ISRIC
World Soil Information

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

Procedures Manual 2020

ISRIC Report 2020/01

Eloi Ribeiro
Niels H. Batjes
Ad van Oostrum

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
GDPR   General Data Protection Regulation
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 and Wageningen University Library
ISO   International Organization for Standardization
ISRIC   ISRIC - World Soil Information (legally registered as: International Soil Reference and Information 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
SOP   Standard Operating Procedures e.g. as developed by GLOSOLAN
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 ISC World Data System (ISC-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 profile data 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)'. This can be indicated using a Creative Commons license.

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, with the ultimate aim to make them 'comparable as if assessed by a single given (reference) method'. The most recent set of standardized soil profile data served from WoSIS, commonly referred to as 'WoSIS Latest', may be accessed freely through our Soil Data Hub (http://data.isric.org). Further, on an irregular basis, we provide static snapshots of the data for consistent citation purposes.

In its capacity as World Data Center (WDC) for Soils, ISRIC also serves its data products to the global user community through auxiliary portals, in particular those of the International Science Council (ISC) World Data System (WDS) 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

ISRIC - World Soil Information manages a centralized PostgreSQL database, known as WoSIS (World Soil Information Service), for the shared benefit of the international community. The original aim, in 2013, was to accommodate any type of soil data (point, polygon, and grid) (Ribeiro et al., 2015; Tempel et al., 2013). Presently, however, the scope is on safeguarding, processing and standardizing geo-referenced soil profile data for the world. Grid and polygon data/maps are now handled using other components of our spatial data infrastructure, such as ISRIC's soil data hub (http://data.isric.org) and the SoilGrids/WoSIS portal (https://soilgrids.org/).

The general procedure for processing profile data in WoSIS is as follows. First, new source data are imported 'as is' into the ISRIC data repository, i.e keeping the original file format as well as the original naming and coding conventions, abbreviations, domains, units of measurement, lineage (provenance) and data licence. Subsequently, these diverse 'source' datasets are converted into PostgreSQL format using an automated procedure. Thereafter, the source data are manually 'mapped' to the standard WoSIS naming conventions, standard values and/or units of measurement. Basically, this corresponds with the first major step of standardization to make the data queryable and usable. The second step of standardization involves importing the standardized data into the WoSIS data model itself; this process also involves automated checks on the plausibility of the reported values.

Methodological changes, since the release of the preceding version (Ribeiro et al., 2018), are described in the present document. Presently, we consider a wider range of soil chemical properties for standardization: organic carbon, total carbon, total carbonate equivalent, total nitrogen, phosphorus (extractable P, total P and P retention), soil pH, cation exchange capacity and electrical conductivity. We also consider the following physical properties: soil texture (sand, silt, and clay), bulk density, coarse fragments and water retention. Both of these sets of properties are grouped according to analytical procedures that are operationally comparable. Further, for each profile we provide the original soil classification (FAO, WRB and USDA, with their version) and original horizon designations, insofar as these have been specified 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, with accompanying flowcharts (Appendix E).

The ultimate third step, data harmonization, aims to make the analytical data comparable that is as 'if assessed by a single given (reference) method'. Full data harmonization in a global context, however, will first become feasible once results of extensive proficiency testing programmes become available, for example in the framework of Pillar 5 of the Global Soil Partnership and GLOSOLAN (Global Soil Laboratory Network), and a common set of 'international' reference methods (e.g. ISO procedures or SOP's developed by GLOSOLAN) has been accepted by the international soil community.

WoSIS is an important component of ISRIC's Spatial Data Infrastructure (SDI). Further developments will consider how soil data, shared by third parties in an inter-operable way, can be ingested and standardized. As such, WoSIS complements the activities of the federated Global Soil Information System (GLOSIS) developed by the Global Soil Partnership (GSP, 2018).

Chapter 1

Introduction

The World Soil Information Service, or WoSIS in short, 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 to the international community. These data, in combination with spatial environmental co-variate layers, can then be used to generate soil property maps using digital soil mapping (Hengl et al., 2017; de Sousa et al., 2020). Subsequently, such maps can be used to address global challenges such as climate change, food security, and the degradation of land and water resources at continental and (sub)regional scale.

This report supersedes the preceding version of the WoSIS procedures manual (Ribeiro et al., 2018). It consists of six Chapters and eight Appendices. Following up on the introduction, and prior to describing the database design (Chapter 3) itself, basic principles for flagging repeated (e.g. duplicate) soil profiles originating from different international data compilations, basic measures for defining data quality, as well as the main steps towards standardization and harmonization of numerical soil data are discussed (Chapter 2). Aspects of data inter-operability and web services (Chapter 4) as well as developments towards a federated soil database, set in the context of ISRIC's Spatial Data Infrastructure (SDI) and contributions to the GSP's Global Soil Information System (GLOSIS, de Sousa et al. (2019)) are discussed in Chapter 5. Finally, future developments are discussed in Chapter 6.

Appendix A explains how the quality-assessed and standardized data served from WoSIS can be accessed by users. Appendix B describes basic principles for compiling soil profile data to facilitate entry into WoSIS; standardized templates are provided 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. An example of a possible approach for harmonising the standardized data, using pH as an example, is provided in Appendix G. The database tables, 39 in total, are described in Appendix H.

Chapter 2

Basic principles for processing data

As indicated, the 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.

This Chapter discusses the main principles for processing new source data into WoSIS upon their registry in the ISRIC WDC-Soils repository (i.e., metadata of the source data, license and provenance). It subsequently discusses the 'flagging of repeated soil profiles' as commonly observed with data compilations, 'checks for completeness of the source information', 'measures for data quality', and 'main steps towards data harmonization'.

2.1 Flagging repeated soil profiles

One of the first tasks in the process of importing data into WoSIS is the search for repeated soil 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), the Africa Soil Profile Database (Leenaars et al., 2014b) and the International Soil Carbon Network (Nave et al., 2017). This elaborate screening process will yield a unique set of soil profiles and thus produce a truthful profile count (Figure 2.1).
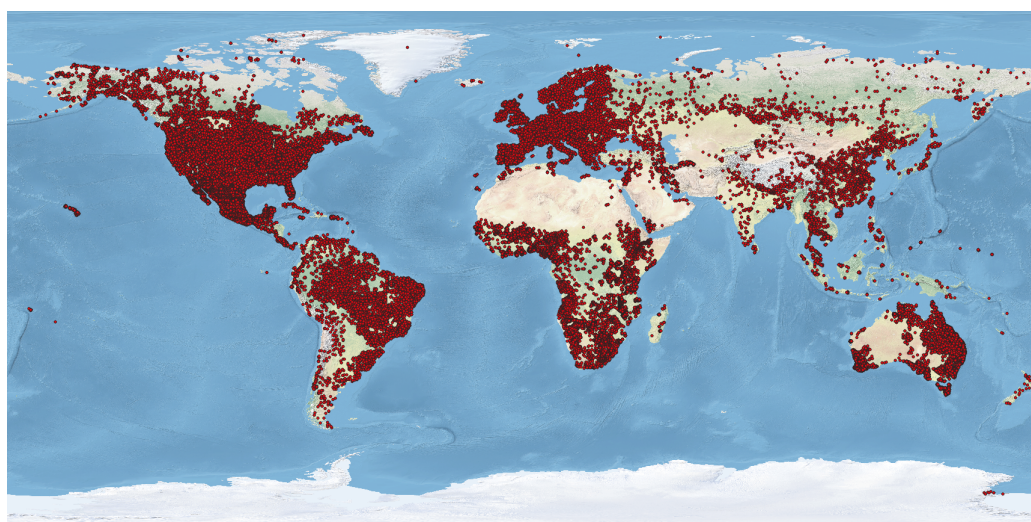
Figure 2.1: Location of unique, shared geo-referenced profiles served from WoSIS (September 2019).

Repeated profiles are identified using checks on lineage and checks on geographical proximity. The first, considers the data source identifiers, uses this information to trace the original data source, and from there looks for duplicates. Alternatively, the proximity check is based on the geographic coordinates. The procedure first identifies profiles that are within <100 m of each other as possible duplicates. Upon additional, semi-automated checks concerning the first three layers (upper and lower depth), i.e. sand, silt and clay content, the duplicates with the least comprehensive component of attribute data are flagged. When still in doubt at this stage, additional visual checks are made with respect to other commonly reported soil properties, such as pH_{water} and organic carbon content. So far, this laborious, yet critical, screening process has led to the exclusion of some 50,000 profiles from the original complement of some 450,000 soil profiles (Batjes et al., 2020).
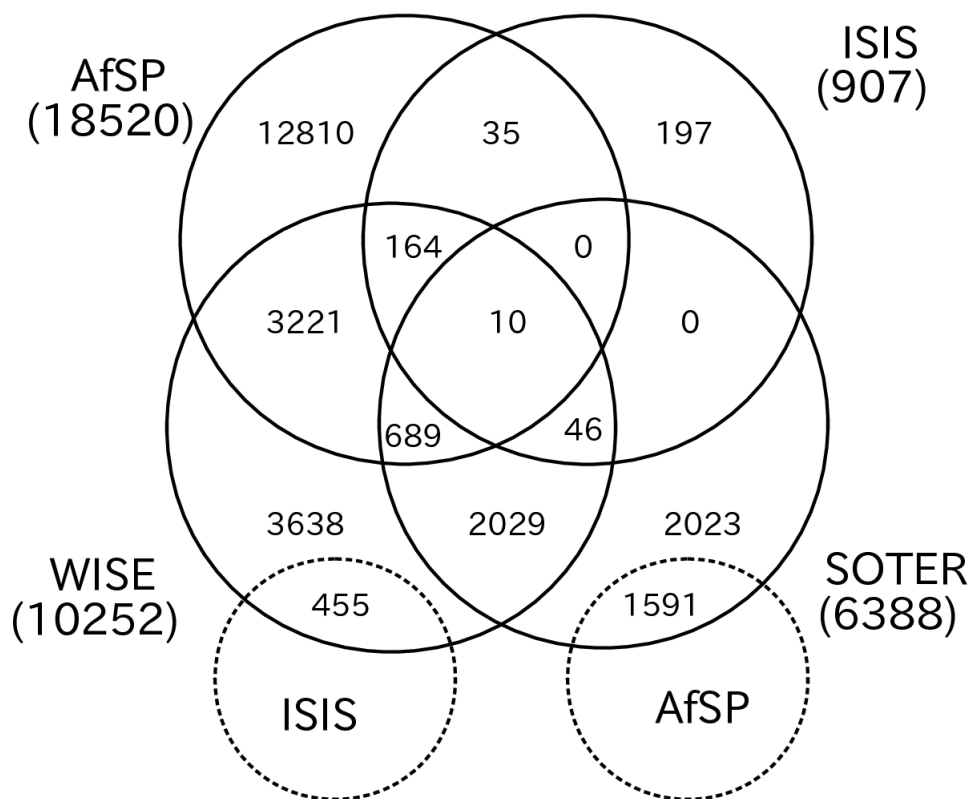
Figure 2.2: Intersection between ISRIC stand-alone profile databases showing the number of overlapping profiles (AfSP, Africa Soil Profile Database; ISIS, ISRIC Soil Information Service, SOTER, Soil and Terrain Database; WISE, World Inventory of Soil Emissions potentials database).

Figure 2.2 illustrates the intersection between four profile databases compiled in the framework of collaborative ISRIC projects: AfSP (Leenaars et al., 2014a), ISIS [1], WISE (Batjes, 2009)) and various national and continental scale SOTER 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 are 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

Table 2.1: Basic requirements for considering soil profiles in the WoSIS standardisation workflow.

Case	(X,Y)	Layer depth	Soil properties	Classification	Keep
1	+	+	+	+	Yes
2	+	+	+	-	Yes
3	+	-	-	+	Yes[3]
4	-	+	+	+	Yes/no[4]
5	-	+	+	-	Yes/no
6	+	+	-	-	No
7	+	-	+	-	No[5]

2.2 Checks for completeness of the source information

To be considered in the actual WoSIS standardization workflow, each profile must meet several criteria (Table 2.1). First, we assess if each profile is geo-referenced, has (consistently) defined upper and lower depths for each layer (or horizon), and data for at least some soil properties (e.g. sand, silt, clay and pH). Having a soil (taxonomic) classification is considered desirable (case 1), though not mandatory (case 2). Georeferenced profiles for which only the classification is specified can still be useful for mapping of soil taxonomic classes (case 3). Alternatively, profiles without any geo-reference may still prove useful to develop pedotransfer functions (case 4 and 5); however, they cannot be served through WFS because there is no geometry (x,y). The remaining cases (6 and 7) are automatically excluded from the WoSIS workflow.

2.3 Measures for data quality

2.3.1 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; Soil Survey Division Staff, 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.
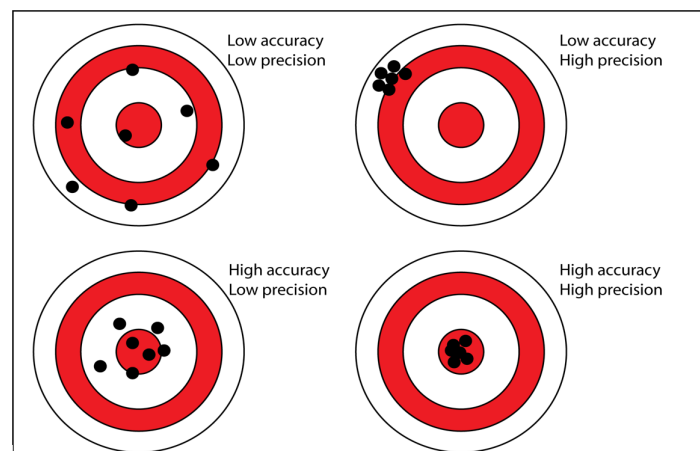

Figure 2.3: Depiction of the accuracy and precision of measurements.

Quality of data can be evaluated against a range of properties, for example positional accuracy, date of observation, 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; Schut and Giller, 2020).

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.

To address and document the above issues, four 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 2.3.2), and

- Accuracy and precision, the *inferred* Laboratory related uncertainty (column: accuracy; see Section 2.3.3).

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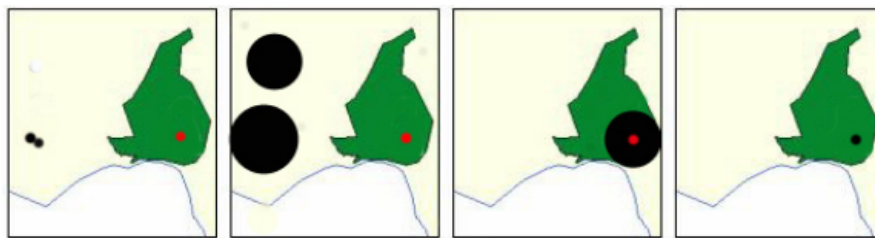
Figure 2.4: Differences between accuracy and precision in a spatial context (From left to right: High precision, low accuracy; Low precision, low accuracy showing random error; Low precision, high accuracy; High precision and high accuracy).

The above indicators were introduced to provide 'flags' that allow the WoSIS database manager to recognize factors that may compromise the quality of certain data, hence their suitability for use. Consideration of these 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 does not allow for a full characterization of all indicators (Appendix C).

2.3.2 Level of trust

Attributes managed in WoSIS are characterized in terms of inferred trust, for which there are three levels. The first level ('A') is used for datasets submitted 'as is'. Such 'A' level data are subsequently standardized to level 'B'. This second step considers the soil property, analytical method and unit of measurement. It also includes automated error-checking for possible inconsistencies with respect to 'plausible' minimum and maximum values, complemented with some visual checks that also consider the mean; these lower and upper limits are based on expert knowledge of field pedologists. These are 'one-dimensional' checks per layer of the attribute values themselves, without considering values reported for other soil attributes for the same layer (e.g. total nitrogen content, irrespective of the organic carbon content or C/N ratio; bulk density irrespective of the content of organic matter). These criteria are similar to those developed to screen soil chemical and physical properties in the WISE (see Batjes (2008), App. 3) and AfSP (see Leenaars et al. (2014a), App. 6) soil profile data compilations.

Ultimately, B-level data can be harmonized ('C') to an agreed reference (target) method, subject to international agreement about the recommended target method as well as the availability of appropriate

(soil type specific) transfer functions (Baritz et al., 2014; GlobalSoiMap, 2015). As such, level 'C' is considered the highest achievable degree of harmonization in WoSIS. Such values have to be approved by a (regional) expert who has performed an in-depth check, ideally considering relations between the available values (e.g. soil carbon content *versus* bulk density) in relation to the soil layer itself as well as the full soil profile, and found no *apparent* anomalies. Users should note that level 'C' requires the different values to be harmonized first (see Section 2.4.2). Therefore, users of the standardized data served from WoSIS (e.g. 'WoSIS Latest') should be aware that the assessment of the 'accuracy and applicability of the data for a particular application is strictly a user responsibility' (Finke, 2006; Schut and Giller, 2020; Guillera-Arroita et al., 2015).

2.3.3 Measurement 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 is not always the case.

Since 2019, we provide a measure for the *inferred* uncertainty associated with the different operationally defined analytical methods. This is a first approximation, which should be revised as more data from laboratory intercomparison programmes, such as WEPAL and GLOSOLAN, become available.

2.3.4 Positional accuracy

All profile coordinates in WoSIS are presented according to the World Geodetic System (i.e. WGS84, EPSG code 4326). These coordinates were converted from a diverse range of national projections. Further, the source referencing may have been in decimal degrees (DD) or expressed in degrees, minutes, and seconds (DMS) for both latitude and longitude. The (approximate) accuracy of geo-referencing in WoSIS is given in decimal degrees (Table 2.2). If the source only provided degrees, minutes, and seconds (DMS) then the geographic accuracy is set at 0.01; if seconds (DM) are missing it is set at 0.1; and if seconds and minutes (D) are missing it is set at 1. For most profiles (86 percent), the approximate accuracy of the point locations, as inferred from the original coordinates given in the source datasets, is less than 10 m (total = 196 498 profiles). Typically, the geo-referencing of soil profiles described and sampled before the advent of GPS (Global Positioning Systems) in the 1970s is less accurate; sometimes we just do not know the 'true' accuracy. Digital soil mappers should duly consider the inferred geometric accuracy of the profile locations in their applications (Grimm and Behrens, 2010), since the soil observations and covariates may not actually correspond in both space and time (Cressie and Kornak, 2003).

Table 2.2: Approximate accuracy of the profile location

Decimal places	Decimal degrees	Approximate precision
7	0.0000001	1 cm
6	0.000001	10 cm
5	0.00001	1 m
4	0.0001	10 m
3	0.001	100 m
2	0.01	1 km
1	0.1	10 km
0	1	100 km

2.4 Main steps towards data harmonization

2.4.1 Data lineage and access conditions

WoSIS aims to facilitate the exchange and use of soil data collated within the context of various 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 common standard they importantly gain in appropriateness for use for broad scale applications.

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. Consequently, data standardization generally occurs a *posteriori*. Such is preferably done by the data providers themselves, as they are best able to correctly interpret the data; this would yield a 'double dataset' holding both the original data as well as the standardized data (Leenaars et al., 2014b). Such would be best be done in the framework of a federated, global soil information system as being developed by GSP (2018). Alternatively, pending availability of an operational federated system, data standardization would largely need to be done by a 'central compiler'. Therefore, any soil dataset shared for consideration in WoSIS should be sufficiently documented, with adequate metadata, to make the data understandable and processable.

Data providers must specify conditions for access to each dataset they submit to the ISRIC Data Repository. This is preferably done using a Creative Commons[6] license, but other licences are also accepted. Importantly, the licence should be provided as part of the metadata, including data lineage or provenance, since it may indicate possible 'inherited restrictions'.

Access conditions for each dataset in the ISRIC Data Repository are enforced through 'access registers' that are managed in the WoSIS database; overall terms and conditions are in accord with the ISRIC Data and Software Policy[7]. Ultimately, only standardized data derived from 'shared' sources will be provided to the international community though our SDI (Appendix A).

Alternatively, some datasets come with more restrictive licences indicating they may only be used at ISRIC for making SoilGrids and similar visualizations. The corresponding profiles, however, are standardized using the same workflow, but the point data themselves are not served to the international community.

2.4.2 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 consideration 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, and facilitate the ingestion of data.

Main steps for processing data in WoSIS are schematized in Figure 2.5. All submitted datasets, pre-screened for completeness (e.g. lineage, licence), 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 diverse 'source' datasets are converted into PostgreSQL format using an automated procedure. Thereafter, the source data are manually 'mapped' to the standard WoSIS naming conventions, standard values and/or units of measurement. Basically, this corresponds with the first major step of standardization to make the data queryable and usable. The second step of standardization involves importing the standardized data into the WoSIS data model itself; this process also involves automated checks on the plausibility of the reported values (e.g. pH should between 1 and 13).

---

[6] http://creativecommons.org/licenses/
[7] http://www.isric.org/about/data-policy

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[8] (Global Soil Laboratory Network) and WEPAL (2015) the 'Wageningen Evaluating Programmes for Analytical Laboratories'.

As indicated by GlobalSoiMap (2015) and others, the necessary pedotransfer functions are likely to be region and soil type specific. Probably, here again, the biggest challenge is the quality of the data used to build the regional transfer functions. Building a transfer function generally is not the issue (i.e. apply an appropriate statistical model to describe the relationship between or among properties). In practice, proficiency testing initiatives have revealed unacceptably variable control charts, suggesting varying methods for the same *named* analysis across laboratories and/or varying quality control measures (and enforcement). Therefore, until analytical methods and quality control are standardized (and enforced), errors of transfer functions built on variously sourced data of varying quality can be expected to be large. In principle, such issues can only be avoided when all soil data are collected and analysed using defined standard methods in a single reference laboratory as is the case with the EU LUCAS Topsoil programme (Orgiazzi et al., 2018) and the NCSS soil database (Soil Survey Division Staff, 2014a; USDA-NCSS, 2018).

[8]http://www.fao.org/global-soil-partnership/resources/events/detail/en/c/1037455/
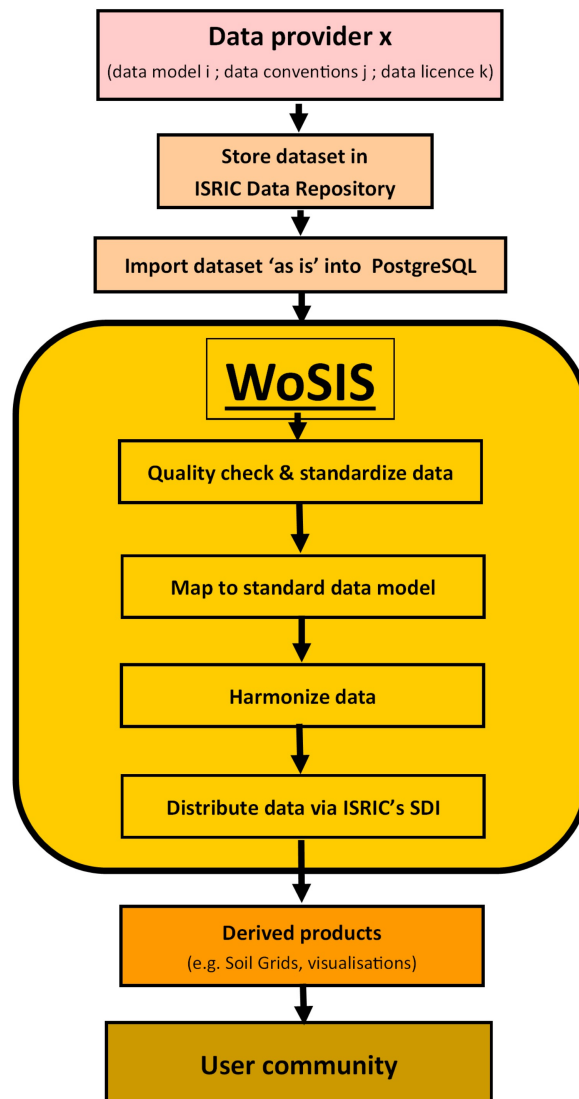
Figure 2.5: Main stages of data standardization and harmonization.

Chapter 3

Database design

3.1 General concept

The database design for WoSIS consists of 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 metadata, auditing and web applications, to enforce role grant access. This will allow other projects or applications, to operate in the same database using different schemas.

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

- Metadata

- Datasets and classification

- Attributes, methods and laboratories

- Profiles, layers and observations

The individual 'components' are described in the following sections, while the structure of each table is presented in Appendix H.

By convention, in the text, table names appear in **bold** and column names in *italic*. Further, for legibiliy, the names of schema's are given in ***italics, bold***.

Figure 3.1: Main reference groups and components of the WoSIS database model, with primary and foreign keys
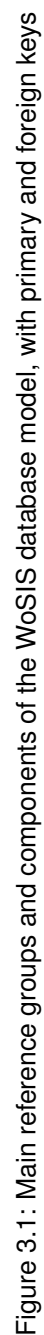
3.2 Main components

3.2.1 Metadata

Metadata are data that succintly define and describe other data. They define the terms and conditions for use of the data and ensure that all data can be properly attributed and cited. At ISRIC, all metadata are handled in GeoNetwork, a catalogue application for spatially referenced resources[1]. GeoNetwork 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 ISC World Data System [2] and GEOSS (Global Earth Observation System of Systems) portal[3].

Tables for storing the metadata are managed in a dedicated database schema that has its own data model. As shown in Figure 3.2, the *dataset_id* in table **dataset** provides the link between the ***wosis*** schema and ***metadata*** schema. Table **version** holds information about the licence conditions and provides a link to the 'data sharing agreement' for each dataset. The metadata themselves are published in the ISRIC Data Hub[4], subject to the permission of the data provider.
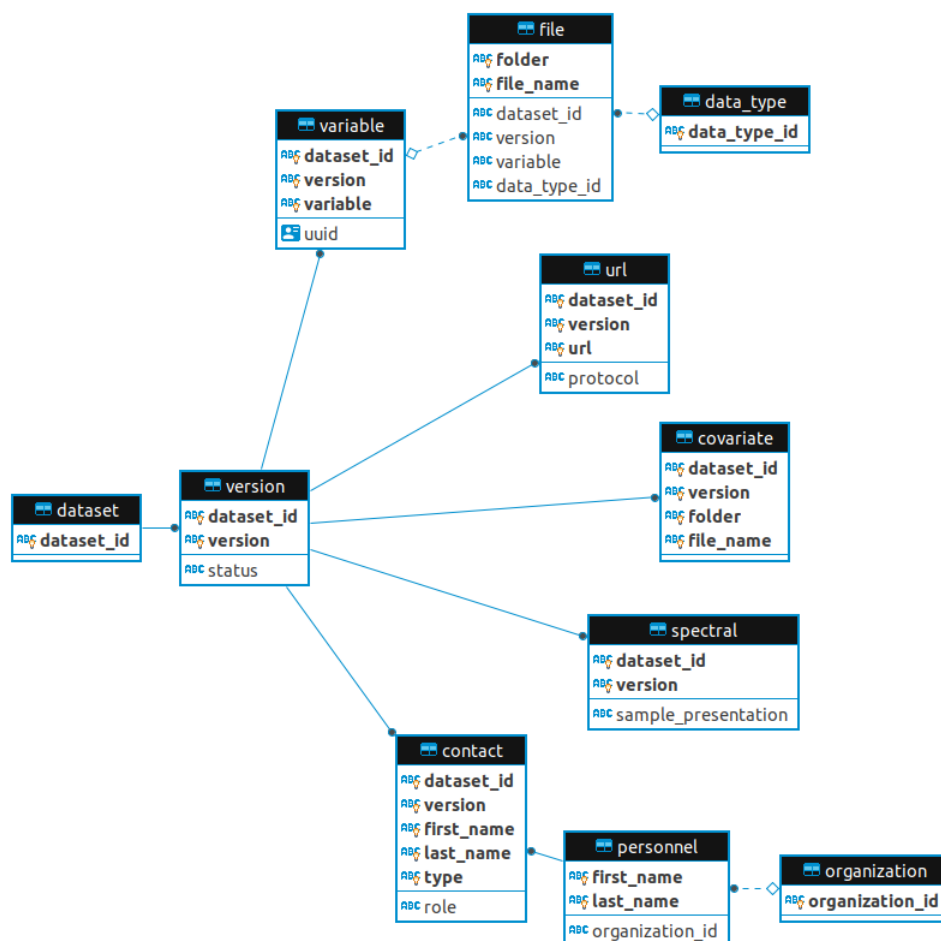
Figure 3.2: Metadata data model, showing tables, primary and foreign keys.

The **dataset** table, at the top of the hierarchy of the ***metadata*** schema, defines where the data come from. There may be several versions for a dataset such as AfSPv1.0 and AfSPv2.0. The table **version**

---

[1] http://geonetwork-opensource.org/
[2] WDC-Soils, see https://www.icsu-wds.org/services/data-portal
[3] GEOSS, see http://www.geoportal.org/
[4] http://data.isric.org/

serves to store information for each version of a dataset, such as the licence (column **legal_constraints**), the collection period of the data or the status (shows whether a project is ongoing or has been completed).

Table **variable** is used when a dataset has multiple variables that require a single metadata record. Similar to SoilGrids, for each soil property considered in WoSIS there is a unique xml document (e.g. an xml file for soil pH measured in water). This variable may consist of several files having different formats (e.g. CSV's and GeoTIFF); these files are documented in table **file** which serves to store all information related to files considered in the system. This information is the output of the gdalinfo of a GIS layer. This way, in a single table, an overview of all spatial and non-spatial data for which metadata are stored in the system are visible. Table **data_type** stores the minimum and maximum values that compromise the different data types in GeoTiff, one of the most common and flexible GIS raster formats.

The **url** table serves to store links (i.e. references) to reports, online documents or web-services, and serves to document the lineage of the data.

Table **covariate** serves to indicate which datasets were used to create derived products such as SoilGrids soil property layers (see de Sousa et al. (2020); de Sousa et al. (2020)).

Table **spectral** serves to store metadata describing how spectral data was collected in the laboratory. So far, two types of spectrally-derived data can be accomodated in WoSIS: Near-infrared (NIR, table **spectral_nir**) and Mid-infrared (MIR, table **spectral_mir**).

The tables **organization**, **personnel** and **contact** provide information about the provenance and authorship of the data.

Information about the data providers themselves is stored in three tables. Table **organization** serves to describe the organizations while table **personnel** lists the staff/persons instrumental in collating or providing the data. Table **contact** is an intermediate table to record what role a given person had in compiling a given dataset.

Using this approach, there is one single entry point to authoritative names and contact information in the database. This ensures that a single spelling is used consistently for the same organization or individual throughout the database (e.g., KIT, Tropen Instituut, Royal Tropical Institute, Koninklijk Instituut voor de Tropen).

The above tables can accommodate all the information required for metadata in compliance with ISO-19115 and INSPIRE standards. This information is then extracted by a Python script to produce an xml document and publish it to `http://data.isric.org/`, subject to the permission of the data provider.

Personal information in WoSIS is handled conform GDPR (General Data Protection Regulation)[5], following principles of 'data minimization'. This information serves to acknowledge our data providers.

3.2.2 Datasets and classification

As indicated, the **dataset** table stays at the top of the hierarchy of the ***wosis*** schema defining where the data come from and pointing to the ***metadata*** schema.

Table **dataset** also serves to make a bridge, as mentioned before, to the ***metadata*** schema, where the metadata is stored.

Sometimes, the same soil profile has been considered/processed in different source databases. As indicated, this regularly happens with compilations such as WISE, SOTER, AfSP and ISCN. This can be indicated in table **dataset_profile**, which allows to assign a given profile (in the **profile** table) to different source datasets (in the **dataset** table).

Possible sources of information for the data managed in WoSIS 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. Table **reference_dataset_profile** allows these sources to be specified per profile and dataset, thus acknowledging the specific origin of a certain profile in a compilation dataset. It also serves to describe the full reference of the source materials; in so far as possible, a link is provided to the actual, scanned document in the ISRIC World Soil Library using the unique *isn* code.

[5]See `https://eur-lex.europa.eu/eli/reg/2016/679/oj`
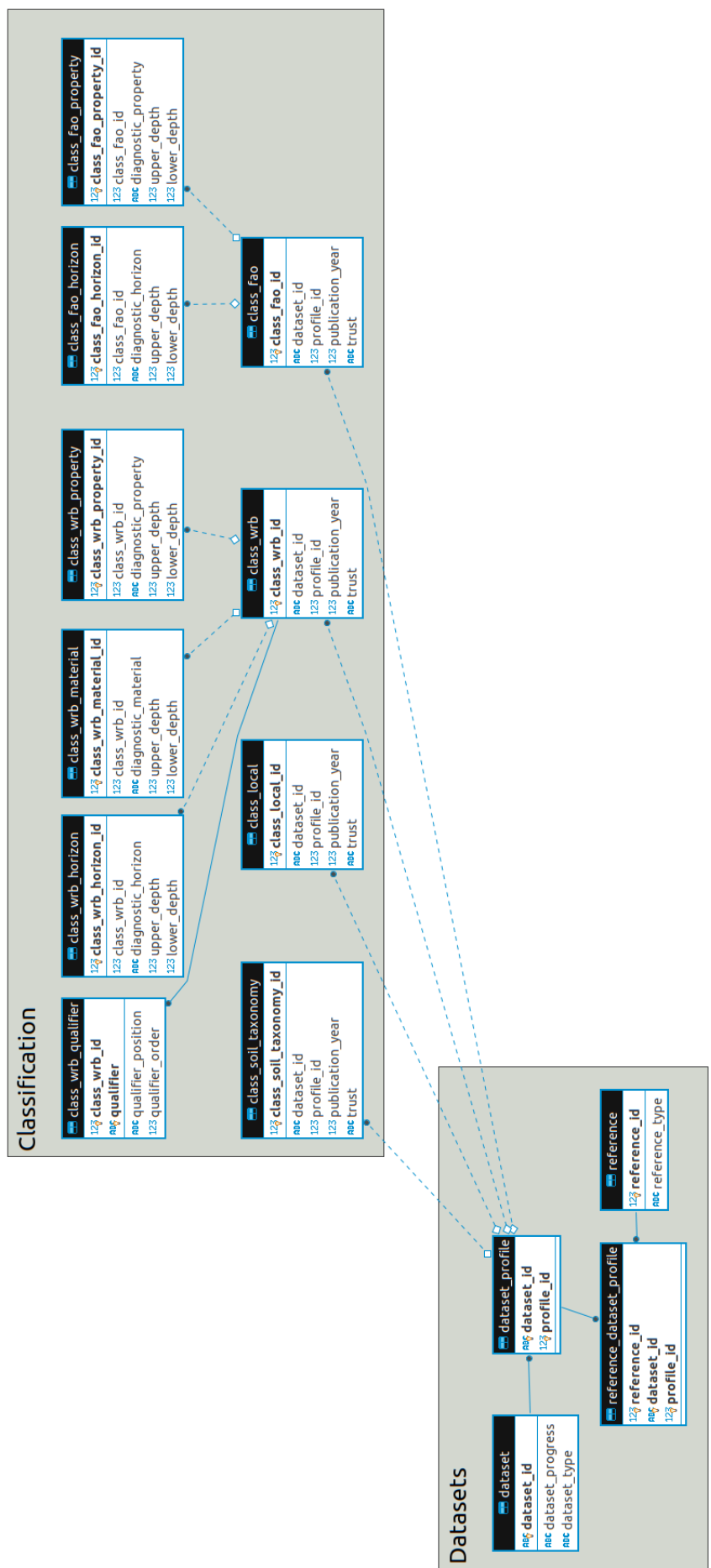
Figure 3.3: Tables describing 'datasets' and 'classification' with primary and foreign keys.

Historically, soils worldwide have been classified according to a wide range of evolving national systems (Krasilnikov et al., 2019; FAO, 2015). International correlation between these diverse systems is currently being addressed by the World Reference Base for Soil Resources (IUSS WG-WRB, 2015), and earlier through the FAO-Unesco Soil Map of the World (FAO-Unesco, 1974; FAO, 1988) system. Another, widely used 'international' soil classification system is USDA Soil Taxonomy (Soil Survey Division Staff, 2014b), currently in its 12th edition.

The 'classification' tables (Figure 3.3) can accommodate three widely used, international 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). FAO classifications are described in tables **class_fao**. Tables **class_fao_horizon** and **class_fao_property** define the domains for the horizons respectively the diagnostic soil properties.

- World Reference Base for Soil Resources (WRB): this international classification system is endorsed by the International Union of Soil Sciences (IUSS WG-WRB, 2015). The version used is shown by the year of publication. WRB-related information is specified four tables: **class_wrb**, **class_wrb_horizon**, **class_wrb_material**, **class_wrb_property** and **class_wrb_qualifier**).

- USDA Soil Taxonomy (Soil Survey Division Staff, 2014b), and earlier approximations as indicated by the year of publication (see table **class_soil_taxonomy**).

Further, the national or local soil classification can be stored *verbatim* in table **class_local**. In the future, once fully developed, a table for the Universal Soil Classification (Michéli et al., 2016) can be created in WoSIS.

Sometimes, as indicated, the same soil profile may have been considered/processed in different data compilations. As a result, the same profile may have been classified/correlated differently in each compilation, using the same classification system (and version), depending on the classifier's perspective (Kauffman, 1987). Therefore, for data compilations, WoSIS contains a link table (**dataset_profile**) to a given profile (**profile** table) to a specific source datasets (**dataset** table). All classifications thus refer to an entry in the **dataset_profile** link table (that is, a profile in a particular dataset), thus allowing for one unique classification per profile and dataset.

In some cases, the USDA Soil Taxonomy coding is inconsistent between editions as the same codes were used for different soil types in successive versions (Soil Survey Division Staff, 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 Soil Map of the World each use a consistent 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 version (i.e. year of publication).

In WoSIS, soil classifications 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 (2006a) soil horizon nomenclature, is considered the responsibility 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; Gerasimova et al., 2013).

3.2.3 Attributes, methods and laboratories

Each dataset (as described in **dataset** table) comes with a list of attributes (or parameters or properties) to express a description or measurement. These source attributes are described in 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, measurement unit, and domain.

For each attribute (**desc_attribute**), the definition, analytical methods and source laboratory must be defined explicitly 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 (Soil Survey Division Staff, 2011; van Reeuwijk, 2002). However, 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 (i.e. analytical method being standardized) is described in table **desc_method_standard** using a defined number of options, as described in table **desc_method_option**. Table **desc_attribute_agg** defines what methods options are available for each standard attribute. Further, details about the laboratory where the measurements were made are stored in table **desc_laboratory_source**. A standardized list of laboratory names is presented in table **desc_laboratory_standard**.

Importantly, table **descriptor** serves to combine the attribute (source and standard), analytical method and laboratory id's into a new, unique id (*descriptor_id*). The *descriptor_id* is later used in tables such as **profile_attribute** and **profile_layer_attribute** in which the measured values respectively descriptions are stored.

Recognizing the broad scope of the domain of knowledge that can be accommodated in WoSIS, 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 domain in **desc_domain_value**. Conversely, a data domain may be used to describe more than one characteristic. For example, in the Guidelines for Soil Description (FAO, 2006a), several surface characteristics are defined using the same surface coverage classes, hence the same data domain. This is illustrated in Figure 3.4.
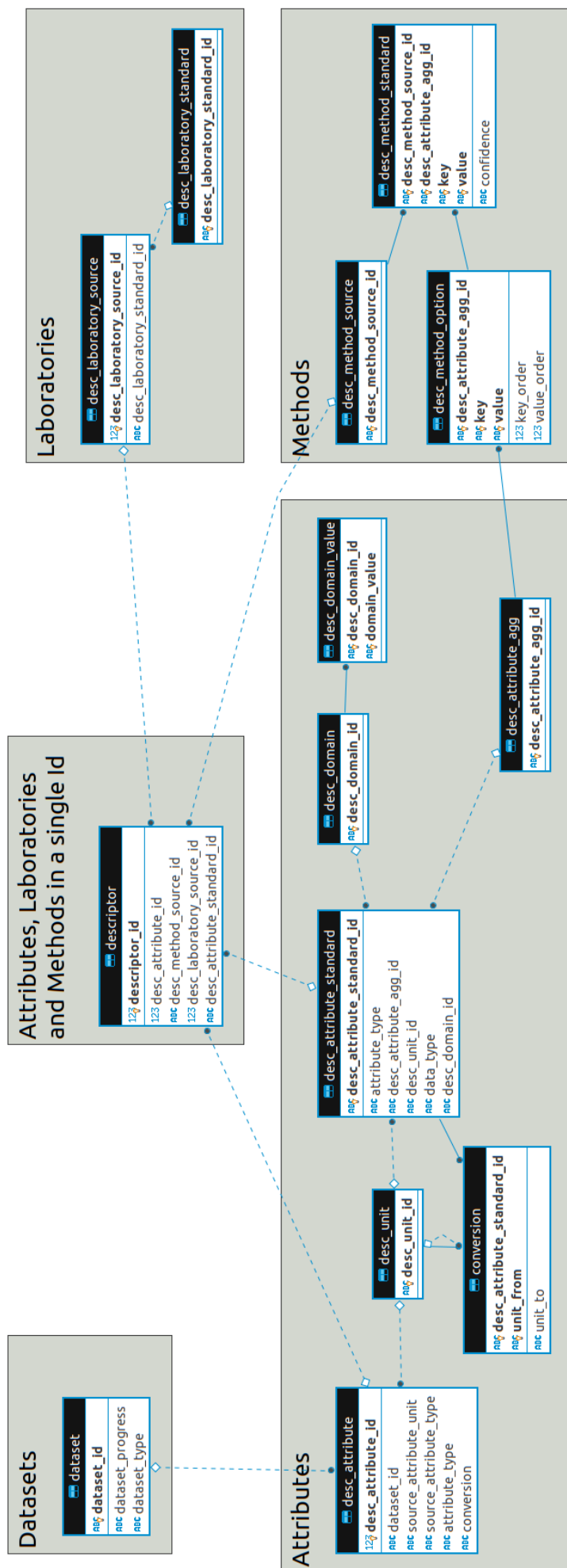
Figure 3.4: Attribute reference group tables showing primary and foreign keys.

3.2.4 Profiles, layers and observations

Tables in this group (Figure 3.5) 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** documents unique soil profiles along with their geometry (x,y). The coordinates are stored in column *geom*, in binary mode, using the PostGIS[6] 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 of origin is registered using the 2 character ISO code (e.g. BE for Belgium) in column *country_id*, that links to the **country** table which defines unique codes for country names, dependent territories and special areas of geographical interest based on ISO-3166 derived from the Global Administrative Units Layer (GAUL)[7], 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. The geometry is stored in table **country_geom** because one (row) country can be composed by multiple (rows) polygons in this table.

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 the metadata records.

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 3.5 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*.

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

- Profile (point 2D): **profile_attribute**
- Layer (point 3D): **profile_layer_attribute**
- Layer spectral (point 3D): **profile_layer_spectral**

Table **profile_attribute** serves to manage the properties describing 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 (or horizon), measured from the surface, including organic layers (O)[9] and mineral covers, downwards in accord with current conventions (FAO, 2006b; Soil Survey Division Staff, 2012). This table links to **profile_layer_sample** to accommodate (possible) multiple samples in the same horizon.

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

---

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

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

[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 (Soil Survey Division Staff, 2012) p. 2-6).
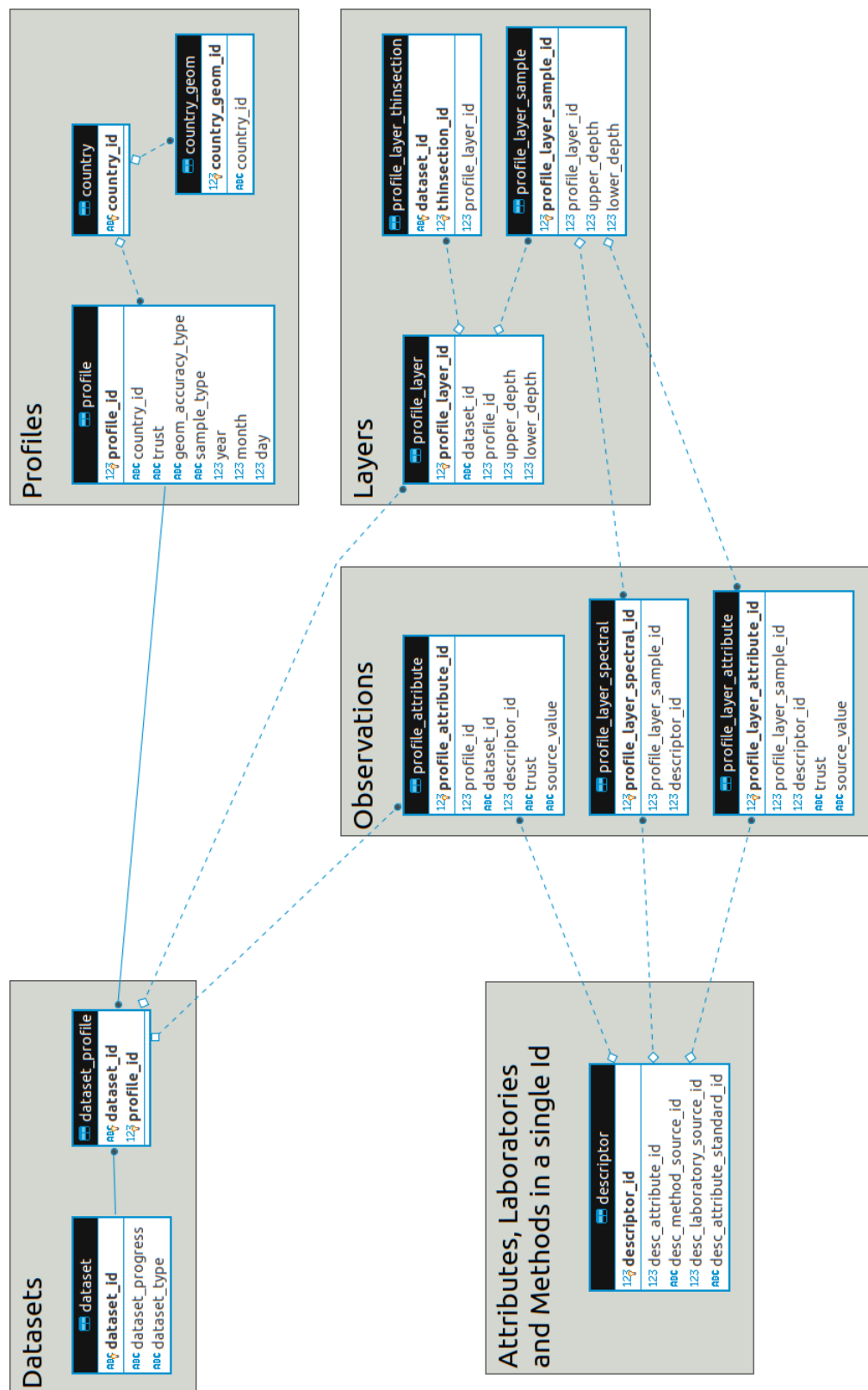
Figure 3.5: Profile, layer and observation-related tables showing primary and foreign keys.

Chapter 4

Interoperability and web services

In this Chapter, 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. Extensible Markup Language (XML) defines a set of rules for encoding documents in a format that is both human-readable and machine-readable. It is widely used in web services as a means (i.e. language) for transferring structured 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 organizations and with clients. The web services, which form an important part of ISRIC's 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 2.3 and 2.4); 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 earlier.

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 compatible with web communication. A Web Feature Service (WFS) is implemented using MapServer that connects to the WoSIS PostgreSQL database, 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 requests. 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, however, there is no common standard for this (Mendes de Jesus et al., 2017; Wilson, 2016; Ritchie, 2016). Developing such a standard is an important objective of GSP Pillar 5 (Baritz et al., 2014) and the IUSS Working Group on Soil Information Standards.

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, WFS-T, transaction).

---

[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 MapServer, which connects to WoSIS reading its views and tables. Further, a Representational State Transfer (REST) service is available that permits downloading/streaming for the web-service, querying based on coordinates (longitude (X) and latitude (Y)) as provided by the client. An ongoing development at ISRIC is to allow a client's web-service to become a data provider to WoSIS, for example in the context of anticipated crowd-sourced projects (Hobley et al., 2017).
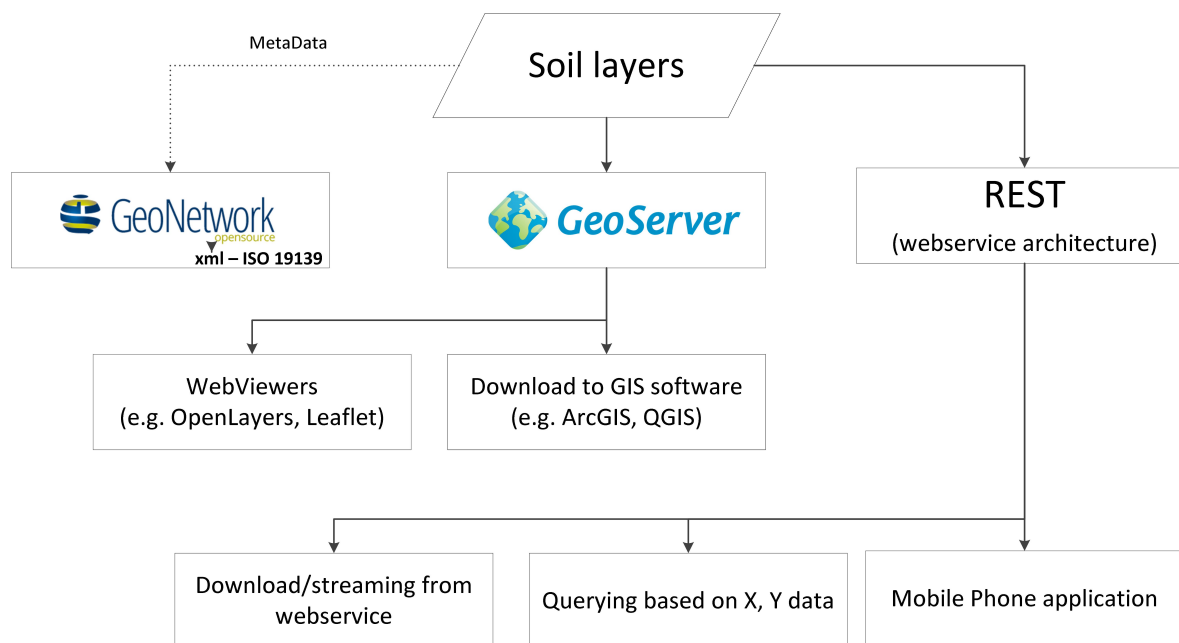
Figure 4.1: Serving soil layers from WoSIS to the user community.

Chapter 5

Towards a federated soil database

Worldwide there are many organizations with valuable soil data (Omuto et al., 2012; Arrouays et al., 2017). Yet, these data are accommodated in different databases using a range of data models and conventions. Alternatively, the data 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.
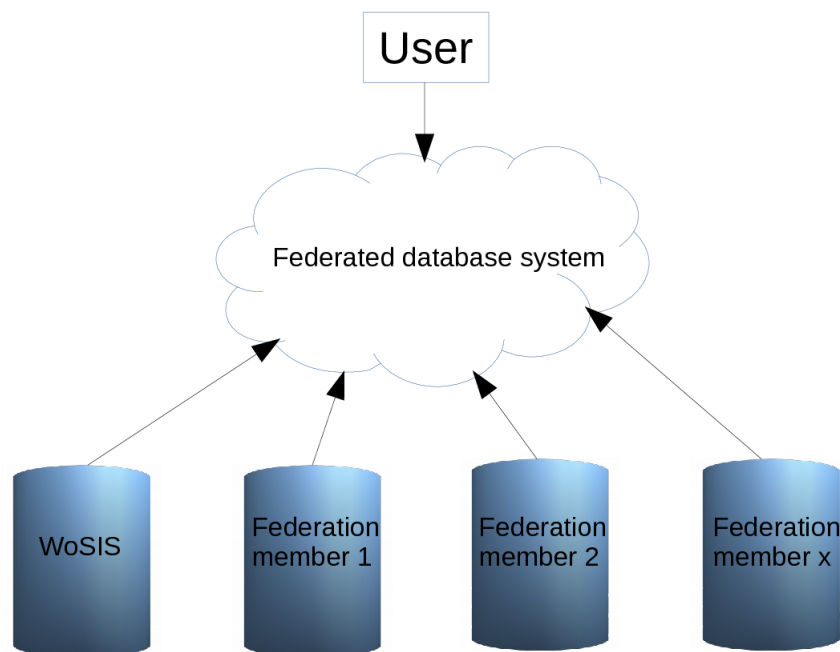
Figure 5.1: Example of a federated database system.

A federated database, also called a virtual database, is a way to view and query several databases on-line 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 themselves; 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 sub-queries 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's that enable PostgreSQL Server to

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

At ISRIC, we 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 database, using a bottom-up-approach, is an important goal of the Global Soil Partnership (GSP, 2016). ISRIC has been coordinating the design of the GSP's Global Soil Information System (GLOSIS) since 2016. GLOSIS is being designed as a decentralized system that interacts with multiple national sources. This is unlike WoSIS, which like its predecessors such as SOTER and WISE, was set up as a centralized database (Tempel et al., 2013) within the broader framework of the ICS World Data System.

Various components of ISRIC's own SDI are being used to develop modules for GLOSIS (GSP, 2018; de Sousa et al., 2019, 2020). Once implemented at a satisfactory level of detail and authority, the newly 'shared' data accessible online through GLOSIS may provide new materials for consideration in WoSIS itself. Alternatively, WoSIS may be envisaged as a support node within the federated GLOSIS framework (Figure 5.2).
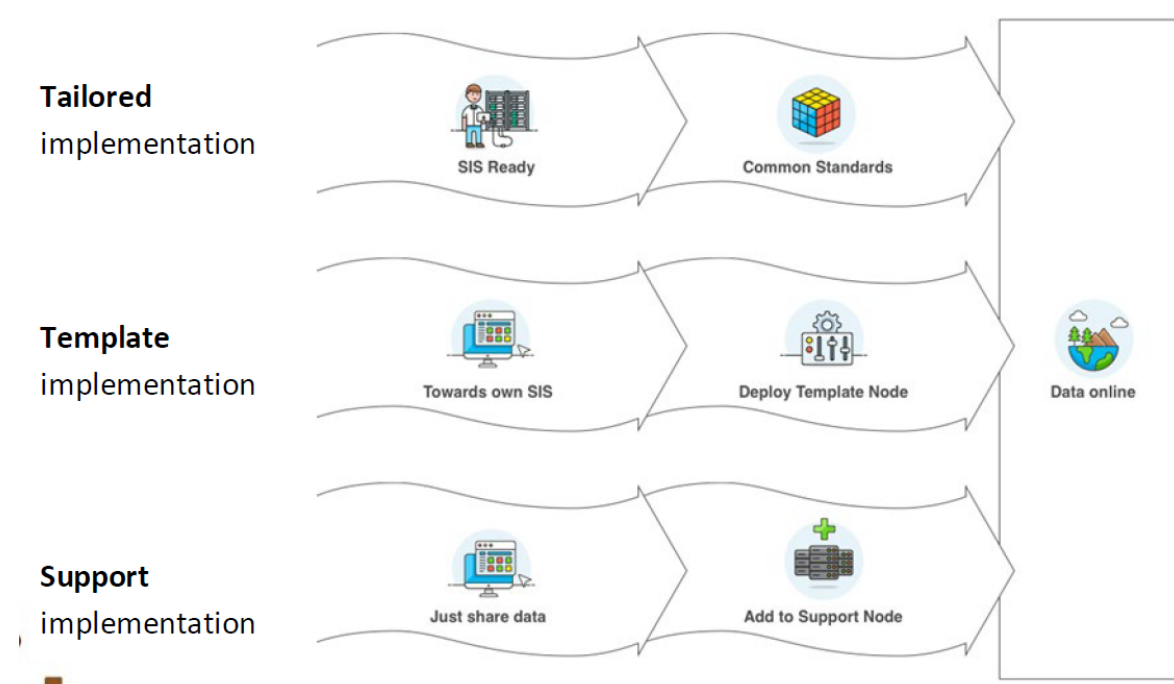
Figure 5.2: Different participation levels foreseen for GLOSIS (Yigini and van Egmond, 2000)
.

Three different levels of participation to the GLOSIS federation are proposed (de Sousa et al., 2019):

- Tailored implementation: data providers with an established soil spatial data infrastructure apply/ implement the GLOSIS data exchange model, i.e. they bring their data to the common GLOSIS standard and then publish their data through the GLOSIS web portal.

- Template implementation: data providers with soil data stored in simple databases or plain tables and with the ambition to establish a (national) soil information system will be trained in setting up a reference node. Once a reference node is deployed, the data will be served through a national web portal and become discoverable through the GLOSIS data hub (if allowed by the data provider).

- Support implementation: data providers with their soil data stored in simple databases or plain tables standardize their data according to the GLOSIS data model and send their data to the GSP who will act as a custodian of the data by storing the data in the support node. Again, the data will be discoverable through the GLOSIS data hub only when allowed by the data provider(s).

Chapter 6

Reflections and future developments

In view of its global scope, WoSIS will always remain 'work in progress' as new source datasets become available for standardization and web technologies for uploading / exchanging/ serving data evolve.

At the time of writing, 'WoSIS Latest' comprised standardized analytical data for over 200,000 globally distributed soil profiles. In addition to this, another 40,000 profiles with more restrictive licences were standardized, but, as indicated, these can only be used by ISRIC for generating SoilGrids layers and other visualizations.

Inherently, there are various gaps (e.g. geographic, taxonomic, soil properties) in WoSIS as not all soil properties were measured routinely in the underpinning source databases. Further, the number of observations generally decreases with depth.

As indicated, the source data are of varying quality and completeness. To facilitate the user or modelling community, we provide three broad measures of uncertainty: date of sampling, geometric accuracy, and a first approximation of measurement uncertainty.

The coming 3-5 years, the following activities will be considered (as realistic within the allocated project time):

- Ingest and process 'new' soil profile datasets into WoSIS, in principle in order of receipt of the various datasets, with priority for 'fully shared' datasets from so far under-represented regions.

- Expand the number of soil properties for which standardized soil analytical method descriptions are developed, gradually working towards the list of soil physical and chemical properties considered in WISE and SOTER.

- Within the framework of collaborative projects, consider how geo-referenced soil biology data may be added to, or accessed from, WoSIS.

- Elaborate procedures for handling data derived from proximal sensing as well as methodologies for standardizing the spectrally-derived soil data themselves within the framework of 'the Global Soil Spectral Calibration Library and Estimation Service' (GLOSOLAN).

- Improve workflows for ingesting, standardising and serving soil data as possible global 'support node' within the broader framework of a federated Global Soil Information System (GLOSIS).

- Refine procedures for quantifying uncertainty in the analytical methods through probability distributions (based on outcome of ongoing PhD research (2019-2024) at ISRIC).

- Reconsider procedures for distributing WOSIS data (in 2019, OGC started moving from OGC WFS 2 with XML to OGC Geo API with ReST functionality), and for 'extraction, transformation and load' (ETL).

- Publish static snapshots on a 2-3 yearly basis after standarization of say 75,000 - 100,000 new soil profiles, for consistent citation purposes. This, besides serving 'WoSIS Latest' with the most recent additions.

- Strive to embed WoSIS as the global 'support node' within the GSP's evolving Global Soil Information System. The ultimate goal is to arrive at a federated system of inter-operable databases where national data providers maintain and update their own data for the greater benefit of the international community.

Acknowledgements

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., Gonçalves, 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. (2008). ISRIC-WISE harmonized global soil profile dataset (version 3.1). Report 2008/02, ISRIC - World Soil Information. http://dx.doi.org/10.13140/2.1.4306.7683.

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., Ribeiro, E., and van Oostrum, A. (2020). Standardised soil profile data to support global mapping and modelling (wosis snapshot 2019). *Earth System Science Data*, 12:299–320. https://doi.org/10.5194/essd-12-299-2020.

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.

CPCS (1967). Classification des sols (edition 1967). Report, Laboratoire de Géo1ogie-Pédologie de l' ENSA (Grignon).

Cressie, N. and Kornak, J. (2003). Spatial statistics in the presence of location error with an application to remote sensing of the environment. *Statist. Sci.*, 18(4):436–456. `https://projecteuclid.org/download/pdf_1/euclid.ss/1081443228`.

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

de Sousa, L., Poggio, L., Batjes, N., Heuvelink, G., Kempen, B., Ribeiro, L., and Rossiter, D. (2020). Soilgrids 2.0: producing quality-assessed soil information for the globe. *SOIL Discussion*. `https://doi.org/10.5194/soil-2020-65`.

de Sousa, L. M., Kempen, B., Mendes de Jesus, J., Yigini, Y., Viatkin, K., Medyckyj-Scott, D., Richie, A., Wilson, P., van Egmond, F., and Baritz, R. (2019). Conceptual desing of the global soil information system infrastructure. Report, ISRIC, FAO, Manaaki Whenua (Landcare Research), CSIRO, Wageningen UR, European Environment Agency.

de Sousa, L. M., Poggio, L., Dawes, G., Kempen, B., and van den Bosch, R. (2020). Computational infrastructure of soilgrids 2.0. In Athanasiadis, I. N., Frysinger, S. P., Schimak, G., and Knibbe, W. J., editors, *Environmental Software Systems. Data Science in Action*, pages 24–31. Springer International Publishing.

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. (2001). Selection of an appropriate phosphorus test for soils. Technical report, Soil Survey Laboratory, USDA, Lincoln (NE). `https://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.

Gerasimova, M., Lebedeva, I., and Khitrov, N. (2013). Soil horizon designation: State of the art, problems, and proposals. *Eurasian Soil Science*, 46(5):599–609. `https://dx.doi.org/10.1134/S1064229313050037`.

GlobalSoiMap (2015). Specifications tiered globalsoilmap products (release 2.4), 52 p. Report, GlobalSoilMap comsortium.

GLOSOLAN (2018). The global soil laboratory network (glosolan) – towards its establishment and implementation, global soil partnership. `http://www.fao.org/global-soil-partnership/pillars-action/5-harmonization/glosolan/en/`.

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.

Grimm, R. and Behrens, T. (2010). Uncertainty analysis of sample locations within digital soil mapping approaches. *Geoderma*, 155(3-4):154–163. `https://doi.org/10.1016/j.geoderma.2009.05.006`.

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`.

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`.

GSP (2018). Towards the implementation of glosis through a country soil information systems (countrysis) framework (concept note, draft). prepared by gsp pillar 4 working group, gsp-soil data facility and the gsp secretariat. `http://www.fao.org/global-soil-partnership/pillars-action/4-information-data/glosis/en/`.

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`.

Guillera-Arroita, G., Lahoz-Monfort, J. J., Elith, J., Gordon, A., Kujala, H., Lentini, P. E., McCarthy, M. A., Tingley, R., and Wintle, B. A. (2015). Is my species distribution model fit for purpose? matching data and models to applications. *Global Ecology and Biogeography*, 24(3):276–292. `https://doi.org/10.1111/geb.12268`.

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`.

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 für 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`.

Hu, B., Bourennane, H., Arrouays, D., Denoroy, P., Lemercier, B., and Saby, N. P. A. (2021). Developing pedotransfer functions to harmonize extractable soil phosphorus content measured with different methods: A case study across the mainland of france. *Geoderma*, 381:114645. `http://www.sciencedirect.com/science/article/pii/S0016706119324437`.

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-SIS (2015). IUSS working group on soil information standards. Technical report, International Union of Soil Sciences (IUSS). `http://www.soilinformationstandards.org/`.

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.

Krasilnikov, P., Ibánez Marti, J., Arnold, R., and Shoba, S. (2019). *Soil Terminology, Correlation and Classification*. Earthscan, London (UK).

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 guide 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.

Minasny, B., McBratney, A. B., Wadoux, A. M. J. C., Akoeb, E. N., and Sabrina, T. (2020). Precocious 19th century soil carbon science. *Geoderma Regional*, 22:e00306. http://www.sciencedirect.com/science/article/pii/S2352009420300559.

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/.

Nave, L., Johnson, K., van Ingen, C., Agarwal, D., Humphrey, M., and Beekwilder, N. (2017). Iscn database v3-1. Report, International Soil Carbon Network (ISCN).

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.

Orgiazzi, A., Ballabio, C., Panagos, P., Jones, A., and Fernández-Ugalde, O. (2018). Lucas soil, the largest expandable soil dataset for europe: a review. *European Journal of Soil Science*, 69(1):140–153. http://dx.doi.org/10.1111/ejss.12499.

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.

Ribeiro, E., Batjes, N., and van Oostrum, A. (2018). World soil information service (wosis) - towards the standardization and harmonization of world soil data: Procedures manual 2018. Technical Report ISRIC Report 2018/01, ISRIC - World Soil Information, Wageningen. https://www.isric.org/sites/default/files/isric_report_2018_01.pdf.

Ritchie, A. (2016). Ogc soil data interoperability experiment. Report, Open Geosptial Consortium.

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.

Sadovski, A. and Ivanova, M. (2020). Transformation of soil texture schemes and determination of water-physical properties of soils. *Eurasian Journal of Soil Science*, 9:306–313. https://doi.org/10.18393/ejss.760201.

Schut, A. G. T. and Giller, K. E. (2020). Soil-based, field-specific fertilizer recommendations are a pipe-dream. *Geoderma*, 380:114680. http://www.sciencedirect.com/science/article/pii/S0016706120314750.

Soil Survey Division Staff (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.

Soil Survey Division Staff (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.

Soil Survey Division Staff (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).

Soil Survey Division Staff (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.

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

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

Soil Survey Division Staff (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.

Soil Survey Division Staff (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.

Soil Survey Division Staff (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.

Soil Survey Division Staff (2014a). *Kellogg Soil Survey Laboratory Methods Manual (Version 5.0. R. Burt and Soil Survey Staff (ed.))*, volume Soil Survey Investigations Report No. 42,. U.S. Department of Agriculture, Natural Resources Conservation Service, Lincoln (Nebraska).

Soil Survey Division Staff (2014b). *Keys to Soil Taxonomy (12th edition)*. United States Department of Agriculture, Natural Resources Conservation Service, Washington, DC. file:///C:/Users/batje001/Downloads/2014_Keys_to_Soil_Taxonomy.pdf.

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.

Suvannang, N., Hartmann, C., Yakimenko, O., Solokha, M., Bertsch, F., and Moody, P. (2018). Evaluation of the first global soil laboratory network (glosolan) online survey for assessing soil laboratory capacities. Report GLOSOLAN/18/Survey Report, Global Soil Partnership (GSP) / Food and Agriculture Organization of the United Nations (FAO). http://www.fao.org/3/CA2852EN/ca2852en.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`.

USDA-NCSS (2018). National cooperative soil survey (ncss) soil characterization database.

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. 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 detpm2002ermining 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.

Yigini, Y. and van Egmond, F. (2000). Glosis development. 6th meeting of the international network of soil information institutions (INSII). `http://www.fao.org/fileadmin/user_upload/GSP/docs/6th%20INSII%20-%20%20Meeting%20Agenda.pdf`.

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 load 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 Latest' in the URL field put **http://maps.isric.org/mapserv?map=/map/wosis_latest.map**, then press the OK button. Press 'Connect' and all available layers will appear on screen (Figure A.1). No authentication is needed, so leave blank the 'User name' and 'Password' boxes.
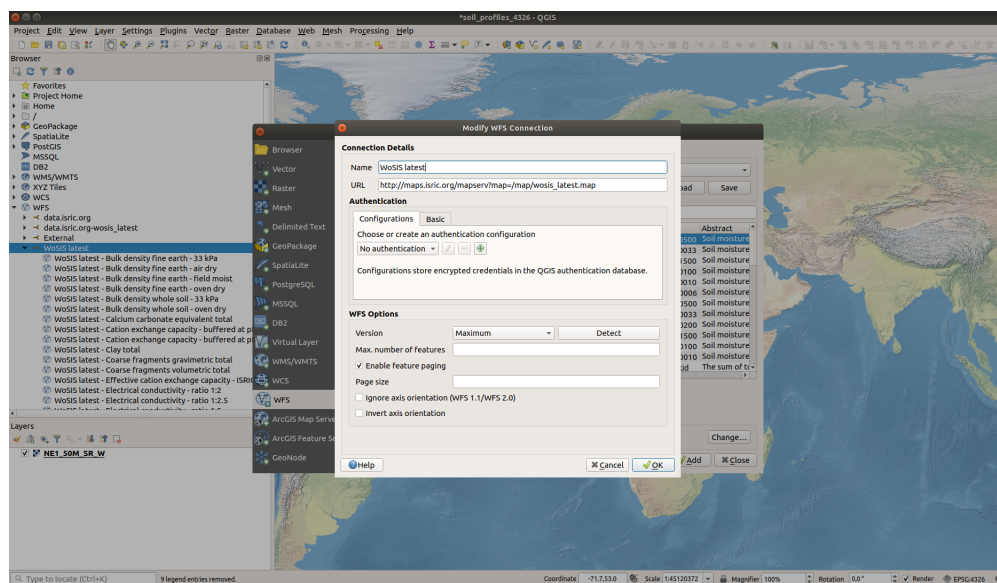
Figure A.1: Adding WoSIS WFS configuration in QGIS.

Next, select the layer of interest listed under 'Title' and press 'Add' button (Figure A.2). Note that it may take some time until the system fetches 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 and, for example, in the SQL statement box add the following

---

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

45

'WHERE' clause, like this: WHERE country_name='Brazil' (Figure A.3). The complete SQL statement should look like this: *SELECT * FROM wosis_latest_phaq WHERE country_name='Brazil'*.
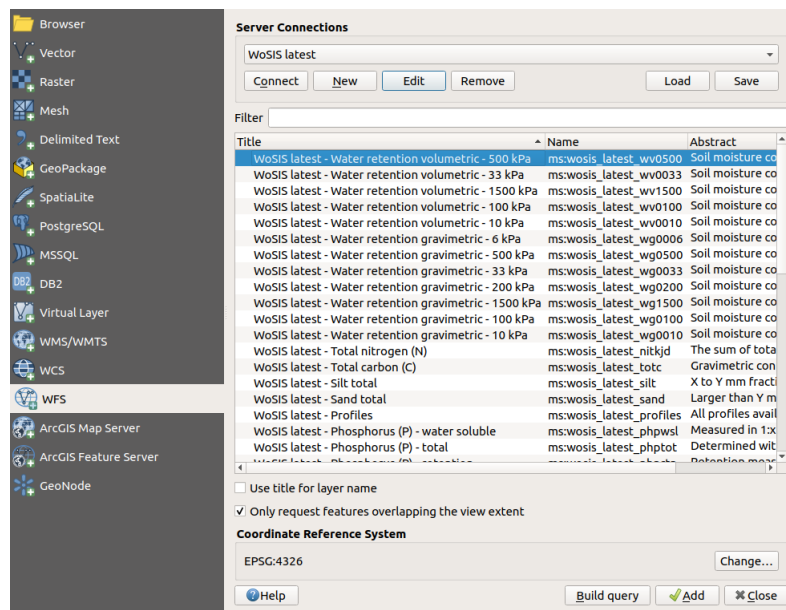
Figure A.2: Selecting WFS available layers.
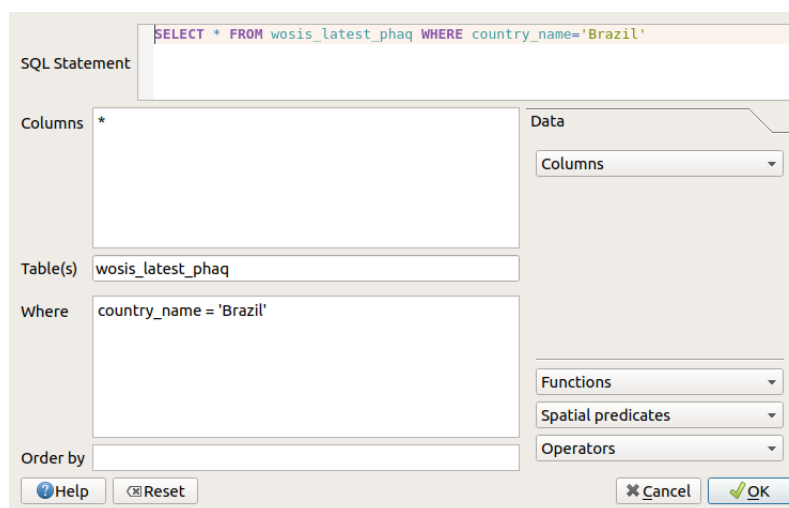
Figure A.3: WFS filter records.

After loading the layer you can save it as a shape file (or other format) on your local machine for further offline use. Check for updates on the WFS every month, as this is a dynamic dataset.

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

Detailed procedures for accessing WoSIS data from R, using R markdown, were prepared by our guest researcher David Rossiter. To run the procedures in your own environment (recommended), please download the following scripts and subsequently load these into your R Studio.

There are two Rmd scripts, one for 'WoSIS Latest' and the other for the last 'WoSIS snapshot':

`https://git.wur.nl/batje001/WoSIS/-/blob/master/R_scripts/`

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; Soil Survey Division Staff, 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 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 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). Those values, either numeric, categorical or descriptive, are expressed according to the associated domain as dictated by the references used for defining units of expression or pick lists. Typically, each profile is characterized by a number of consecutive layers or horizons, with defined upper and lower boundaries[2]. Alternatively, for soil fertility assessments only the upper 10 to 30 cm will be sampled.

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. However, the general principles of describing the data should be followed so that they can be readily understood by the WoSIS database managers.

Below, we present a suggested template for compiling and submitting datasets. It consists of one spreadsheet with seven different sheets:

- **read_me_first**
- **dataset**
- **profile**
- **layer**
- **column_definition**
- **standard_attribute**

---

[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.

- **standard_domain**.

The corresponding templates and a worked example may be downloaded from:

`https://git.wur.nl/batje001/WoSIS/-/blob/master/Templates/`.

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

Sheet **read_me_first** briefly explains the purpose and lay out of the of the template.

Sheet **dataset** serves to describe the main characteristics of your dataset. It has the following rows:

- *title*: Dataset title, project or thesis title.
- *version*: Dataset version.
- *publication_date*: Publication date (yyyy-mm-dd).
- *abstract*: Concise description of the dataset.
- *license*: Access and use constraints of the dataset; please provide an equivalent to Creative Commons licence (`https://creativecommons.org/choose/`).
- *organization_name*: Organization name.
- *organization_url*: Link to organization or project web page.
- *organization_country*: Organization country.
- *organization_city*: Organization city.
- *organization_postal_code*: Organization postal code.
- *organization_delivery_point*: Organization address.
- *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).
- *laboratory_name*: Laboratory name.
- *laboratory_country*: Laboratory country.
- *laboratory_city*: Laboratory city.
- *laboratory_postal_code*: Laboratory postal code.
- *laboratory_delivery_point*: Laboratory address.

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

Sheet **profile** serves to describe the main characteristics of each profile. It starts with the following columns:

- *profile_code*: 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.

- *classification_system_name*: The soil classification system used to classify the profile.

- *classification_system_year*: The publication year of the soil classification system used.

- *profile_classification_name*: Classification according to system and year previously defined.

- *site_attribute_1*: Add name for first attribute in given column (these are so-called site properties, e.g. drainage conditions).

- *site_attribute_2*: Add name for second attribute in given column.

- *site_attribute_x*: Add name for next attribute.

- (...)

In case a profile has been classified according to several systems, for example CPCS (1967), FAO (1988) and IUSS WG-SIS (2015), new columns for this can be added to sheet **profile**, for example as 'classification_system_name_2', 'classification_system_year_2' and 'profile_classification_name_2'.

Importantly, each row may only contain data for a given soil profile.

Sheet **layer** serves to describe properties for each (taxonomic) horizon respectively fixed-depth layer, for a given profile. It starts with the following columns:

- *profile_code*: Unique identifier of the profile. Provides the reference to data in sheet **profile**.

- *layer_name*: Horizon designation, for example A, B or C.

- *sample_code*: Laboratory sample code.

- *upper_depth*: Depth of upper layer or horizon (cm).

- *lower_depth*: Depth of lower layer or horizon (cm).

- *layer_attribute_1*: insert name for first attribute here (e.g., ph_cacl2).

- *layer_attribute_2*: Insert name for second attribute.

- *layer_attribute_x*: Insert name for next attributes.

- (...)

Any other layer description attributes can follow after these columns. For example, ph_kcl, organic_carbon, clay, silt, sand or bulk_density. Note that each column can only contain values measured using one single analytical method, expressed using one uniform unit of measurement (to be specified in sheet **column_definition**).

Importantly, each row in the **layer** sheet can only contain data for defined combinations of profile and layer (e.g., profile_code and layer_name).

Sheet **column_definition** serves to describe all the columns (attributes) that have been specified earlier in the **profile** and **layer** sheets. It starts with the following columns:

- *sheet_name*: Either 'profile' or 'layer'.

- *column_name*: The exact name of the column added after the default (mandatory) ones.

- *description*: Description of the attribute.

- *unit*: Units used (e.g. cm), if not used, 'unitless' should be indicated (e.g. for soil pH).

- *data_type*: Data type, use one out of ('Text', 'Integer', 'Real', 'Boolean', 'Date').

- *analytical_method*: Analytical method used in the laboratory. If if none is given enter 'Not applicable'.

- *desc_attribute_standard_id*: Name for standard attribute as defined for WoSIS, see sheet **standard_attribute**.

- *domain_name*: Domain name from the **standard_domain** sheet, when applicable (for categorical attributes only). *Note:* For regional data sets it may be necessary to create look up tables that permit to relate the 'region specific' domains to the standard domains adopted for WoSIS. For example, in case of a data set from China, the Chinese terms (characters) should be 'correlated' to their equivalent in English.

New rows can be added to sheet **column_definition**, as necessary for the given dataset

Table **standard_attribute** contains the standard attributes definitions as used in WoSIS. The exact names as listed in column *desc_attribute_standard_id* are needed to compile the look up tables mentioned earlier. Providing this crucial information, will make it possible for ISRIC to easily incorporate any newly submitted data into WoSIS. Otherwise, we will have to consult the data provider for additional information.

Table **standard_domain** contains the variables and categories (domains) as used in WoSIS; these are derived from the 4th edition of the 'FAO - Guidelines for Soil Description' (FAO, 2006a).

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 predefined 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; GLOSOLAN, 2018).

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 publishes a Soil Survey Laboratory Methods Manual (Soil Survey Division Staff, 2011), which is the reference source for the National Cooperative Soil Survey Soil Characterization Database (USDA-NCSS, 2018) and widely referred to internationally as reference. More recently, GLOSOLAN has been harmonizing standard operation procedures (SOPs) within the framework of the GSP.

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 Programme 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' (Soil Survey Division Staff, 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, 2001). Hence, 'vague' descriptions for available-P methods are essentially 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 GlobalSoiMap (2015) 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; Sadovski and Ivanova, 2020; Hu et al., 2021). The work of the Global Soil Laboratory Network (GLOSOLAN), established by the Global Soil Partnership [1] in 2017, deserves particular attention in this respect.

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 themselves, 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 for GLOSIS, within the framework 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 subsets that are analysed according to defined (and comparable) methods and may be considered as having equal quality or subsets that are considered suitable for a specific use. When new data are entered, the table is used for describing (coding) the added data. If necessary, values / options not yet considered 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. 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.

---

[1] http://http://www.fao.org/global-soil-partnership/pillars-action/5-harmonization/glosolan/en/

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' (Soil Survey Division Staff, 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	not specified
sample pretreatment	1	sieved over 1 mm sieve
sample pretreatment	2	sieved over 2 mm sieve
solution	0	not specified
solution	1	Water [H_2O]
solution	2	Calcium chloride [$CaCl_2$]
solution	3	Potassium chloride [KCl]
solution	4	Sodium fluoride [NaF]
solution	5	Ammonium chloride [NH4CL]
concentration	0	not specified
concentration	1	not applied
concentration	2	0.01 M
concentration	3	0.02 M
concentration	4	0.1 M
concentration	5	0.2 M
concentration	6	1 M
ratio	0	not specified
ratio	1	1:1
ratio	2	1:2
ratio	3	1:2.5
ratio	4	1:5
ratio	5	1:10
ratio	6	1:20
ratio	7	1:25
ratio	8	1:40
ratio	9	1:50
ratio	10	saturated paste
ratio	11	slurry
ratio base	0	not specified
ratio base	1	weight / volume
ratio base	2	volume / volume
instrument	0	not specified
instrument	1	electrode
instrument	2	electrode (field measured)
instrument	3	indicator paper (field measured)
monitoring	0	not specified
monitoring	1	not applied
monitoring	2	oxidizible sulfur compounds; initial pH
monitoring	3	oxidizible sulfur compounds; pH stabilized in $\geq$ 10 days, pH $\leq$ 0.1 unit, for two days
spectral	0	not specified
spectral	1	false
spectral	2	true

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 standardization process (Table D.1), starting with a short introductory description of main aspects of each analytical method (Section D.2 to D.14). 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 with standardized analytical method descriptions.

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

---

[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 3.2.4).

[2]Water holding capacity may be calculated as the amount of water held between 1/3 bar and 15 bar (USDA conventions) (Soil Survey Division Staff, 2014b). 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.14 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 populated with data coming 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 data base. Nonetheless, in some instances, the expertise of a soil laboratory specialist may be required for a complete characterization.

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$_4$OAc'.

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 situations it may be necessary to contact the data providers for additional information. Alternatively, for some datasets it may be unrealistic to aim for a full characterization of the analytical methods. 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 (Soil Survey Division Staff, 2014b). 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 and difficult to harmonize unless large data sets are available for the comparisons (Sadovski and Ivanova, 2020). 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.
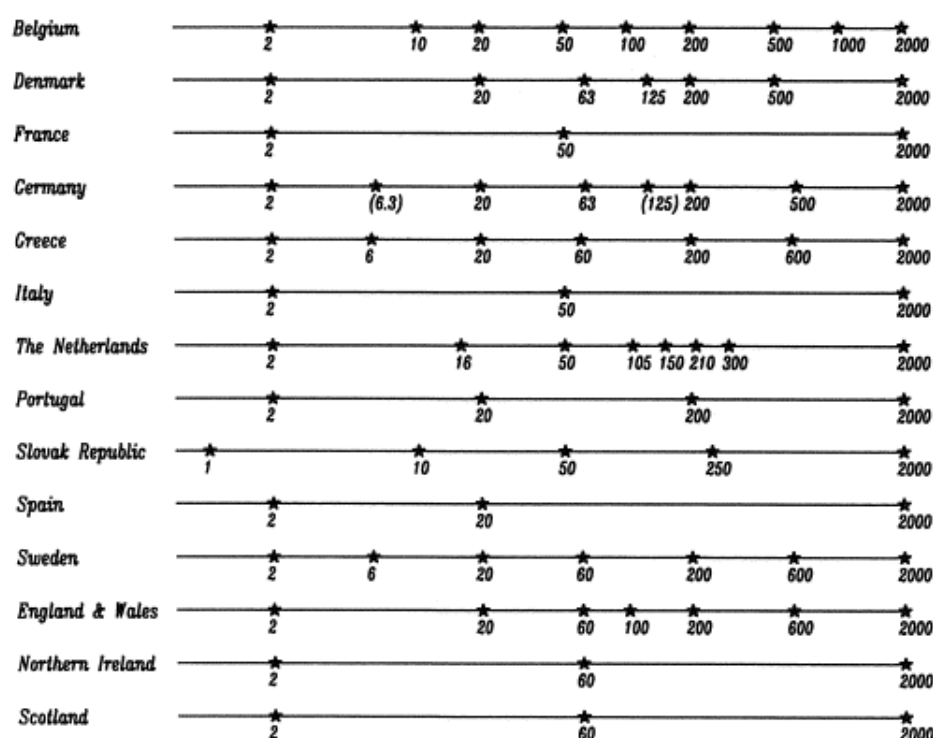
Figure D.1: Range in textural definitions as used in Europe.

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 five different reference methods for pH KCl: ISO 10390:2005 (ISO, 2005), USDA (Soil Survey Division Staff, 2014b), ISRIC (van Reeuwijk, 2002), WEPAL (2015) and Arrouays et al. (2014).

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

Key	ISO[5]	ISRIC[6]	USDA[7]	WEPAL[8]	GLOSOLAN
			Procedure		
Pretreatment	<2 mm	<2 mm	<2 mm	<2 mm	<2 mm
Solution	KCl	KCl	KCl	KCl	KCL
Concentration	1 M	1 M	1 M	1 M	1 M
Ratio	1 : 5	1 : 2.5	1 : 1	1 : 5	1 : 5
Ratio base	v/v	w/v	w/v	v/v	w/v
Instrument	Electrode	Electrode	Electrode	Electrode	Electrode
Monitoring	Not applied	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 (Soil Survey Division Staff, 2014b). 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 bulk density of a soil 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 (Soil Survey Division Staff, 2014b).

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 fragments is known or can be approximated, corrections can be included in the calculations of the bulk density (Soil Survey Division Staff, 2014b).

---

[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 (Soil Survey Division Staff, 2014b).

[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 four times a year, four 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.

Sometimes bulk density is expressed on the basis of the whole soil. Such measurement refer to all particle-size fractions, including boulders with maximum horizontal dimensions less than those of the pedon (Grossman and Reinsch, 2002; Soil Survey Division Staff, 2014b).

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 do 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 Soil Survey Division Staff (2014b).

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 (Soil Survey Division Staff, 2014b).

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 elsewhere (Baize (1993); Pansu and Gautheyrou (2006); Soil Survey Division Staff (2014b); 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; Soil Survey Division Staff, 2014b).

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; Soil Survey Division Staff, 2014b).

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 such materials 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 (i.e. volume (V) basis and 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 (Soil Survey Division Staff, 2014b).

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 (Soil Survey Division Staff, 2014b).

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$_2$O 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; Soil Survey Division Staff, 2014b).

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', or total 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).

Note: See also determination of 'soil organic matter' (Section D.8) and 'total carbon' (Section D.11).

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 also contains 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; Soil Survey Division Staff, 2014b; 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 preferably be removed by prior acid treatment.

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; Soil Survey Division Staff, 2014b) 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).

D.8 Organic matter

D.8.1 Background

Soil organic matter has been defined as the organic fraction of the soil exclusive of undecayed plant and animal residues. It has been used synonymously with "humus". For laboratory analyses, however, the soil organic matter generally includes only those organic materials that accompany soil particles in the fine earth fraction (Soil Survey Division Staff, 2014a). Organic matter content can be determined directly by ignition (400 °C) after which the loss in weight of the soil sample is taken as a measure of the organic matter content as well as *indirectly*.

D.8.2 Method

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.

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

Calculation: No conversion factor is applied for soil organic matter (i.e. LOI values are reported as is). According to Soil Survey Division Staff (2014a), the percent organic matter lost on ignition (400 °C) can be used in place of *indirect* organic matter estimates by the Walkley-Black organic C method.

Note: Generally, soil organic matter is assumed to contain 1.724 times the amount of soil organic carbon (TOC) (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; Minasny et al., 2020). The appropriate factor can be used to convert soil organic matter (SOM) content to content of organic carbon (TOC): TOC (g/kg) = 1/*factor* * SOM (g/kg).

D.9 Soil pH

D.9.1 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.9.2 Method

Methods for the determination of the pH of soil samples are characterized using seven features. The 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' (Soil Survey Division Staff, 2014b). Such conditions commonly occur in intra-tidal zones adjacent to oceans.

D.10 Sand, silt and clay fractions

D.10.1 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; Soil Survey Division Staff, 2014b); 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 (Soil Survey Division Staff, 2014b). At one time, the term rock fragments was differentiated from the term coarse fragments, which excluded stones and boulders with diameters >250 mm (Soil Survey Division Staff, 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 (Soil Survey Division Staff, 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.10.2 Method

Analytical methods for the determination of the sand, silt and clay (size) fraction in soil samples are described using three options for the sand and silt size fraction, and four 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 are excluded from the laboratory procedure, measurement will be for 'water dispersable' or 'natural clay' (see below).

Dispersion: During the dispersion of clay size particles, effects of the electric charges at the exchange surfaces have to be controlled carefully (Soil Survey Division Staff, 2014b). 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 (Soil Survey Division Staff, 2014b). 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.11   Total carbon

D.11.1   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 (TIC, 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.11.2   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; Soil Survey Division Staff, 2014b). 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.12   Total Nitrogen

D.12.1   Background

Nitrogen (N) in soils is mainly present as NO_3^- and NH_4^{+2} with negligible amounts as NO_2; together they form the mineral N-fraction. N in organic matter forms the organic fraction. N_2O and N_2 are present

as gases. Due to microbial activity and other soil processes, the size of the different fractions will vary with time and management conditions (including sample pre-treatment). The individual N-fractions present at the time of sampling are estimated by extractions. For the determination of Total Nitrogen, two methods are commonly used: combustion and digestion. A detailed account is provided in soil laboratory manuals (van Reeuwijk, 2002; Soil Survey Division Staff, 2014a).

D.12.2 Method

Analytical methods to determine Total Nitrogen are divided over four possible techniques and described using three more features. The corresponding flowchart and feature table are presented in Appendix E and F.

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

Technique: Traditional determinations for Total Nitrogen can be grouped into two categories: combustion of organic matter in a furnace (Dumas) and digestion with acid treatment and heating (Kjeldahl). NIR / MIR spectroscopic methods allow for non-destructive determinations of Nitrogen, yet these methods need to be calibrated against conventional wet-chemistry measurements.

- Digestion: Since the introduction of the Kjeldahl method for the determination of N, the method has been modified to accelerate the procedure and include all forms of nitrogen in soils. Salts like Potassium or Sodium-sulphate are added to sulphuric acid to raise the boiling point of the digestion mixture. A catalyst (i.e. $CuSO_4$) is used to accelerate the reactions. An addition of $KMnO_4$ or $HClO_4$ may be needed to complete the digestion.

To include NO_3 and NO_2, the Kjeldahl digestion is preceded by an oxidation step with i.e. H_2O_2. After excess H_2O_2 is removed, NO_3 is coupled with salicylic acid in the sulphuric acid medium. Some procedures use sodium thiosulfate to reduce the formed nitro-compounds into amino compounds. Fixed NH_4^+ is liberated with HF-treatment and combined into the Kjeldahl digestion.

- Combustion: In the Dumas combustion method, Nitrogen is oxidized with copper(I) oxide in a stream of CO_2 or with O_2, in a high temperature furnace at 900 - 1350 ^oC. The flushed gas stream is purified and after conversion of NO_x to N_2 total Nitrogen is estimated.

Detection: With sodium hydroxide, Nitrogen as NH3 is released from the Kjeldahl digest and distilled into a solution of boric acid and titrated. For quantification by semi-automatic systems, the digest is diluted and an aliquot used for the colorimetric determination of phosphorus by nitroprusside reaction at 660 nm.

Modern Element Analyzers use thermal conductivity for quantification. Other systems measure N_2 manometrically after absorption of CO_2.

D.13 Total Phosphorus

D.13.1 Background

D.13.2 Method

Analytical methods to determine total Phosphorus are divided over 33 possible techniques and described using three more features. The corresponding flowchart and feature table are presented in Appendix E and F.

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

Technique: To determine total phosphorus, sample material has to be decomposed with acids and oxidizing reagents, and destructed at high temperature. Fractions of this total phosphorus are determined by selective extractants. Spectroscopic methods like MIR/NIR may be used, but these non-destructive procedures need calibration against conventional laboratory methods to allow for interpretation of the P-measurements (Viscarra Rossel et al., 2016).

Extractants (for extractable P): Soil-P is probably one of the most cumbersome properties to measure and interprete (Soil Survey Division, 2014). The pH of a soil is a useful criterion for selecting the appropriate extractant for 'extractable-P' (Elrashidi, 2001). Such extractants differ in pH, buffer capacity, and ionic strength which makes them selective for certain P types and fractions. As a result, measures of extractable P for a given soil sample (soil type) as obtained using different methods, such as P-Olsen (for basic soils) and P-Bray (for acid soils), are not interchangeable.

As indicated, the 'chemistry' of the analytical methods is complex due to the different (types) of components in the extractant. Soil legacy data thus can best be grouped first using the source (name) of the extractant (i.e. P-Olsen or Mehlich III), and subsequently considering key characteristics of the extractant as given in the description.

Decomposition (for Total P): For total P methods, the sample has to be decomposed by treatments with destructive and oxidizing capacities. Several mixtures of acids can be used. In the same process organic matter is oxidized.

Detection: For quantification of Phosphorus, the colorimetric 'molybdate blue' and 'molybdate yellow' method are considered standard methods. Organic matter in the extract may give an interference by its colour and so has to be removed. Adsorption to charcoal and oxidation of dissolved P are frequently used. These treatments of the extract and the acids used towards the end of the procedure may influence the fractionation of phosphate in the extract. Ascorbic acid is often preferred over $SnCl_2$ as a reductant for colour development in the Molydate-blue method.

D.14 Water retention

D.14.1 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 (Soil Survey Division Staff, 2014b).

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 (Soil Survey Division Staff, 2014b). 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 (Soil Survey Division Staff, 2014a).

D.14.2 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)

$\qquad$ = 1020 cm H2O $\approx$ 1000 cm H2O

$\qquad$ = 75.01 cm Hg

$\qquad$ = 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 (Soil Survey Division Staff, 2014b).

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.
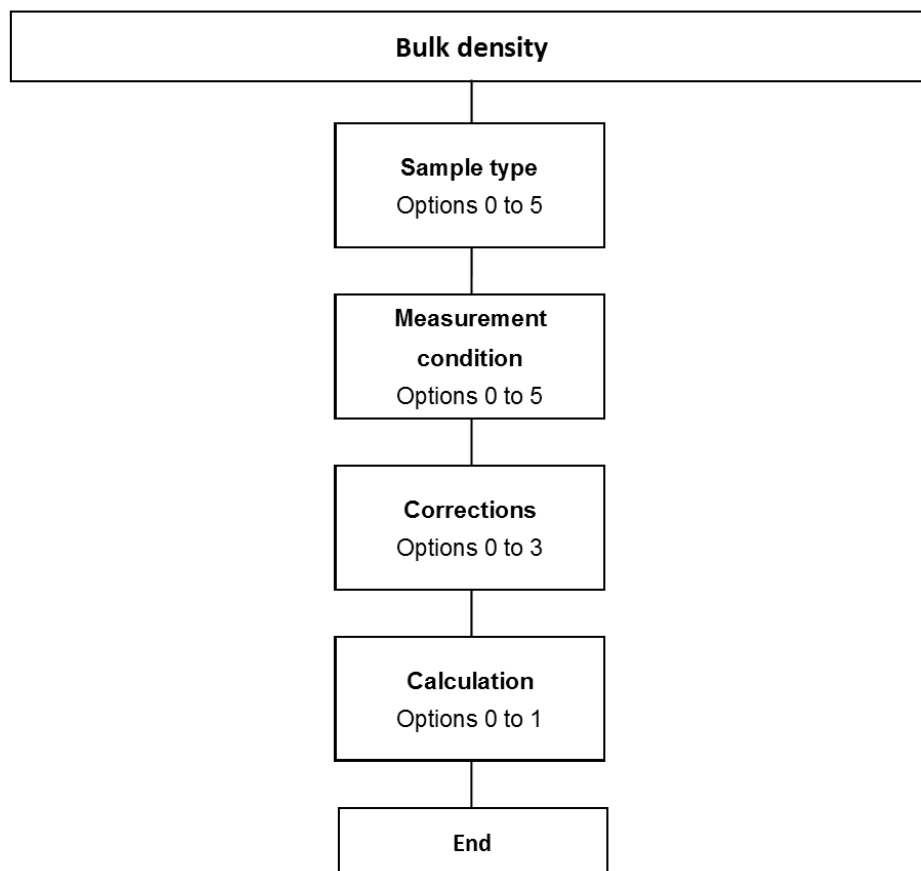
Figure E.1: Flowchart for standardizing bulk density methods.
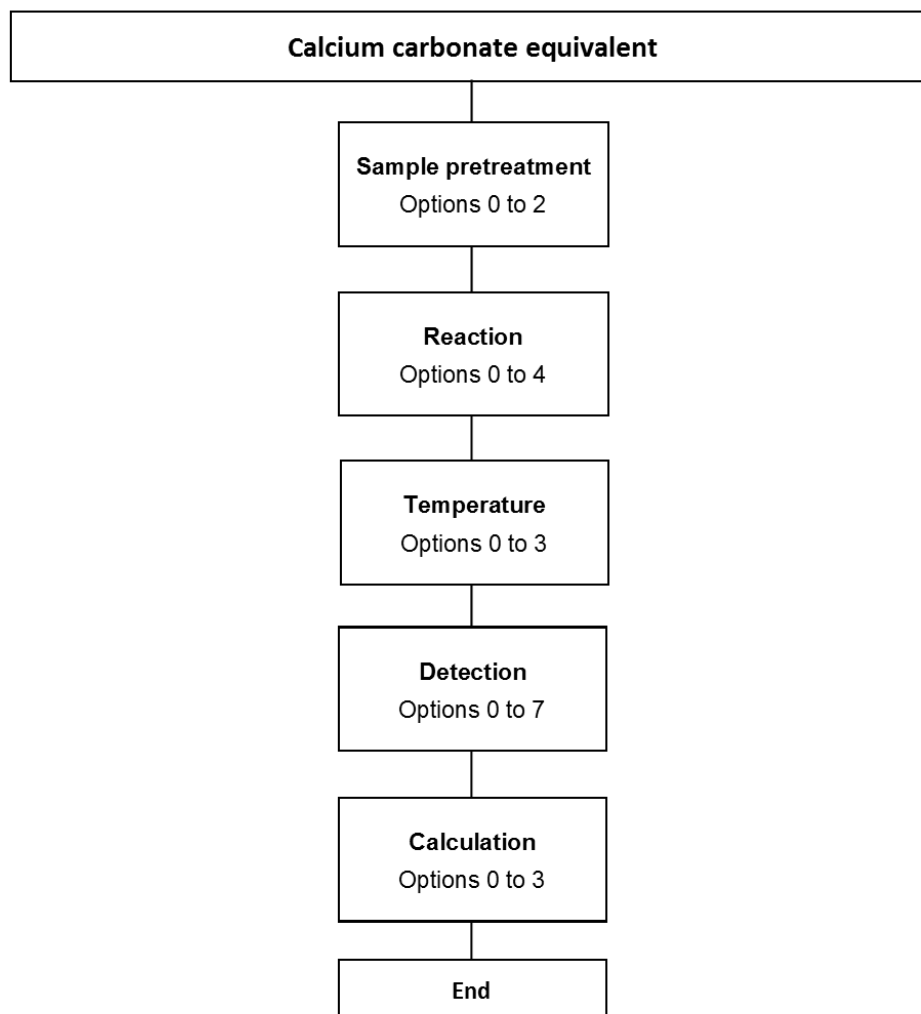
Figure E.2: Flowchart for standardizing calcium carbonate equivalent methods.
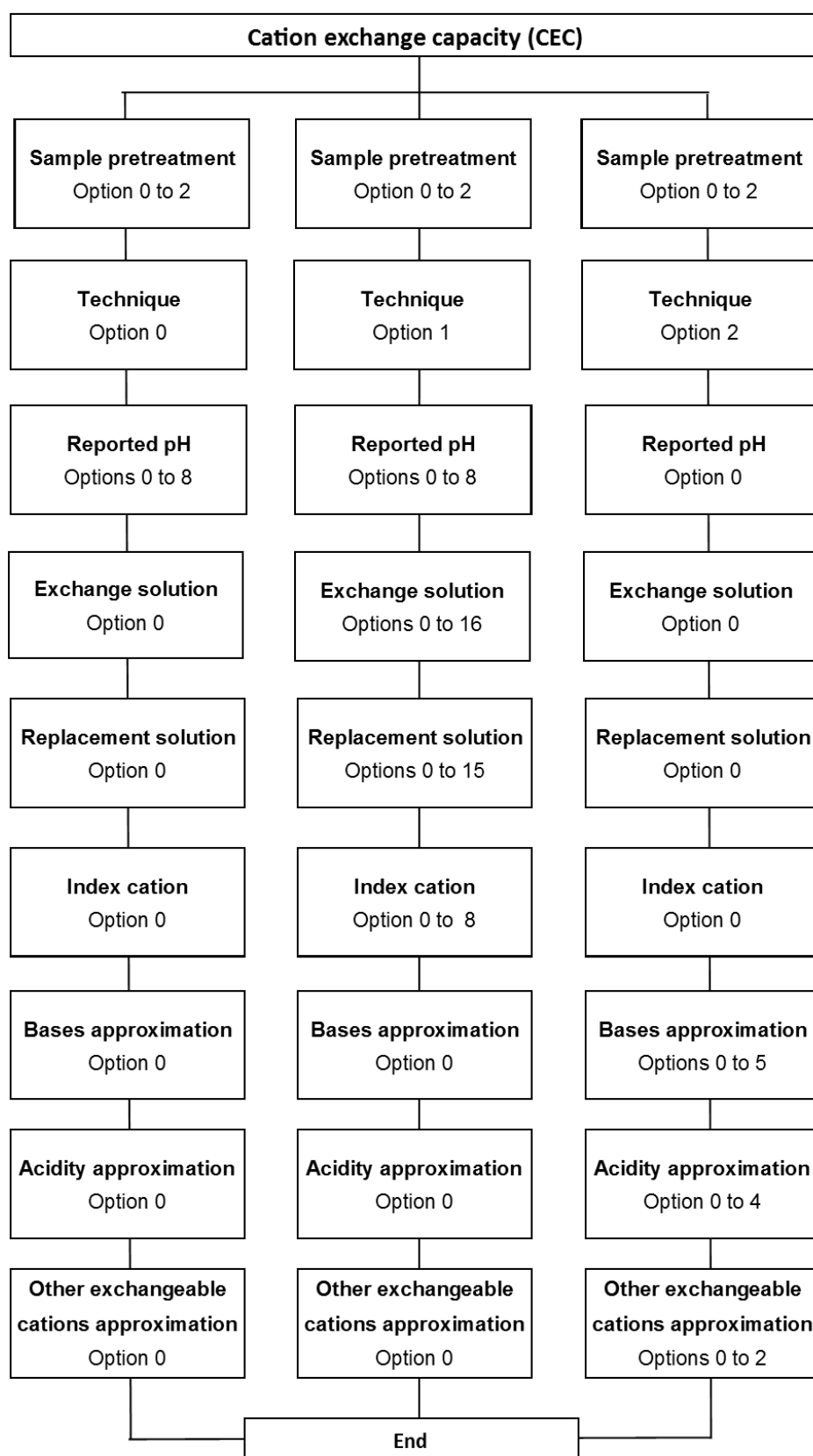
Figure E.3: Flowchart for standardizing cation exchange capacity methods.
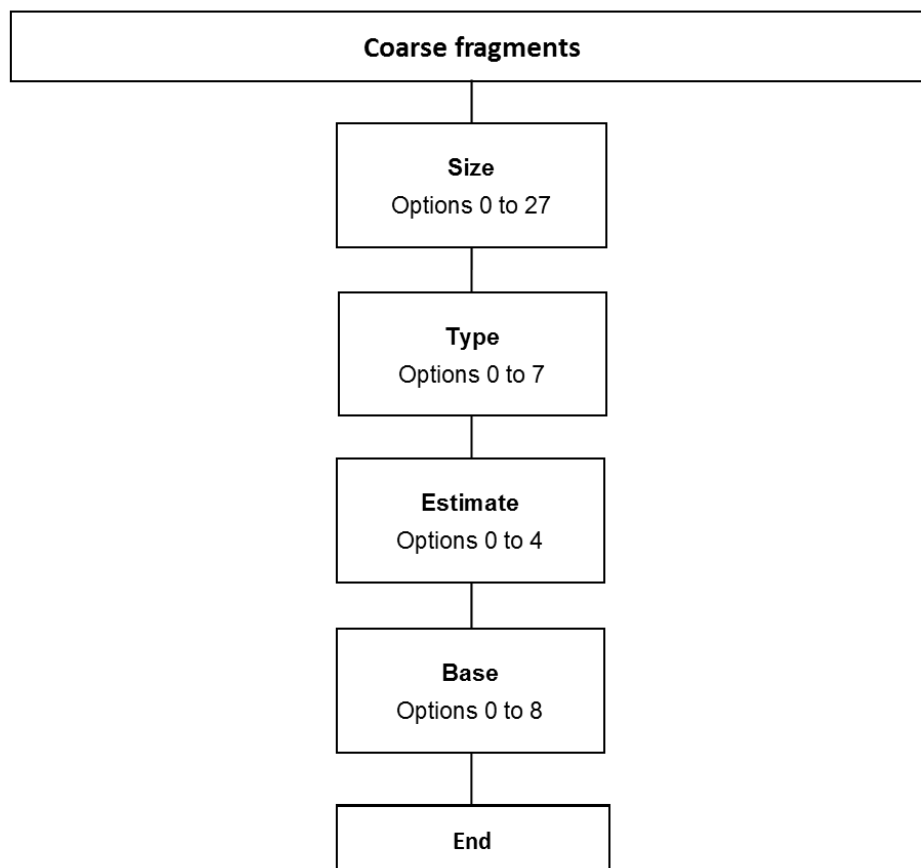
Figure E.4: Flowchart for standardizing coarse fragments methods.
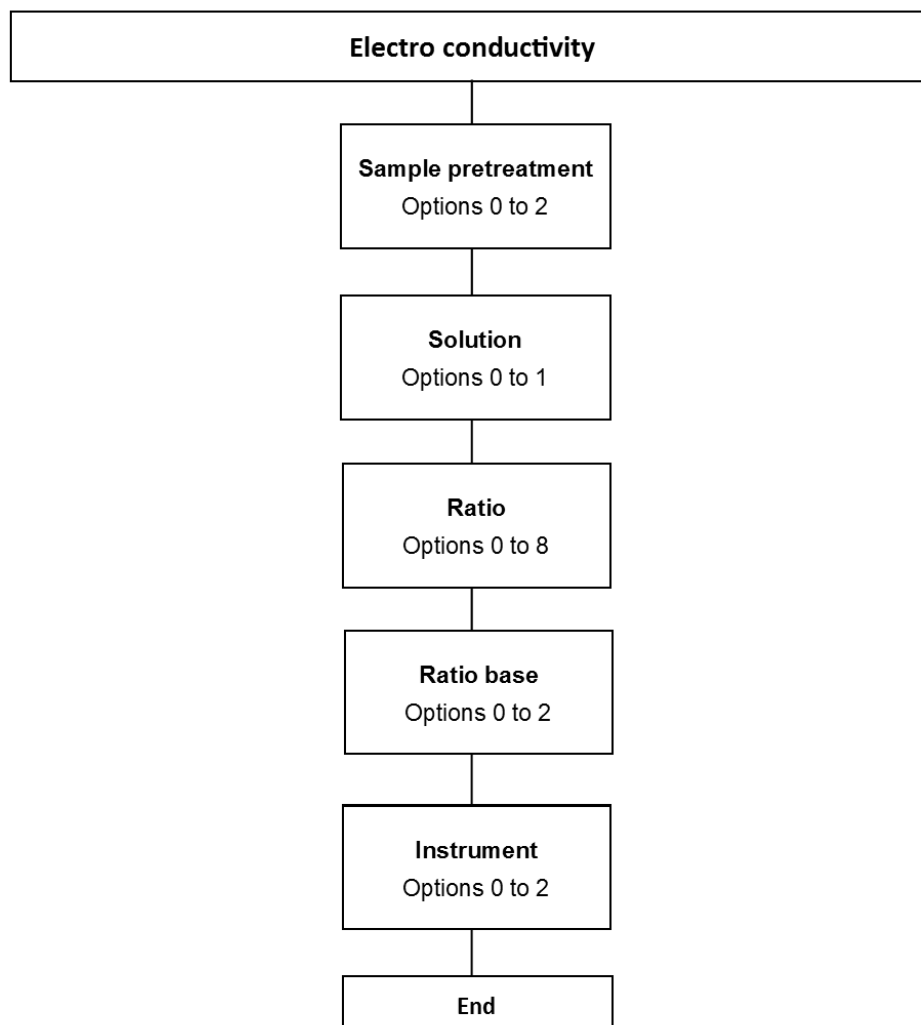
Figure E.5: Flowchart for standardizing electrical conductivity methods.
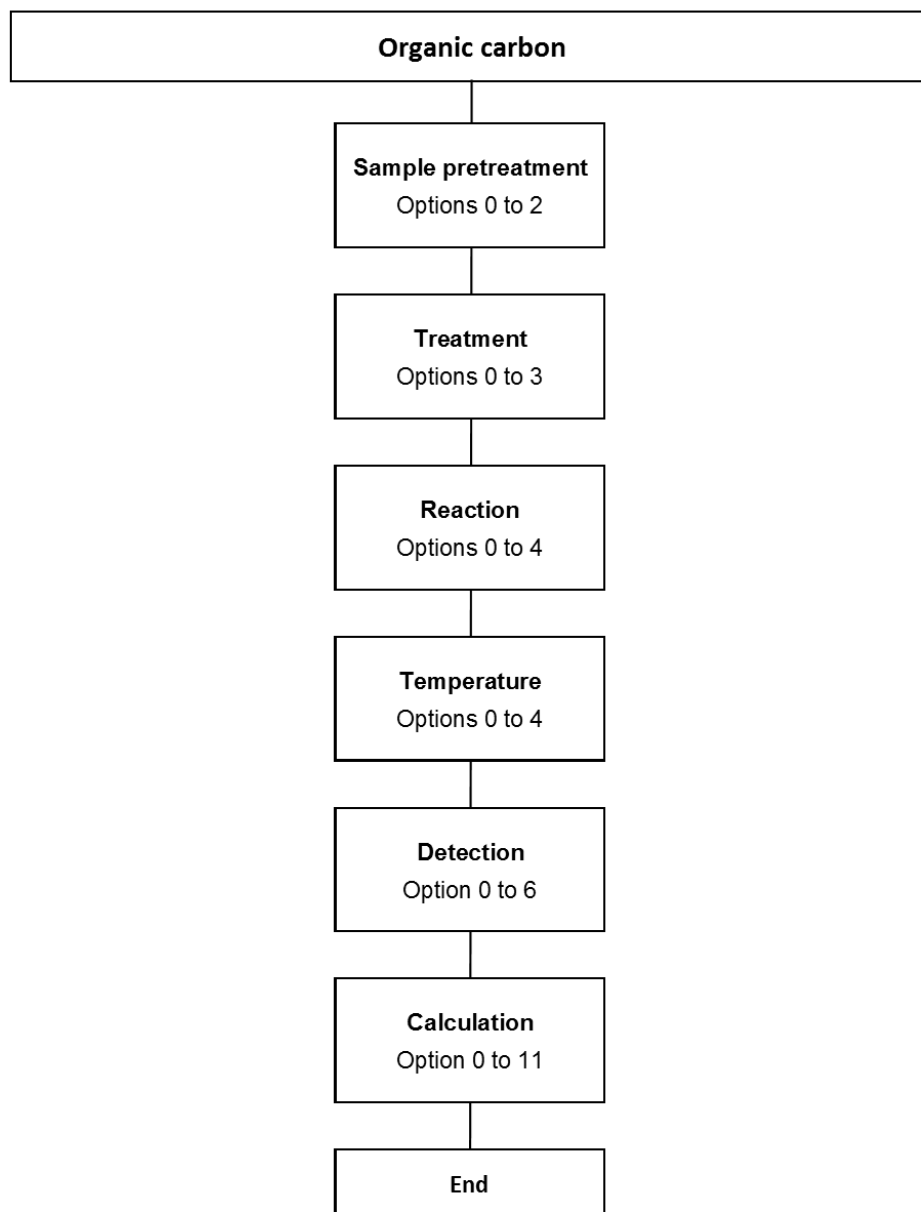
Figure E.6: Flowchart for standardizing organic carbon methods.
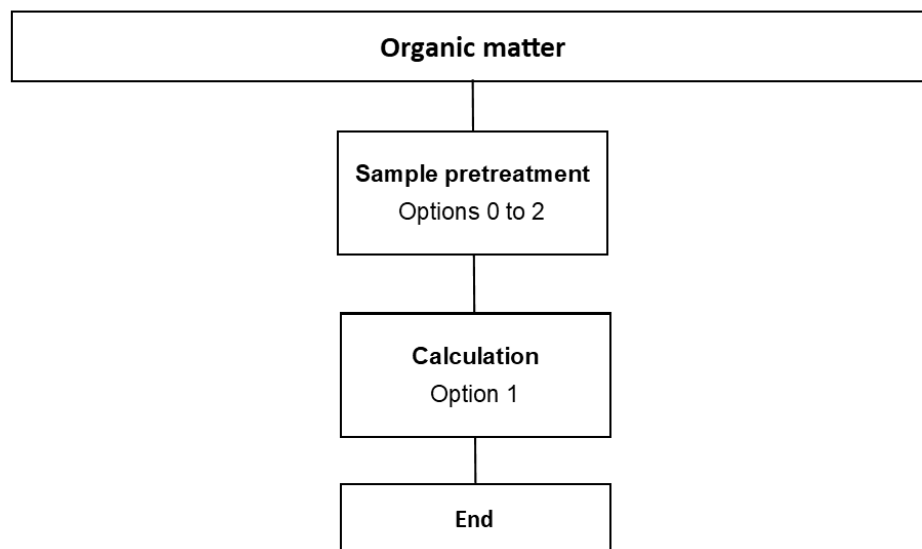
Figure E.7: Flowchart for standardizing organic matter methods.
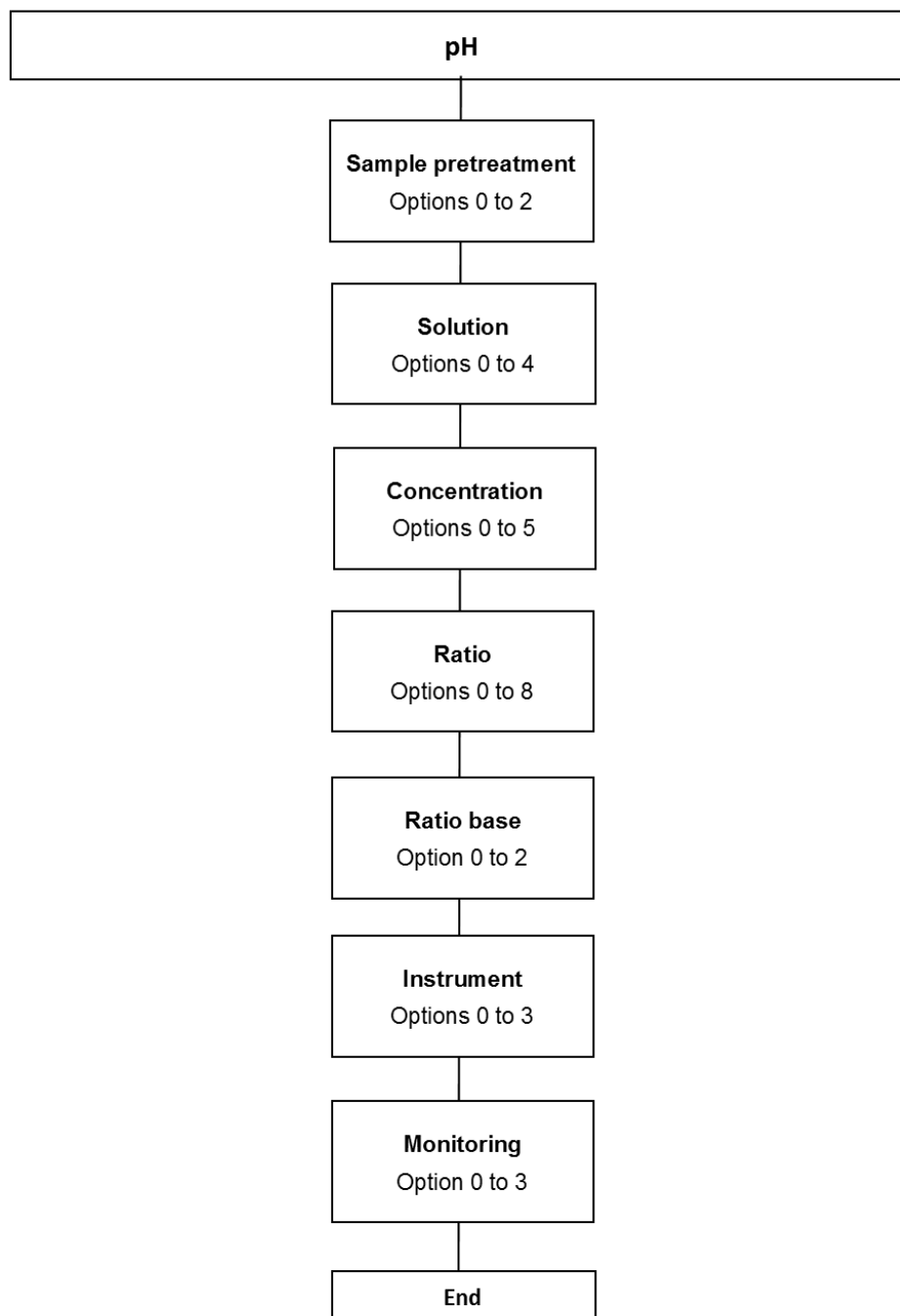
Figure E.8: Flowchart for standardizing pH methods.
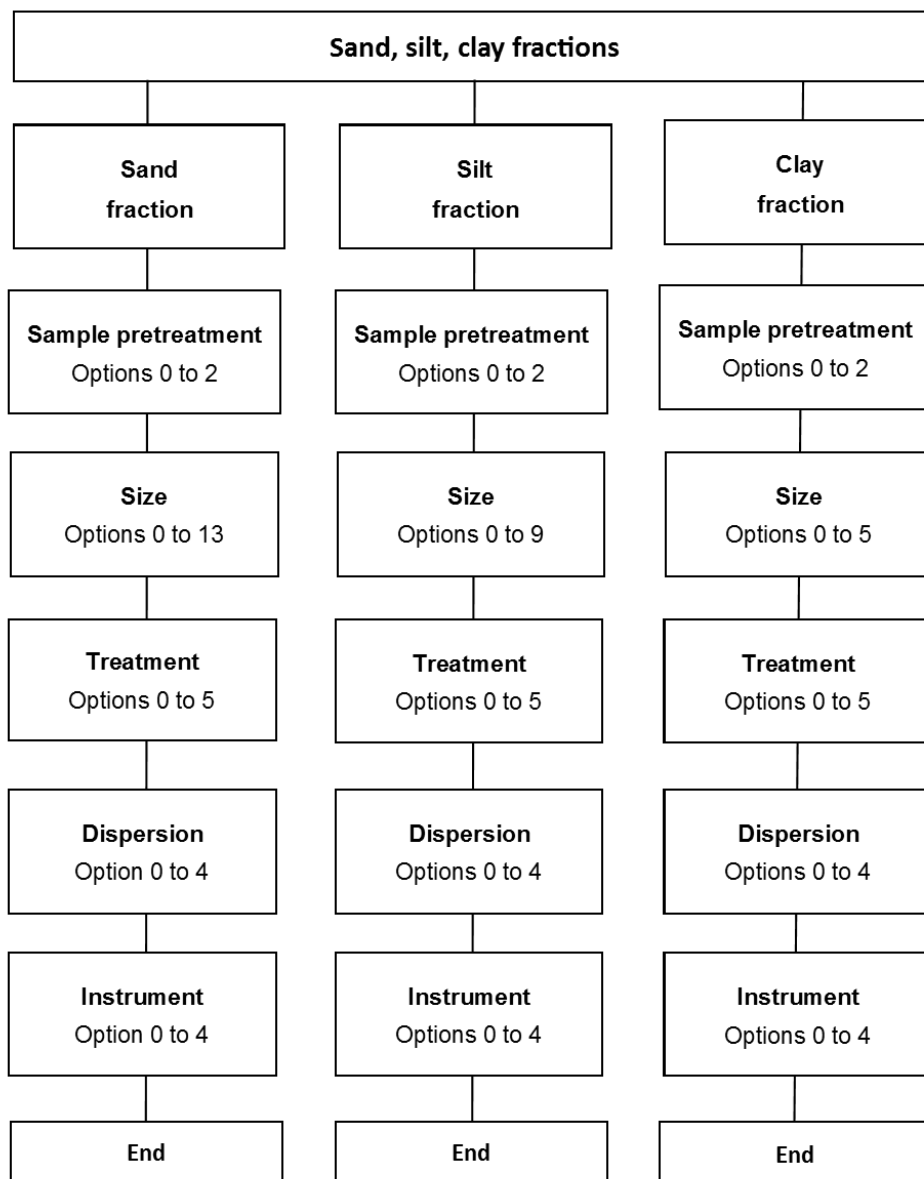
Figure E.9: Flowchart for standardizing sand, silt and clay fractions methods.
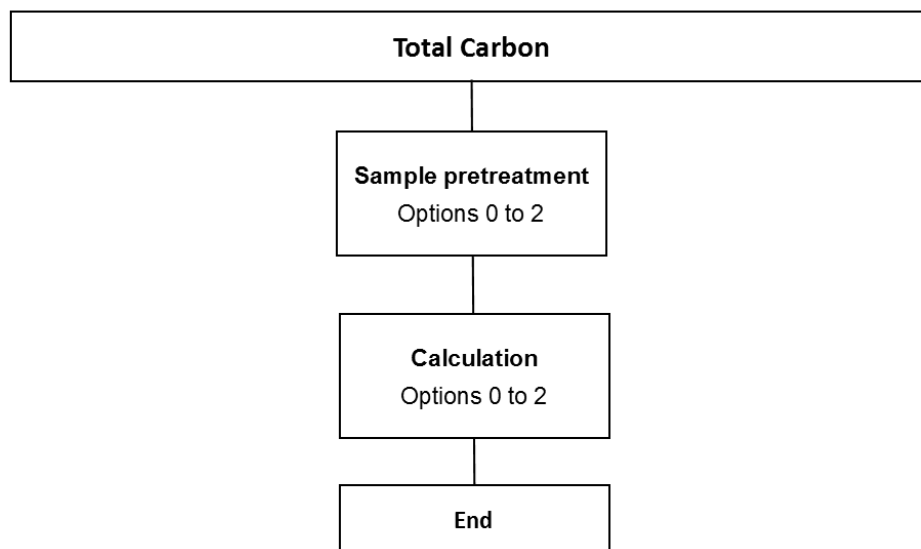
Figure E.10: Flowchart for standardizing total carbon methods.
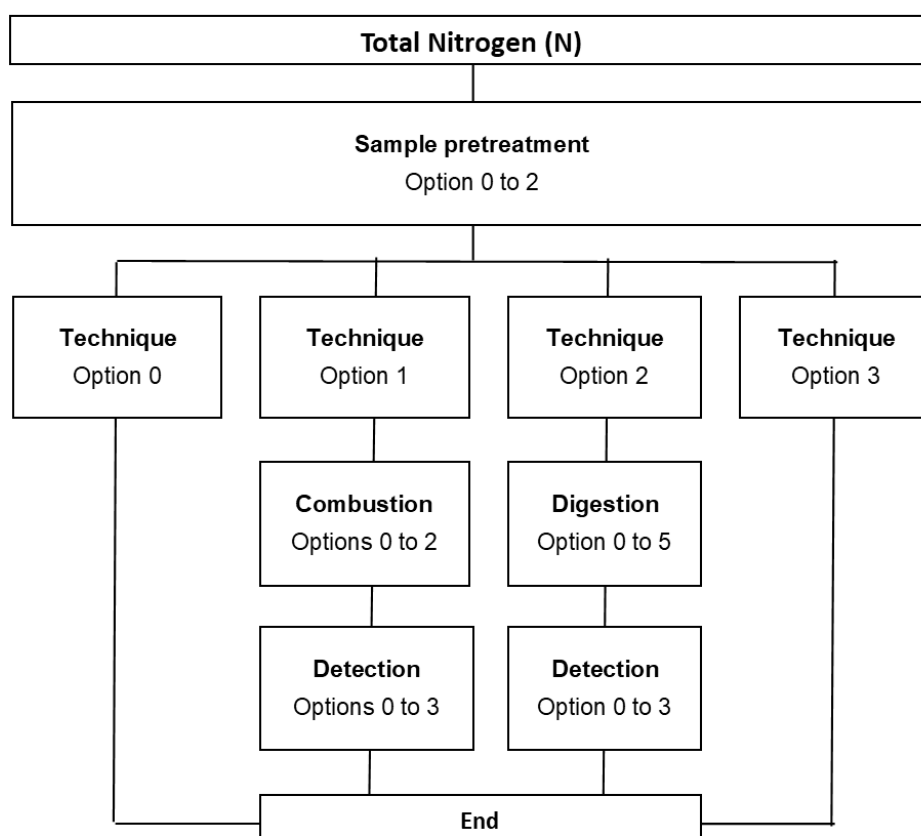
Figure E.11: Flowchart for standardizing total nitrogen methods.
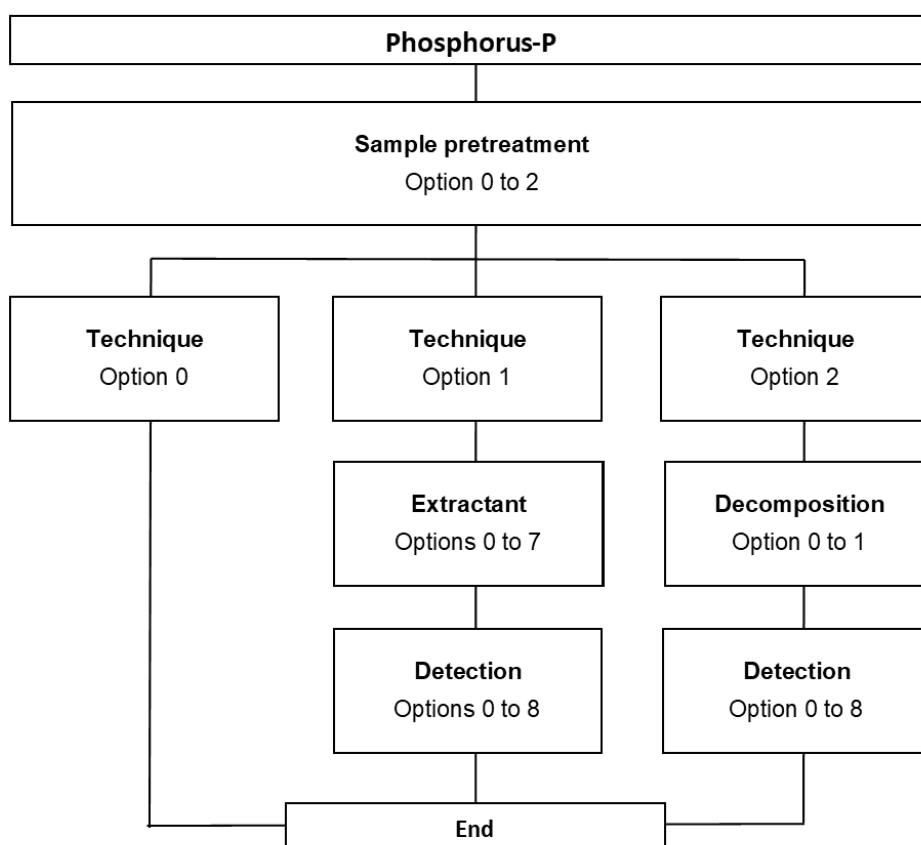
Figure E.12: Flowchart for standardizing total phosphorus methods.
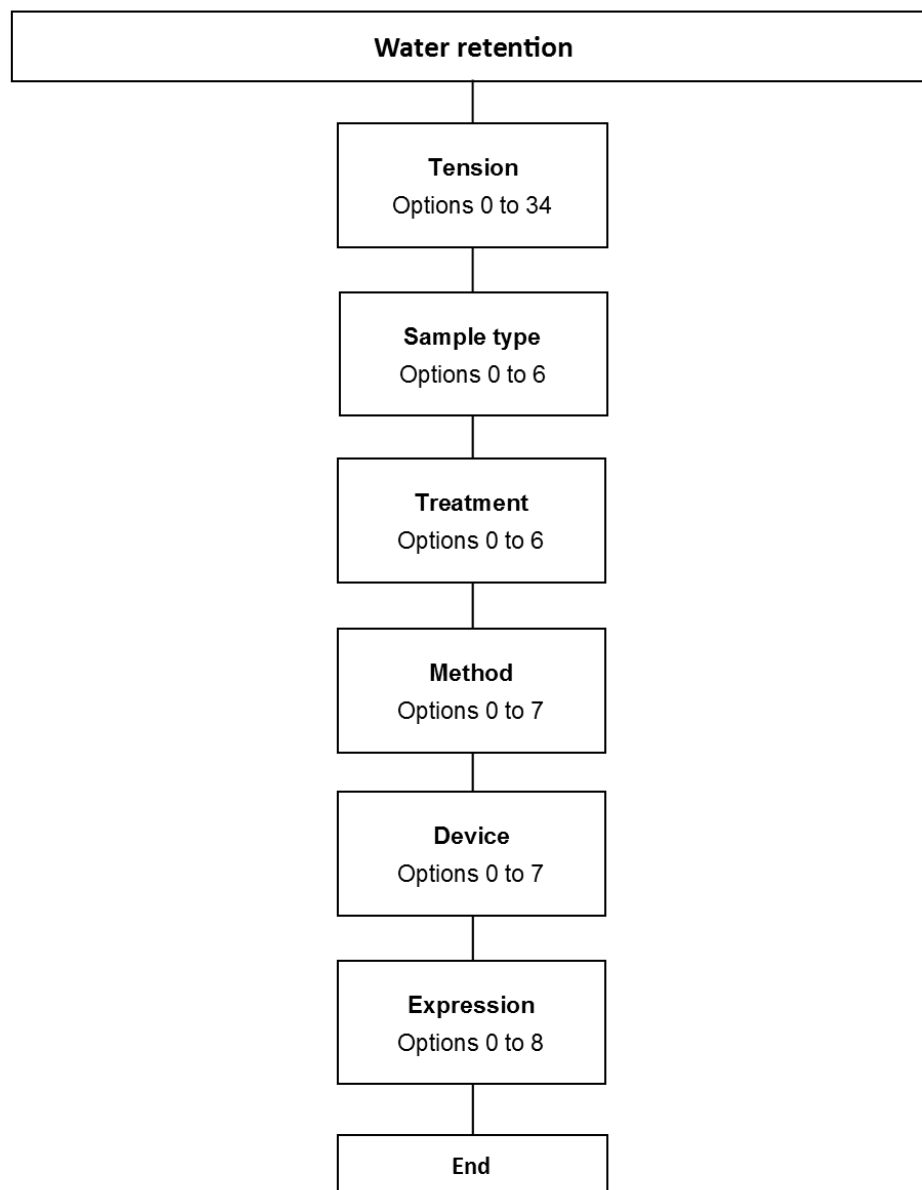
Figure E.13: Flowchart for standardizing water retention methods.

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	not specified
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	not specified
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	not specified
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	not specified
calculation	1	guessed value, expert field estimate

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

Key	Code	Value
sample pretreatment	0	not specified
sample pretreatment	1	sieved over 1 mm sieve
sample pretreatment	2	sieved over 2 mm sieve
reaction	0	not specified
reaction	1	dissolution of Carbonates by Hydrochloric acid [HCl], or Perchloric acid [HClO$_4$]
reaction	2	dissolution of Carbonates by Sulfuric acid [H$_2$SO$_4$]
reaction	3	dissolution of Carbonates by Phosphoric acid [H$_3$PO$_4$]
reaction	4	dissolution of Carbonates by Acetic acid [CH$_3$COOH]
reaction	5	dissolution by other acid solutions
temperature	0	not specified
temperature	1	no external heat
temperature	2	external heat, elevated temperature; ignition $\leq$ 400 $^{\circ}$C
temperature	3	external heat, combustion (element analyzer)
detection	0	not specified
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	not specified
calculation	1	(in)direct estimates of Carbonates [XXCO$_3$.xxH$_2$O] 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.

Key	Code	Value
sample pretreatment	0	not specified
sample pretreatment	1	sieved over 1 mm sieve
sample pretreatment	2	sieved over 2 mm sieve
technique	0	not specified
technique	1	determination by lab procedure
technique	2	approximated by summation exchangeable cations
reported pH	0	not specified
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	not specified
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	not specified
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	not specified
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	not specified
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	not specified
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	not specified
other exchangeable cations	1	not applied
other exchangeable cations	2	(sum of) extractable Fe, Mn (NH_4Cl, 0.05M, unbuffered)
spectral	0	not specified
spectral	1	false
spectral	2	true

Table F.4: Procedure for coding Clay.

Key	Code	Value
sample pretreatment	0	not specified
sample pretreatment	1	sieved over 1 mm sieve
sample pretreatment	2	sieved over 2 mm sieve
size	0	not specified
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	not specified
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	not specified
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	not specified
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	not specified
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 3/4 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	not specified
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	not specified
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	not specified
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	not specified
sample pretreatment	1	sieved over 1 mm sieve
sample pretreatment	2	sieved over 2 mm sieve
solution	0	not specified
solution	1	water [H_2O]
ratio	0	not specified
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	not specified
ratio base	1	weight / volume
ratio base	2	volume / volume
instrument	0	not specified
instrument	1	electrode
instrument	2	electrode (field measured)

Table F.7: Procedure for coding Organic carbon.

Key	Code	Value
sample pretreatment	0	not specified
sample pretreatment	1	sieved over 1 mm sieve
sample pretreatment	2	sieved over 2 mm sieve
treatment	0	not specified
treatment	1	not applied
treatment	2	inorganic carbon removed; Hydrochloric acid [HCl]
treatment	3	inorganic carbon removed; Phosphoric acid [H_3PO_4]
reaction	0	not specified
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	not specified
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	not specified
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	not specified
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
spectral	0	not specified
spectral	1	false
spectral	2	true

Table F.8: Procedure for coding Organic matter.

Key	Code	Value
sample pretreatment	0	not specified
sample pretreatment	1	sieved over 1 mm sieve
sample pretreatment	2	sieved over 2 mm sieve
calculation	1	No conversion

Table F.9: Procedure for coding pH.

Key	Code	Value
sample pretreatment	0	not specified
sample pretreatment	1	sieved over 1 mm sieve
sample pretreatment	2	sieved over 2 mm sieve
solution	0	not specified
solution	1	Water [H_2O]
solution	2	Calcium chloride [$CaCl_2$]
solution	3	Potassium chloride [KCl]
solution	4	Sodium fluoride [NaF]
solution	5	Ammonium chloride [NH_4CL]
concentration	0	not specified
concentration	1	not applied
concentration	2	0.01 M
concentration	3	0.02 M
concentration	4	0.1 M
concentration	5	0.2 M
concentration	6	1 M
ratio	0	not specified
ratio	1	1:1
ratio	2	1:2
ratio	3	1:2.5
ratio	4	1:5
ratio	5	1:10
ratio	6	1:20
ratio	7	1:25
ratio	8	1:40
ratio	9	1:50
ratio	10	saturated paste
ratio	11	slurry
ratio base	0	not specified
ratio base	1	weight / volume
ratio base	2	volume / volume
instrument	0	not specified
instrument	1	electrode
instrument	2	electrode (field measured)
instrument	3	indicator paper (field measured)
monitoring	0	not specified
monitoring	1	not applied
monitoring	2	oxidizible sulfur compounds; initial pH
monitoring	3	oxidizible sulfur compounds; pH stabilized in $\geq$ 10 days, pH $\leq$ 0.1 unit, for two days
spectral	0	not specified
spectral	1	false
spectral	2	true

Table F.10: Procedure for coding Phosphorus.

Key	Code	Value
sample pretreatment	0	not specified
sample pretreatment	1	sieved over 1 mm sieve
sample pretreatment	2	sieved over 2 mm sieve

Table F.11: Procedure for coding Sand.

Key	Code	Value
sample pretreatment	0	not specified
sample pretreatment	1	sieved over 1 mm sieve
sample pretreatment	2	sieved over 2 mm sieve
size	0	not specified
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	not specified
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	not specified
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	not specified
instrument	1	sieve
instrument	2	hydrometer
instrument	3	analyzer
instrument	4	field hand estimate

Table F.12: Procedure for coding Silt.

Key	Code	Value
sample pretreatment	0	not specified
sample pretreatment	1	sieved over 1 mm sieve
sample pretreatment	2	sieved over 2 mm sieve
size	0	not specified
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	not specified
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	not specified
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	not specified
instrument	1	pipette
instrument	2	hydrometer
instrument	3	analyzer
instrument	4	field hand estimate

Table F.13: Procedure for coding Total carbon.

Key	Code	Value
sample pretreatment	0	not specified
sample pretreatment	1	sieved over 1 mm sieve
sample pretreatment	2	sieved over 2 mm sieve
calculation	0	not specified
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.14: Procedure for coding Total nitrogen.

Key	Code	Value
sample pretreatment	0	not specified
sample pretreatment	1	sieved over 1 mm sieve
sample pretreatment	2	sieved over 2 mm sieve
spectral	0	not specified
spectral	1	false
spectral	2	true

Table F.15: Procedure for coding Water retention.

Key	Code	Value
tension	0	not specified
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	not specified
sample type	1	natural clod
sample type	2	clod, reconstituted / disturbed (sieved material <2 mm)
sample type	3	undisturbed soil in metal/PVC-ring (soil core)
sample type	4	clod, reconstituted / disturbed (sieved material <1 mm)
sample type	5	<2 mm (sieved) disturbed samples
sample type	6	<1 mm (sieved) disturbed samples
treatment	0	not specified
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	not specified
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	not specified
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	not specified
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

Towards the harmonization of standardized soil analytical data

As indicated earlier, a desired final step in data processing, full data harmonization, would involve making similar data comparable, that is as if assessed by a commonly endorsed single reference method (e.g. for pH, CEC, and organic carbon, see Baritz et al. (2014). So far, however, no consensus has been reached as to the nature of these methods (e.g. ISO standards (not free) or GLOSOLAN SOP's (freely available)).

Various proficiency testing (PT) programmes are working towards full harmonization, for example GLOSOLAN (Suvannang et al., 2018) and WEPAL (2015). In such PT programmes, standardized soil samples are analyzed according to various 'national' methods (X) as well as the selected standard method Y for say soil pH (e.g. pH_{water}, 1:2.5). In practice, however, there are few open data sets that permit such comparisons. The available transfer functions, as reviewed by (GlobalSoiMap, 2015), often are soil type or region-specific, limiting their widespread use for 'global gap filling'.

With WoSIS, we can only compare pH results obtained in water (phaq), a KCl solution (phkc) and a CaCl$_2$ solution (phca), see Table G.1.

Table G.1: Number of samples analysed according to three distinct pH methods.

Property	All ratio's	1:1	1:2.5	1:5	1:10	NS
phaq	612359	301520	37360	106192	11313	156270
phca	313660	316	11197	41530	248	10024
phkc	150238	43903	12775	1011	0	85518

Note: NS ratio not specified. phaq stands for pH_{water}, pkkc for pH measured in a KCL solution, and phca for pH measured in a CaCl$_2$ solution. Molarity of the KCl and CaCl$_2$ solutions were not taken into consideration at this stage. Table based on data for 'WoSIS latest 2018'.

Overall, the regressions explained 92-98 percent of the observed variation in the case of 'phaq ~phca' and from 57-85 percent of the variation for 'phaq ~phkc' (Table G.2). Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE), which provide a measure for the accuracy of the predictions, are some 0.3 pH units for 'phaq ~phca' and 0.5 pH units for 'phaq ~phkc'. More detailed, follow up analyses should cluster the available data by e.g. main soil types (e.g. low activity versus high activity soils) and take the molarity of the solution into account for pH KCL and pH CaCl$_2$.

Table G.2: Linear regressions between phaq and phca respectively phkc.

function	ratio	n	regression	R^2_{adj}	RMSE	MAE	PERR
phaq ˜phca	1:X	265808	Y=0.9609241 + 0.9295786*Y	0.95	0.293	0.220	0.046
	1:1	242	Y=1.111025 + 0.900101*Y	0.96	0.353	0.250	0.056
	1:2.5	4550	Y=1.211000 + 0.928085*Y	0.92	0.383	0.289	0.058
	1:5	19447	Y=0.868627 + 0.991284*Y	0.92	0.390	0.298	0.061
	1:10	225	Y=0.999619 + 0.925271*Y	0.98	0.272	0.219	0.042
phkc ˜phkc	1:X	142384	Y=0.254422 + 0.005394*Y	0.85	0.483	0.357	0.081
	1:1	43243	Y=1.1441411 + 0.908604*Y	0.85	0.458	0.347	0.079
	1:2.5	12586	Y=1.172626 + 0.977563*Y	0.84	0.520	0.397	0.085
	1:5	782	Y=0.67227 + 0.82341*Y	0.57	0.603	0.487	0.108
	1:10	0	n/a	n/a	n/a	n/a	n/a

Note: Repeated K-fold cross validation (10 fold, repeated 5 time); PERR is the prediction error, taken as the mean error from the number of repeats. Cases such as (phaq $< 2*$ phkc) or (phaq $> 2*$ phkc) (377 cases), respectively (phaq $> 2*$ phca) or (phaq $< 2*$ phca) (6 cases) were removed from the analyses as likely blatant 'lab errors'.

Appendix H

Database model

This appendix describes the structure of all PostgreSQL tables considered in WoSIS version 2020. Pragmatically, each table starts on a new page.

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

Column	Type	Default	Comment
class_fao_id	integer	Sequence	Primary key
dataset_id	character varying		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		Major soil group code
major_group	text		Major soil group name
soil_unit_code	character varying		Soil unit code
soil_unit	text		Soil unit name
subunit_code	character varying		Soil unit code
subunit	text		Soil unit name
phase_code	character varying		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	'A'::bpchar	Level of trust: "A" as entered, no validation; "B" standardized; "C" harmonized

108

Table H.2: class_fao_horizon - Diagnostic horizon according to the FAO Soil Map of the World Legend.

Column	Type	Default	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		Soil horizon
upper_depth	smallint		Depth of upper horizon boundary (cm)
lower_depth	smallint		Depth of lower horizon boundary (cm)

109

Table H.3: class_fao_property - Diagnostic property according to the FAO Soil Map of the World Legend.

Column	Type	Default	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 H.4: class_local - Soil name according to the national soil classification system.

Column	Type	Default	Comment
class_local_id	integer	Sequence	Primary key
dataset_id	character varying		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		Full name according to national classification system
common_name	text		Common (local) soil name
note	text		Comments field
trust	character	'A'::bpchar	Level of trust: "A" as entered, no validation; "B" standardized; "C" harmonized

Table H.5: class_soil_taxonomy - Soil name according to USDA Soil Taxonomy with defined version.

Column	Type	Default	Comment
class_soil_taxonomy_id	integer	Sequence	Primary key
dataset_id	character varying		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	'A'::bpchar	Level of trust: "A" as entered, no validation; "B" standardized; "C" harmonized

112

Table H.6: class_wrb - Soil name according to World Reference Base for Soil Resources with defined version.

Column	Type	Default	Comment
class_wrb_id	integer	Sequence	Primary key
dataset_id	character varying		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
reference_soil_group_code	character varying		Reference soil group code
reference_soil_group	text		Reference soil group 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	'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

113

Table H.7: class_wrb_horizon - Diagnostic horizons according to the World Reference Base for Soil Resources.

Column	Type	Default	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 H.8: class_wrb_material - Diagnostic materials according to the World Reference Base for Soil Resources.

Column	Type	Default	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 H.9: class_wrb_property - Diagnostic properties according to the World Reference Base for Soil Resources.

Column	Type	Default	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 H.10: class_wrb_qualifier - Qualifiers according to the World Reference Base for Soil Resources.

Column	Type	Default	Comment
class_wrb_id	integer		Foreign key that refers to table "class_wrb"
qualifier	text		Qualifier name
qualifier_position	character		Qualifier position prefix(WRB 2006)/principal(WRB 2015) or suffix(WRB 2006)/suplementary(WRB 2015)
qualifier_order	smallint		Qualifier position, ordering according to WRB version
specifier	text		Specifier, according to WRB version

Table H.11: conversion - Conversion factor between different units within the same attribute.

Column	Type	Default	Comment
desc_attribute_standard_id	text		Standard attribute name
unit_from	text		Source unit
operation	text		Operation to be used in the convertion
value	real		Value to be used in the convertion
unit_to	text		Target unit

Table H.12: country - Global Administrative Unit Layers (GAUL) from FAO and ISO 3166 International Standard country codes.

Column	Type	Default	Comment
country_id	character		Primary key and ISO 3166-1 alpha-2, two-letter country code
iso3_code	character		ISO 3166-1 alpha-2, three-letter country code
gaul_code	integer		Global Administrative Unit Layers (GAUL) country code
color_code	character		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		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

119

Table H.13: 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	Default	Comment
dataset_id	character varying		Primary key
version	text		Dataset version
dataset_progress	text		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 standardized 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 standardized records from the dataset describing the layers
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 smaller the number the higher the confidence
note	text		Comments field

Table H.14: 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	Default	Comment
dataset_id	character varying		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

121

Table H.15: desc_attribute - Description of all the soil properties for each dataset that has been imported in WoSIS.

Column	Type	Default	Comment
desc_attribute_id	integer	Sequence	Primary key
dataset_id	character varying		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
conversion	text		Conversion applied to standardize the values
sql_insert	text		SQL code for moving the data from the source into WoSIS
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_domain	ARRAY		Array of conflict values from source compared with the allowed domain values
user_name	text	"current_user"()	Who imported the soil propwety
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
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 H.16: desc_attribute_agg - Used to aggregate desc_attribute_standard_id and asign analytical method descriptions to multiple desc_attribute_standard_id.

Column	Type	Default	Comment
desc_attribute_agg_id	text		Standard attribute aggregation name
counter	smallint		Shows how many desc_attribute_standard_id are being aggregated here
method	boolean		Is it being used in desc_method_feature table

Table H.17: desc_attribute_standard - Description of standard attributes and standardisation progress.

Column	Type	Default	Comment
desc_attribute_standard_id	character varying		Standard attribute name
attribute_type	text		Attribute type (Site/Horizon/Database)
desc_attribute_agg_id	text		Standard attribute aggregation name
desc_unit_id	character varying		Standard attribute unit
data_type	character varying		Standard attribute type
decimals	smallint		Number of decimals
minimum	numeric		Minimum value
maximum	numeric		Maximum value
desc_domain_id	text		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 attribute is from Global Soil Map
gfsd	boolean		If the attribute is from Guide lines for soil description FAO (2006)
note	text		Comments field
observed_minimum	numeric		Minimum observed value
observed_maximum	numeric		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	text		Standard attribute code
attribute_description	text		Standard attribute description
attribute_common_name	text		Attribute common name
sg_variable	boolean		If the attribute is used in SoilGrids project
accuracy	numeric		Attribute accuracy
phys_chem	text		Attribute category: Physical, Chemical, Categorical or Quantitative
uuid	uuid	uuid_generate_v1()	Universal unique identifier of the attribute

Table H.18: 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	Default	Comment
desc_domain.id	text		Primary key
document	text		Reference document of the domain
year	smallint		Publication year of the reference document
page	smallint		Page number of the reference document
figure	smallint		Figure number of the reference document

125

Table H.19: 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	Default	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 H.20: desc_laboratory_source - Listing of laboratories where soil samples have been analysed.

Column	Type	Default	Comment
desc_laboratory_source_id	integer	Sequence	Primary key
source_laboratory_name	text		Comments field
desc_laboratory_standard_id	text		Foreign key that refers to table "desc_laboratory_standard"

Table H.21: desc_laboratory_standard - Listing of laboratories where soil samples have been analysed (clean).

Column	Type	Default	Comment
desc_laboratory_standard_id	text		Primary key
laboratory	text		Laboratory name
country_id	character		Country code where the laboratory is located, ISO 3166-1 alpha-2
city	text		Laboratory city
postal_code	text		Laboratory postal code
street_name	text		Laboratory street name
street_number	text		Laboratory street number
reference_id	integer		Foreign key that refers to table "reference"

Table H.22: desc_method_option - Criteria used to standardise disparate soil analytical method descriptions, for a given analytical method, according to a defined set of features.

Column	Type	Default	Comment
desc_attribute_agg_id	character varying		Standard attribute aggregation name
key	character varying		Standard analytical method feature name. Combined primary key ("standard_attribute_name","feature_name","feature_option")
value	text		Standard analytical method feature option. Combined primary key ("standard_attribute_name","feature_name","feature_option")
value_desc	text		Description of the standard analytical method option
key_order	smallint		Numeric value to order the keys
value_order	smallint		Numeric value to order the values

Table H.23: desc_method_source - Analytical methods descriptions as defined in the respective source databases (i.e. prior to standardization).

Column	Type	Default	Comment
desc_method_source_id	text	Sequence	Primary key
source_analytical_method_name	text		Source analytical method name
source_analytical_method_code	text		Other datasets analytical method code

Table H.24: desc_method_standard - Results of the standardization of the Soil Analytical Methods descriptions to WoSIS 2 standards.

Column	Type	Default	Comment
desc_method_source_id	text		Foreign key that refers to table "descriptor"
desc_attribute_agg_id	text		Standard soil property name. Combined foreign key ("standard_attribute_name","feature_name","feature_option")
key	text		Standard analytical method feature name. Combined foreign key ("standard_attribute_name","feature_name","feature_option")
value	text		Standard analytical method feature option. Combined foreign key ("standard_attribute_name","feature_name","feature_option")
confidence	character varying		Confidence in the standardization procedure

Table H.25: desc_unit - Units used for measurement of soil, site, and terrain characteristics.

Column	Type	Default	Comment
desc_unit_id	character varying		Primary key
unit_description	character varying		Unit description

Table H.26: descriptor - Unique combinations of Attribute, Analytical Method and Laboratory definition that define a measured value.

Column	Type	Default	Comment
descriptor_id	integer	Sequence	Primary key
desc_attribute_id	smallint		Foreign key that refers to table "desc_attribute"
desc_method_source_id	text		Foreign key that refers to table "desc_method_source"
desc_laboratory_source_id	smallint		Foreign key that refers to table "desc_laboratory"
exclude	boolean	false	When TRUE excludes that set of values to be distributed throughout web-services
desc_attribute_standard_id	text		Foreign key that refers to table "desc_attribute_standard"
exclude_note	text		Notes for when "exclude"=TRUE

Table H.27: profile - List of soil profiles, with their location (geometry).

Column	Type	Default	Comment
profile_id	integer	Sequence	Primary key
country_id	character		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	uuid_generate_v1()	Universal unique identifier of the profile
geom	USER-DEFINED		Point geometry of the location of the profile
trust	character	'A'::bpchar	Level of trust: 'A' as entered, no validation; 'B' standardized; 'C' harmonized
geom_accuracy_type	character varying	'UNK'::bpchar	GPS - Coordinates come from a GPS; MAP - Coordinates come from a map; UNK - Unknown coordinates source
sample_type	character varying	'single'::character varying	Either single or composite sample
sample_number	integer	1	Number of samples
sample_area	integer	1	Area sampled (m2)
country_intersect	character		Closest country to profile, based on geometry
country_note	text		Distance of the closest country to profile, based on geometry
hide	boolean	false	If it is a hidden profile or not
year	smallint		Sampling year
month	smallint		Sampling month
day	smallint		Sampling day
continent	text		Continent where the profile is located

Table H.28: profile_attribute - Values of characteristics that are associated with the profiles site.

Column	Type	Default	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		Foreign key that together with "dataset_id" refers to table "dataset_profile"
descriptor_id	integer		Foreign key that refers to table "descriptor"
accuracy	text		Accuracy of an observation or measurement precision
trust	character	'A'::bpchar	Level of trust: 'A' as entered, no validation; 'B' standardized; 'C' harmonized
source_value	text		Profile attribute, source value
standard_value	text		Profile attribute, standard value

Table H.29: profile_covariate - Profiles covariates from diverse sources.

Column	Type	Default	Comment
profile_id	integer		
geom	USER-DEFINED		
biome_name	text		Biome from "Terrestrial Ecoregions of the World (TEOW)" - https://www.worldwildlife.org/publications/terrestrial-ecoregions-of-the-world
eco_name	text		Ecoregion from "Terrestrial Ecoregions of the World (TEOW)" - https://www.worldwildlife.org/publications/terrestrial-ecoregions-of-the-world
gelu_ef	text		Ecological Facets from "Global Ecological Land Units (2015)" - https://www.usgs.gov/centers/gecsc/science/global-ecosystems
gelu_ef_bio	text		Ecological Facets Bioclimate from "Global Ecological Land Units (2015)" - https://www.usgs.gov/centers/gecsc/science/global-ecosystems
gelu_ef_lf	text		Ecological Facets Landform from "Global Ecological Land Units (2015)" - https://www.usgs.gov/centers/gecsc/science/global-ecosystems
gelu_ef_lit	text		Ecological Facets Lithology from "Global Ecological Land Units (2015)" - https://www.usgs.gov/centers/gecsc/science/global-ecosystems
gelu_ef_glc	text		Ecological Facets Land Cover from "Global Ecological Land Units (2015)" - https://www.usgs.gov/centers/gecsc/science/global-ecosystems
gelu_elu	text		Ecological Land Units from "Global Ecological Land Units (2015)" - https://www.usgs.gov/centers/gecsc/science/global-ecosystems
gelu_elu_bio	text		Ecological Land Units Bioclimate from "Global Ecological Land Units (2015)" - https://www.usgs.gov/centers/gecsc/science/global-ecosystems
gelu_elu_lf	text		Ecological Land Units Landform from "Global Ecological Land Units (2015)" - https://www.usgs.gov/centers/gecsc/science/global-ecosystems
gelu_elu_lit	text		Ecological Land Units Lithology from "Global Ecological Land Units (2015)" - https://www.usgs.gov/centers/gecsc/science/global-ecosystems
gelu_elu_glc	text		Ecological Land Units Land Cover from "Global Ecological Land Units (2015)" - https://www.usgs.gov/centers/gecsc/science/global-ecosystems

Column	Type	Default	Comment
gelu_elu_site	text		Ecological Land Units from "Global Ecological Land Units (2015)" - https://www.usgs.gov/centers/gecsc/science/global-ecosystems

Table H.30: profile_layer - Lists of Layers depths and samples definition per profile and dataset.

Column	Type	Default	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
layer_name_source	text		Horizon designation from source
layer_name	text		Adapted horizon designation
upper_depth	smallint		Depth of upper horizon boundary (cm)
lower_depth	smallint		Depth of lower horizon boundary (cm)
upper_depth_source	text		Depth of upper horizon boundary (cm) from source
lower_depth_source	text		Depth of lower horizon boundary (cm) from source
litter	boolean	false	Litter layer, TRUE or FALSE
suspicious	boolean	false	Profile layer depths need to be cheked
uuid	uuid	uuid_generate_v1()	Universally unique identifier profile_layer code

Table H.31: profile_layer_attribute - Lists values of any characteristic associated with a profile layer.

Column	Type	Default	Comment
profile_layer_attribute_id	integer	Sequence	Primary key
profile_layer_sample_id	integer		Foreign key that refers to table "profile_layer_sample"
descriptor_id	integer		Foreign key that refers to table "descriptor"
accuracy	text		Accuracy of an observation or measurement precision
trust	character	'A'::bpchar	Level of trust: 'A' as entered, no validation; 'B' standardized; 'C' harmonized
source_value	text		Layer attribute, source value
standard_value	text		Layer attribute, standard value

139

Table H.32: profile_layer_sample - Lists of samples definition per layer.

Column	Type	Default	Comment
profile_layer_sample_id	integer	Sequence	Primary key
profile_layer_id	integer		Foreign key that refers to table "profile_layer"
type	text		Type of sample taken
code	text		Sample code
composition	text		Sample composition
available	text		Sample availability
upper_depth	smallint		Depth of upper sample boundary (cm)
lower_depth	smallint		Depth of lower sample boundary (cm)

Table H.33: profile_layer_spectral - Spectral data.

Column	Type	Default	Comment
profile_layer_spectral_id	integer	Sequence	Primary key
profile_layer_sample_id	integer		Foreign key that refers to table "profile_layer_sample"
source_sample_id	text		Source sample identifier
descriptor_id	integer		Foreign key that refers to table "descriptor"
payload	jsonb		Observed reflectance (NIR) or absorbance (MIR) for a specific wave (length or number)

Table H.34: profile_layer_thinsection - Lists information related with thin sections.

Column	Type	Default	Comment
dataset_id	character varying		Foreign key that refers to table "dataset"
profile_code	character varying		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		Thin section author
horizon	character varying		Soil horizon
upper_depth	integer		Depth of upper horizon boundary (cm)
lower_depth	integer		Depth of lower horizon boundary (cm)
sample_size	character varying		Sample size
missing	boolean		Missing thin section from collection
note	text		Comments field

142

Table H.35: reference - List of references to source materials managed in WoSIS.

Column	Type	Default	Comment
reference_id	integer	Sequence	Primary key
reference_type	text		Type of reference (Publication/Map/URL/Digital media)
isbn	text		ISBN number
isn	text		ISRIC library code
title	text		Title
issue	text		Issue
publication_year	smallint		Publication year
publisher	text		Publisher
url	text		URL

143

Table H.36: reference_dataset_profile - List of references by imported dataset and their profiles.

Column	Type	Default	Comment
reference_id	integer		Primary key that together with dataset and profile
dataset_id	character varying		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		Reference page where the profile description appears
reference_profile_code	character varying		Profile code in such reference

144

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.