numeric, categorical or descriptive, are expressed according to the associated value domain as dictated by the references used for defining units of expression or pick lists. Note that the profile itself is a depth interval spanning the layer or horizon depth intervals[2].

Often, soil profile data need to be prepared to facilitate their standardization in WoSIS. As indicated, sharing soil data for consideration in WoSIS does not require the use of a specific data entry template with a priori standards, nor is there a minimum dataset size.

A template recommended for compiling and submitting datasets for ‘small’ projects for consideration in WoSIS is presented below. The standardized template consists of 1 spreadsheet with 5 different sheets (**dataset**, **profile**, **horizon**, **table_definition** and **lookup_table**)[3].

By convention, the sheet and column names should not contain diacritical marks, symbols, spaces, upper-case characters and not start with a number.

Sheet **dataset** has the following rows:

- *dataset_title*: Dataset title, project or thesis title.

---

[1] `https://en.wikipedia.org/wiki/Data_lineage`

[2] If the bedrock or an impenetrable layer is observed, this should be specified in the dataset to make the observation explicit.

[3] Templates for this may be downloaded from: LibreOffice - `http://www.isric.org/sites/default/files/template.ods`; MicrosofOffice `http://www.isric.org/sites/default/files/template.xlsx`

- *dataset_description*: Brief description of the dataset.
- *publication_date*: Publication date.
- *dataset_version*: Dataset version.
- *dataset_license*: Access and use constraints of the dataset (Preferably expressed in terms of a Creative Commons licence).
- *organization_name*: Organization name.
- *organization_url*: Organization website (URL).
- *organization_country*: Organization country.
- *organization_city*: Organization city.
- *organization_postal_code*: Organization postal code.
- *organization_street_name*: Organization street name.
- *organization_street_number*: Organization street number.
- *author1_first_name*: Author first name (1st author).
- *author1_last_name*: Author last name (1st author).
- *author1_email*: Author email (1st author).
- *authorx_first_name*: Author first name (Xth author, repeat as needed).
- *authorx_last_name*: Author last name (Xth author, repeat as needed).
- *authorx_email*: Author email (Xth author, repeat as needed).
- *dataset_reference1*: Reference to a document, scientific paper or dataset. A DOI is preferred, but URLs may be used also.
- *dataset_referencex*: Another reference, in so far as needed just increment the last number of the column name and provide the DOI or URL.
- *laboratory_name*: Laboratory name.
- *laboratory_country*: Laboratory country.
- *laboratory_city*: Laboratory city.
- *laboratory_postal_code*: Laboratory postal code.
- *laboratory_street_name*: Laboratory street name.
- *laboratory_street_number*: Laboratory street number.
- *soil_classification_system_name*: The soil classification system used to classify the profile.
- *soil_classification_system_year*: The publication year of the soil classification system used.

Any further items may be entered after these rows when necessary.

Sheet **profile** starts with the following columns:

- *profile_id*: Unique identifier of the profile as used in the source dataset.
- *observation_date*: Date of the observation in format (yyyy-mm-dd).
- *coordinate_system*: Coordinate system used. Please indicate the correspondent EPSG code (e.g. WGS 84, EPSG: 4326).
- *x_coord*: X coordinate, if in geographic coordinates (degrees), the same as Longitude.
- *y_coord*: Y coordinate, if in geographic coordinates (degrees), the same as Latitude.
- *country_id*: ISO 3166-1 alpha-2 country code.

- *profile_classification_name*: The name of the soil profile classification.
- *profile_classification_code*: The code of the soil profile classification.
- *attribute_1*: Add site, first attribute name here.
- *attribute_2*: Add site, second attribute name here.
- *attribute_x*: Add site, x attribute name here.
- *(...)*

Any other profile description attribute may follow after these columns.

Each row may only contain a given profile (description).

Sheet **horizon** starts with the following columns:

- *profile_id*: Unique identifier of the profile. Refers to sheet 'profile'.
- *horizon_number*: Consecutive layer number rated from top to bottom.
- *horizon_name*: Horizon designation. A, B, etc.
- *sample_code*: Laboratory sample code.
- *upper_depth*: Depth of upper horizon boundary (cm).
- *lower_depth*: Depth of lower horizon boundary (cm).
- *attribute_1*: Add horizon, first attribute name here.
- *attribute_2*: Add horizon, second attribute name here.
- *attribute_x*: Add horizon, x attribute name here.
- *(...)*

Any other horizon description attributes can follow after these columns. For example, ph_h2o, ph_kcl, clay, silt, sand, ...). Each row should contain data for a different combination of profile and horizon descriptions.

Sheet **table_definition** serves to describe all the columns (attributes) in profile and horizon sheets. It starts with the following columns:

- *sheet_name*: Either 'profile' or 'horizon'.
- *column_name*: The exact name of the column added after the default ones.
- *description*: Description of the attribute.
- *unit*: Units used, if not used, 'unitless' should be indicated.
- *data_type*: Data type, use one out of ('Text', 'Integer', 'Real', 'Boolean', 'Date').
- *analytical_method*: Analytical method used in the laboratory, if none is used enter 'Not applicable'.
- *domain_name*: Domain name from the **lookup_table** when applicable (for categorical attributes).

Any other attribute description can follow after these rows. Each row should contain a different combination of sheet and column description.

Appendix C

Quality aspects related to laboratory data

C.1 Context

WoSIS is being populated using data produced for different types of studies; the corresponding data were sampled and analysed in a range of laboratories according to a wide range of methods. By implication, the quality of the standardized / harmonized data in WoSIS will be determined by the quality of all preceding steps of data processing. Typically, a quality management system comprises measures necessary to arrive at a pre-defined and constant quality at agreed costs (based on user-specific requirements for use). For instance, (certified) laboratories develop / use protocols for each sub-process, use validated methods for laboratory investigations, and participate in round robin tests to monitor their performance over time with respect to certified or consensus reference materials (van Reeuwijk, 1998; Motsara and Roy, 2008; WHO, 2011; US-EPA, 2015).

ISRIC, for example, published reference procedures for soil analysis as a step towards standardization of analytical methods in soil laboratories (van Reeuwijk, 2002). These procedures cover the range of analytical methods required for soil characterization according to the Revised FAO Legend (FAO, 1988) and the World Reference Base (IUSS WG-WRB, 2015). The Natural Resources Conservation Service of the United States Department of Agriculture published a Soil Survey Laboratory Methods Manual (SSDS, 2011), which is the reference source for the National Cooperative Soil Survey Soil Characterization Database (NCSS, 2010) and widely referred to internationally as reference.

Although adoption of such reference methods at different laboratories contributes to a common quality level, it does not rule out that the quality of individual data held in compiled datasets, such as WoSIS, may differ considerably in quality as discussed below.

C.2 Laboratory error

Important quality characteristics for any measured data are the random and systematic error (Magnusson and Örnemark, 2014). Random errors in experimental measurements are caused by unknown and unpredictable changes in the experiment; such changes may occur in the measuring instruments or in the environmental conditions. Systematic errors in experimental observations usually come from the measuring instruments themselves. Both error components will contribute to a varying extent to the total error as shown earlier. In practice, however, in reports and publications these essential laboratory error characteristics are generally not presented along with the actual data produced. In such cases, error characteristics can only be extracted afterwards from quality management systems or estimated in special experimental designs. Laboratories participating in inter-laboratory studies such as ring tests or round robin tests receive feedback on their quality performance with the particular methods by comparing their results with those from other participants. Examples are WEPAL (2015), the Wageningen Evaluating Programmes for Analytical Laboratories, the Australasia inter-laboratory proficiency programme (ASPAC) (Rayment et al., 2017) and the North American Proficiency Testing Programme (NAPT, 2015). These programmes often are certified according to ISO/IEC 17043. However, they do not consider the influence of differences in sampling procedure and pretreatment at individual laboratories as these programmes use pretreated and homogenized materials. Further, the reference materials need to be relevant / representative for the soil types analysed at a given laboratory. Ross et al. (2015), for example, in studying the inter-laboratory variation in the chemical analysis of acidic forest soil reference samples from eastern North America, stressed the importance of using sample materials representative for the (types of) samples in the batches processed. When a new, or revised, analytical method is introduced, laboratories should do a validation study to compare the quality performance with other (similar) methods, previous versions of the procedure and materials with reference and consensus results.

An extended guide to the validation of methods, consistent with international standards such as ISO/IEC 17025, is given by EURACHEM (2015). It includes validation and verification methods as well as a number of performance characteristics including random and systematic error, limits of detection, and limit of quantification. For laboratory procedures, the latter two characteristics are used to indicate the limit below which the detection of an analyte becomes problematic, respectively the lowest level of analyte that can be determined with acceptable performance. Unfortunately, many laboratories do not include these measures in their quality statements with the data they distribute even though detailed validation reports may be available. These aspects complicate the processing of soil information obtained from different data providers in databases such as WoSIS, hence sometimes necessitating adoption of pragmatic solutions when processing the source data.

Adequate quality management in a laboratory is a prerequisite for reliable results and ‘data fit for use’. However, it should be noted that the contribution of laboratory error is not necessarily the major component of the total error in derived interpretations; spatial variability can contribute even more (Goodchild, 1994; Goodchild and Gopal, 1989; Heuvelink, 2014). An indication for the presence of other error sources can be found in the difference between the nugget in a variogram and the smaller values for lab error from validation and comparable experiments (Heuvelink, 1996).

While cost-efficient and cost-effective procedures for field sampling are often well described (De Gruijter et al., 2006; Louis et al., 2014), less attention is paid to quality requirements for laboratory investigations. They are often copied from previous and similar studies by applying the same methods. If for practical reasons alternative methods have to be selected, it should be remembered that numerous soil properties are based on ‘operational definitions’ (SSDS, 2011) and may apply only for specific user groups. That is, the property is best described by the details of the (laboratory) procedure applied. An example is the ‘pH of the soil’, which needs information on sample pretreatment, soil/solution ratio and description of solution (e.g. water, KCl 1M) to be fully understood. In WoSIS, soil properties also are defined by the analytical methods and the terminology used, based on common practice in soil science. As noted before, if highest laboratory accuracy is important it should be included in the selection criteria as well.

Two other examples where the description of soil analytical methods is particularly important for selection of alternative methods are cation exchange capacity (CEC) and available Phosphorus. The capacity of a soil to adsorb and exchange cations from exchange sites depends importantly on the actual pH and the ionic strength of the solution. However, the need for a sufficiently detailed description of analytical procedures is particularly reflected in the case of so-called plant ‘available phosphate’, where the choice of the appropriate laboratory methods is largely determined by soil pH as a proxy for soil mineralogy and soil type (Elrashidi, 2010). Hence, ‘vague’ descriptions for available-P methods are basically useless, unless used in a specific context such as a (local) fertilizer recommendation scheme. For example, correlation studies have shown that only in specific cases (i.e. soils and intended use) region-specific conversions can be made for available-P values determined according to different analytical procedures, such as P-Olsen and P-Bray (Mallarino, 1995), modified P-Morgan and Mehlich III (Ketterings et al., 2002), making international harmonization of results of such methods cumbersome or possibly at best ‘broad brush’. Examples of such efforts include the work of Ciampalini et al. (2013) in Tunisia and those of Maire et al. (2015) at the global level. According to GlobalSoilMap (GSM, 2013) there is generally no universal equation for converting from one analytical method to another in all situations. Within the framework of the Global Soil Partnership (Baritz et al., 2014), for example, this would imply that each regional node would need to develop and apply node-specific conversions (towards the adopted standard methods and soils), building on comparative analyses of say archived samples (Jankauskas et al., 2006). In this context, it is encouraging to note that in 2017 the Global Soil Laboratory Network (GLOSOLAN) was established by the Global Soil Partnership [1].

C.3 Standardization of soil analytical method descriptions

Lacking detailed quantitative information on the quality of the soil analytical data held in the diverse source databases shared for use in WoSIS, it was necessary to develop a qualitative procedure to describe the analytical methods in a flexible, yet comprehensive and consistent way. For all source data, as indicated earlier, it is assumed that the quality requirements of the (first) user are met and basic quality checks and screening have taken place and soil-relevant options in the procedure are applied in the source laboratory. This allows users of WoSIS to make their own judgement on the quality of individual data, for instance by the assumption that selected data have comparable quality characteristics or an acceptable (inferred) quality compared to their requirements.

For practical reasons, the options selected for the lab methods in WoSIS are assigned on basis of the descriptions provided in the respective (database) sources. This implies that information interpreted from the original report (source materials) is used here. At a later stage, however, some refinements may be possible if the original data can be consulted again; typically, such would be the task of the original data providers. Such a mechanism would be realisable once an inter-operable, federated system is in place as foreseen in the broader framework of SDF of the Global Soil Partnership.

The WoSIS method for the qualitative description of analytical methods can be seen as complementary to method descriptions used in reports from proficiency tests. In these cases, results from participants are coded to provide details of the methods applied within a particular grouping (WEPAL, 2015). As explained above, the spread of these results may be an indication for the spread in a compiled database.

Soil property ‘pH KCl’ will be used as example here. The selected options within WoSIS are sample pretreatment, the soil/solution ratio, the molarity of the KCl solution, and the measurement technique (see Appendix C.4). It is assumed that each laboratory, for the particular soils investigated, uses a shaking method and an equilibrium time long enough for the measurement to get a stable reading. These conditions may differ per soil type and (pair of) electrode(s) used, but these are considered of minor importance for differentiating methods in the WoSIS database (Table C.1). Once an option is identifiable, based on the available (source) information, the appropriate option / value is added (i.e. 0.1, 0.5, 1 M). Such a grouping allows users of the database to select those data that are analysed according to defined (and comparable) methods and may be judged as having equal quality as well as those that are suited for specific use. When new data are entered, the table is used for describing (coding) the added data. If necessary, values / options not yet considere so far can be added. As indicated, additional soil properties and options for methods will be added gradually in future versions of WoSIS.

In addition to the method description according to the standardized coding system, values have been allocated for the inferred confidence in the conversion; this qualitative assessment is based solely on the information embedded in the ‘summarized’ method descriptions as provided in the various source databases. As indicated, these descriptions were often generalized by the data providers themselves from a more detailed source, such as ‘their’ laboratory manual. Importantly, the present confidence flags should not be seen as a measure for the quality of a particular laboratory. The rationale and criteria for coding ‘standardized analytical methods’ in WoSIS, as developed so far, are presented in Appendix D with the corresponding flowcharts presented in Appendix E and coding in Appendix F.

Future versions of the WoSIS Procedures Manual will extend the description of analytical methods to accommodate data sets derived from proximal sensing methods.

Soil sensing methods rely strongly on conventional laboratory methods for their calibration as expressed in region-specific soil spectral libraries (Viscarra Rossel et al., 2016) with the accompanying statistical models. To allow for international comparison of such ‘derived’ data, the underpinning soil analytical methods must also be described in WoSIS.

---

[1] `http://http://www.fao.org/global-soil-partnership/resources/events/detail/en/c/1037455//`

Criteria for this will be visualized as a ‘branch’ at the top level of the present flow charts in Appendix E and further extension of the tables presented in Appendix F.

C.4 Worked out example (soil pH)

As indicated, when selecting (alternative) laboratory methods for specific uses or data for further use, it should be remembered that numerous soil properties are based on ‘operational definitions’ (SSDS, 2011). That is, the property is best described by key elements of the (laboratory) procedure applied. Such an approach has been developed for WoSIS; the procedure is illustrated below using pH as an example.

Table C.1: Procedure for coding standardized analytical methods using pH as an example.

Key	*Code*	*Value*
sample pretreatment	0	unknown
sample pretreatment	1	sieved over 1 mm sieve
sample pretreatment	2	sieved over 2 mm sieve
solution	0	unknown
solution	1	water [H_2O]
solution	2	Calcium chloride [$CaCl_2$]
solution	3	Potassium chloride [KCl]
solution	4	Sodium fluoride [NaF]
concentration	0	unknown
concentration	1	not applied
concentration	2	0.01 M
concentration	3	0.02 M
concentration	4	0.2 M
concentration	5	1 M
ratio	0	unknown
ratio	1	1:1
ratio	2	1:2
ratio	3	1:2.5
ratio	4	1:5
ratio	5	1:10
ratio	6	1:50
ratio	7	saturated paste
ratio	8	slurry
ratio base	0	unknown
ratio base	1	weight / volume
ratio base	2	volume / volume
instrument	0	unknown
instrument	1	electrode
instrument	2	electrode (field measured)
instrument	3	indicator paper (field measured)
monitoring	0	unknown
monitoring	1	not applied
monitoring	2	oxidizible sulfur compounds; initial pH
monitoring	3	oxidizible sulfur compounds; pH stabilized in 10 days, pH 0.1 unit, for two days

Major characteristics of commonly used methods for determining a given soil property are identified first, based on a detailed review of available soil laboratory procedures manuals. For soil property pH, for example, these are the sample pretreatment, extractant solution (water or salt solution), and in case of salt solutions the salt concentration (molarity), the ratio base (e.g. v/v or w/w) and the soil/solution ratio. A further descriptive element is the type of instrument used for the actual laboratory measurement.

Next, for each of the options per method, specifications that are used in data descriptions or known from reference laboratory manuals are tabled. For soil property ‘pH’ and *feature_name* soil/solution ‘ratio’, the available options range from ‘unknown’ to ‘saturated paste’ and slurry (Table C.1).

The above approach for describing laboratory methods (data) in WoSIS allows for flexible and straightforward database queries, as required in the framework of interoperable, distributed systems (Chapter 4).

Appendix D

Rationale and criteria for standardizing soil analytical method descriptions

D.1 General

D.1.1 Background

This appendix explains procedures and WoSIS conventions for describing and coding soil analytical method descriptions. First, the general procedures are explained. Subsequently, details are provided for each soil property considered so far in the WoSIS standardisation process (Table D.1), starting with a short introductory description of main aspects of each analytical method (Section D.2 to D.11). All measurement values in WoSIS are expressed using SI units or non-SI units accepted for use with the International Systems of Units.

Table D.1: List of soil properties for which soil analytical methods descriptions have been standardized.

Soil property	*Standard units* [1]	*Decimals*
Bulk density	kg/dm^3	2
Calcium carbonate equivalent	g/kg	1
CEC	cmol(c)/kg	1
Coarse fragments	100 cm^3 / cm^3	1
Electrical conductivity	dS/m	1
Organic carbon	g/kg	1
pH	unitless	1
Sand, silt and clay fractions	g/100 g	1
Total carbon	g/kg	1
Water holding capacity [2]	100 cm^3 / cm^3	2

---

[1] Conversions: g kg^{-1} or promille (1 = 0.1%); vol% is equivalent to 100 x cm^3 cm^{-3}; wt% is equivalent to 100 x g g^{-1}; kg dm^{-3} is equivalent to g cm^{-3} or Mg m^{-3}; dS m^{-1} is equivalent to mS cm^{-1}, originally mmho cm^{-1}, at 25°C; cmol(c) kg^{-1} is equivalent to meq 100g^{-1}. Layer depth (top resp. bottom) expressed in cm, measured from the surface, including organic layers and mineral covers (see Section 3.2.5).

[2] Water holding capacity may be calculated as the amount of water held between 1/3 bar and 15 bar (USDA conventions) (SSDS, 2014). At a later stage, in case of missing measured data, this may be done using a range of pedotransfer functions (Botula et al., 2014); procedures for this have been tested by ISRIC in the framework of the soilIE experiment (OGC, 2015).

D.1.2 Guiding principles

In WoSIS, the description of analytical results considers key aspects of methods that are widely used in soil laboratories for the particular soil property. These are tabled along with available options with a concise description for each option; the coding process is visualized in a series of flowcharts or Figures (Appendix E). The building blocks of a given flowchart make reference to both the feature and the number of possible options. Additional information for each method is given in sections Section D.2 to Section D.11 and in Appendix E and Appendix F.

Laboratory results in publications and reports are generally labeled with information that can be found in laboratory manuals, yet the full information is seldom provided in the databases themselves. It is assumed here that all source data (as shared for consideration in WoSIS) have passed routine quality control procedures at the source laboratory and a rigorous check by the first user. If during that process the decision was taken to include or exclude a special treatment, the assumption in WoSIS is that the original decision was made correctly (i.e. that the given option was deliberately left out from the selected criteria). Examples for this are washing out of water soluble salts in the procedure for the determination of the CEC, and use of non-dried materials in case of samples with volcanic ash. It is also assumed that when a laboratory method has characteristics that restrict its application to specific soils, these have been duly pondered upon by the source laboratory. In principle, such checks form an integral part of quality management procedures in a laboratory (van Reeuwijk and Houba, 1998).

As indicated, WoSIS is underpinned by data from many sources and these consider different laboratory standards and approaches; as such, it is not possible here to include all possible error components from these sources. Therefore, in WoSIS the number of features has been limited to a practical level keeping in mind the intended use of the WoSIS data base.

D.1.3 Methodology

Soil analytical method descriptions are described using a sequential approach. First, the soil property matching the label of the source data to be entered in WoSIS needs to be selected. This must be done with utmost care: some soil property 'labels' in the source database can include information on the laboratory method applied; they are 'operational' definitions, for example 'CEC-NH_4OAc'.

Once the appropriate soil property is identified, method characterization should be according to the corresponding option table in Appendix F.

Next, the option that best describes the particular element of the laboratory method is identified and coded. This step may be 'tricky' when incomplete information has been provided in the source metadata; in such cases it may be necessary to contact the data providers for additional information. In such cases a lower confidence level is assigned based on an evaluation of all features (from high (1) to low (3)) [3]. This step, however, is not shown in the option table and the flowcharts.

Results of soil analytical analyses are expressed with reference to a base, known as 'fine earth fraction' of a soil sample. This fraction is prepared during sample pretreatment at the source laboratory upon air drying. Next, part of the air-dried material is sieved. The actual sieve size determines both the upper limit for the fine earth fraction and the lower limit for the coarse fraction. Most countries use 2 mm as the upper limit for the fine earth fraction (SSDS, 2014). Alternatively, in the former Soviet Union and satellite countries, laboratories used 1 mm for this (Katschinski, 1956). Analytical results expressed with respect to the <1 mm and <2 mm fractions as base are essentially different. Such aspects need special attention where international datasets are to be fully harmonized, for example using transformation functions (Nemes et al., 1999; Rousseva, 1997), and are worthy of consideration in international laboratory intercomparison programmes such as GLOSOLAN[4].

Similarly, sub-divisions (name and size limits) within the defined coarse and fine fractions can vary markedly between countries (Figure D.1) (Nemes et al., 1999). Further, as observed by Loveland (2016),

---

[3] In case a low confidence rating is specified in the standardized data sets, national data providers are encouraged to help ISRIC WDC-Soils updating the corresponding descriptions as they may have access to more detailed sources, such as the actual laboratory manuals.

[4] `http://www.fao.org/global-soil-partnership/resources/events/detail/en/c/1037455/`

interchangeability of sub fractions also is restricted by differences in quality of say the pipet-method and the grainsize distribution determined using diffraction.

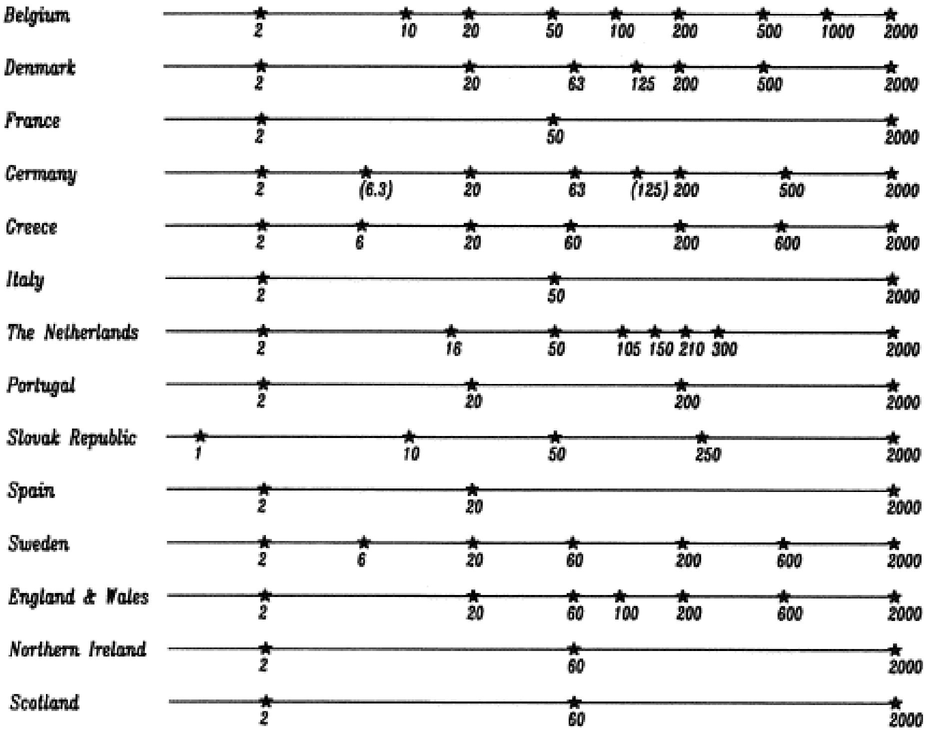


Figure D.1: Range in textural definitions as used in Europe.

The sample pretreatment is defined in the feature-option table for relevant soil properties only. For example, in the case of bulk density methods, the pretreatment is considered only in the 'clod' method.

D.1.4 Example for the description of analytical and laboratory methods

The flexibility of the WoSIS approach for coding soil analytical method descriptions is illustrated in Table D.2, using soil pH as an example. The approach has been used to code or describe four different reference methods for pH KCl: ISO 10390:2005 (ISO, 2005), USDA (SSDS, 2014), ISRIC (van Reeuwijk, 2002), and (WEPAL, 2015).

Table D.2: Grouping of soil analytical methods for soil pH according to key criteria considered in ISO, ISRIC, USDA and WEPAL laboratory protocols (Example for KCl solutions).

	Procedure			
Key	*ISO*[5]	*ISRIC*[6]	*USDA*[7]	*WEPAL*[8]
Pretreatment	<2 mm	<2 mm	<2 mm	<2 mm
Solution	KCl	KCl	KCl	KCl
Concentration	1 M	1 M	1 M	1 M
Ratio	1 : 5	1 : 2.5	1 : 1	1 : 5
Ratio base	v/v	w/v	w/v	v/v
Instrument	Electrode	Electrode	Electrode	Electrode
Monitoring	Not applied	Not applied	Not applied	Not applied

D.2 Bulk density

D.2.1 Background

Bulk density provides a measure for soil compaction. It is defined as the ratio of the mass of oven dried solids to the total or bulk volume. In contrast to particle size density, soil bulk density includes both the volume of solids and pore space. Pore space depends on the soil‘s extensibility and so the volume of soil changes almost dynamically with soil water content. Bulk density is needed to convert data from weight to a volume basis and *vice versa*.

D.2.2 Method

Sampling techniques for bulk density are determined by the natural setting of the solid soil components and the pores. Therefore, the moisture condition has to be defined at the time the bulk density is estimated. Measurements can be based on weight and volume.

Methods for determining bulk density are described using four features (see also Appendix E and F):

Sample type: In one group of methods, a sample is taken by pressing cores with known volume (rings) into the soil. Other methods use clods taken in the field; after sealing, the volume of the coated clod is measured by submerging into water (nowadays, this can also be done using 3D scanning). Alternatively, in some cases, clods are reconstituted from soil material <2 mm (SSDS, 2014). If soil coherence is too weak to prepare a clod or take a soil core, the volume of excavated soil is measured instead.

Measurement condition: The moisture condition of the sample determines whether swelling or shrinking has taken place (e.g. air dried, rewetted).

Base: The soil bulk density of a sample is the ratio of the mass of solids to the total or bulk volume. The total volume includes both solids and pore space. Bulk density is usually reported for the ‘<2 mm’ soil fabric, the mass and volume of rock fragments are subtracted from the total mass and volume (SSDS, 2014).

Corrections: Bulk density is calculated for soil materials <2 mm in g cm^{-3} (i.e. fine earth fraction). Mineral materials larger than 2 mm and vegetal materials are removed from the sample. If the density of rock

---

[5] ISO 10390:2005 specifies an instrumental method for the routine determination of pH using a glass electrode in a 1:5 (volume fraction) suspension of soil in water (pH in H2O), in 1 mol/l potassium chloride solution (pH in KCl) or in 0.01 mol/l calcium chloride solution (pH in CaCl2) (ISO, 2005); this coding example is for pH KCl.

[6] ISRIC: Method 4-1 for pH-KCl (van Reeuwijk, 2002).

[7] USDA: Method 4C1a2a3 (SSDS, 2014).

[8] WEPAL Wageningen Evaluating Programs for Analytical Laboratories `www.wepal.nl`. WEPAL is an accredited world-leading organiser of proficiency testing programmes in the field’s of plants, soil, sediments and organic waste. Participants in the International Soil-Analytical Exchange programme receive 4 times a year, 4 samples to be analysed for comparison of results. Participants describe the applied extraction / digestion, and the method of detection of the particular element of their method applied.

fragments is known or can be approximated, corrections can be included in the calculations of the bulk density (SSDS, 2014).

Sometimes bulk density is expressed on the basis of the whole soil. This refers to all particle-size fractions, including boulders with maximum horizontal dimensions less than those of the pedon (Grossman and Reinsch, 2002; SSDS, 2014).

Calculation: When not actually measured, bulk density can be based on expert knowledge. In such cases, 'calculation' is flagged as 'unknown'.

D.3 Calcium carbonate equivalent

D.3.1 Background

Carbon in soils can be subdivided into organic Carbon and inorganic Carbon. Inorganic carbon is present in various mineralogical materials such as carbonates (e.g. $CaCO_3$, $MgCO_3$, $CaMg(CO_3)_2$. A convenient solution in the laboratory is to express results for inorganic carbon forms as if originating from calcium carbonate, i.e. as 'Total Calcium Carbonate Equivalent'. Most soils with a pH less than 7 will not contain any significant amounts of carbonates.

Inorganic carbon can be quantified by dissolution of carbonates in acid and titration of residual acid, or measurement of evolved CO_2 by volume, pressure or absorption on a solid or precipitation in a solution. Instrumental methods use sensors for CO_2 in purified streams of evolved gases with acid treatment or burning at elevated temperatures. An alternative is to measure the weight loss with burning at selected temperatures. For a detailed description of available options see SSDS (2014).

D.3.2 Method

Analytical methods to determine calcium carbonate equivalent in soils are described using 4 features; the corresponding flowchart and feature table are presented in Appendix E and F.

Sample pretreatment: Analyses are for the fine earth fraction using the limits defined by the source laboratory, that is <1 mm (Katschinski, 1956) or <2 mm.

Reaction: Quantification of carbonates often is based on dissolution in acids. The choice of the acid depends amongst others on the expected amount of carbonates, and the subsequent detection technique. The concentrations should be strong enough to get dissolution in a reasonable time, but exclude the risk of and error by acid fumes.

Temperature: At higher temperature dissociation of acids and so dissolution of carbonates increases. Possible options are described in the feature table, for example 'dissolution of carbonates by sulphuric acid [H_2SO_4]'.

Detection: Where large amounts of carbonates are present, titrimetric methods for the excess of acid after dissolution or the volume of CO_2 evolved at dissolution can be used. Evolved CO_2 can also be quantified by increased pressure in a fixed volume. Alternatives are gravimetric methods with precipitation of CO_2 or weight loss from the sample. If the evolved stream of gases is purified, sensors sensitive to Carbon forms can be applied as well. These instrumental methods (element analyzers) usually are applied only with smaller amounts of (inorganic) carbon / carbonates in the sample (SSDS, 2014).

Calculation: Many laboratories for soil analyses do not have methods that can discriminate inorganic C from different sources. Most methods are based on acid consumption with dissolution of carbonates, or quantification of CO_2 evolved. An indirect method is the subtraction of total organic carbon (TOC) from total carbon (TC). These forms are for instance determined by element analysers with and without prior acid treatment for removal of carbonates. In all cases the quantified amounts have to be converted to the same molecular amounts of $CaCO_3$.

D.4 Cation exchange capacity

D.4.1 Background

Micro- and macro-nutrients and heavy metals are held and released into the soil solution from exchange sites associated with clay and organic matter fractions; cations are held by the negative charges at the exchange sites. The electric charge that can develop at these sites varies with clay mineralogy, type of organic matter and the pH (Dixon et al., 1977). The cation exchange capacity (CEC) is an estimate of the total amount of cations held at these exchange sites. The 'effective CEC' is the CEC determined at the pH of the soil.

Salts present in the soil are considered a separate resource of ions and have to be removed by washing prior to analyses of the saturation status of the exchange complexes. If present in high amounts, sodium and calcium from carbonates, and gypsum can dominate the saturation of the exchange sites. Part of these soil components may dissolve in the solutions used (pH $\leq$ 8.2 to 8.5) preventing complete saturation of exchange sites in the lab procedures as described below.

Usually, before the determination of the CEC, the amount and type of exchangeable bases are determined first. These 'exchangeable bases' are: Calcium, Magnesium, Sodium and Potassium. From these, the bases saturation can be calculated as the ratio of exchangeable bases and the cation exchange capacity (at pH 7).

D.4.2 Method

Various methods are used to determine CEC. In practice, these methods are selected considering available lab facilities, soil types to be investigated and specific (research) requirements. CEC methods are described here according to their main characteristics; detailed information about specific methods, including their limitations, may be found in underpinning laboratory manuals. In the standardization process for WoSIS, CEC methods are first subdivided on the basis of the technique applied (i.e. direct and approximate (indirect) methods). These categories, as visualized by options 1 (middle column) and 2 (right column) in Appendix E are subsequently refined using 3 respectively and 5 features.

Sample pretreatment: This feature indicates whether data are reported for a fine earth fraction defined as having an upper limit of 1 mm (Katschinski, 1956) or 2 mm (most common).

Technique: As a preceding step to the determination of the total amount of cations that can be held in soil often the actual saturation of the exchange sites with basic cations is determined. This amount is known as 'exchangeable bases'. Typically, these cations are exchanged by a (buffered) solution with a cation not present in the soil (e.g. NH_4^+ which is not considered a 'base cation'). If exchangeable acidity (sum of exchangeable Hydrogen and Alumina) is also needed, the actual saturation of the exchange complex with cations can be estimated (indirect method), although minor fractions of manganese and iron are ignored in the procedure. These techniques with summation (of groups of) exchangeable cations are considered here as indirect methods for the determination of the CEC.

Reported pH: The release of cations from and the saturation of the exchange sites depend on the charge of the exchange surfaces, as determined by clay mineralogy and type of organic matter. These charges will vary with the pH of the exchange solution (e.g. buffer) as applied in the laboratory procedure. This effect is particularly for soils with variable charge components (e.g. volcanic materials) and high organic matter content. Conversely, soils with a permanent charge show only small differences in CEC with increasing lab pH (Dixon et al., 1977). Conventionally, the 'CEC' is determined at pH 7 ('operational definition') that is with use of solutions buffered at pH 7; this CEC value then is considered representative for most agricultural soils. If the CEC at field pH is needed, then unbuffered solutions are used for saturation.

Exchange solution: 1 N KCl and 1 N NH_4Cl are considered un-buffered solutions; they can be used for the determination of exchangeable acidity in soils with pH less than 5.5. Alternatively, the potentially available acidity is determined by extraction with a solution buffered at pH 8.2 such as $BaCl_2$; the change in acid 'consumption by the buffer' is measured by back titration of the solution. This amount (i.e. potential acidity) is usually higher than the amount of field exchangeable acidity.

The total amount of cations that can be held in soil can also be calculated in a *direct* way: the soil sample is saturated by applying an excess of a suitable index cation with subsequent determination of the excess or replacement of the index cation by another cation.

In the direct methods, used to achieve complete saturation of exchange sites at a particular pH, buffered solutions that contain a cation with high selectivity are used to saturate all exchange sites. The cation can be regarded as an index cation if the cation is used to quantify the exchange sites. The index cation can be a cation from a buffer component (i.e. Na^+ or NH_{4+}). Cations that can be considered absent in the soil (i.e. Ba^{2+}, Ag-thiourea, and Li^+) or have been exchanged in previous treatments (i.e. Na^+) can also be index cations. To improve exchange often solutions with a high concentration of the index cation and replacing cation have to be used. The selection of an index cation, is not only based on the selectivity of the ions at the exchange sites, but also on the instrumentation available for quantification.

To enhance the saturation with an index cation, samples can be pre-leached with a buffered solution as well. The small amount of buffer solution and cation(s) left improves the selectivity of the analysis.

Further information may be found in Baize (1993); Pansu and Gautheyrou (2006); SSDS (2014); ISO (2005).

Replacement solution: Once the exchange sites are saturated the excess of the solution used for the exchange is washed out. In a next step the index cation is brought into solution by a replacement solution (buffered) at the particular pH. This is used for instance in the well-known CEC procedures with $BaCl_2$ Triethanolamine (TEA) for (Ba^{2+}) buffered at pH 8.2, and NH4-acetate/Na-acetate at pH 7 and similar methods.

This (washing and) replacement step can be omitted if the amount of index cation removed from the (buffered) index cation solution is estimated. For instance if solutions with AgThiourea (+) unbuffered or buffered from 4.0 to 8.0 are used.

Index cation: The index cation criterion further specifies the procedure for the determination of the CEC. The index cation can be a cation from a buffer solution or a cation added to the buffer solution. The latter group of cations are generally not common in soils or cations with a high selectivity. Their selection may depend on the clay minerals and type of organic matter present in a given sample. For some of these cations sophisticated instruments like flame atomic absorption and inductively coupled plasma spectrometers are needed for quantification.

Indirect method (CEC by Approximation): Cations that may be present on the exchange sites can be subdivided into three groups. Through summation of the cations in these three groups, the cation exchange capacity can be approximated in an indirect way. In most agricultural soils, the exchangeable bases Calcium, Magnesium, Sodium and Potassium dominate. In case of soil with a pH <5 part of the exchange sites will be saturated with Alumina and Hydrogen. Iron and Manganese are often the most prominent in the 'rest group'. The cation exchange capacity can be approximated by summation of these groups of cations (Pansu and Gautheyrou, 2006; SSDS, 2014).

Exchangeable bases and exchangeable acidity can be determined in buffered and unbuffered solutions. If determined in unbuffered solutions the CEC at (field) soil pH is approximated. It should be noted that the CEC based on summation of individual cations may be less accurate than values for CEC determined using a direct determination.

D.5 Coarse fragments

D.5.1 Background

As indicated, the solid part of the soil mass is subdivided by sieving into a fraction smaller and a fraction larger than 2 mm; these fractions are often denoted as the fine earth and coarse earth fraction. Data sets from the former Soviet Union and its satellites, however, often use 1 mm as upper the limit (Katschinski, 1956) for the fine earth fraction.

The amount of coarse fragments is needed in calculations that consider the total soil mass as reference. Internationally, different systems are used for naming subclasses within the coarse fraction, using a range of class limits and names (Nemes et al., 1999; FAO, 2006a; ISO, 2016; SSDS, 2014).

Alternatively, the fine earth fraction is most relevant for soil processes. This fraction provides the base material for most soil physical, chemical, and mineralogical analyses; see section on particle size distribution and analysis (sand, silt and clay fractions).

D.5.2 Method

Methods for the determination of mass of coarse fragments are described using the following features (see Appendix E and F):

Sample pretreatment: This feature indicates whether data are reported with reference to an upper limit of 1 mm (Katschinski, 1956) or 2 mm for the fine earth fraction.

Size: The coarse fraction can be subdivided into several subclasses. Typically, their respective amounts can best be estimated or weighed in the field or laboratory.

Type: The coarse fraction can comprise different types of materials which can be described according to their size and nature, as e.g. determined by their resistance against disintegration in water or sodium-hexametaphosphate. During sample pretreatment, pedogenetic materials have to be carefully removed from the sample and treated as a separate sub fraction. Breaking up of may lead to the release of smaller ferromanganese and calcium carbonates (and barium sulphate or gypsum concentrations) nodules.

Estimates: Basically there are two ways to express the amount of coarse fragments: on volume (V) basis and on weight (W) basis. For conversions from V to W the bulk density is required. In the absence of such data, default data for bulk density (1.45 g cm^3) and mineral particle size density (2.65 g cm^3) are often assumed (SSDS, 2014).

Base: The amount of respective coarse fractions can be expressed as part of the whole soil or related to e.g. the fine earth fraction or other fractions determined.

D.6 Electrical conductivity

D.6.1 Background

The electrical conductivity (EC) of a soil extract is a measure for the salt content in the solution (salinity of soil). EC is a measurement that correlates with soil properties that affect crop productivity, including soil texture, cation exchange capacity (CEC), drainage conditions, organic matter level, salinity, and subsoil characteristics. The Saturation extract (ECe) is considered to give a better representation of actual soil conditions with respect to plant environments (SSDS, 2014).

D.6.2 Method

Electrical conductivity is described using four features: sample preparation, solution, ratio and instrument (see corresponding flowchart and feature table in Appendix E and F).

Sample pretreatment: The analyses apply to fine earth fraction as defined in the source laboratory (i.e. $<$1 mm (Katschinski, 1956) or $<$2 mm.

Solution: Electrical conductivity is measured in soil extracts prepared with distilled or demineralized water (i.e. with very low electro conductivity).

Ratio: The amount of salts that can be dissolved depends on the soil / water ratio used (e.g. 1:5). These ratios are often predefined in soil classification systems and soil fertility evaluating schemes. With smaller ratios, EC can easily be determined in the supernatant solution after sedimentation of coarser materials. Saturated paste solutions are close to the soil solution. Often the pH H_2O and EC are measured in the same soil extract (soil/water ratio).

Instrument: EC is measured using an electro conductivity electrode. This can be done in the laboratory and in the field using hand held instruments (Rhoades et al., 1999; SSDS, 2014).

D.7 Organic carbon

D.7.1 Background

Carbon in soils is divided over several fractions and components. Major fractions are organic and inorganic carbon. Inorganic carbon is present in mineralogical materials such as carbonates. Cells of microorganisms, plant and animal residues at various stages of decomposition, stable 'humus' synthesized from residues, and nearly inert and highly carbonized compounds, such as charcoal, graphite, and coal ((Nelson and Sommers, 1982), p. 347), all contain organic C. Organic materials coarser than 2 mm are considered residues. The 'fine earth' fraction, is the basis for the laboratory analyses for soil carbon.

The preferred procedure for determining total organic and inorganic carbon should consider an element analyser that allows for controlled temperature and detection based on CO_2 evolution. Other methods often require a correction for lower recovery (Lettens et al., 2007).

D.7.2 Method

Methods used for the determination of organic carbon are described with 6 features, as described in the corresponding feature table and flowchart (see Appendix E and Appendix F).

Sample pretreatment: This feature indicates how a laboratory has defined the upper limit for the fine earth fraction, as discussed earlier: <1 mm (Katschinski, 1956) or <2 mm. These values also set the lower limit for the coarse fraction.

pretreatment: Carbon in the fine earth fraction can be determined by oxidation and quantification of CO_2 released. If the sample contains also carbonates (i.e. inorganic carbon) this fraction can liberate CO_2 as well. This inorganic fraction can be removed by acid treatment (without oxidation power) prior to the actual determination of organic carbon.

Reaction: Generally, organic carbon can be determined by wet or dry oxidation (ISO, 1995; SSDS, 2014; van Reeuwijk, 2002). For wet oxidation, mixtures of bichromate and sulphuric acid are frequently used. Determination can be based on excess bichromate or CO_2 development. Methods based on CO_2 detection need sample pretreatment to remove carbonates. CO_2 is also liberated with dry oxidation and quantified in element analyzers.

Temperature: In both wet and dry oxidation methods the completeness of the reaction depends on the temperature. When the temperature is too low during wet oxidation (should be about 125 °C), the oxidation of organic materials is not complete; external heating is applied to achieve higher recovery fractions for organic Carbon. Thermal decomposition of bichromate may occur when external heating cannot be limited and controlled during wet oxidation. Dry oxidation requires a furnace with controlled temperature of over 900 °C.

Detection: Excess of bichromate can be determined by titration with e.g. Fe^{2+} and colorimetric techniques. Wet oxidation methods with detection based on excess bichromate are hampered by interferences from charcoal, Fe^{2+}, Mn^{2+} where present.

In a furnace, with extra oxygen and catalysts added to complete oxidation to CO_2 and purification of the gas stream evolved, total C (organic C and inorganic C) can be detected. Gas chromatography, coulometric methods and/or infrared-based methods can be used to quantify CO_2 evolution. Inorganic Carbon should preferbly be removed by prior acid treatment.

Weight loss on ignition (LOI), which involves ignition at 400 C in a furnace, is a measure for the content of organic matter present in the fine earth fraction.

Calculation:

- Dry oxidation: Total carbon (TC) is defined as the sum of total organic carbon (TOC) and total inorganic carbon (TIC), where TIC is derived from the calcium carbonate equivalent or from a direct determination (as measured for basic soils). Methods to assess TIC by prior treatment with acid or separate determination are covered in the section D.3.
- Wet oxidation: Procedures based on the original method of Walkley and Black (1934) often involve incomplete oxidation (Lettens et al., 2007; SSDS, 2014) and consider a recovery factor. Generally, this factor is based on the assumption that 77% of the total carbon present is determined during the analyses. However, in practice, the actual factor will depend on the type of soil organic matter present in the sample and soil type (Grewal et al., 1991; Lettens et al., 2007). Use of a correction factor that is not representative for the soil material under consideration will introduce a systematic error. Only few methods with wet-oxidation, with or without external heating, claim 100% recovery (for all soil types); for a critical review see Pribyl (2010).
- Loss on ignition (LOI): Generally, soil organic matter is assumed to contain 1.724 times the amount of soil organic carbon (SOC) (van Bemmelen, 1890), but the actual factor may be up to 2.0 for topsoil's and even 2.5 for subsoils depending on the type of soil organic matter (Pribyl, 2010).

D.8 Soil pH

D.8.1 Background

The soil solution reflects the composition of the soil. Therefore, the pH and electro-conductivity of a standardized soil water mixture are often used for the first discrimination of soils and soil layers. The difference in pH measured in a soil/water extract and extract with 1 N KCl (pH delta value) for instance is an indication for the potential acidity. This 'delta pH' should be measured with equal conditions for both measurements. For some analytical methods, the pH determines the sample pretreatment for that method (i.e. presence of carbonates) or is a criterion for additional analyses. pH in saturated paste of soil, for example, is used in cross checking salinity data (Rhoades et al., 1999).

D.8.2 Method

Methods for the determination of the pH of soil samples are characterized using 7 features. The corresponding procedure is visualized in a flowchart in Appendix E with details provided in the corresponding option table in Appendix F.

Sample pretreatment: Refers here to the upper size limit adopted by a laboratory to define the 'fine earth fraction'. Two physical pretreatments are widely used internationally for this: a) sieve crushed, dried soil material through a 1 mm sieve, and b) idem, but using a 2 mm sieve. By implication, these values will determine the corresponding lower limit for the 'coarse fraction'.

Solution: To determine the pH of a soil sample, H^+ ions need to be brought into solution. Distilled water and solutions with low ionic strength are mainly used to stimulate this process. 'Stronger' salt solutions are needed to force exchange of i.e. Al^{3+}. Measurements in a saturated paste are aimed to represent the natural environmental conditions (Pansu and Gautheyrou, 2006).

Ratio: pH methods consider different 'soil / solution' ratios. The specific ratio determines the composition of the supernatant solution, leading to different pH values. More ions can dissolve in a larger volume until maximum solubility is reached for the particular combination of exchanged and dissolved ions. Agitation time and method of shaking, as well as place of measurement (i.e. 'in the rest' or 'actively stirred' portion) have to be standardized in a laboratory to establish consistent measurement conditions. To obtain reliable measurements for pH H_2O in soils with high organic matter content, usually a higher water: soil ratio is used (Pansu and Gautheyrou, 2006). Measurements in which electrodes are in contact with the sediment, may show a 'suspension' effect; this effect can modify results by +/- 1 pH unit.

Ratio base: Results can be expressed on a weight or volume basis; as such, ratios can be expressed as weight / volume (w/v) and volume / volume (v/v).

Concentration: Solutions with a high salt concentration enforce ion exchange processes. For instance, a 1 M KCl solution is used to release hydrogen ions and Al^{3+} ions from the exchange complex. With 1 M NaF, OH^- is released in the solution and Al_{3+} complexed; the observed increase in pH is an indication for 'active aluminum' (van Reeuwijk, 2002).

Instrument: In the laboratory, usually the pH is measured with a pH-meter. The (combination of) electrode(s) and their maintenance are controlled by the laboratory; they are checked before the measurement procedure starts with calibration of the instrument. As the position of the electrode in the solution is considered most important, the type of electrode used is not considered a key-criterion here. pH test strips and hand-held pH meters are also used to measure soil pH in the field; however, such pH values may differ widely from those recorded under standard laboratory conditions.

Monitoring: This term is used here solely with respect to observing (monitoring) changes in oxidizable sulfur compounds over time for which changes in soil pH are measured until equilibrium is reached, providing a measure for 'sulfidic soil materials' (SSDS, 2014). Such conditions commonly occur in intra-tidal zones adjacent to oceans.

D.9 Sand, silt and clay fractions

D.9.1 Background

The particle size distribution (PSD) in soils is important in a wide range of agricultural, geological, and technical and engineering studies. It influences soil properties such as the workability, water holding capacity and drainage, and the capacity to hold (micro) nutrients as a reservoir for plant growth. For most laboratory investigations samples are first passed through a sieve to determine the 'fine earth' fraction (Nemes et al., 1999; SSDS, 2014); the upper limit of 2 mm is widely used for this. Alternatively, in the former Soviet Union and its satellite countries, the upper limit for fine earth fraction has been set at 1 mm (Katschinski, 1956).

The 'coarse fraction' is always defined with reference to the upper limit of the 'fine earth fraction' in the corresponding scheme. Whole soil refers to all particle-size fractions, including boulders with maximum horizontal dimensions less than those of the pedon. In general, the term texture is used to describe the fine-earth fraction (SSDS, 2014). At one time, the term rock fragments was differentiated from the term coarse fragments, which excluded stones and boulders with diameters $>$250 mm (SSDS, 1975). The rationale for this distinction was that particles $<$250 mm were generally regarded as part of the 'soil mass' as they affect moisture storage, infiltration, runoff, root growth, and tillage (SSDS, 1951).

Soil particles in the fine earth fraction can be subdivided into fractions of different size classes, as illustrated in the corresponding flowchart in Appendix E. In most cases, the following (nominal) classes are defined: 'sand', 'clay', and an intermediate class named 'silt'. However, the actual size limits for these classes can vary between countries and laboratories, for example an upper limit for the 'silt' fraction given as 0.050, 0.060 or 0.63 mm. Such fractions, in turn, can be subdivided to provide additional detail; again, subclasses are defined using different size limits and (local) names, complicating overall standardization and harmonization efforts (Loveland and Whalley, 2000).

In view of the above, in WoSIS descriptive criteria for 'sand', 'silt', and 'clay' are presented in one single feature table and one flowchart (see Appendix E and Appendix F). During laboratory analyses, these fractions are often determined simultaneously using the same sample.

D.9.2 Method

Analytical methods for the determination of the sand, silt and clay (size) fraction in soil samples are described using 3 options for the sand and silt size fraction, and 4 for the clay size fraction. Brief explanatory texts for the particular features are provided as separate table (Appendix F). When the exact fraction sizes are not reported in the source metadata, of necessity the inferred class limits are inferred (with a flag for lower confidence).

Sample pretreatment: Refers here to the upper size limit adopted by a laboratory to define the 'fine earth fraction'. Two physical pretreatments are used internationally: a) sieve crushed, air dry soil material through a 1 mm sieve (Katschinski, 1956), and b) idem, but using a 2 mm sieve. By implication, these values will determine the corresponding lower limit for the 'coarse fraction', as well as the actual 'base' to which results of soil chemical analyses apply.

Size: Fractions for a given particle size distribution should be distinguished according to their size and not according to their allocated names as used in different (inter)national systems. This practice, facilitates logical grouping of size fractions in WoSIS according to the particle size distribution schemes as adopted in the various source databases.

Treatment: Soil particles are often bound into aggregates by cementing materials such as organic matter, calcium carbonate and iron oxides. Such aggregates are broken up during the pretreatment step of the analytical procedure. For example, hydrogen peroxide (H_2O_2) is used to remove organic matter by oxidation. HCl, HAc (acetic acid) or buffer solutions (Na acetate) are generally used to remove carbonates in basic pH soils (pH $>$7). Ultrasonic treatment is needed for samples that contain kaolinite, micas (e.g. illite and muscovite) and allophanic materials; the energy generated by the vibrations is transferred into the suspension and breaks the cohesion of the aggregates. In some laboratories, sesquioxides are removed in an optional deferration step. Many of these treatments are soil type specific. When such (pre)treatments and dispersion procedures (see below) are excluded from the laboratory procedure, measurement will be for 'water dispersable' or 'natural clay'.

Dispersion: During the dispersion of clay size particles, effects of the electric charges at the exchange surfaces have to be controlled carefully (SSDS, 2014). A clay suspension is most stable when the particles have large electrical double layers which is the case with monovalent cations and diluted solutions (Bolt and Bruggenwert, 1978). Chemisorption of phosphate at pH values far from the isoelectric point, can also reduce repulsion forces by charges on the clay surface. Therefore, a mixture of sodium hexa-metaphosphate buffered at pH 10 by Na_2CO_3 (also known as 'Calgon' type) is frequently used as dispersant. Alternative dispersants include ammonia and soda. Well-dispersed soil solutions remain turbid for a longer time than those that have not been submitted to any pretreatment.

Instrument: Subsequent to pretreatment and dispersion of the fine earth fraction, this fraction is divided into so-called sand, silt and clay fractions (see above discussion about class size limits and naming issues). In the '$<$2 mm' scheme, the sand fraction is separated from the silt and clay fraction by sieving; for example using 0.050 or 0.063 mm as the lower limit. Typically, the silt and clay fractions are determined after sedimentation in a cylinder based on Stokes' law, which assumes silt and clay particles to be perfect spheres. The rate of sedimentation depends amongst others on the radius, shape, density of the different particles, density of the liquid used for dispersion, and temperature (SSDS, 2014). Detailed descriptions of possible error components in results from particle size analyses can be found elsewhere (Loveland, 2016).

The main instruments for particle size analysis include:

- Pipette method: In the (Robinson) pipette method the relevant (i.e. $<$0.050 or 0.063 mm) fraction is sampled at a predefined depth and time after sedimentation started. The actual depth is calculated from Stokes' law for a specific settling time interval, size fraction and temperature. The pipette method is considered to be the reference method.
- Hydrometer: This method (often referred to as Bouyoucos hydrometer method) is commonly used as an alternative to the 'pipette' method. It is based on the density of the suspension, which is a function of the concentration and kind of particles present (after a certain time of settling). Results are considered to be less reliable than those obtained using the pipette method.
- Laser beam: These methods for grain size analyses are based on the measurement of the intensity of diffracted laser beams of near infrared / infrared light on suspended particles. The scattered light is detected at an angle (often 90°) of the incident beam. The number of detectors used, determines the number of fractions in the measurement range. With a high number of detectors, this results in a 'continuous' spectrum. Wavelength, size, refractive index of surface of particles are important factors in the calibration of the instrument. Results may differ from those obtained with the pipette method for the same samples (Buurman and van Doesburg, 2007; Loveland and Whalley, 2000).

- Field hand estimates: Trained personnel with field experience in a specific region or with specific soil types may estimate the clay content in a semi quantitative way by ‘hand, finger’ sensing. Texture classes assigned from field texture and laboratory particle size analyses are not always equivalent, especially where the clay content is considered to be high.

D.10 Total carbon

D.10.1 Background

Carbon in soils can be divided in an organic and inorganic fraction. Inorganic carbon is present in carbonate minerals such as calcite ($CaCO_3$) and dolomite ($MgCO_3$). Organic carbon is found in organic materials such as plant residues, humus, and charcoal. Total carbon (TC) is defined as the sum of total organic carbon (TOC) plus total inorganic carbon (TC, expressed as Calcium carbonate equivalent). From an analytical point of view, TOC and TIC are defined measurement categories, also referred to as ‘sum parameters’, whereas each part can contain several substances in varying proportions (Hilke, 2015).

D.10.2 Method

Analytical methods for total carbon are described using two options; additional details are provided in Appendix E and F.

Sample pretreatment: Analyses are for the fine earth fraction using the limits defined by the source laboratory. This can be $<$1 mm (Katschinski, 1956) or $<$2 mm, which may be seen as the present international standard for defining the coarse fraction.

Calculation: When small amounts of inorganic carbon are present, total carbon can be estimated by dry combustion at higher temperature using an elemental analyser (Kuhlbusch et al., 2009; SSDS, 2014). However, in practice inorganic carbon is removed first after which the organic carbon fraction is determined. Larger amounts of inorganic carbon usually are usually determined after dissolution of the mineral fraction with acids and release of CO_2. Results are expressed as total carbonate equivalent. Total carbon can be calculated using a range of methods for both inorganic and organic carbon.

D.11 Water retention

D.11.1 Background

The capacity of a soil to hold water (and air) depends on the amounts and types of organic matter, content of sand, silt and clay, as well as soil structure or physical arrangement of the particles. Water and air are held in the inter-connected pore spaces between the solid materials. If all these spaces are filled with water, the soil is at its maximum water holding capacity. The corresponding volume of water can be gradually depleted by plant growth, transpiration, evaporation and drainage. Water is less readily released by and transported through smaller sized pores. In some soils types, such as shrink-swell clays, the water holding capacity can change with desorption or absorption of water (hysteresis effect).

In the laboratory, water holding capacity at predefined suctions is determined by suction or application pressure on a sample. A detailed account is provided in soil physics handbooks (Jury et al., 1991; Koorevaar et al., 1983) and laboratory manuals (SSDS, 2014).

The nature of the soil material may preclude some sampling types. For example, water retention for expanding clays is overestimated when sieved samples are used instead of natural soil fabric at tensions of 6, 10 and 33 kPa (SSDS, 2014). Cores cannot be used when the soil material is not coherent enough, such as in very sandy soils.

When reporting values for water retention, it is very important to note whether results are expressed on a w/w basis or w/v basis. Conventionally, with pF curves the moisture content is expressed in volume % (w/v) rather than weight % (w/w); conversion between these units require the availability of bulk density data.

D.11.2 Method

Analytical methods for water retention are described using six options; additional details are provided in Appendix E and F.

Tension: The capacity of a soil to hold water is often expressed as the water content determined after equilibration of samples, from different depth layers, with water at various suction values (kPa). For practical reasons, high suction values are often controlled by tension in pressure pans. Desorbed or absorbed water between these equilibrium points and the initial saturated condition is generally measured on a mass basis (W). Water content at selected tension points can be expressed in different measurement units, as follows:

1 bar = 100 kPa (kilo Pascal)

= 1020 cm H2O $\approx$ 1000 cm H2O

= 75.01 cm Hg

= 0.9869 atm $\approx$ 1 atm

Sample type: Water holding capacity can be derived from: air dry soil material sieved (disturbed samples, fine earth fraction as defined in given laboratory e.g. '1 mm' or '<2 mm'); natural clods; reconstituted clods of soil material; and soil cores taken at a particular depth.

Treatment: For desorption methods, irrespective of the initial moisture status, all samples are first saturated on a tension table by capillarity; subsequently, they are equilibrated at pre-defined water retention levels. Methods that involve wetting of oven dry soils may cause irreversible changes in pore size. To study changes resulting from wetting and drying, the gravimetric water content can be determined after a second equilibration.

In the absorption methods, the sample is initially very dry and subsequently equilibrated at the pre-defined retentions levels without prior saturation.

To facilitate desorption by tension or pressure in lab experiments samples have to be in close contact with a supporting surface. Silt or kaolin can be used as an intermediate layer to improve the contact of the sample with a ceramic or porous plate or, in a sandbox for multiple samples.

Clods are generally sealed by a plastic lacquer (e.g. 'Saran F 310 resin'). A flat surface is cut to enhance contact on the ceramic plate or to allow clods to get wet on a tension table. At higher tension levels water is removed from the smallest pores by pressure. Here the original arrangement of soil materials is no longer important; bulk material (sieved <2 mm, air dry, e.g. removed from cores), is placed in retainer rings on the ceramic plate.

Desorption method: Equilibrium values for water desorption can be set by a hanging water column or application of 'over pressure'. A hanging water column with a water manometer is generally used for lower retention levels (<250 kPa). Hanging water columns should not exceed a length of 10 meters unless vacuum conditions are created and controlled by an Hg-manometer at the 'open end' of the column. With the pressure method, multiple samples are placed on a ceramic plate in a 'pressure pan'. In a series of equilibrium points the above methods are often combined; also there is an overlap in potential use of the respective desorption methods.

Device: Within the range of equilibrium values for the determination of water holding capacity (e.g. from pF 1.5 to pF 4.2) several 'overlapping' devices can be applied. Saturation is achieved on a tension table at 5 cm tension (5 kPa) applied at the base of the sample. Other equilibrium points may be set in a pressure plate extractor, on a porous plate with burette to measure the volume of desorbed water, or in a kaolin bath. Tensiometers are used with the wind evaporation method.

Expression: Water content may be expressed on a volume basis (volume of water per unit volume of moist soil), on a dry mass basis (mass of water per unit mass of soil solids), or on a wet mass basis (mass of water per unit mass of wet soil). Water holding capacity conventionally is expressed in volume % (weight per volume, w/v) rather than in weight % (weight per weight, w/w). To arrive at volume %, weight% has to be multiplied by bulk density. Where present, a correction needs to be applied for coarse fragments (SSDS, 2014).

Appendix E

Flowcharts for standardizing soil analytical method descriptions

The following flowcharts serve to visualize and complement Appendix D, which describes the rationale and criteria for standardizing soil analytical method descriptions, and Appendix F, which describes the domains for the various options.

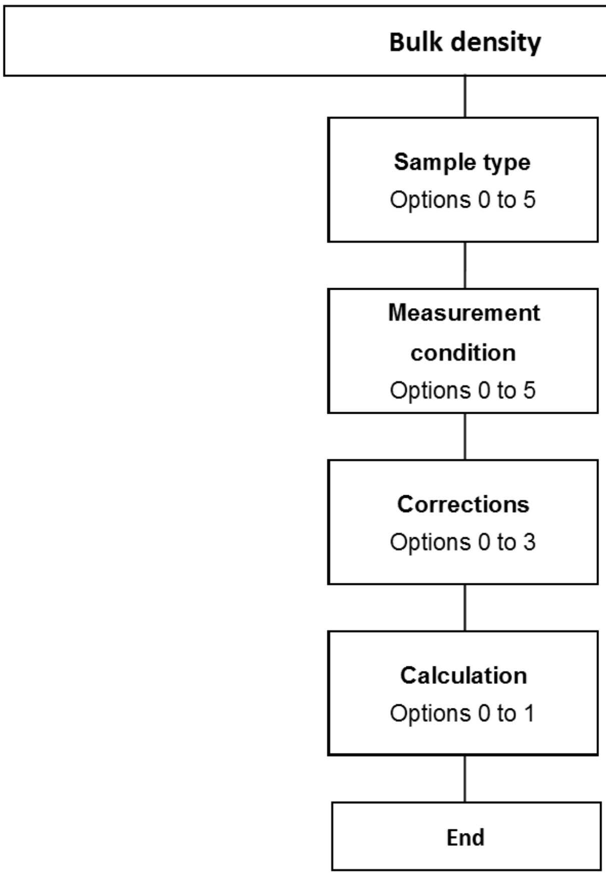


Figure E.1: Flowchart for standardizing bulk density methods.

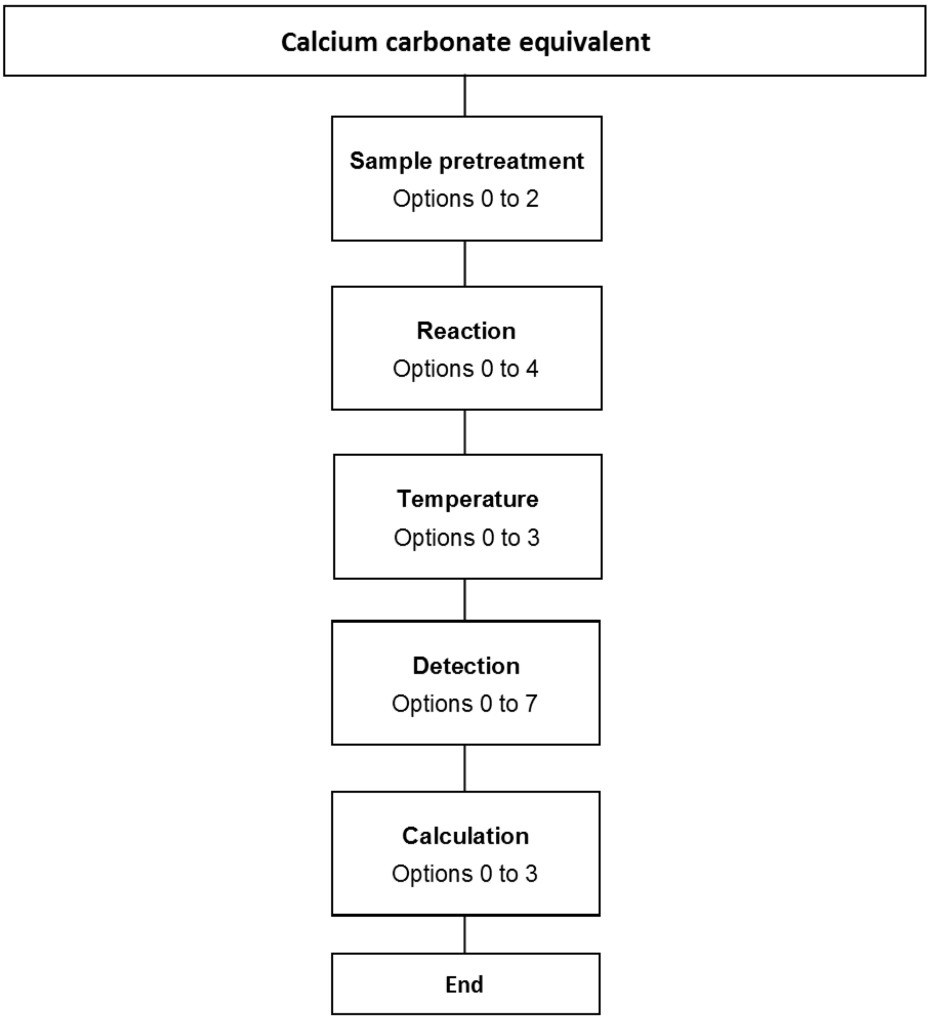


Figure E.2: Flowchart for standardizing calcium carbonate equivalent methods.

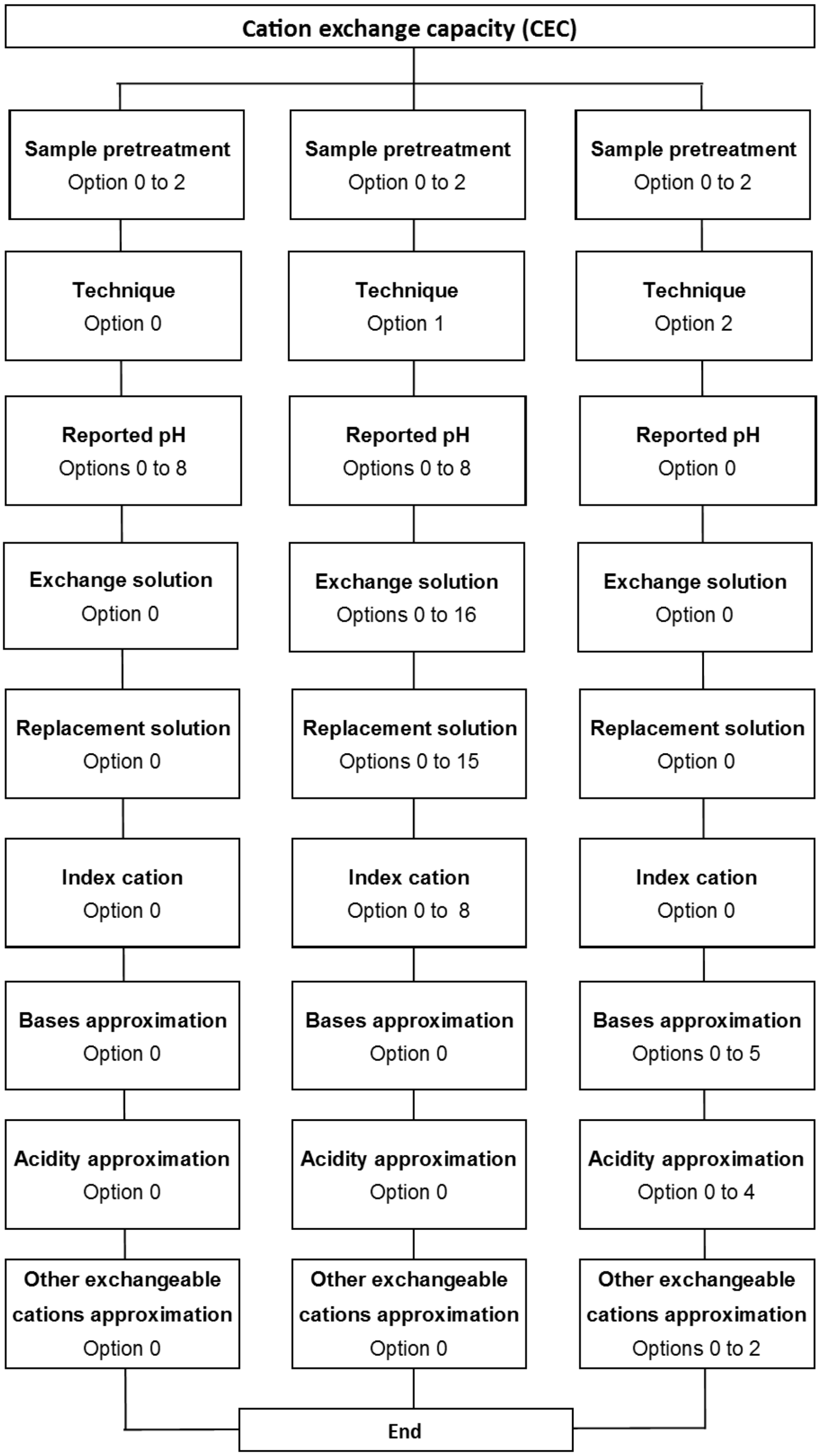


Figure E.3: Flowchart for standardizing cation exchange capacity methods.

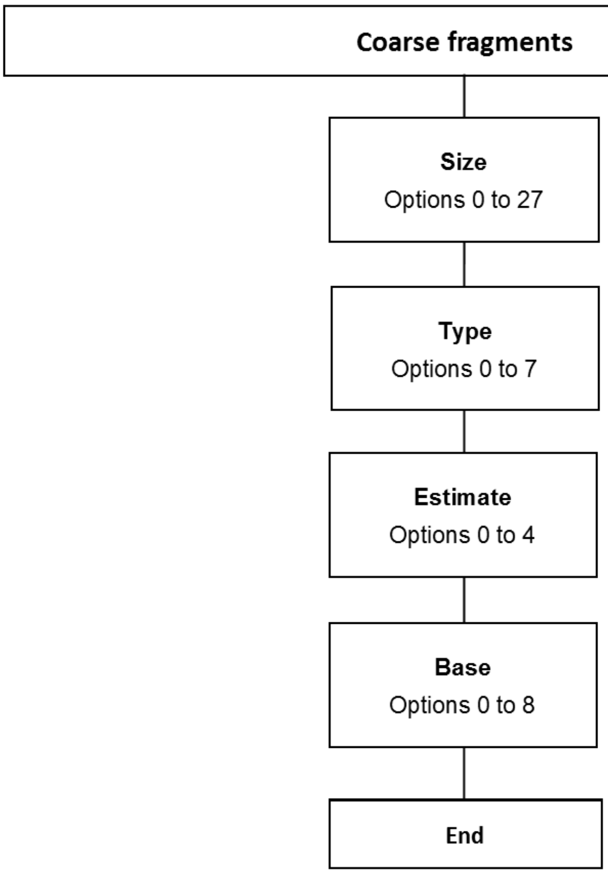


Figure E.4: Flowchart for standardizing coarse fragments methods.

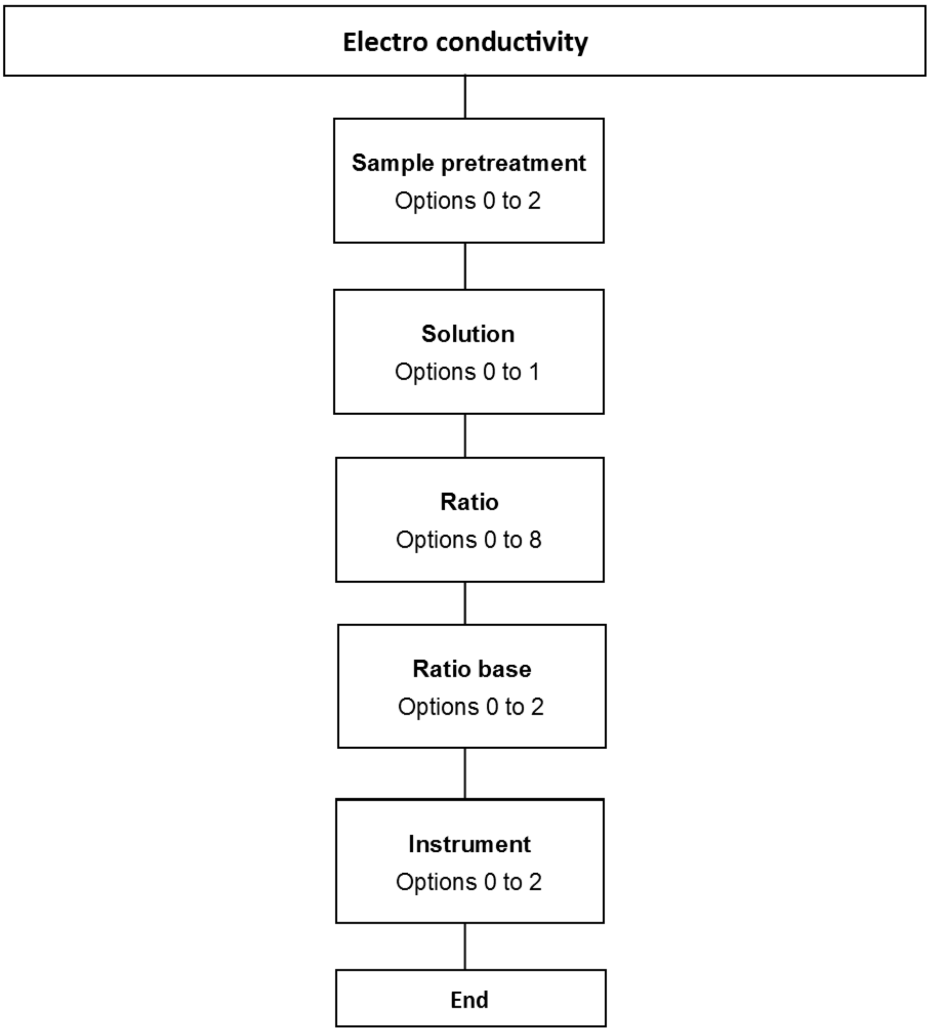


Figure E.5: Flowchart for standardizing electrical conductivity methods.

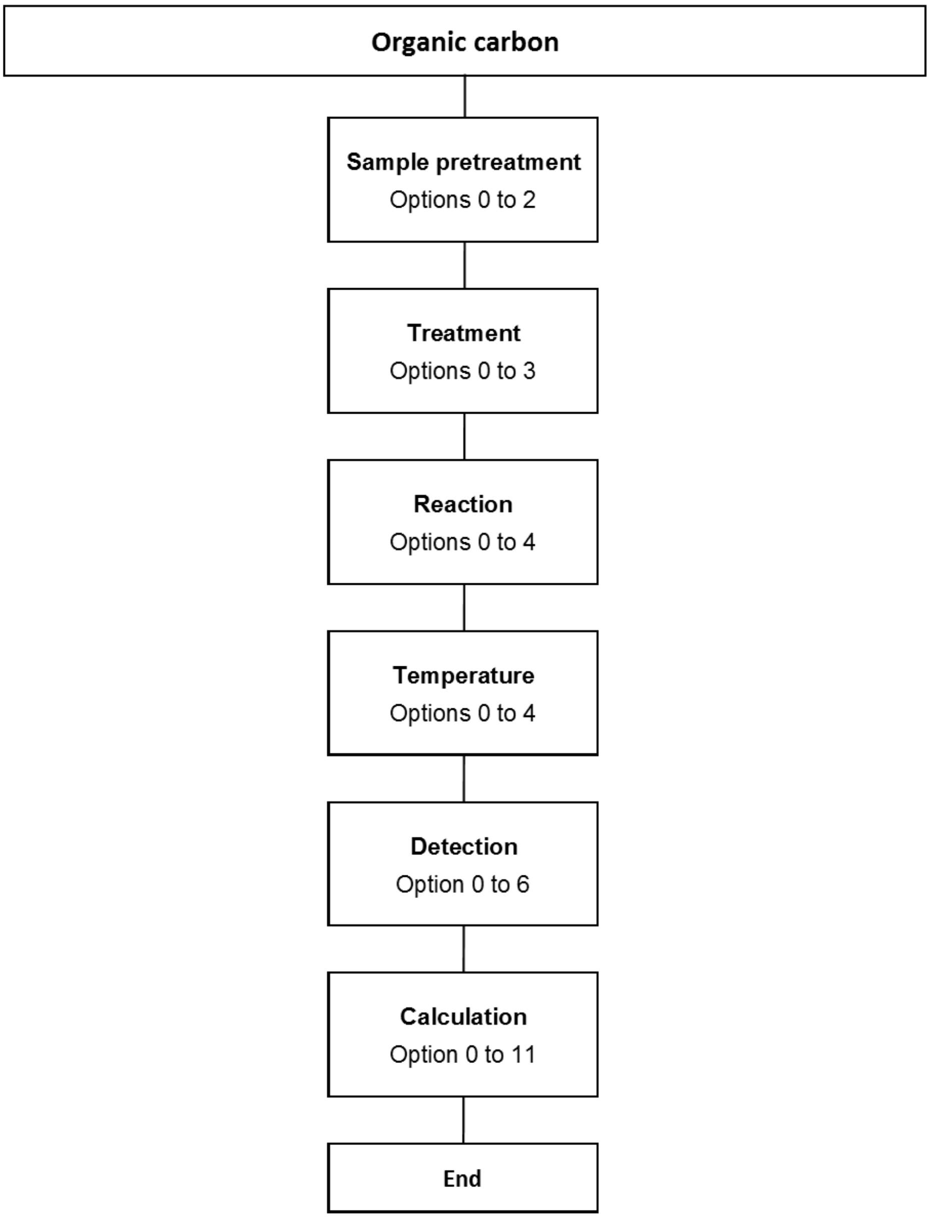


Figure E.6: Flowchart for standardizing organic carbon methods.

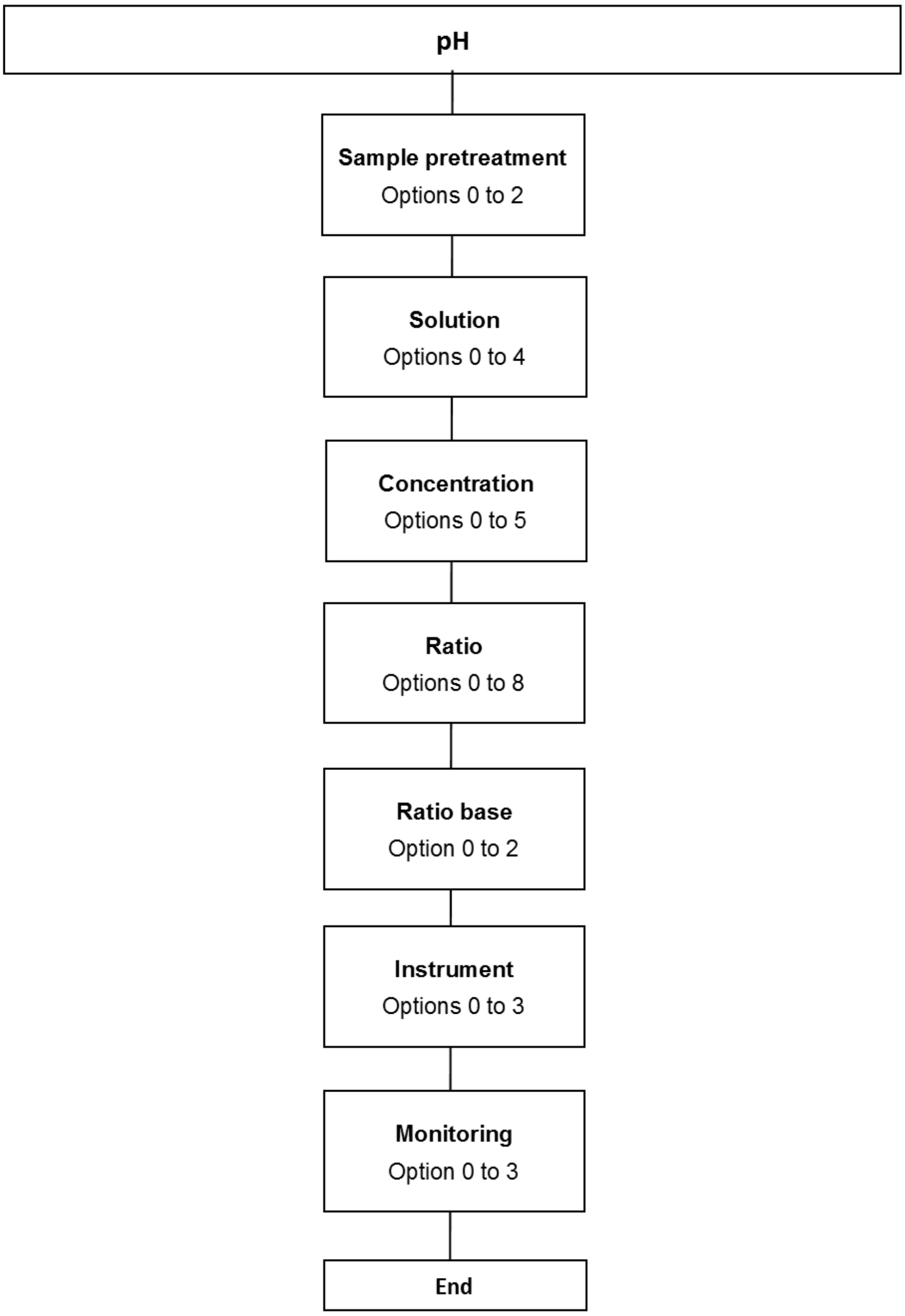


Figure E.7: Flowchart for standardizing pH methods.

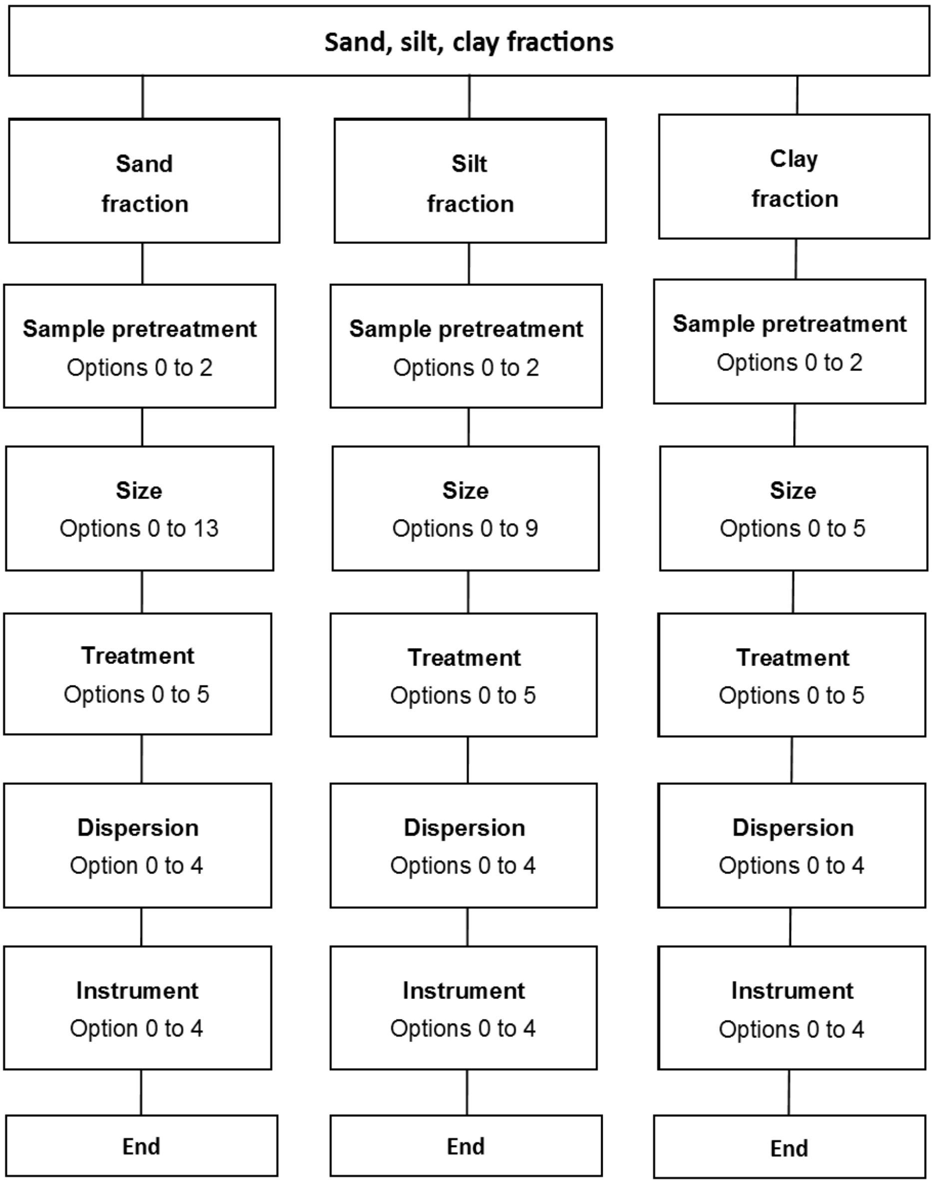


Figure E.8: Flowchart for standardizing sand, silt and clay fractions methods.

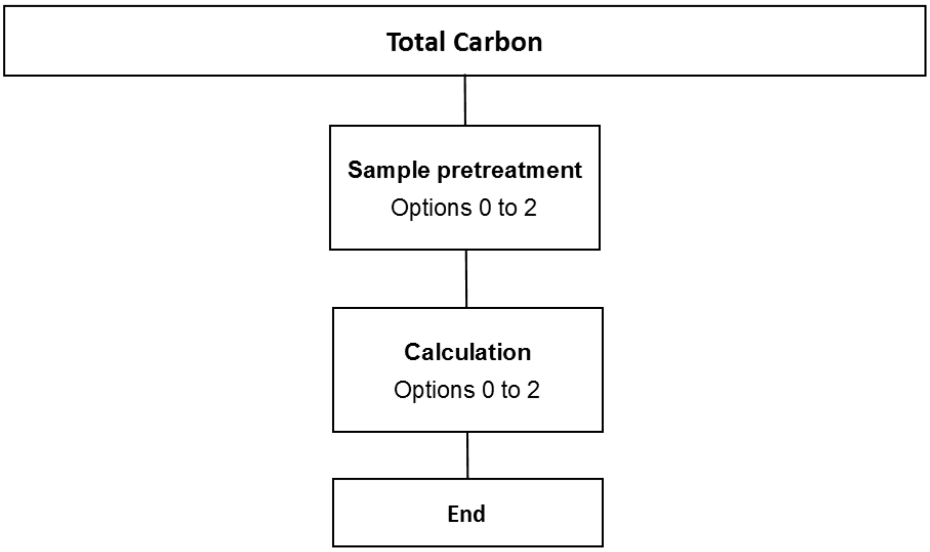


Figure E.9: Flowchart for standardizing total carbon methods.

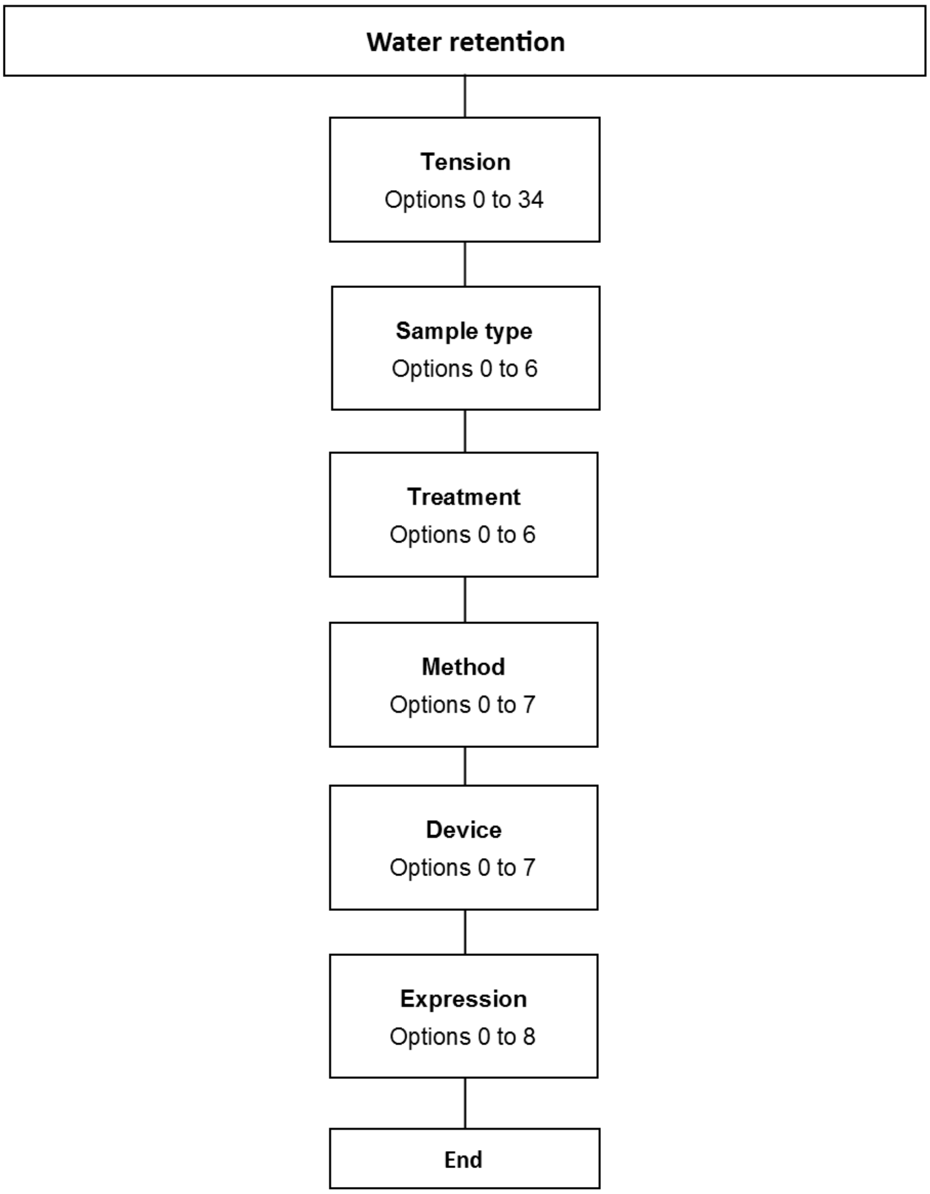


Figure E.10: Flowchart for standardizing water retention methods.

Appendix F

Option tables for soil analytical method descriptions

This appendix lists the criteria used for standardizing different analytical method descriptions to the WoSIS standard. To facilitate data entry (i.e. standardization of soil analytical method descriptions by third parties) the recommended sequence (1,2, .., n) for describing attribute-specific options is listed below. For each soil property, the workflow is visualized in Appendix E.

Table F.1: Procedure for coding bulk density.

Key	*Code*	*Value*
sample type	0	unknown
sample type	1	excavation (i.e. soils too fragile to remove a sample); compliant cavity, ring excavation, frame excavation)
sample type	2	undisturbed soil in metal/PVC-ring (soil core) (soil sufficiently coherent)
sample type	3	natural clod
sample type	4	clod reconstituted from $<$2 mm sample formed by wetting and dessication cycles that stimulate reconsolidating by water in a field setting
sample type	5	volume by 3D scanning
measurement condition	0	unknown
measurement condition	1	field moist
measurement condition	2	equilibrated at 33 kPa ($\sim$1/3 bar)
measurement condition	3	oven dry ($\sim$ 105-110 °C)
measurement condition	4	air dry
measurement condition	5	air dried and re-equilibrated (rewet)
corrections	0	unknown
corrections	1	in calculation, for $>$2 mm (rock, coarse) fragments removed from sample, density of fragments not reported
corrections	2	in calculation, for $>$2 mm (rock, coarse) fragments removed from sample; density fragments default value 2.65 g cm^{-3}
corrections	3	in calculation, for $>$2 mm (rock, coarse) fragments; correction for weight and volume
calculation	0	unknown
calculation	1	guessed value, expert field estimate

Table F.2: Procedure for coding calcium carbonate equivalent.

Key	*Code*	*Value*
sample pretreatment	0	unkown
sample pretreatment	1	sieved over 1 mm sieve
sample pretreatment	2	sieved over 2 mm sieve
reaction	0	unknown
reaction	1	dissolution of Carbonates by Hydrochloric acid [HCl], or Perchloric acid [$HClO_4$]
reaction	2	dissolution of Carbonates by Sulfuric acid [H_2SO_4]
reaction	3	dissolution of Carbonates by Phosphoric acid [H_3PO_4]
reaction	4	dissolution of Carbonates by Acetic acid [CH_3COOH]
temperature	0	unknown
temperature	1	no external heat
temperature	2	external heat, elevated temperature; ignition $\leq$ 400 °C
temperature	3	external heat, combustion (element analyzer)
detection	0	unknown
detection	1	titrimetric (for example titration excess acid)
detection	2	gravimetric - weight increase (from trapped Carbon dioxide [CO_2] evolved)
detection	3	volumetric (i.e. volume of Carbon dioxide [CO_2] evolved) (1 Pa, room temperature)
detection	4	pressure (i.e. pressure build bij Carbon dioxide [CO_2] evolved, manometric)
detection	5	gravimetric - weight loss (from Carbon dioxide [CO_2] evolved)
detection	6	sensoric (as in element analyzer)
detection	7	change of pH with dissolution
calculation	0	unknown
calculation	1	(in)direct estimates of Carbonates [$XXCO_3.xxH_2O$] or Inorganic Carbon, expressed as Calcium carbonate equivalent
calculation	2	subtraction; (Total C - Organic C) expressed as Calcium carbonate equivalent
calculation	3	emperically; standard (neutralization) curve relating pH to known concentrations of $CaCO_3$

Table F.3: Procedure for coding cation exchange capacity (cec).

Key	*Code*	*Value*
sample pretreatment	0	unkown
sample pretreatment	1	sieved over 1 mm sieve
sample pretreatment	2	sieved over 2 mm sieve
technique	0	unknown
technique	1	determination by lab procedure
technique	2	approximated by summation exchangeable cations
reported pH	0	unknown
reported pH	1	reported as (no buffer applied); field pH
reported pH	2	buffered at 7.0
reported pH	3	buffered at 8.0
reported pH	4	buffered at 8.1
reported pH	5	buffered at 8.2
reported pH	6	buffered at 8.3
reported pH	7	buffered at 8.4
reported pH	8	buffered at 8.5
exchange solution	0	unknown
exchange solution	1	not applied
exchange solution	2	1 M KCl
exchange solution	3	1 M NH_4-acetate
exchange solution	4	1 M NH_4-chloride
exchange solution	5	0.5 M Li-acetate
exchange solution	6	0.5 M $BaCl_2$-TEA
exchange solution	7	0.1 M $BaCl_2$-TEA
exchange solution	8	1 M Na-acetate
exchange solution	9	0.01 M Ag-thioura
exchange solution	10	0.01 M Ag-thioura + buffer
exchange solution	11	0.25 M NH_4Ac
exchange solution	12	0.1 M $CaCl_2$
exchange solution	13	0.5 M Li-chloride
exchange solution	14	0.1 M Li-EDTA
exchange solution	15	Ba-acetate
exchange solution	16	$BaCl_2$
replacement	0	unknown
replacement	1	not applied
replacement	2	NaOH
replacement	3	KCl
replacement	4	NaCl
replacement	5	KNO_3 $Ca(NO_3)^2$
replacement	6	NH_4-acetate
replacement	7	Ca-Acetate
replacement	8	$Mg(NO_3)_2$
replacement	9	$Mg(SO_4)_2$
replacement	10	NH_4Cl
replacement	11	$NaNO_3$
replacement	12	10% NaCl + HCl
replacement	13	$CaCl_2$
replacement	14	K-EDTA
replacement	15	Na-acetate
index cation	0	unknown
index cation	1	not applied
index cation	2	NH_4+
index cation	3	Na+
index cation	4	0.01 M Ag-thioura$^+$
index cation	5	Ba^{2+}
index cation	6	Li^+

Key	*Code*	*Value*
index cation	7	Mg^{2+}
index cation	8	Ca^{2+}
bases approximation	0	unknown
bases approximation	1	not applied
bases approximation	2	sum exchangeable bases (NH_4Cl, unbuffered)
bases approximation	3	sum exchangeable bases (NH_4OAc, buffered pH 7)
bases approximation	4	sum exchangeable bases ($BaCl_2$-TEA pH 8.1)
bases approximation	5	sum of extractable Na, K, Ca, Mg (NH_4Cl, 0.05M, unbuffered)
acidity approximation	0	unknown
acidity approximation	1	not applied
acidity approximation	2	exchangeable acidity (KCl extract), (sum of) H, Al
acidity approximation	3	extractable / potential acidity ($BaCl_2$ - TEA, pH 8.2)
acidity approximation	4	(sum of) extractable H, Al (NH_4Cl, 0.05M, unbuffered)
other exchangeable cations	0	unknown
other exchangeable cations	1	not applied
other exchangeable cations	2	(sum of) extractable Fe, Mn (NH_4Cl, 0.05M, unbuffered)

Table F.4: Procedure for coding clay.

Key	*Code*	*Value*
sample pretreatment	0	unknown
sample pretreatment	1	sieved over 1 mm sieve
sample pretreatment	2	sieved over 2 mm sieve
size	0	unknown
size	1	0 - 0.002 mm
size	2	0 - 0.005 mm
size	3	0 - 0.001 mm
size	4	0 - 0.0002 mm
size	5	0.002 -0.05 mm
treatment	0	unknown
treatment	1	no pretreatment
treatment	2	Hydrogen peroxide [H_2O_2] plus mild Acetic acid [CH_3COOH] / Sodium acetate [CH_3COONa] buffer treatments (if pH-H_2O $>$6.5)
treatment	3	Hydrogen peroxide [H_2O_2] plus Hydrochloric acid [HCl] or Acetic acid [CH_3COOH] (if pH-H_2O $>$6.5)
treatment	4	pretreatment, deferration included
treatment	5	Hydrogen peroxide [H_2O_2]
dispersion	0	unknown
dispersion	1	no dispersion
dispersion	2	Sodium hexametaphosphate [($NaPO_3$)6] - Calgon type (ultrasonic treatment might be included)
dispersion	3	Ammonium hydroxide [NH_4OH]
dispersion	4	Sodium hydroxide [NaOH]
instrument	0	unknown
instrument	1	pipette
instrument	2	hydrometer
instrument	3	analyzer
instrument	4	field hand estimate

Table F.5: Procedure for coding coarse fragments.

Key	*Code*	*Value*
size	0	unknown
size	1	4.76 mm / Mesh 4 / 0.187 inches
size	2	2.00 mm / Mesh 10 / 0.0787 inches
size	3	0.841 mm / Mesh 20 / 0.0331 inches
size	4	0.420 mm / Mesh 40 / 0.0165 inches
size	5	0.250 mm / Mesh 60 / 0.0098 inches
size	6	0.149 mm / Mesh 100 / 0.0059 inches
size	7	0.105 mm / Mesh 140 / 0.0041 inches
size	8	0.074 mm / Mesh 200 / 0.0029 inches
size	9	>2 mm
size	10	2 - 5 mm
size	11	5 - 20 mm
size	12	2 - 20 mm
size	13	20 - 75 mm
size	14	2 - 75 mm
size	15	>75 mm
size	16	20 - 250 mm
size	17	75 - 250 mm
size	18	>250 mm
size	19	0.002 mm
size	20	0.005 mm
size	21	0.02 mm / Mesh 635
size	22	9.51 mm / Mesh 3/8 in / 0.375 inches
size	23	19 mm / Mesh in / 0.75 inches
size	24	25.4 mm / Mesh 1 / 1 inches
size	25	37.5 mm / 1.5 inches
size	26	50 mm / 2 inches
size	27	>1 mm
type	0	unknown
type	1	no pretreatment
type	2	rock fragments, coarse concretions, roots and adhering finer particles >2 mm (removed from field sample, with sample pretreatment in the lab)
type	3	in the >2 mm fraction: mineral coarse fragments, as well as wood fragments that are >20 mm in cross section and cannot be crushed and shredded with fingers
type	4	rock fragments that resist abrupt immersion in tap water
type	5	from the 20 - 2 mm field sub sample; coarse fragments in the 2- 5 mm fraction that do not slake with Sodium hexametaphosphate [($NaPO_3$)6] treatment
type	6	coarse fragments with hard carbonate- or gypsum-indurated material and not easily pass a 2 mm sieve
type	7	in the >2 mm fraction: Cr or R material
estimate	0	unknown
estimate	1	w/w% weighing (lab, field) procedure
estimate	2	v/v%, visual (field) estimate (presumed)
estimate	3	w/w% after conversion from v/v% for material >20 mm, by particle density 2.65 g cc-1, bulk density fine earth fraction 1.45 g cc-1
estimate	4	w/w% after conversion from v/v% for material >20 mm, by or particle density or bulk density fine earth fraction or both ”measured” g cc-1
base	0	unknown
base	1	<2mm, fine earth

Key	*Code*	*Value*
base	2	>2mm; rock and pararock fragments; rock fragments means particles of the whole soil that are >2 mm in diameter and includes all particles with horizontal dimensions smaller than the size of the pedon
base	3	2 - 20 mm
base	4	20 - 75 mm
base	5	2 - 75 mm
base	6	>75 mm
base	7	whole soil; all particle-size fractions, including boulders with maximum horizontal dimensions less than those of the pedon
base	8	whole soil at 1/3 Bar

Table F.6: Procedure for coding electrical conductivity.

Key	*Code*	*Value*
sample pretreatment	0	unknown
sample pretreatment	1	sieved over 1 mm sieve
sample pretreatment	2	sieved over 2 mm sieve
solution	0	unknown
solution	1	water [H_2O]
ratio	0	unknown
ratio	1	1:1
ratio	2	1:2
ratio	3	1:2.5
ratio	4	1:5
ratio	5	1:10
ratio	6	1:50
ratio	7	saturated paste
ratio	8	slurry
ratio base	0	unknown
ratio base	1	weight / volume
ratio base	2	volume / volume
instrument	0	unknown
instrument	1	electrode
instrument	2	electrode (field measured)

Table F.7: Procedure for coding organic carbon.

Key	*Code*	*Value*
sample pretreatment	0	unknown
sample pretreatment	1	sieved over 1 mm sieve
sample pretreatment	2	sieved over 2 mm sieve
treatment	0	unknown
treatment	1	not applied
treatment	2	inorganic carbon removed; Hydrochloric acid [HCl]
treatment	3	inorganic carbon removed; Phosphoric acid [H_3PO_4]
reaction	0	unknown
reaction	1	wet oxidation with Sulphuric acid [H_2SO_4] - Potassiumbichromate [$K_2Cr_2O_7$] (and Phosphoric acid [H_3PO_4]) mixture
reaction	2	wet oxidation - other methods
reaction	3	dry oxidation (i.e loss on ignition)
reaction	4	dry oxidation (such as element analyzer)
temperature	0	unknown
temperature	1	no external heat
temperature	2	controlled, temperature range 350 - 550 °C (assumed; loss on ignition, muffle furnace)
temperature	3	controlled, at elevated temperature (wet oxidation, temperature (not) specified)
temperature	4	controlled, at 960 °C and higher (assumed: element analyzer)
detection	0	unknown
detection	1	titrimetric
detection	2	colorimetry (i.e. by graphing a standard curve)
detection	3	gravimetric; increase weight by trapping evolved Carbon dioxide [CO_2]
detection	4	volumetric
detection	5	sensoric (in element analyzer)
detection	6	weight loss (i.e. ”loss on ignition” method)
calculation	0	unknown
calculation	1	not applied
calculation	2	Total Carbon minus Total inorganic Carbon
calculation	3	conversion factor ”organic matter to total carbon” = 1/1.7 (1.7 = Van Bemmelen factor)
calculation	4	complete recovery (assumed)
calculation	5	correction factor for recovery not specified
calculation	6	default (Walkley and Black) correction factor for recovery of 1.3 applied
calculation	7	default correction factor for recovery of 1.3 - assumed
calculation	8	correction factor = 1.25
calculation	9	correction factor = 1.18
calculation	10	correction factor = 1.03
calculation	11	correction factor = 1.15

Table F.8: Procedure for coding ph.

Key	*Code*	*Value*
sample pretreatment	0	unknown
sample pretreatment	1	sieved over 1 mm sieve
sample pretreatment	2	sieved over 2 mm sieve
solution	0	unknown
solution	1	water [H_2O]
solution	2	Calcium chloride [$CaCl_2$]
solution	3	Potassium chloride [KCl]
solution	4	Sodium fluoride [NaF]
concentration	0	unknown
concentration	1	not applied
concentration	2	0.01 M
concentration	3	0.02 M
concentration	4	0.2 M
concentration	5	1 M
ratio	0	unknown
ratio	1	1:1
ratio	2	1:2
ratio	3	1:2.5
ratio	4	1:5
ratio	5	1:10
ratio	6	1:50
ratio	7	saturated paste
ratio	8	slurry
ratio base	0	unknown
ratio base	1	weight / volume
ratio base	2	volume / volume
instrument	0	unknown
instrument	1	electrode
instrument	2	electrode (field measured)
instrument	3	indicator paper (field measured)
monitoring	0	unknown
monitoring	1	not applied
monitoring	2	oxidizible sulfur compounds; initial pH
monitoring	3	oxidizible sulfur compounds; pH stabilized in 10 days, pH $\leq$ 0.1 unit, for two days

Table F.9: Procedure for coding sand.

Key	*Code*	*Value*
sample pretreatment	0	unknown
sample pretreatment	1	sieved over 1 mm sieve
sample pretreatment	2	sieved over 2 mm sieve
size	0	unknown
size	1	0.05 - 2 mm
size	2	0.063 - 2 mm
size	3	0.06 - 2 mm
size	4	0.02 - 2 mm
size	5	1 - 2 mm
size	6	0.05 - 1 mm
size	7	0.25 - 0.5 mm
size	8	0.10 - 0.25 mm
size	9	0.05 - 1.7 mm
size	10	0.05 - 0.1 mm
size	11	0.2 - 2 mm
size	12	0.5 - 1.0 mm
size	13	0.1 - 2.0 mm
treatment	0	unknown
treatment	1	no pretreatment
treatment	2	Hydrogen peroxide [H_2O_2] plus mild Acetic acid [CH_3COOH] / Sodium acetate [CH_3COONa] buffer treatments (if pH-H_2O $>$6.5)
treatment	3	Hydrogen peroxide [H_2O_2] plus Hydrochloric acid [HCl] or Acetic acid [CH_3COOH] (if pH-H_2O $>$6.5)
treatment	4	pretreatment, deferration included
treatment	5	Hydrogen peroxide [H_2O_2]
dispersion	0	unknown
dispersion	1	no dispersion
dispersion	2	Sodium hexametaphosphate [($NaPO_3$)6] - Calgon type (ultrasonic treatment might be included)
dispersion	3	Ammonium hydroxide [NH_4OH]
dispersion	4	Sodium hydroxide [NaOH]
instrument	0	unknown
instrument	1	sieve
instrument	2	hydrometer
instrument	3	analyzer
instrument	4	field hand estimate

Table F.10: Procedure for coding silt.

Key	*Code*	*Value*
sample pretreatment	0	unknown
sample pretreatment	1	sieved over 1 mm sieve
sample pretreatment	2	sieved over 2 mm sieve
size	0	unknown
size	1	0.02 - 0.063 mm
size	2	0.002 - 0.06 mm
size	3	0.02 - 0.05 mm
size	4	0.005 - 0.05 mm
size	5	0.001 - 0.05 mm
size	6	0.002 - 0.05 mm
size	7	0.002 - 0.02 mm
size	8	0.002 - 0.074 mm
size	9	0.05 - 0.074 mm
treatment	0	unknown
treatment	1	no pretreatment
treatment	2	Hydrogen peroxide [H_2O_2] plus mild Acetic acid [CH_3COOH] / Sodium acetate [CH_3COONa] buffer treatments (if pH-H_2O $>$6.5)
treatment	3	Hydrogen peroxide [H_2O_2] plus Hydrochloric acid [HCl] or Acetic acid [CH_3COOH] (if pH-H_2O $>$6.5)
treatment	4	pretreatment, deferration included
treatment	5	Hydrogen peroxide [H_2O_2]
dispersion	0	unknown
dispersion	1	no dispersion
dispersion	2	Sodium hexametaphosphate [($NaPO_3$)6] - Calgon type (ultrasonic treatment might be included)
dispersion	3	Ammonium hydroxide [NH_4OH]
dispersion	4	Sodium hydroxide [NaOH]
instrument	0	unknown
instrument	1	pipette
instrument	2	hydrometer
instrument	3	analyzer
instrument	4	field hand estimate

Table F.11: Procedure for coding total carbon.

Key	*Code*	*Value*
sample pretreatment	0	unknown
sample pretreatment	1	sieved over 1 mm sieve
sample pretreatment	2	sieved over 2 mm sieve
calculation	0	unknown
calculation	1	dry combustion at high controlled temperature, complete recovery, sensoric detection (as in element analyzer) no pretreatment to dissolve inorganic carbon (if present) applied
calculation	2	calculated by sum of Organic carbon and Calcium carbonate equivalent

Table F.12: Procedure for coding water retention.

Key	*Code*	*Value*
tension	0	unknown
tension	1	not applied
tension	2	kPa=0.1, cm water head=1.0, bar=0.001, pF=0.0
tension	3	kPa=0.3, cm water head=3.2, bar=0.003, pF=0.5
tension	4	kPa=0.5, cm water head=5.0, bar=0.005, pF=0.7
tension	5	kPa=1, cm water head=10.2, bar=0.01, pF=1.0
tension	6	kPa=3, cm water head=30.6, bar=0.03, pF=1.5
tension	7	kPa=5, cm water head=51.1, bar=0.05, pF=1.7
tension	8	kPa=6, cm water head=61.3, bar=0.06, pF=1.8
tension	9	kPa=7, cm water head=75.0, bar=0.07, pF=1.9
tension	10	kPa=10, cm water head=102.2, bar=0.10, pF=2.0
tension	11	kPa=12, cm water head=125.0, bar=0.12, pF=2.1
tension	12	kPa=15, cm water head=150.0, bar=0.15, pF=2.2
tension	13	kPa=20, cm water head=204.3, bar=0.20, pF=2.3
tension	14	kPa=24, cm water head=250.0, bar=0.24, pF=2.4
tension	15	kPa=33, cm water head=337.1, bar=0.33, pF=2.5
tension	16	kPa=40, cm water head=408.6, bar=0.40, pF=2.6
tension	17	kPa=50, cm water head=510.8, bar=0.50, pF=2.7
tension	18	kPa=60, cm water head=613.0, bar=0.60, pF=2.8
tension	19	kPa=70, cm water head=715.1, bar=0.70, pF=2.9
tension	20	kPa=80, cm water head=817.3, bar=0.80, pF=2.9
tension	21	kPa=90, cm water head=919.4, bar=0.90, pF=3.0
tension	22	kPa=100, cm water head=1021.6, bar=1.00, pF=3.0
tension	23	kPa=200, cm water head=2043.2, bar=2.00, pF=3.3
tension	24	kPa=250, cm water head=2554.0, bar=2.50, pF=3.4
tension	25	kPa=400, cm water head=4086.4, bar=40.90, pF=3.6
tension	26	kPa=500, cm water head=5108.0, bar=51.10, pF=3.7
tension	27	kPa=1500, cm water head=15324.0, bar=15.00, pF=4.2
tension	28	kPa=500, cm water head=5085.0, bar=5.00, pF=3.7
tension	29	kPa=580, cm water head=5998.6, bar=5.80, pF=3.8
tension	30	kPa=, cm water head=, bar=, pF=4.1
tension	31	kPa=, cm water head=, bar=, pF=3.5
tension	32	kPa=, cm water head=, bar=, pF=5.1
tension	33	kPa=, cm water head=, bar=, pF=5.8
tension	34	kPa=1000, cm water head=, bar=, pF=
sample type	0	unknown
sample type	1	natural clod
sample type	2	clod, reconstituted / disturbed (sieved material <2 mm)
sample type	3	clod, reconstituted / disturbed (sieved material <1 mm)
sample type	3	undisturbed soil in metal/PVC-ring (soil core)
sample type	4	<2 mm (sieved) disturbed samples
sample type	6	<1 mm (sieved) disturbed samples
treatment	0	unknown
treatment	1	not applied
treatment	2	oven dried, no saturation applied (i.e.: absorption curve)
treatment	3	field moist condition, then saturated
treatment	4	air dry, then saturated
treatment	5	oven dry, then saturated
treatment	6	saturated, desorbed, rewetted and desorbed again
method	0	unknown
method	1	saturation (pF 0)
method	2	desorption, pressure
method	3	desorption, suction (hanging water column, water manometer)
method	4	desorption, suction (hanging water column + Hg manometer)

Key	*Code*	*Value*
method	5	desorption, evaporation
method	6	desorption, oven drying
method	7	absorption into oven dry sample (curve, dry to wet)
device	0	unknown
device	1	not applied
device	2	tension table
device	3	pressure plate extractor
device	4	porous plate and burette
device	5	fine textured medium; (presumed) sandbox
device	6	fine textured medium; (pressumed) kaolin box
device	7	balans, tensiometers (wind evaporation method)
expression	0	unknown
expression	1	dry mass basis; mass water per unit mass of soil solids (w/w, gravimetric water content)
expression	2	wet mass basis; mass of water per unit mass of wet soil (w/w)
expression	3	volume base; volume of water per unit volume of moist soil (v/v, volumetric water content)
expression	4	volume base; volume of water per unit volume of moist soil (v/v, volumetric water content). w/w % converted by Unknown bulk density
expression	5	volume base; volume of water per unit volume of moist soil (v/v, volumetric water content). Presumed; w/w % converted by bulk density if presented
expression	6	volume base; volume of water per unit volume of moist soil (v/v, volumetric water content). w/w % converted by bulk density oven dry
expression	7	volume base; volume of water per unit volume of moist soil (v/v, volumetric water content). w/w % converted by bulk density pKa 33
expression	8	volume base; volume of water per unit volume of moist soil (v/v, volumetric water content). w/w % converted by bulk density rewet

Appendix G

Database model

This appendix describes the structure of all PostgreSQL tables considered in WoSIS. Pragmatically, each table starts on a new page.

Table G.1: class_cpcs - Soil name according to the French soil classification system (Commission de pedologie et de cartographie des sols).

Column	*Type*	*Modifiers*	*Comment*
class_cpcs_id	integer	sequence	Primary key
dataset_id	character varying(20)		Foreign key that together with ”profile_id” refers to table ”dataset_profile”
profile_id	integer		Foreign key that together with ”dataset_id” refers to table ”profile_profile”
publication_year	smallint		Year of publication of the version used for the characterization
group_code	character varying(5)		Soil group code
group_name	text		Soil group name
unit_code	character varying(5)		Soil unit code
unit_name	text		Soil unit name
note	text		Comments field
trust	character(1)	’A’::bpchar	Level of trust: ”A” as entered, no validation; ”B” standardized; ”C” harmonized

Table G.2: class_fao - Soil name according to the Legend of the 1:5M scale FAO Soil Map of the World.

Column	*Type*	*Modifiers*	*Comment*
class_fao_id	integer	sequence	Primary key
dataset_id	character varying(20)		Foreign key that together with ”profile_id” refers to table ”dataset_profile”
profile_id	integer		Foreign key that together with ”dataset_id” refers to table ”dataset_profile”
publication_year	smallint		Year of publication of the version used for the characterization
major_group_code	character varying(2)		Major soil group code
major_group	text		Major soil group name
soil_unit_code	character varying(1)		Soil unit code
soil_unit	text		Soil unit name
subunit_code	character varying(5)		Soil unit code
subunit	text		Soil unit name
phase_code	character varying(2)		Phase code - limiting factor related to surface or subsurface features of the land
phase	text		Phase name - limiting factor related to surface or sub surface features of the land
verified	boolean		Was the classification verified (True/False)
verified_user	text		Who made the verification?
verified_date	date		Date when was the classification verified?
note	text		Comments field
trust	character(1)	’A’::bpchar	Level of trust: ”A” as entered, no validation; ”B” standardized; ”C” harmonized

Table G.3: class_fao_horizon - Diagnostic horizon according to the FAO Soil Map of the World Legend.

Column	*Type*	*Modifiers*	*Comment*
class_fao_horizon_id	integer	sequence	Primary key
class_fao_id	integer		Foreign key that refers to table ”class_fao”
diagnostic_horizon	text		Name of the diagnostic horizon
horizon	character varying(20)		Soil horizon
upper_depth	smallint		Depth of upper horizon boundary (cm)
lower_depth	smallint		Depth of lower horizon boundary (cm)

Table G.4: class_fao_property - Diagnostic property according to the FAO Soil Map of the World Legend.

Column	*Type*	*Modifiers*	*Comment*
class_fao_property_id	integer	sequence	Primary key
class_fao_id	integer		Foreign key that refers to table ”class_fao”
diagnostic_property	text		Name of the diagnostic property
upper_depth	smallint		Depth of upper horizon boundary (cm)
lower_depth	smallint		Depth of lower horizon boundary (cm)

Table G.5: class_local - Soil name according to the national soil classification system.

Column	*Type*	*Modifiers*	*Comment*
class_local_id	integer	sequence	Primary key
dataset_id	character varying(20)		Foreign key that together with ”profile_id” refers to table ”dataset_profile”
profile_id	integer		Foreign key that together with ”dataset_id” refers to table ”dataset_profile”
system_name	text		Name of local soil classification system
publication_year	smallint		Year of publication of the version used for the characterization
classification_name	text		Taxon name
common_name	text		Taxon common name
note	text		Comments field
trust	character(1)	’A’::bpchar	Level of trust: ”A” as entered, no validation; ”B” standardized; ”C” harmonized

Table G.6: class_soil_taxonomy - Soil name according to USDA Soil Taxonomy with defined version.

Column	*Type*	*Modifiers*	*Comment*
class_soil_taxonomy_id	integer	sequence	Primary key
dataset_id	character varying(20)		Foreign key that together with ”profile_id” to table ”dataset_profile”
profile_id	integer		Foreign key that together with ”dataset_id” refers to table ”dataset_profile”
publication_year	smallint		Year of publication of the version used for the characterization
order_name	text		Order name
suborder	text		Suborder name
great_group	text		Great group name
subgroup	text		Subgroup name
temperature_regime	text		Temperature regime
moisture_regime	text		Moisture regime
mineralogy	text		Mineralogy
textural_class	text		Textural class
other	text		Other information
note	text		Comments field
trust	character(1)	’A’::bpchar	Level of trust: ”A” as entered, no validation; ”B” standardized; ”C” harmonized

Table G.7: class_wrb - Soil name according to World Reference Base for Soil Resources with defined version.

Column	*Type*	*Modifiers*	*Comment*
class_wrb_id	integer	sequence	Primary key
dataset_id	character varying(20)		Foreign key that together with ”profile_id” refers to table ”dataset_profile”
profile_id	integer		Foreign key that together with ”dataset_id” refers to table ”dataset_profile”
publication_year	smallint		Year of publication of version used for the characterization
prefix_qualifier_code	text[]		Prefix qualifier code
prefix_qualifier	text[]		Prefix qualifier name
reference_soil_group_code	character varying(4)		Reference soil group code
reference_soil_group	text		Reference soil group name
suffix_qualifier_code	text[]		Sufix qualifier code
suffix_qualifier	text[]		Sufix qualifier name
verified	boolean		Is the classification verified (True/False)
verified_user	text		Who made the verification?
verified_date	date		Date when the classification was verified
note	text		Comments field
trust	character(1)	’A’::bpchar	Level of trust: ”A” as entered, no validation; ”B” standardized; ”C” harmonized
correlation	boolean	false	True if this classification comes from a correlation
correlation_system	text		Name of the original Soil Classification System
correlation_year	smallint		Year of the original Soil Classification System
soil_unit_code	character varying(3)		Soil unit adjectives codes (WRB 1998 only)
soil_unit	text		Soil unit adjectives (WRB 1998 only)
soil_unit_specifier_code	character varying(1)		Soil unit specifier codes (WRB 1998 only)
soil_unit_specifier	character varying(6)		Soil unit specifier (WRB 1998 only)

Table G.8: class_wrb_horizon - Diagnostic horizons according to the World Reference Base for Soil Resources.

Column	*Type*	*Modifiers*	*Comment*
class_wrb_horizon_id	integer	sequence	Primary key
class_wrb_id	integer		Foreign key that refers to table ”class_wrb”
diagnostic_horizon	text		Name of the diagnostic horizon
upper_depth	smallint		Depth of upper horizon boundary (cm)
lower_depth	smallint		Depth to lower horizon boundary (cm)

Table G.9: class_wrb_material - Diagnostic materials according to the World Reference Base for Soil Resources.

Column	*Type*	*Modifiers*	*Comment*
class_wrb_material_id	integer	sequence	Primary key
class_wrb_id	integer		Foreign key that refers to table ”class_wrb”
diagnostic_material	text		Name of the diagnostic material
upper_depth	smallint		Depth of upper horizon boundary (cm)
lower_depth	smallint		Depth of lower horizon boundary (cm)

Table G.10: class_wrb_property - Diagnostic properties according to the World Reference Base for Soil Resources.

Column	*Type*	*Modifiers*	*Comment*
class_wrb_property_id	integer	sequence	Primary key
class_wrb_id	integer		Foreign key that refers to table ”class_wrb”
diagnostic_property	text		Name of diagnostic property
upper_depth	smallint		Depth of upper horizon boundary (cm)
lower_depth	smallint		Depth of lower horizon boundary (cm)

Table G.11: class_wrb_qualifier - Qualifiers according to the World Reference Base for Soil Resources.

Column	*Type*	*Modifiers*	*Comment*
class_wrb_qualifier_id	integer	sequence	Primary key
class_wrb_id	integer		Foreign key that refers to table ”class_wrb”
qualifier_position	character varying(6)		Qualifier position (prefix/suffix)
qualifier	text		Qualifier name
specifier	text		Specifier name
qualifier_order	smallint		Qualifier ordering number

Table G.12: contact - Holds contact information about people that in some capacity have played a role in the creation, gathering, management, or dissemination of data held in WoSIS.

Column	*Type*	*Modifiers*	*Comment*
first_name	text		Contact first name and Primary key together with ”last_name” and ”contact_organization_id” column
last_name	text		Contact last name and Primary key together with ”first_name” and ”contact_organization_id” column
contact_organization_id	character varying(20)		Organisation acronym or abbreviation and Primary key together with ”first_name” and ”last_name” column
department	text		Department
job_title	text		Job title
international_prefix	character varying(5)		Telephone international prefix
telephone	character varying(20)		Work telephone
email	text		Work E-mail
note	text		Comments field

Table G.13: contact_organization - Holds contact information about organizations that in some capacity have played a role in the creation, gathering, management, or dissemination of data held in WoSIS.

Column	*Type*	*Modifiers*	*Comment*
contact_organization_id	character varying(20)		Primary key and organisation acronym or abbreviation
name_main	text		Organization main name
name_sub	text		Organization sub name
international_prefix	character varying(5)		Telephone international prefix
telephone	character varying(20)		Telephone number
email	text		E-mail
url	text		Web site
country_id	character(2)		Foreign key that refers to table ”country”
city	text		City
street_name	text		Street name
street_number	character varying(20)		Street number
postal_code	character varying(20)		Postal (ZIP) code
note	text		Comments field

Table G.14: country - Global Administrative Unit Layers (GAUL) from FAO and ISO 3166 International Standard country codes.

Column	*Type*	*Modifiers*	*Comment*
country_id	character(2)		Primary key and ISO 3166-1 alpha-2, two-letter country code
iso3_code	character(3)		ISO 3166-1 alpha-2, three-letter country code
gaul_code	integer		Global Administrative Unit Layers (GAUL) country code
color_code	character(3)		Country map colour by GAUL
ar	text		Country name in Arabic
en	text		Country name in English
es	text		Country name in Spanish
fr	text		Country name in French
pt	text		Country name in Portuguese
ru	text		Country name in Russian
zh	text		Country name in Chinese
status	text		Status of the country
disp_area	character varying(3)		Unsettled Territory (True/False)
capital	text		Country capital name
continent	text		Continent name
un_reg	text		UN region name
unreg_note	text		Note about UN region
continent_custom	text		Customized Continent name

Table G.15: dataset - Describe datasets imported to the WoSIS database or datasets that are known to contain some of the profiles in the imported datasets.

Column	*Type*	*Modifiers*	*Comment*
dataset_id	character varying(20)		Primary key
dataset_title	text		Dataset title
summary	text		Dataset summary
dataset_version	text		Dataset version
publication_date	text		Publication date
dataset_folder	text		File system folder with the source dataset
licence_type	text		Type of CC license for the dataset
licence_wosis	boolean		If the dataset can be shared via WoSIS web-services to the public
licence_soilgrids	boolean		If the dataset can be used to produce derived data as SoilGrids
licence_file	text		Path to the License document in the file system
dataset_priority	smallint		Dataset import priority. Zero (0) means it was already imported
dataset_progress	character varying(20)		Dataset import progress (Planned/In progress/Complete/Not planned)
n_profiles	integer		Number of profiles in the dataset
n_layers	integer		Number of layers in the dataset
n_profile_attr	integer		Number of different attributes in the dataset describing the site where the profile was taken
n_layer_attr	integer		Number of different attributes in the dataset describing layers
n_profile_rows_inserted	integer		Number of inserted records from the dataset describing the site
n_profile_rows_standard	integer		Number of standerdized records from the dataset describing the site
n_layer_rows_inserted	integer		Number of inserted records from the dataset describing the layers
n_layer_rows_standard	integer		Number of standerdized records from the dataset describing the layers
n_profiles_estimate	integer		Estimated number of soil profiles descriptions. Before dataset import, useded to evaluate import priority
dataset_type	text		Type of dataset (Source/Compilation/Note)
dataset_rank	smallint		Ranking of dataset based on expert knowledge; a measure for the inferred degree of confidence in given dataset. The smallest the number the higher the confidence
uuid	text		Foreign key that refers to table ”geonetwork.metadata” for further detailed metadata about the dataset
licence_check	boolean		If the license agreement was checked

Column	*Type*	*Modifiers*	*Comment*
note	text		Comments field
geom	geometry(Polygon,4326)		Bounding box polygon geometry of the points in the dataset

Table G.16: dataset_contact - Holds contact information about people that in some capacity have played a role in the creation, gathering, management, or dissemination of data held in WoSIS.

Column	*Type*	*Modifiers*	*Comment*
dataset_id	character varying(20)		Foreign key that refers to table ”dataset”
first_name	text		Contact first name and Foreign key together with ”last_name” and ”contact_organization_id” column
last_name	text		Contact last name and Foreign key together with ”first_name” and ”contact_organization_id” column
role	character varying(20)		Role played in the elaboration of the dataset
contact_organization_id	character varying(20)		Organisation acronym or abbreviation and Foreign key together with ”first_name” and ”last_name” column

Table G.17: dataset_profile - Links soil profiles to one or more (source) datasets and stores the original code of each profile as used in the respective source databases.

Column	*Type*	*Modifiers*	*Comment*
dataset_id	character varying(20)		Primary key that together with ”profile_id” and Foreign key refers to table ”dataset”
profile_id	integer		Primary key that together with ”dataset_id” and Foreign key refers to table ”profile”
profile_code	text		Code for soil profile as used in the source database
note	text		Comments field

Table G.18: desc_attribute - Description of all the soil properties for each dataset that has been imported in WoSIS.

Column	*Type*	*Modifiers*	*Comment*
desc_attribute_id	integer	sequence	Primary key
dataset_id	character varying(20)		Foreign key that refers to table ”dataset”
schema_name	text		Database schema where the dataset was stored in preparation for import into WoSIS
table_name	text		Table name for the source where de soil property comes from
column_name	text		Column name for the source where de soil property comes from
source_attribute_name	text		Source soil property name
source_attribute_description	text		Source soil property description
source_attribute_unit	text		Source soil property unit
source_attribute_type	text		Source soil property type
source_attribute_domain	text		Source soil property domain
attribute_type	text		If the attribute refers to Site or Horizon
desc_attribute_standard_id	character varying(100)		Standard soil property name and Foreign key that refers to table ”desc_attribute_standard”
conversion	text		Conversion applied to standerdize the values
sql_insert	text		SQL code for moving the data from the source into WoSIS
sql_standard	text		SQL code used to standardize the data
number_rows_source	integer		Number of records in the source database
number_rows_inserted	integer		Number of records inserted into WoSIS
number_rows_standard	integer		Number of records that have been standardized
conflict_source_min_value	text[]		Array of conflict values from source compared with the allowed minimum value
conflict_source_max_value	text[]		Array of conflict values from source compared with the allowed maximum value
conflict_source_domain	text[]		Array of conflict values from source compared with the allowed domain values
conflict_standard_min_value	text[]		Array of conflict values from standard compared with the allowed minimum value
conflict_standard_max_value	text[]		Array of conflict values from standard compared with the allowed maximum value
conflict_standard_domain	text[]		Array of conflict values from standard compared with the allowed domain values
user_name	text	”current_user”()	Who imported the soil proprety
start_date	date	(now())::date	Date when this soil property was imported
start_time	time without time zone	(now())::time without time zone	Start time when this soil property was imported

Column	*Type*	*Modifiers*	*Comment*
proce_time	time without time zone		Duration of the import of the soil property
client_addr	text	inet_client_addr()	IP address from where the soil property was imported
server_addr	text	inet_server_addr()	IP address to where the soil property was exported
note	text		Comments field

Table G.19: desc_attribute_standard - Discription of standard attributes and standardisation progress.

Column	*Type*	*Modifiers*	*Comment*
desc_attribute_standard_id	character varying(100)		Standard attribute name
attribute_type	character varying(7)		Attribute type (Site/Horizon)
attribute_agg_name	text		Standard attribute aggregation name
desc_unit_id	character varying(15)		Standard attribute unit
data_type	character varying(10)		Standard attribute type
decimals	smallint		Number of decimals
minimum	numeric(5,2)		Minimum value
maximum	numeric(9,2)		Maximum value
desc_domain_id	character varying(255)		Standard attribute domain
priority	smallint		Standardisation priority
progress_name	boolean		If the attribute name has been standardised
progress_value	boolean		If the attribute measured values have been standardised
progress_method	boolean		If the attribute analytical methods have been standardised
distribute	boolean		If distribution is intended
gsm	boolean		If the ttribute is from Global Soil Map
gfsd	boolean		If the ttribute is from Guide lines for soil description FAO (2006)
note	text		Comments field
observed_minimum	numeric(6,2)		Minimum observed value
observed_maximum	numeric(6,2)		Maximum observed value
n_attributes	integer		Number attributes that have been standardized
n_profiles	integer		Number of profiles
n_rows_inserted	integer		Number of rows inserted
n_rows_standard	integer		Number of rows that have been standardized
attribute_code	character(4)		Standard attribute code
attribute_description	text		Standard attribute description
attribute_common_name	text		Attribute common name

Table G.20: desc_domain - Data domains that are available for the categorical soil properties; a data domain refers to all unique values which a given data element (attribute) may contain.

Column	*Type*	*Modifiers*	*Comment*
desc_domain_id	text		Primary key
reference_id	bigint		Foreign key that refers to table ”reference”
note	text		Comments field
page_num	smallint		Page number, from the referenced document
object_num	smallint		Figure or table number from the referenced document

Table G.21: desc_domain_value - Description per domain (as defined in the ”desc_domain” table) of all unique values which a site, soil, or terrain characteristic may contain.

Column	*Type*	*Modifiers*	*Comment*
desc_domain_id	text		Foreign key that refers to table ”desc_domain”
domain_value	text		Domain class code
description	text		Domain class description
explanation	text		Domain class explanation

Table G.22: desc_laboratory - Listing of laboratories where soil samples have been analysed.

Column	*Type*	*Modifiers*	*Comment*
desc_laboratory_id	integer	sequence	Primary key
acronym	character varying(25)		Laboratory acronym or abbreviation
lab_name_1	text		Laboratory name, level 1
lab_name_2	text		Laboratory name, level 2
country_id	character(2)		Foreign key that refers to table ”country”
city	text		The city where the laboratory is located
postal_code	character varying(20)		Laboratory postal code
street_name	text		Laboratory street name
street_number	character varying(10)		Laboratory street number
note	text		Comments field

Table G.23: desc_method_feature - Criteria used to standardise disparate soil analytical method descriptions, for a given analytical method, according to a defined set of features.

Column	*Type*	*Modifiers*	*Comment*
attribute_agg_name	character varying(100)		Standard attribute aggregation name
feature_name	character varying(50)		Standard analytical method feature name. Combined primary key (”standard_attribute_name”,”feature_name”,”feature_option”)
feature_option	text		Standard analytical method feature option. Combined primary key (”standard_attribute_name”,”feature_name”,”feature_option”)
feature_description	text		Description of the standard analytical method option
feature_option_code	smallint		Code of the standard analytical method option
feature_option_order	smallint		Order of the standard analytical method options

Table G.24: desc_method_source - Analytical methods descriptions as defined in the respective source databases (i.e. prior to standardisation).

Column	*Type*	*Modifiers*	*Comment*
desc_method_source_id	integer	sequence	Primary key
source_analytical_method_name	text		Source analytical method name
note	text		Comments field
afsp_code	character varying(7)		AfSP analytical method code
isis_code	character varying(9)		ISIS analytical method code
wise_code	character varying(4)		WISE analytical method code
soter_code	character varying(16)		SOTER analytical method code
other_code	character varying(15)		Other datasets analytical method code
method_description	text		Name of analytical method description

Table G.25: desc_method_standard - Results of the standardization of the Soil Analytical Methods descriptions to WoSIS 2 standards.

Column	*Type*	*Modifiers*	*Comment*
descriptor_id	smallint		Foreign key that refers to table ”descriptor”
attribute_agg_name	text		Standard soil property name. Combined foreign key (”standard_attribute_name”,”feature_name”,”feature_option”)
feature_name	text		Standard analytical method feature name. Combined foreign key (”standard_attribute_name”,”feature_name”,”feature_option”)
feature_option	text		Standard analytical method feature option. Combined foreign key (”standard_attribute_name”,”feature_name”,”feature_option”)
confidence	character varying(10)		Confidence in the standardization procedure
reference_id	smallint		Foreign key that refers to table ”reference”

Table G.26: desc_unit - Units used for measurement of soil, site, and terrain characteristics.

Column	*Type*	*Modifiers*	*Comment*
desc_unit_id	character varying(15)		Primary key
unit_description	character varying(50)		Unit description

Table G.27: descriptor - Unique combinations of Attribute, Analytical Method and Laboratory definition that define a measured value.

Column	*Type*	*Modifiers*	*Comment*
descriptor_id	integer	sequence	Primary key
desc_attribute_id	smallint		Foreign key that refers to table ”desc_attribute”
desc_method_source_id	smallint		Foreign key that refers to table ”desc_method_source”
desc_laboratory_id	smallint		Foreign key that refers to table ”desc_laboratory”
note	text		Comments field

Table G.28: image - Images that ilustrate soil profiles or the site of their provenance.

Column	*Type*	*Modifiers*	*Comment*
image_id	integer	sequence	Primary key
country_id	character(2)		Foreign key that refers to table ”country”
identification	text		Image internal code
image_year	smallint		Year in which the image was produced
image_timestamp	timestamp without time zone		Timestamp of the image
author	text		Image author
image_medium_type	text		Image medium type
extent_height	integer		Image height extent
extent_width	integer		Image width extent
extent_unit	character varying(10)		Image height and width units
file_path	text		Path to image file
file_path_description	text		File path description
description	text		Image description
description_location	text		Image location description
note	text		Comments field
internal_note	text		Image internal note
coordinates	text		Image coordinates (WGS 84)
copyright	boolean		The image is copyrighted (True/False)
qualified_for_web_publication	boolean		Image resolution is suited for web publication (True/False)
geom	geometry(Point,4326)		Point geometry of the location of the picture

Table G.29: image_profile - Links between images and the corresponding profiles.

Column	*Type*	*Modifiers*	*Comment*
profile_id	integer		Primary key
image_id	integer		Foreign key that refers to table ”image”
note	text		Comments field

Table G.30: image_subject - Images listed by subject (meant to store pictures tags).

Column	*Type*	*Modifiers*	*Comment*
image_id	integer		Primary key that together with ”subject” and Foreign key refers to table ”image”
subject	text		Image subject tag and Primary key that together with ”image_id”

Table G.31: map_attribute - Holds values of any characteristic associated with ”map_unit”,”map_unit_component” and ”map_unit_soil_component”.

Column	*Type*	*Modifiers*	*Comment*
map_attribute_id	integer	sequence	Primary key
map_unit_id	integer		Foreign key that refers to table ”map_unit”
map_unit_component_id	smallint		Foreign key that refers to table ”map_unit_component”
map_unit_soil_component_id	smallint		Foreign key that refers to table ”map_unit_soil_component”
descriptor_id	integer		Foreign key that refers to table ”descriptor”
observation_date	time with time zone		Date of observation or measurement
source_value	text		Attribute source value
standard_value	text		Attribute standardised value
accuracy	character varying(20)		Accuracy of an observation or measurement precision
trust	character(1)	’A’::bpchar	Level of trust: ’A’ as entered, no validation; ’B’ standardized; ’C’ harmonized

Table G.32: map_unit - Polygon geometry of homogeneous map unit features.

Column	*Type*	*Modifiers*	*Comment*
map_unit_id	integer	sequence	Primary key
dataset_id	character varying(20)		Foreign key that refers to table ”dataset”
country_id	character(2)		Foreign key that refers to table ”country”
source_id	character varying(20)		Original map unit code
geom	geometry(MultiPolygon,4326)		Map unit geometry

Table G.33: map_unit_component - Information about mapping unit component.

Column	*Type*	*Modifiers*	*Comment*
map_unit_component_id	integer	sequence	Primary key
map_unit_id	integer		Foreign key that refers to table ”map_unit”
unit_number	smallint		Map unit component number
proportion	smallint		Map unit component proportion

Table G.34: map_unit_soil_component - Coding of soil components.

Column	*Type*	*Modifiers*	*Comment*
map_unit_soil_component_id	integer	sequence	Primary key
map_unit_component_id	integer		Foreign key that refers to table ”map_unit_component”
unit_number	smallint		Map unit soil component, number
proportion	smallint		Map unit soil component, proportion

Table G.35: map_unit_soil_component_x_profile - Links soil components, listed in table ”map_unit_soil_component”, to one or more reference profiles (table still void in this version of WoSIS).

Column	*Type*	*Modifiers*	*Comment*
map_unit_soil_component_x_profile_id	integer	sequence	Primary key
map_unit_soil_component_id	integer		Foreign key that refers to table ”map_unit_soil_component”
dataset_id	character varying(20)		Foreign key that together with ”profile_id” refers to table ”dataset_profile”
profile_id	integer		Foreign key that together with ”dataset_id” refers to table ”dataset_profile”

Table G.36: profile - list of soil profiles, with their location (geometry).

Column	*Type*	*Modifiers*	*Comment*
profile_id	integer	sequence	Primary key
country_id	character(2)		Foreign key that refers to table ”country”
geom_accuracy	real		Accuracy of the geometry in degrees. Example: If degree, minutes and seconds are provided then geom_accuracy is asigned with the value 0.01, if seconds are missing then 0.1, if seconds and minutes are missing then 1
uuid	uuid		Universally unique identifier profile code
geom	geometry(Point,4326)		Point geometry of the location of the profile
trust	character(1)	’A’::bpchar	Level of trust: ’A’ as entered, no validation; ’B’ standardized; ’C’ harmonized
date_time	timestamp without time zone		Timestamp of the GPS reading. May be used for differential correction purposes
geom_accuracy_type	character varying(3)	’UNK’::bpchar	GPS - Coordinates come from a GPS; MAP - Coordinates come from a map; UNK - Unknown coordinates source
sample_type	character varying(10)	’single’::character varying	Either single or composite sample
sample_number	integer	1	Number of samples
sample_area	integer	1	Area sampled (m2)
country_geom_closest	character(2)		Closest country to profile, based on geometry
country_geom_distance	integer		Distance of the closest country to profile, based on geometry
hide	boolean	false	If it is a hidden profile or not

Table G.37: profile_attribute - Values of characteristics that are associated with the profiles site.

Column	*Type*	*Modifiers*	*Comment*
profile_attribute_id	integer	sequence	Primary key
profile_id	integer		Foreign key that together with ”profile_id” refers to table ”dataset_profile”
dataset_id	character varying(20)		Foreign key that together with ”dataset_id” refers to table ”dataset_profile”
descriptor_id	integer		Foreign key that refers to table ”descriptor”
accuracy	character varying(20)		Accuracy of an observation or measurement precision
trust	character(1)	’A’::bpchar	Level of trust: ’A’ as entered, no validation; ’B’ standardized; ’C’ harmonized
observation_date	date		Date of observation or measurement
source_value	text		Profile attribute, source value
standard_value	text		Profile attribute, standard value

Table G.38: profile_layer - Lists of Layers depths and samples definition per profile and dataset.

Column	*Type*	*Modifiers*	*Comment*
profile_layer_id	integer	sequence	Primary key
dataset_id	text		Foreign key that together with ”profile_id” refers to table ”dataset_profile”
profile_id	integer		Foreign key that together with ”dataset_id” refers to table ”dataset_profile”
layer_number	smallint		Consecutive layer number rated from top to bottom
sample_type	text		Type of sample taken
sample_code	text		Sample code
sample_composition	text		Sample composition
sample_available	text		Sample availability
upper_depth	smallint		Depth of upper horizon boundary (cm)
lower_depth	smallint		Depth of lower horizon boundary (cm)
upper_depth_source	real		Depth of upper horizon boundary (cm) from source
lower_depth_source	real		Depth of lower horizon boundary (cm) from source
layer_name_source	text		Horizon designation from source
layer_name	text		Adapted horizon designation
note	text		Comments field
u_tmp	smallint		
l_tmp	smallint		
upper_sample_depth	smallint		Depth of upper sample boundary (cm)
lower_sample_depth	smallint		Depth of lower sample boundary (cm)

Table G.39: profile_layer_attribute - Lists values of any characteristic associated with a profile layer.

Column	*Type*	*Modifiers*	*Comment*
profile_layer_attribute_id	integer	sequence	Primary key
profile_layer_id	integer		Foreign key that refers to table ”profile_layer”
descriptor_id	integer		Foreign key that refers to table ”descriptor”
accuracy	character varying(20)		Accuracy of an observation or measurement precision
trust	character(1)	’A’::bpchar	Level of trust: ’A’ as entered, no validation; ’B’ standardized; ’C’ harmonized
observation_date	date		Date of observation or measurement
source_value	text		Layer attribute, source value
standard_value	text		Layer attribute, standard value

Table G.40: profile_layer_thinsection - Lists information related with thin sections.

Column	*Type*	*Modifiers*	*Comment*
dataset_id	character varying(15)		Foreign key that refers to table ”dataset”
profile_code	character varying(20)		Profile code used in the source database
profile_id	integer		Foreign key that refers to table ”profile”
profile_layer_id	integer		Foreign key that refers to table ”profile_layer”
thinsection_id	integer		Thinsection collection identifier
sample_date	date		Date of sample
author	character varying(10)		Thin section author
horizon	character varying(10)		Soil horizon
upper_depth	integer		Depth of upper horizon boundary (cm)
lower_depth	integer		Depth of lower horizon boundary (cm)
sample_size	character varying(20)		Sample size
missing	boolean		Missing thin section from collection
note	text		Comments field

Table G.41: raster - Registed raster layers.

Column	*Type*	*Modifiers*	*Comment*
rid	integer	sequence	Primary key
descriptor_id	smallint		Foreign key that refers to table ”descriptor”
rast	raster		Raster binary data
filename	text		Path to raster file

Table G.42: reference - List of references to source materials managed in WoSIS.

Column	*Type*	*Modifiers*	*Comment*
reference_id	integer	sequence	Primary key
reference_type	character varying(13)		Type of reference (Publication/Map/URL/Digital media)
isbn	character varying(20)		ISBN number
isn	integer		ISRIC library code
title	text		Title
subtitle	text		Subtitle
issue	text		Issue
serie	text		Series
page	character varying(10)		Page
publication_year	smallint		Publication year
publisher	text		Publisher
url	text		URL
map_type	text		Map type
map_sheet	character varying(20)		Map sheet
map_scale	integer		Map scale
min_latitude	numeric(7,5)		Min latitude
max_latitude	numeric(7,5)		Max latitude
min_longitude	numeric(8,5)		Min longitude
max_longitude	numeric(8,5)		Max longitude
digital_format	character varying(20)		Digital format
content	text		Content description
note	text		Comments field

Table G.43: reference_author - List of author names considered in table references.

Column	*Type*	*Modifiers*	*Comment*
reference_id	smallint		Foreign key that refers to table ”reference”
first_name	text		Contact first name and Foreign key together with ”last_name” and ”contact_organization_id” column
last_name	text		Contact last name and Foreign key together with ”first_name” and ”contact_organization_id” column
contact_organization_id	character varying(255)		Organisation acronym or abbreviation and Foreign key together with ”first_name” and ”last_name” column

Table G.44: reference_dataset - List of references by imported dataset.

Column	*Type*	*Modifiers*	*Comment*
reference_id	integer		Primary key
dataset_id	character varying(20)		Foreign key that refers to table ”dataset”

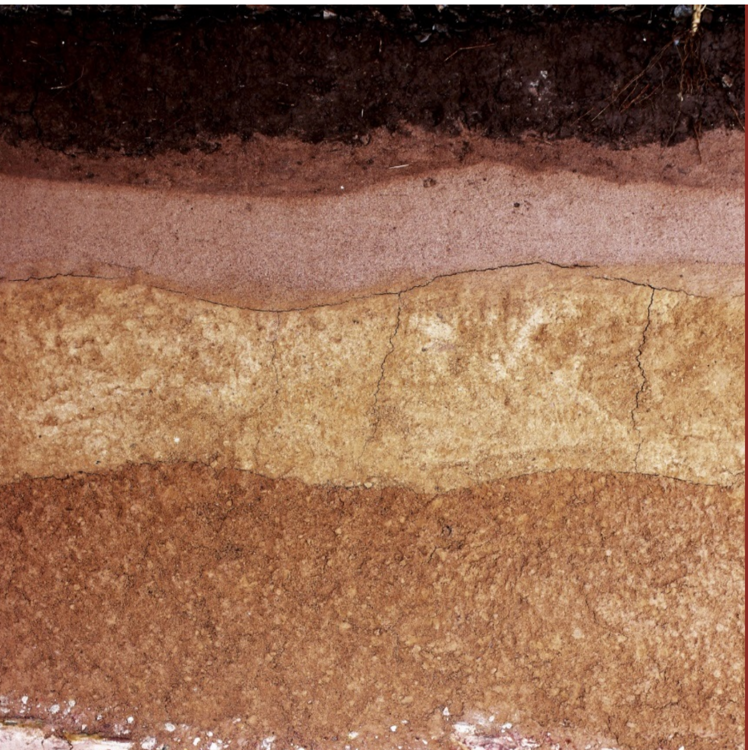
Table G.45: reference_dataset_profile - List of references by imported dataset and their profiles.

Column	*Type*	*Modifiers*	*Comment*
reference_id	integer		Primary key that together with dataset and profile
dataset_id	character varying(20)		Foreign key that together with ”profile_id” refers to table ”dataset_profile”
profile_id	integer		Foreign key that together with ”dataset_id” refers to table ”dataset_profile”
reference_page	character varying(50)		Reference page where the profile description appears
reference_profile_code	character varying(50)		Profile code in such reference

Table G.46: reference_file - List of references that have a file as support.

Column	*Type*	*Modifiers*	*Comment*
reference_file_id	bigint	sequence	Primary key
reference_id	integer		Foreign key that refers to table ”reference”
file_type	text		File type
file_name	text		File name
file_description	text		File description
file_path	text		File path

Together with our partners we produce, gather, compile and serve quality-assessed soil information at global, national and regional levels. We stimulate the use of this information to address global challenges through capacity building, awareness raising and direct cooperation with users and clients.



www.isric.org