Towards the standardization and harmonization of world soil data

Procedures manual

ISRIC World Soil Information Service (WoSIS version 2.0)



Eloi Ribeiro, Niels H. Batjes, Johan G.B. Leenaars, Ad van Oostrum and Jorge Mendes de Jesus

Towards the standardization and harmonization of world soil data

Procedures manual:
ISRIC World Soil Information Service (WoSIS version 2.0)
Eloi Ribeiro, Niels H. Batjes, Johan G.B. Leenaars, Ad van Oostrum and Jorge Mendes de Jesus
ISRIC Report 2015/03

Wageningen, November 2015

All rights reserved. Reproduction and dissemination for educational or non-commercial purposes are permitted without any prior written permission provided the source is fully acknowledged. Reproduction of materials for resale or other commercial purposes is prohibited without prior written permission from ISRIC. Applications for such permission should be addressed to:

Director, ISRIC-World Soil Information PO BOX 353 6700 AJ Wageningen The Netherlands E-mail: soil.isric@wur.nl

Disclaimer:

The designations employed and the presentation of materials do not imply the expression of any opinion whatsoever on the part of ISRIC concerning the legal status of any country, territory, city or area or of is authorities, or concerning the delimitation of its frontiers or boundaries.

Despite the fact that this publication is created with utmost care, the authors(s) and/or publisher(s) and/or ISRIC cannot be held liable for any damage caused by the use of this publication or any content therein in whatever form, whether or not caused by possible errors or faults nor for any consequences thereof.

Additional information on ISRIC-World Soil Information can be accessed through http://www.isric.org

Citation:

Ribeiro E, Batjes NH, Leenaars JGB, van Oostrum AJM, Mendes de Jesus J 2015. Towards the standardization and harmonization of world soil data: Procedures manual ISRIC World Soil Information Service (WoSIS version 2.0). Report 2015/03, ISRIC - World Soil Information, Wageningen.

Note:

This report describes ongoing work with the WoSIS 2 database; as such it will be regularly expanded (with clear time stamps) and only be distributed as an online PDF version. Being based on a LATEX template, the lay out differs from that used for printed reports in the ISRIC series.



Contents

Pr	eface	1								
Su	ımmary	2								
Ac	cronyms and abbreviations	3								
1	Introduction	4								
2	Basic principles for processing data 2.1 Flagging repeated profiles 2.2 Measures for data quality 2.2.1 General considerations 2.2.2 Level of trust 2.2.3 Data quality rating 2.2.4 Accuracy 2.3 Main steps towards data harmonization 2.3.1 Data lineage and access conditions 2.3.2 Data standardization and harmonization	6 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8								
3	3.2 Main components 3.2.1 Metadata 3.2.2 Soil classification 3.2.3 Attribute definition 3.2.4 Source materials (Reference) 3.2.5 Profile data 3.2.6 Map unit (polygon) data	11 14 14 17 20 22 24 26								
4	Interoperability and web services	27								
5	Future developments	29								
Ac	Acknowledgements 30									
ВІ	iography	32								
A	Basic principles for compiling a soil profile dataset	37								
В	B.2 Laboratory error	41 41 43 44								
С	Background information soil analytical procedures C.1 General	45								

	C.2 Soil pH	45
D	Procedures for coding standardized analytical methods	47
Ε	Database design	54

List of Tables

B.1	Procedure for coding standardized analytical methods using pH as an example	43
	Initial list of soil properties for standardization (i.e. for which standardized descriptions are being developed)	45
	Unique option codes for describing 'pH KCl' as determined according to ISO, ISRIC, and	
	USDA laboratory protocols	46
D.1	Propodure for coding Pulk density	47
	Procedure for coding Bulk density	47
	Procedure for coding Carbon	48
	Procedure for coding Clay	49
	Procedure for coding Coarse fragments	49
	Procedure for coding pH	50
	Procedure for coding Sand	50
	Procedure for coding Salta	51
	Procedure for coding Water retention	51
ט.פ	Procedure for coding water retention	51
E.1	class_cpcs	55
	class_fao	56
	class_fao_horizon	57
	class_fao_property	58
	class_local	59
_	class_soil_taxonomy	60
	class_wrb	61
	class_wrb_horizon	62
	class_wrb_material	63
	class_wrb_property	64
	class_wrb_qualifier	65
	country	66
	dataset	67
	dataset_organization	68
	dataset_organization_contact	69
	dataset_profile	70
	desc_attribute	71
	desc_attribute_standard	73
	desc_domain	74
	desc_domain_value	75
	desc_laboratory	_
	desc_method_feature	77
	desc_method_source	78
	desc_method_standard	79
	desc_unit	80
	descriptor	81
	image	82
	image_profile	83
	image_subject	84
	man attribute	95 95

E.31	map_unit								 										 	 86
E.32	map_unit_componer	nt							 											 87
E.33	map_unit_soil_comp	onen	t.						 											 88
E.34	map_unit_soil_comp	onen	t_x.	p۱	of	ile	٠.		 											 89
E.35	profile								 											 90
E.36	profile_attribute								 											 91
E.37	profile_layer								 											 92
E.38	profile_layer_attribu	te							 											 93
E.39	profile_layer_thinsed	ction							 											 94
E.40	raster								 											 95
E.41	reference								 											 96
E.42	reference_author .								 											 97
E.43	reference_dataset .								 											 98
E.44	reference_dataset_p	rofile							 											 99
E.45	reference_file								 											 100

List of Figures

2.1	Location of shared, unique, geo-referenced profiles held in WoSIS 2 (about 76,000,	
	status "received as is" i.e. prior to further standardization/harmonization)	5
2.2	Intersection between ISRIC stand-alone profile databases showing the number of overlapping	g
	profiles (AfSP, Africa Soil Profile Database; ISIS, ISRIC Soil Information Service, SOTER,	
	Soil and Terrain Database; WISE, World Inventory of Soil Emissions potentials database).	6
2.3	Depiction of the accuracy and precision of measurements (Credit: EH&E, 2001)	7
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 (Credit: Chapman, 2005)	7
2.5	Main stages of data standardization and harmonization	9
3.1	Main reference groups and components of the WoSIS 2 database model	c
3.2	Metadata reference group tables	
3.3	Classification tables	
3.4	Attribute reference group tables	-
3.5	Reference data group tables	
3.6	Profile data group tables	
3.7	Map unit tables	
3.8	Raster tables	
4.1	Serving soil layers from WoSIS 2 to the user community	3

Preface

ISRIC - World Soil Information has a mission to serve the international community as custodian of global soil data and information, and to increase awareness and understanding of soils in major global issues. We have developed a centralized enterprise database known as WoSIS (World Soil Information Service) to safeguard and share soil data (point, polygons and grids) upon their standardization and harmonization.

Everybody may contribute data for inclusion in WoSIS. Data providers must indicate how their data may be distributed through WoSIS. This may be "please safeguard a copy of our national dataset", "you may distribute derived soil data but not the actual profile data" or "welcome, there are no restrictions (open access)". Conditions for use are stored in WoSIS together with the full data lineage to ensure that data providers are properly acknowledged. In accord with these conditions, the submitted data will be gradually standardized and harmonized, ultimately to make them "comparable as if assessed by a single given (reference) method".

At this early stage, our focus has been on the standardization of disparate soil analytical method descriptions; the initial set of quality-assessed data has been served through a web feature service. At a later stage, other aspects will be considered such as the harmonization of soil classifications and soil profile descriptions. Ultimately, all quality-assessed and harmonized 'shared' data managed in WoSIS will be queryable using various web services embedded in ISRIC's upcoming GeoNode.

Recommended procedures are being developed and tested to align with the frameworks of on-going international activities such as the Global Soil Partnership, GlobalSoilMap, and the Open Geospatial Consortium. As the International Council for Science accredited World Data Centre for Soils, ISRIC will also serve WoSIS-derived products to the global soil observation system as part of the Global Earth Observing System of Systems.

WoSIS is the result of collaboration with a steadily growing number of partners and data providers, whose contributions are gratefully acknowledged. Successive releases of WoSIS, that accommodate a broader range of quality-assessed soil data, will gradually be released by ISRIC for the shared benefit of the international community and national stakeholders.

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

Summary

To better address the growing demand for soil information ISRIC - World Soil Information has developed a centralized and user-focused database for the shared benefit of the international community. This relational database, referred to as World Soil Information Service (WoSIS), will only serve quality-assessed and authorized data with detailed information on data lineage. Such information may be used to address pressing challenges of our time including food security, land degradation, water resources, and climate change.

Following the release of WoSIS version 1.0 in 2013, several needs for modification were identified by the user community. The consolidated feedback led to a revised version, WoSIS 2, which has been designed in such a way that, in principle, any type of soil data (point, polygon, and grid) can be accommodated. Basic principles for this are described in this report.

Seen the magnitude of the task, the initial focus has been on processing disparate soil profile data sets obtained from various data providers. These data sets are imported 'as is' into a PostgreSQL database, with the original naming and coding conventions, abbreviations, domains, lineage and data licence; thereby copies of the source materials are safeguarded at ISRIC.

Next the source databases are imported into WoSIS, forming the first major step of data standardization (into a single data model). The second step of data standardization, applied to the values for the various soil properties as well as to the naming conventions, is needed to make the data gueryable and useable. Special attention has been paid here to the standardization of analytical method descriptions, in first instance focusing on the list of soil attributes considered in the GlobalSoilMap specifications (e.g. organic carbon, soil pH, soil texture (sand, silt, and clay), coarse fragments (<2mm), cation exchange capacity, bulk density, and water holding capacity). Major characteristics of commonly used methods for determining a given soil property are identified first. For soil pH, for example, these are the 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. The standardized procedures developed so far have been applied to the 'shared' profile data, forming the first set of 'standardized' data generated from WoSIS. A possible third step in the standardization / harmonization process, not yet undertaken in this version of WoSIS, will require data harmonization to make the data comparable that is as if assessed by a single given (reference) method. Such work will require further international collaboration and data sharing to the benefit of the international user community.

So far, WoSIS 2 contains over 98,000 unique 'shared' soil profiles of which some 76,000 are georeferenced within defined limits. In total this corresponds with over 28 million soil records, some 45% of which have been quality-assessed and standardized using the procedures discussed in this report. This initial queryable dataset has been made available through a web feature service, in accord with the license specified by each data provider. In the coming years, ISRIC will serve a wider range of soil information derived from WoSIS, including spatial data sets, through ISRIC's upcoming GeoNode facility. WoSIS forms an important building block for ISRIC's evolving Spatial Data Infrastructure (SDI).

Instrumental to enhanced usability and accessibility of data managed in WoSIS will be the continued harmonization of soil property values as well as the further standardization of soil analytical procedure descriptions. Development of such procedures/interfaces will allow for the fulfilment of future demands for global soil information, and enable further incorporation of soil data shared by third parties.

Acronyms and abbreviations

AfSP Africa Soil Profiles database (a compilation of soil legacy data for Africa)

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
GML Geography Markup Language

INSPIRE Infrastructure for Spatial Information in the European Community

ISIS ISRIC Soil Information System (holds the world soil reference collection)

ISO International Organization for Standardization

ISRIC - World Soil Information(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

SOAP Simple Object Access Protocol

SoilML Soil Markup Language

SOTER Soil and Terrain database programme

UNEP United Nations Environmental Program

USDA United States Department of Agriculture

UUID Universally unique identifier (for soil profiles)

WDC World Data Center of the ICSU World Data System (ICSU-WDS)

WISE World Inventory of Soil Emission potentials (harmonized soil profile data for the world)

WoSIS World Soil Information Service (server database)

WSDL Web Services Description Language

XML Extensible Markup Language
XSL Extensible Stylesheet Language

Chapter 1

Introduction

ISRIC's mission is to "serve the international community with quality-assessed information about the worlds soil resources to help addressing major global issues". Since the 1980's, ISRIC has developed and managed a number of soil-related databases that are freely accessible and available for use to the scientific community and other non-commercial groups. However, dissemination opportunities have changed drastically in the past decades permitting faster and more efficient forms of information delivery. Strategies adjusted to these opportunities led to the development of "A centralized and userfocused database containing only validated and authorized data with a known and registered accuracy and quality"; this system is now known as World Soil Information Service (WoSIS; Tempel et al., 2013). WoSIS forms a part of the Global Soil Information Facility (GSIF), ISRIC's overarching framework for production of gridded soil property maps for the world (Hengl et al., 2014). GSIF is a key component of ISRIC's evolving Spatial Data Infrastructure (SDI), through which quality-assessed data about soils can be made accessible and shared across disciplines to address global challenges such as climate change, food security, and the degradation of land and water resources (Batjes et al. 2013).

Testing of the initial version (WoSIS 1), which accommodated data for some 30,000 soil profiles, soon pointed at the need for improving several elements of the initial approach. The resulting revised procedures and database structure for WoSIS version 2.0, hereafter referred to as WoSIS 2, are described in this report.

WoSIS 2 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 and possible restrictions for use (licences). Subject to the terms of these licences, the available information will be served to third parties using standardized interfaces. In first instance, the focus has been on developing procedures for serving quality-assessed soil data for those properties mentioned in the GlobalSoilMap specifications (GSM, 2013).

This report consists of five Chapters and five Appendices. Following the Introduction, and prior to describing the new database structure (Chapter 3, with details in Appendix E), basic principles for flagging repeated (e.g. duplicate) soil profiles originating from disparate international databases, measures for defining data quality (i.e. level of trust, data quality rating, and accuracy), and the three main steps towards standardization and harmonization of numerical soil data are discussed in Chapter 2, with supplemental information provided in Appendix A, B and C. Conclusions and an outlook concerning future work are presented in Chapter 5. Appendix A describes basic principles for compiling soil profile data to facilitate entry into WoSIS; Appendix B focuses on quality aspects related to soil laboratory data, and Appendix C serves to provide some background information about analytical methods that have been considered in the scheme for standardizing soil analytical descriptions to a common (emerging) WoSIS standard (Appendix D). Finally, the (steadily growing number of) quality-assessed data managed in WoSIS will be served to the user community; aspects of interoperability and required web services are discussed in Chapter 4. The initial data is served via Web Features service (WFS)¹.

¹http://www.isric.org/data/wosis

Chapter 2

Basic principles for processing data

2.1 Flagging repeated profiles

One of the first tasks in the process of importing data into WoSIS 2 is the search for repeated profiles. This is necessary as the same profile may have been described in multiple source databases, albeit using different procedures and profile identifiers. Such a situation is likely to arise with stand-alone databases that have been developed for specific projects such as SoTER (Van Engelen et al., 2013) or the Africa Soil Profile Database (Leenaars et al., 2014b). The outcome of the screening process will be a unique set of soil profiles and thus produce a truthful profile count.

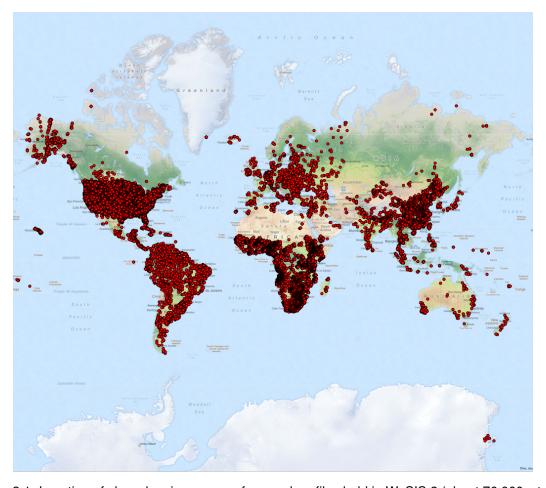


Figure 2.1: Location of shared, unique, geo-referenced profiles held in WoSIS 2 (about 76,000, status "received as is" i.e. prior to further standardization/harmonization).

Repeated profiles may be identified using various approaches. So far, two approaches or checks are applied in WoSIS: lineage and geographical proximity. The lineage check 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 suspiciously close to another (e.g. <10 m). Subsequently, the information for these profiles is compared and the database manager assesses the likelihood of such profiles being identical. In case of duplicates, the coordinates of the profile that originates from the most detailed source database are visualized (resp. shared) in WoSIS; presently, some 76,000 georeferenced profiles (Figure 2.1) out of a total of some 98,000 profiles. Inherently, the screening process is a very exhaustive and time consuming task as it cannot be automated fully; additional techniques for identifying possible duplicates are under investigation.

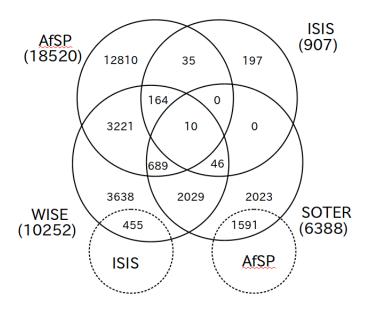


Figure 2.2: Intersection between ISRIC stand-alone profile databases showing the number of overlapping profiles (AfSP, Africa Soil Profile Database; ISIS, ISRIC Soil Information Service, SOTER, Soil and Terrain Database; WISE, World Inventory of Soil Emissions potentials database).

Figure 2.2 illustrates the intersection between profile databases compiled in the framework of collaborative ISRIC projects: AfSP (Africa Soil Profile Database; Leenaars et al., 2014a), ISIS (ISRIC Soil Information System)¹, WISE (World Inventory of Soil Emission potentials; Batjes, 2009) and various SOTERs (Soil and Terrain Databases)². Except for ISIS, which holds profile data for the ISRIC World Soil Reference Collection, the other datasets are project-specific compilations from various (possibly overlapping) data sources. As shown in Figure 2.2, 12,810 profiles are exclusively present in AfSP; 35 are shared among AfSP and ISIS; 164 are shared between AfSP, ISIS and WISE; 10 profiles are present in the 4 databases, and so on. Ultimately, in case of duplicate profiles, only the profile with the most complete data set and most detailed lineage will be shared in accord with the corresponding data licence.

2.2 Measures for data quality

2.2.1 General considerations

"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"

http://isis.isric.org/

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

(Chapman, 2005). As indicated, WoSIS is being populated using datasets produced for different types of studies ranging from routine soil surveys to more specific assessments. The corresponding samples were analysed in a range of laboratories or in the field according to a wide range of methods (e.g. wet chemistry or spectroscopy), each with their own uncertainty. As discussed by Kroll (2008), issues of soil data quality are not restricted to uncertainty issues, they also include aspects like completeness and accessibility of data. The systematic treatment of uncertainty in model-based decision support activities has been discussed by various researchers, including Walker et al. (2015) and Raupach et al. (2015). Quality aspects related to soil laboratory data are discussed in Appendix B to facilitate further contributions of data, by an anticipated growing number of international partners, to WoSIS. This in order to provide a sound and consistent basis for the subsequent standardization and harmonization stages being implemented in WoSIS (Section 2.3) and, ultimately, to allow servicing of those data using interoperable web services (Chapter 5).

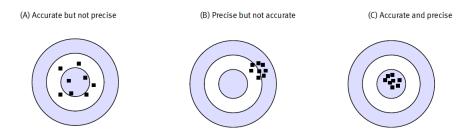


Figure 2.3: Depiction of the accuracy and precision of measurements (Credit: EH&E, 2001).

Quality of data can be evaluated against a range of properties, for example positional accuracy, attribute accuracy, logical consistency, completeness and lineage. Underlying these properties are always the two central themes in data quality assessment, the concepts of accuracy and precision. In the case of environmental data, accuracy can be defined as "the degree of correctness with which a measurement reflects the true value of the property being assessed", and precision as "the degree of variation in repeated measurements of the same quantity of a property" (EH&E, 2001). A high degree of precision and accuracy need not occur simultaneously in a process (Figure 2.3), thereby determining attribute uncertainty. When results are both precise and accurate (situation C), confidence in data quality is maximized. Similarly, differences between accuracy and precision in a positional context can be visualized (Figure 2.4; a red spot shows the true location, a black spot, represents the locations as reported by a collector). For point data, the aspect of positional accuracy, in the context of digital soil mapping, has been discussed in detail with respect to legacy soil profiles collated for the Africa Soil Profile Database (Leenaars et al., 2014a).

In order to address and document the above issues, quality indicators can be applied throughout the WoSIS database. These are:

- Observation date: date of observation or measurement,
- Level of Trust, a subjective measure based on soil expert knowledge (column: trust; see Section 2.2.2),
- Data Quality rating, based on expert judgement (column: quality; see Section 2.2.3),
- Accuracy, the Laboratory/Field/Location related uncertainty (column: accuracy; see Section 2.2.4).



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 (Credit: Chapman, 2005).

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

2.2.2 Level of trust

Different attributes held in the WoSIS database need to be characterized in terms of inferred trust. The lowest level 'A' is used for data entered "AS IS". Subsequently, such 'A' level data can be standardized ('B'), this step considers the soil property, method, and unit of measurement), respectively harmonized ('C') to an agreed reference method; these steps include error-checking for possible inconsistencies. Flagging data at level 'A' to 'C' can be performed automatically without any human interaction, while the final step to level 'D' requires certain skills and experience. Level 'D' data are those that have been approved by an expert (e.g., a soil scientist) who has performed an in-depth check, considering the value in relation to the full soil profile, the given natural conditions and the surrounding profiles and found no apparent anomalies.

2.2.3 Data quality rating

The data quality rating for a given record serves to characterize the (inferred) data quality in numerical space. Any new value entered in WoSIS initially receives a zero value, which is the lowest rank or rating value. Subsequently, this value may increase based on whether it passes specific tests. At present, the 0-4 range is considered adequate to characterize the data quality in numerical space.

2.2.4 Accuracy

Any given measurement has a specific measurement error, which can be determined with a variety of methods. The accuracy for values derived in a laboratory can be characterized using blind samples or based on repeated measurements on reference materials. Any laboratory should be able to provide these parameters according to good laboratory practice (OECD, 1998; Van Reeuwijk and Houba, 1998), but in practice this need not be the case. For measurements that use other devices, such as GPS and soil moisture sensors, the accuracy can be determined by obtaining the necessary information from manufacturers, literature and even expert knowledge. For example, the accuracy for GPS locations depends on several factors. Before the year 2000, a normal, commercial GPS had a nominal accuracy coarser than 100 m. Nowadays (anno 2015), the accuracy of commercial products can be less than 10 m; with the release of the Galileo system the accuracy will increase to below 1 m (ESA 2015).

2.3 Main steps towards data harmonization

2.3.1 Data lineage and access conditions

The aim of WoSIS is to facilitate the exchange and use of soil data shared within the context of collaborative efforts of institutes, organizations and individuals operating at global, regional, national and local level, and otherwise. Soil data providers over the world, however, have been and are generating soil data following numerous approaches and procedures in accord with their 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 the data for wider use may imply a loss of appropriateness for originally intended purposes. However, once compiled under a global common standard they importantly gain in appropriateness for use for international purposes.

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

Data providers who submit data for possible inclusion in WoSIS must specify conditions for access to the data they deposit. This may be done using a Creative Commons³ license or other existing licence. In practice, this information is provided as part of the data lineage (i.e. possible 'inherited restrictions'). Access conditions for third parties to each dataset managed in WoSIS are enforced through 'access registers'; overall conditions are in accord with the ISRIC Data Policy⁴.

2.3.2 Data standardization and harmonization

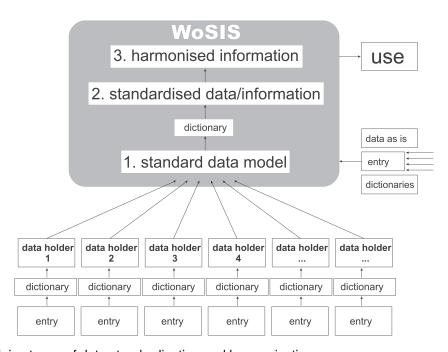


Figure 2.5: Main stages of data standardization and harmonization.

WoSIS 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), through any data entry template. The source soil datasets are imported 'as is', with the original naming and coding conventions, abbreviations, domains and so on; basically, this import is the first major step of data standardization (into a single data model). The second step of data standardization, applied to the values for the various soil properties as well as to the naming conventions etc., is required to make the data queryable and usable. A possible third step, not yet undertaken in this version of WoSIS, implies data harmonization. This refers to making the data comparable, as if assessed by a single given (reference) method. Soil datasets compiled following a few basic principles, as described by Leenaars et al. (2014b), meet the minimum criteria for being processed into and shared through WoSIS (Batjes et al., 2015).

Adoption of the above basic principles will permit to: a) keep track of data sources and identify uniqueness of profile records (through their lineage); b) understand the full (source) data so that they may be correctly collated into WoSIS (first step of standardization, including basic quality control); c) standardize the data according to common standard conventions, thus making the data queryable (second step of

³http://creativecommons.org/licenses/

⁴http://www.isric.org/data/data-policy

standardization, including routine quality control); d) harmonize the data, based on the laboratory or field methods originally applied, making the data comparable (third step of standardization, including full quality control).

The main soil data processing steps developed for WoSIS to generate globally, coherently harmonized soil profile data and derived soil information maps are schematized in Figure 2.5; additional information is provided in Appendix A.

Chapter 3

Database design

3.1 General concept

The database design for WoSIS 2 consists of 45 interrelated tables following a standard relational model implemented in PostgreSQL, a powerful, open source object-relational database system (PostgreSQL, 2015). An important development has been to reduce the number of tables (from 76 to 45) and schemas (from 15 to 1) in the database model in order to improve robustness and usability.

Every table has an unique identifier. Existing unique identifier/attributes were used as Primary key (Natural key). 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. In WoSIS 2, all these objects were renamed using the following 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 object's, 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>

Unlike for the preceding version, WoSIS 2 makes use of one single database schema to logically group objects such as tables, views, triggers and so on. Other schemas will be used for other purposes, such as Database Auditing and Web Applications, to enforce role grant access and use different tablespaces. This allows other systems such as GeoNetwork and GeoNode to run using the same database.

To better understand how the tables in WoSIS are related, they have been grouped according to their functions, as shown with different colours in Figure 3.1:

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

The above components are described in Sections 3.2.1 to 3.2.7; the structure of each table is documented in Appendix E. By convention, in the text, table names appear in **bold** and column names in *italic*.

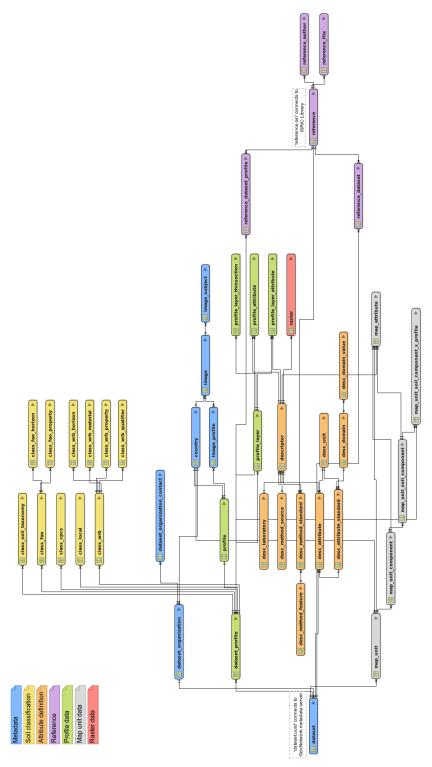


Figure 3.1: Main reference groups and components of the WoSIS 2 database model.

3.2 Main components

3.2.1 Metadata

GeoNetwork is a catalogue application for spatially referenced resources¹, providing metadata information and services. Metadata are data that define and describe other data. They also define the terms and conditions for use of the data and ensure that all data can be properly attributed and cited. 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). It is interoperable with standards used by the ICSU World Data System of which ISRIC is a regular member².

GeoNetwork is currently used by ISRIC; WoSIS links its **dataset** table with ISRIC's Geonetwork database³; Figure 3.2).

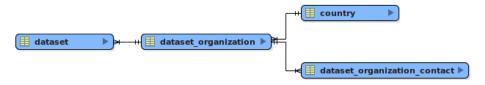


Figure 3.2: Metadata reference group tables

The **dataset** table stays at the top of the hierarchy of the database defining where the data come from. Most importantly, the **dataset** table is used to manage and enforce the access rights, for example whether the associated data may be shared freely with the general public or not; conditions for this are specified by each data provider in accordance with the ISRIC conditions for data use and citation⁴ and Data Policy⁵. It also makes a bridge, as mentioned before, to the Geonetwork database and links to table **dataset_organization** and **dataset_organization_contact**. The latter describe organizations and/or persons that have been instrumental in collating or providing the observation results (either descriptive or measured) that are stored in the database. It is the single entry point to authoritative names and contact information in the overall database. This is to prevent the use of different names or spellings for the same organization or individual in various parts of the database (e.g., KIT, Tropen Instituut, Royal Tropical Institute, Koninklijk Instituut voor de Tropen).

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

3.2.2 Soil classification

Soil classification involves the systematic categorization of soils based on distinguishing characteristics as well as criteria that dictate choices in use. It is probably one of the most controversial soil science subjects. Unlike plant taxonomy, there is no truly universally accepted classification system for soil, and the principles of soil systematics have been varied and polemical (Strzeminsky, 1975). Many countries have therefore developed their own classification systems (see FAO, 2015); international correlation

¹http://geonetwork-opensource.org/

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

³http://meta2.isric.org/geonetwork/

⁴http://www.isric.org/content/data-usage-and-citation

⁵http://www.isric.org/data/data-policy

⁶See http://www.fao.org/geonetwork/srv/en/metadata.show?id=12691

of the various systems is being addressed by the World Reference Base for Soil Resources (IUSS Working Group WRB, 2014) and earlier through the FAO-Unesco Soil Map of the World (FAO-Unesco 1974; FAO 1988).

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

- FAO Soil Map of the World: This system was originally intended as legend for the Soil Map of the World, 1:5M, but in the course of time it has been used increasingly as a classification system (FAO-Unesco, 1974; FAO 1988); the FAO system has now been subsumed into the WRB (tables class_fao_class_fao_horizon and class_fao_property).
- World Reference Base for Soil Resources: the international, standard soil classification system endorsed by the International Union of Soil Sciences (IUSS WG-WRB, 2014),and earlier versions as indicated by the year of publication (tables class_wrb, class_wrb_horizon, class_wrb_material, class_wrb_property and class_wrb_qualifier).
- USDA Soil Taxonomy (Soil Survey Staff, 2010), and earlier approximations as indicated by the year of publication (table **class_soil_taxonomy**).

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

Sometimes, as discussed earlier (Section 2.2.1), the same soil profile may have been considered/processed in different ISRIC datasets. As a result, this profile may have been classified/correlated differently in each source dataset, based on the same soil classification system, depending on the classifiers perspectives (Kauffman, 1987). Therefore, WoSIS contains a link dataset_profile table to assign profiles (profile table) to specific source datasets (dataset table). All classifications refer to an entry in the dataset_profile link table (that is, a profile in a particular dataset), thus enabling one classification per profile and dataset. In some cases, the USDA Soil Taxonomy coding is inconsistent between editions as different standard notations have been used in successive versions (e.g., Soil Survey Staff 1975,1992, 1998, 2003 and 2010); examples are given elsewhere (Spaargaren and Batjes, 1995). Alternatively, the original (FAO-Unesco, 1974) and revised Legend (FAO, 1988) to the FAO Soil Map of the World use a well-established coding scheme. However, there is no agreed coding scheme yet for the WRB Legend (IUSS WG-WRB, 2014), as WRB is not a hierarchical system. Therefore, to avoid any ambiguity in soil classification names, for any soil classification system full descriptive names are stored in the database, together with the edition (year) of the classification system.

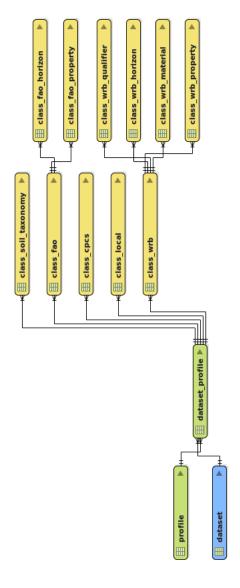


Figure 3.3: Classification tables

3.2.3 Attribute definition

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

Together with the attribute definition (**desc_attribute**), the analytical methods and the laboratory where the soil analyses have been carried out must be defined. Soil analytical methods, and their description, are a complex topic as many of these analyses are soil type specific (Soil Survey Staff, 2011; van Reeuwijk, 2006). Often they are poorly defined in the source materials; alternatively, the same analytical method may have different ways of expression. Therefore, the analytical method(s), as defined in the source materials, is preserved 'as is' in table **desc_method_source**.

Each standard method is described in table **desc_method_standard** where it is described by a number of standardized features, as documented in table **desc_method_feature**. Similarly, details about the laboratory where the measurement took place are stored in table **desc_laboratory**. Next in the data model is the table **descriptor** where the attribute, analytical method and laboratory ids are combined into a single id (*descriptor_id*). The *descriptor_id* is later used in tables such as **profile_attribute**, **profile_layer_attribute**, **map_attribute** and raster in which the measured respectively description values are stored.

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

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

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

In data management and database analysis, a data domain refers to all unique values that a data element may contain. The rule for determining the domain boundary may be as simple as a data type with an enumerated list of values. For example, a table that has information about soil drainage with one record per spatial soil feature might have a 'drainage class'. This class might 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, etc. The data domain for the drainage class is: V, P, I, M, W, S, E. Other datasets with information about soil drainage, however, 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 to link a phenomenon to more than one data domain is required. 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 its available data domains in desc_domain_value. Conversely, a data domain may be used to describe more than one characteristic. For example, in the FAO Guidelines for Soil Description (FAO, 2006), several surface characteristics are defined using the same surface coverage classes, ergo the same data domain.

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

Less simple domain boundary rules, when database-enforced, are implemented through a check constraint or, in more complex cases, a database trigger. For example, a column requiring positive numeric values

may have a check (i.e. validation) constraint declaring that the values must be greater than zero.

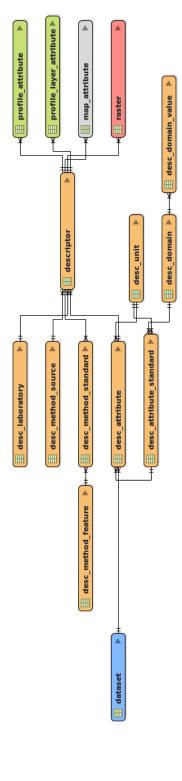


Figure 3.4: Attribute reference group tables

3.2.4 Source materials (Reference)

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

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

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

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

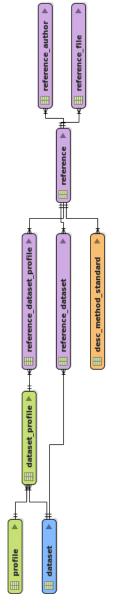


Figure 3.5: Reference data group tables

3.2.5 Profile data

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

Table **profile** holds unique soil profiles along with their geometry (x,y). The coordinates are stored in column geom, in binary mode, using the PostGIS⁷ spatial extension for PostgreSQL. The default coordinate system used in WoSIS is WGS84, EPSG code 4326. The accuracy of the profile coordinates is stored in column **geom_accuracy** in decimal degrees. Further, the country in which a profile is located is registered.

Each soil profile in WoSIS is given a specific integer ID as well as an UUID⁸ (universally unique identifier), 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. UUIDs allow for easy profile identification in diverse computer systems like harvesting environment, web services, broadcasting in social networks (e.g. Twitter, Facebook), and integration with GeoNetwork.

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

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

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

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

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

⁹Prior 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-ISRIC, 1986). Organic horizons were recorded as above and mineral horizons recorded as below, relative to the mineral surface (Soil Survey Staff, 2012 p. 2-6).

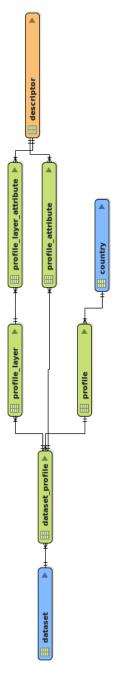


Figure 3.6: Profile data group tables

3.2.6 Map unit (polygon) data

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

In WoSIS, each map unit is stored as a single or multi polygon geometry. All polygon maps, and therefore their mapping units, are stored in a single table called **map_unit**. The *map_unit_id* identifies each individual map unit within the table. Hence, every map unit must refer to a dataset as uniquely defined in the **dataset** table. As indicated, in WoSIS, the reference datum for any point or polygon on the Earths surface is WGS84¹⁰.

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

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

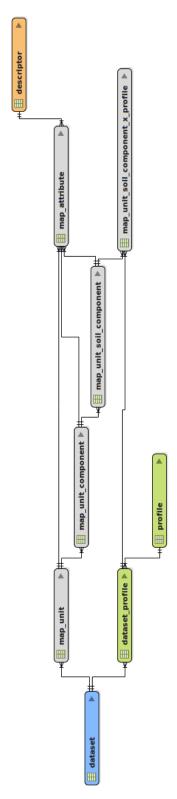


Figure 3.7: Map unit tables

3.2.7 Raster data

With the release of PostGIS 2.0 it has become possible to store raster data in PostgreSQL. Taking advantage of this improvement, WoSIS 2 can accommodate raster data such as those being produced by SoilGrids (Hengl et al., 2014; Hengel et al., 2015), the Africa Soil Information System (AfSIS), GlobalSoilMap (Arrouays et al., 2014) and other projects. The source of the product can be described in the **dataset** table, while **desc_attribute** describes the raster image or the attribute they represent (Figure 3.8). The **descriptor** is an intermediate table used to define the attribute plus the method and source laboratory using one single id (*descriptor_id*); the **raster** table is where all rasters are registered in the database.

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



Figure 3.8: Raster tables

¹¹http://spatialreference.org/ref/epsg/4326/.

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

Chapter 4

Interoperability and web services

Web services are interfaces that allow machine-to-machine communication where data, from various data holders, are exchanged and/or processed. This approach to communication and infrastructure between computers is normally referred to as a Service Oriented Architecture (SOA) approach/strategy, where multiple web services are orchestrated or choreographed with the objective of assembling bigger structures (e.g. SDI - Spatial Data Infrastructures). The web service interface normally follows predefined standards like REST (Representational State Transfer), OGC (Open Geospatial Consortium)¹ standards or WSDL (Web Services Description Language) / SOAP (Simple Object Access Protocol). The interfaces, but also the data being transmitted, have to follow precise data model standards such as XML (Extensible Markup Language) or GML (Geography Markup Language), JSON (JavaScript Object Notation) and SOA messages. The SOA approach can only be successful if data models and web services are compatible and easy to implement. Interoperability of the data exchanged or processed by the web services is achieved through a priori standardization of the data themselves (see Section 2.3); the latter is done according to agreed upon data conventions which express the (soil) data in a (machine) understandable 'soil-vocabulary'.

WoSIS 2 can store multiple soil data types and sources. For this, the soil data have first to be modelled respecting the schema, tables and relationships of the WoSIS 2 structure. Standardized data (i.e. known modelled data) are of extreme importance since web services have to translate the database data model into a form more compatible with web communication, such as XML/GML and JSON. First, a Web Feature Service (WFS) has been implemented using Geoserver that connects to WoSIS 2 reading its views and tables, see http://www.isric.org/data/wosis. Subsequently, a REST web-service will be developed; this will require a more custom-made solution based on programming the needs of the REST clients to the data in the database.

The client's web services consuming the data are totally independent from WoSIS 2, as these clients are located in a very broad range range of platforms, from mobile phones to GIS softwares. The client web services are mainly oriented as data providers to humans (or systems being used by humans).

The approach of using OGC web services and model data in XML is necessary for fulfilment of INSPIRE requirements (GS Soil 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.

The 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 (Figure 4.1). The request objectives can be:

- Determine capabilities of the providing service,
- · Get data based on query,
- Submit data from the client into the provider (the client becoming himself a provider).

 $^{^{1} \}verb|http://www.opengeospatial.org/standards/common.$

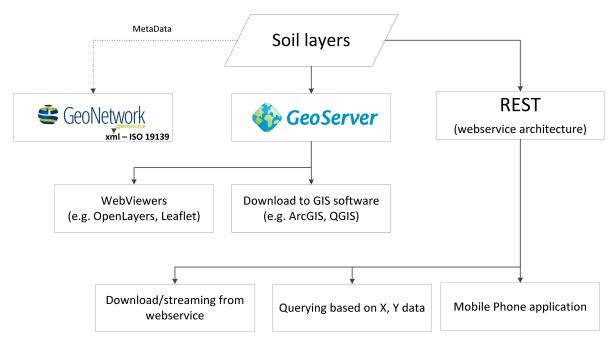


Figure 4.1: Serving soil layers from WoSIS 2 to the user community.

The possibility of the client web-service becoming a data provider to WoSIS 2 is of extreme importance for future crowdsourced projects, where users can contribute data into the database. Once implemented at a satisfactory level of detail and authority, this will add value to WoSIS as a quality-assessed, queryable source of soil terms and analytical procedures for soil properties.

Chapter 5

Future developments

During 2015, the following activities were undertaken by the WoSIS team:

- The database structure and technical documentation for WoSIS have been updated.
- Some 98,000 profiles from disparate soil databases have been imported into WoSIS 2; some 76,000 of these are georeferenced within defined limits.
- Special attention has been paid to the standardization of soil analytical method descriptions with focus on the set of nine soil properties considered in the GlobalSoilMap specifications.
- Newly developed procedures for the above, that consider the soil property, analytical method and unit of measurement, have been applied to the present set of geo-referenced soil profile data.
- The resulting standardized datasets will be made available through web services¹ hosted by ISRIC World Soil Information, pending the release of ISRIC's evolving GeoNode facility.

In view of its global scope, WoSIS will always remain 'work in progress'. For the future, the following activities will be considered gradually (as realistic within the allocated project time):

- Expand the number of procedures for which standardized soil analytical method descriptions are developed, in first instance working towards the list of eightteen soil properties considered in recent WISE-derived soil property databases².
- Development of procedures for handling soil profile data derived from spectral analysis.
- Processing of 'new' soil profile datasets into WoSIS when such are shared by new data providers (in principle in order of receipt of the various datasets/permissions, with priority for 'fully shared' datasets from so far under-represented regions).
- Add map unit based soil datasets to WoSIS, starting with those derived from ISRIC-related projects.
- Distribute WoSIS-derived data through ISRIC's upcoming GeoNode facility.
- Develop a facility to upload 'raw' soil data into WoSIS in accord with SoilnfoApp³ developments.
- The documentation for WoSIS will be regularly expanded and made available online as PDF's with clear time stamps.

Capacity building and collaboration with (inter)national soil institutes around the world on data collection and sharing, data screening, standardization and harmonization as well as mapping and the subsequent dissemination of the derived information will be essential to create ownership of the newly standardized / harmonized soil information as well as to create the necessary expertise and capacity to further develop and test the system worldwide.

http://www.isric.org/data/wosis

²http://www.isric.org/sites/default/files/ISRIC_Report_2015_01.pdf.

³http://soilinfo.isric.org/

Acknowledgements

The authors would like to thank the following experts that have contributed in various ways to the technical development of WoSIS since its conceptualization in 2010, in particular: Piet Tempel (2009-2012), Koos Dijkshoorn (2010-2012), Vincent van Engelen (2010-2012), Bob MacMillan (2010-2013), Hannes I. Reuter (2011-2013), Daniel van Kraalingen (2010-2012), and Rob Lokers (2010-2012); ISRIC Directors Prem Bindraban (2009-2013), Hein van Holsteijn (2013) and Rik van den Bosch (2014-present) were instrumental in these developments. Alessandro Samuel-Rosa, Bas Kempen, Ingrid Haas, Marcos Angelini, Thomas Caspari, Tom Hengl, and Piet van Reeuwijk provided expert advice on specific WoSIS-related topics for which they are gratefully thanked. We also thank our colleague Gerard Heuvelink for his thorough review of an earlier version of this document.

Importantly, our special thanks go to the steadily growing number of data providers, including soil survey organizations, research institutes and individual experts that have contributed data for consideration in WoSIS⁴:

- Africa Soil Profiles Database (for contributors⁵ see: http://www.isric.org/content/africa-soil-profiles-database)
- Food and Agriculture Organization of the United Nations (FAO): various inputs through SOTER and WISE-related projects (see: http://www.fao.org/3/a-i3161e.pdf).
- National Cooperative Soil Survey, National Cooperative Soil Characterization Database (NCSS; see: http://ncsslabdatamart.sc.egov.usda.gov/)
- National Soil Profile database for Canada, Canadian Soil Information Service (CanSIS; see: http://wms1.agr.gc.ca/xgeng/pedon.zip)
- International Soil Carbon Network (ISCN, see http://iscn.fluxdata.org/Pages/default.aspx)
- ISRIC Soil Information System (ISIS, for contributors see: http://www.isric.org/projects/establishment-national-soil-reference-collections-nasrec)
- ISRIC World Inventory of Soil Emission Potentials (WISE) soil profile database (for contributors see: http://www.isric.org/content/cooperating-institutions-and-experts-wise)
- SOTER-AR, Soil and Terrain Database for Argentina (for contributors see http://www.isric.org/projects/soil-and-terrain-database-soter-argentina)
- SOTER-CN, Soil and Terrain Database for China (for contributors, see http://www.isric.org/projects/soter-china)
- SOTER-TN, Soil and Terrain Database for Tunisia (for contributors see http://www.isric.org/projects/soter-tunisia)
- SOTER-SAF, Soil and Terrain Database for Southern Africa (for contributors see: http://www.isric.org/projects/soter-southern-africa-sotersaf)
- SOTER-CAF, Soil and Terrain Database for Central Africa (for contributors see: http://www.isric.org/projects/soter-central-africa-sotercaf)

⁴Some of these datasets will be incorporated in the next release of the WoSIS database.

⁵Names of collaborating institutions and individual experts may be found in the Acknowledgement section of the Technical Reports that accompany the various databases.

- SOTER-SN&GM, Soil and Terrain Database for Senegal and the Gambia (for contributors see: http://www.isric.org/projects/soter-senegal-and-gambia)
- SOTER-CU, Soil and Terrain Database for Cuba (for contributors see: http://www.isric.org/projects/soter-cuba)
- SOTER-EUR, Soil and Terrain Database for Central and Eastern Europe (for contributors see: http://www.isric.org/content/soveur-partners)
- SOTER-ZA, Soil and Terrain Database for South Africa (for contributors see: http://www.isric.org/projects/soter-south-africa)
- SOTER-KE, Soil and terrain database for Kenya (for contributors see: http://www.isric.org/projects/soter-kenya-kensoter)
- SOTER-KET, Soil and Terrain Database for Upper Tana River Catchment (for contributors see http://www.isric.org/projects/soil-and-terrain-soter-database-upper-tana-kenya)
- SOTER-MW, Soil and Terrain Database for Malawi (for contributors see: http://www.isric.org/data/soil-and-terrain-database-malawi-ver-10-sotermalawi)
- SOTER-NP; Soil and Terrain Database for Nepal (for contributors see: http://www.isric.org/projects/soter-nepal)
- SOTER-LAC, Soil and Terrain Database for Latin America and the Caribbean (for contributors see: http://www.isric.org/projects/soter-latin-america-and-caribbean-soterlac)
- United Nations Environment Programme (UNEP, various inputs mainly through regional SOTER

To our regret, at this stage, we can only list the major data contributors here. In the near future, however, the full overview of contributors will be made available through the WoSIS-related web page.

activities)

Bibliography

Arrouays D, Grundy MG, Hartemink AE, Hempel JW, Heuvelink GBM, Hong SY, Lagacherie P, Lelyk G, McBratney AB, McKenzie NJ, Mendonca-Santos M de L, Minasny B, Montanarella L, Odeh IOA, Sanchez PA, Thompson JA, Zhang G-L and Donald LS, 2014. *GlobalSoilMap: Toward a Fine-Resolution Global Grid of Soil Properties*. Advances in Agronomy 125, 93-13

Baritz R, Erdogan H, Fujii K, Takata Y, Nocita M, Bussian B, Batjes NH, 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)*. Global Soil Partnership, FAO, 44 p. http://www.fao.org/fileadmin/user_upload/GSP/docs/ITPS_Pillars/annexV_pillar5.pdf

Batjes NH, 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 NH, Bindraban PS and Reuter HI, 2013. *Towards improved soil information for quantification of environmental, societal and economic sustainability*. Report 2013/05, ISRIC - World Soil Information, Wageningen, 44 p. http://dx.doi.org/10.13140/2.1.4831.0560

Batjes NH, Mendes de Jesus J, Heuvelink GBM, Ribeiro E, Kempen B, Leenaars JGB, Hengl T, Ruiperez Gonzalez M, Van Oostrum A and Van den Bosch R, 2015. *New gridded datasets for global sustainability studies (WISE30sec and SoilGrids)*. In: Keestra S, G Mol, A Zaal, J Wallinga and B Jansen (editors), Book of Abstracts, Wageningen Soil Conference 2015 'Soil Science in a Changing World' (23-27 Aug. 2015), Wageningen, pp 159. http://dx.doi.org/10.13140/RG.2.1.1225.8009

Blakemore LC, Searle PL and Daly BK, 1987. *Methods for chemical analysis of soils*. New Zealand Soil Bureau Scientific Report 80, Wellington.

Botula YD, Van Ranst E and Cornelis WM, 2014. *Pedotransfer functions to predict water retention for soils of the humid tropics: A review.* Revista Brasileira de Ciencia do Solo 38, 679-698.

Ciampalini R, Lagacherie P, Gomez C, Grünberger O, Hamrouni MH, 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.

Chapman AD, 2005. *Principles of Data Quality, version 1.0. Global Biodiversity Information Facility.* Copenhagen, 58 p. http://www.gbif.org/orc/?doc_id=1229

De Gruijter JJ, Brus DJ, Bierkens MFP and Knottters M (editors), 2006. *Sampling for natural resource monitoring*. Springer, Heidelberg, 338 p.

EH&E, 2001. The precision and accuracy of environmental measurements for the building assessment survey and evaluation program. EH&E Report #11995, Environmental Health and Engineering Inc. (EH&E) for US Environmental Protection Agency, Washington DC, 25 p. http://www.epa.gov/iaq/base/pdfs/precisionandaccuracy.pdf

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

ESA, 2015. *Galileo - Europe's own global navigation satellite system*. European Space Agency. http://www.esa.int/Our_Activities/Navigation/The_future_-_Galileo/What_is_Galileo

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. https://www.eurachem.org/

FAO, 1988. *FAO-Unesco Soil Map of the World, Revised Legend, with corrections and updates.* World Soil Resources Report 60, FAO, Rome; Reprinted with updates as Technical Paper 20 by ISRIC, Wageningen, 1997, pp. 140.

FAO, 2006. *Guidelines for soil description (Fourth ed.)*. FAO, Rome, pp. 97. ftp://ftp.fao.org/agl/agll/docs/guidel_soil_descr.pdf

FAO, 2008 Motsara MR and Roy RN, 2008. *Guide to laboratory establishment for plant nutrient analysis*. Fertilizer and Plant Nutrition Bulletin 19. Food and Agriculture Organization of the United Nations. Rome.

FAO, 2015. Soil classification. FAO Soils Portal. http://www.fao.org/soils-portal/soil-survey/soil-classification/en/

FAO-ISRIC, 1986. Guidelines for soil description (3rd Edition, Rev.). FAO, Rome, 70 p.

FAO-Unesco, 1974. *Soil Map of the World, 1:5,000,000.* Vol. 1 Legend. United Nations Educational, Scientific, and Cultural Organization, Paris.

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

GS Soil, 2008. Discover INSPIRE compliant harmonized soil data and services. GS Soil 06/2009 -05-2012, GS Soil Consortium, Paris-Lodron University Salzburg, 60 p. http://www.bgr.bund.de/DE/Themen/Boden/Projekte/Informationsgrundlagen-abgeschlossen/GSSoil/Bod_GSSoil_Broschuere.pdf?__blob=publicationFile&v=2

Goodchild MF, 1994. *Sharing imperfect data*. In: Singh A (editor), Proceedings UNEP and IUFPRO international workshop in cooperation with FAO in developing large environmental databases for sustainable development. UNEP/GRID, Sioux Falls, pp 102-11.

Goodchild MF and Gopal S, 1989. *The accuracy of spatial databases*. Taylor & Francis, London, 312 n

Hengl T, Mendes de Jesus J, MacMillan RA, Batjes NH, Heuvelink GBM, Ribeiro E, Samuel-Rosa A, Kempen B, Leenaars JGB, Walsh MG, Ruiperez Gonzalez M, 2014. SoilGrids1km - Global Soil Information Based on Automated Mapping. PLoS ONE 9(8): e105992. doi:10.1371/journal.pone.0105992.

Hengl T, Heuvelink GBM, Kempen B, Leenaars JGB, Walsh MG, Shepherd KD, Sila A, MacMillan RA, Mendes de Jesus J, Tamene L and Tondoh JE, 2015. *Mapping Soil Properties of Africa at 250 m Resolution: Random Forests Significantly Improve Current Predictions*. PLoS ONE 10, e0125814, doi.org: 10.1371/journal.pone.0125814.

Heuvelink GBM, 1996. *Identification of field attribute error under different models of spatial.* International Journal of Geographical Information Systems 10, 921-935

Heuvelink GBM, 2014. *Uncertainty quantification of GlobalSoilMap products*. In: Arrouays N, McKenzie N, Hempel J, Richer de Forges A, and McBratney A (eds.), GlobalSoilMap. Basis of the Global Spatial Soil Information System, pp. 335-340.

INSPIRE, 2015. Data Specifications - Infrastructure for spatial information in the European Community (INSPIRE). European Commission Joint Research Centre. http://inspire.ec.europa.eu/index.cfm/pageid/2

ISO, 2015. Soil quality - Determination of pH (ISO 10390:2005). International Organization for standardization. http://www.iso.org/iso/catalogue_detail.htm?csnumber=40879

IUSS Working Group WRB, 2006. *World Reference Base for soil resources - A framework for international classification, correlation and communication*. World Soil Resources Reports 103, International Union of Soil Sciences, ISRIC - World Soil Information and Food and Agriculture

Organization of the United Nations, Rome, pp. 128. ftp://ftp.fao.org/agl/agll/docs/wsrr103e.pdf

IUSS Working Group WRB, 2014. World Reference Base for soil resources 2014 - International soil classification system for naming soils and creating legends for soil maps. World Soil Resources Reports 106, Global Soil Partnership, International Union of Soil Sciences, and Food and Agriculture Organization of the United Nations, Rome, 182 pp. http://www.fao.org/3/a-i3794e.pdf

IUSS WG-SIS, 2015. *IUSS Working Group on Soil Information Standards*. International Union of Soil Sciences (IUSS). http://www.soilinformationstandards.org/

Kauffman JH, 1987. Comparative classification of some deep, well-drained red clay soils of Mozambique. Technical Paper 16, International Soil Reference and Information Centre (ISRIC), pp. 68. Wageningen. http://www.isric.org/isric/webdocs/Docs/ISRIC_TechPap16.pdf

Ketterings, QM, Czymmek KJ, Reid WS and Wildman RF, 2002. *Conversion of Modified Morgan and Mehlich III Soil Tests To Morgan Soil Test Values*. Soil Science 167(12), 830-837.

Kroll BGCM, 2008. *Towards a data management framework for digital soil mapping with limited data.* In: Hartemink AE, A McBratney and ML Mendoca Santos (editors), Digital soil mapping with limited data. Springer, pp 137-149.

Leenaars JGB, Kempen B, van Oostrum AJM and Batjes NH, 2014a. *Africa Soil Profiles Database: a compilation of georeferenced and standardised legacy soil profile data for Sub-Saharan Africa*. In: Arrouays D, N McKenzie, J Hempel, AR de Forge and A McBratney, (editors), GlobalSoilMap - Basis of the global spatial information system. Taylor & Francis Group, Orléans pp 51-57. 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 JGB, van Oostrum AJM and Ruiperez Gonzalez M, 2014b. *Africa Soil Profiles Database Version 1.2. A compilation of georeferenced and standardised legacy soil profile data for Sub-Saharan Africa (with dataset).* ISRIC Report 2014/01. Africa Soil Information Service (AfSIS) project. ISRIC- World Soil Information, Wageningen. http://www.isric.org/data/africa-soil-profiles-database-version-01-2

Louis BP, Saby NPA, Orton TG, Lacarce E, Boulonne L, Jolivet C, Rati C and Arrouays D, 2014. Statistical sampling design impact on predictive quality of harmonization functions between soil monitoring networks. Geoderma 213, 133-143.

Mallarino P, 1995. Comparison of Mehlich-3, Olsen, and Bray-P1 procedures for phosphorus in calcareous soils. In: Rehm G (editor), Proceedings of the twenty-fifth North Central extension-industry soil fertility conference. Potash & Phosphate Institute, Manhattan (KS) pp 96-101. http://www.agronext.iastate.edu/soilfertility/info/ComparisonofMehlich-301senandBray-P1Procedures.pdf

Michéli E, Lng V, Owens PR, 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, doi: 10.1016/j.geoderma.2015.09.008.

NATP, 2015. North American Proficiency Testing (NAPT) Program. http://www.naptprogram.org/

NCSS, 2010. National Cooperative Soil Survey. National Cooperative Soil Characterization Database. http://ssldata.nrcs.usda.gov/

OECD, 1998. *OECD Principles of Good Laboratory Practice (as revised in 1997)*. OECD Environmental Health and Safety Publications.

OGC, 2013a. Welcome to the OGC Website, Open Geospatial Consortium. Viewed April 2015. http://www.opengeospatial.org

OGC, 2013b. OGC Abstract Specification Geographic information - Observations and measurements (ver. 2.0). Open Geospatial Consortium. https://portal.opengeospatial.org/files/?artifact_id=41579

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?. FAO, Italy, 69 p. http://typo3.fao.org/fileadmin/templates/nr/kagera/Documents/Soil_information_Report.pdf

Pansu M and Gautheyrou J, 2006. *Handbook of Soil Analysis: Mineralogical, Organic and Inorganic Methods*. Springer, 993 p.

PostgresSQL, 2015. PostgreSQL 9.2.13 Documentation, The PostgreSQL Global Development Group. http://www.postgresql.org/docs/9.2/static/index.html

Raupach MR, Rayner PJ, Barrett DJ, DeFries RS, Heimann M, Ojima DS, Quegan S and Schmullius CC, 2005. *Model-data synthesis in terrestrial carbon observation: methods, data requirements and data uncertainty specifications*. Global Change Biology 11, 378-397.

Ross DS, Bailey SW, Briggs RD, Curry J, Fernandez IJ, Fredriksen G, Goodale CL, Hazlett PW, Heine PR, Johnson CE, Larson JT, Lawrence GB, Kolka RK, Ouimet R, Par D, Richter DD, Schirmer CD and Warby RA, 2015. *Inter-laboratory variation in the chemical analysis of acidic forest soil reference samples from eastern North America*. Ecosphere 6, art. 73.

Soil Survey Staff, 1975. *Soil Taxonomy: A basic system of soil classification for making and interpreting soil surveys.* Soil Conservation Service, USDA Agriculture Handbook 436, USDA, Washington.

Soil Survey Staff, 1983. *Soil Survey Staff 1983. Soil Survey Manual (rev. ed.)*. United States Agriculture Handbook 18, USDA, Washington.

Soil Survey Staff, 1992. *Keys to Soil Taxonomy (5th. ed.)*. 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 Staff, 1998. *Keys to Soil Taxonomy (8th. ed.)*. United States Department of Agriculture, Natural Resources Conservation Service, Blacksburg (Virginia), 326 p.

Soil Survey Staff, 2003. *Keys to Soil Taxonomy (9th. ed.)*. United States Department of Agriculture, Natural Resources Conservation Service, Blacksburg (Virginia), 332 p.

Soil Survey Staff, 2010. Keys to Soil Taxonomy, (11th ed.). USDA-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 Staff, 2011. Soil Survey Laboratory Information Manual (Ver. 2.0). Soil Survey Investigation Report No. 45, National Soil Survey Center, Soil Survey Laboratory, USDA-NRCS, Lincoln (NE), 506 p ftp://ftp-fc.sc.egov.usda.gov/NSSC/Lab_Info_Manual/SSIR_45.pdf

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

Spaargaren OC and Batjes NH, 1995. Report on the classification into FAO-Unesco soil units of profiles selected from the NRCS pedon database for IGBP-DIS. Work. Pap. 95/01, ISRIC, Wageningen, 15 p. http://www.isric.org/sites/all/modules/pubdlcnt/pubdlcnt.php?file=/isric/webdocs/docs/ISRIC_Report_1995-01.pdf&nid=332

Strzeminsky P, 1975. *Ideas Underlying Soil Systematics*. Foreign Scientific Publication, Department of the National Centre for Scientific, Technical and Economic Information, Warsaw, Poland.

Tempel P, van Kraalingen D, Mendes de Jesus J, Reuter HI, 2013. *Towards an ISRIC World Soil Information Service (WoSIS Version 1.0)*. Report 2013/02, ISRIC World Soil Information, Wageningen, 207 pp. http://www.isric.org/sites/default/files/ISRIC_Report_2013_02.pdf

US-EPA, 2015. *Quality Management Tools - Best Practices for Laboratory Quality Systems*. U.S. Envrionmental Protection Agency (US-EPA). http://www.epa.gov/quality/bestlabs.html

Van Engelen VWP, 2011. Standardizing soil data (e-SOTER regional pilot platform as EU contribution to a Global Soil Information System). International Innovation, June, 48-49.

van Engelen VWP and Dijkshoorn JA, 2013. *Global and National Soils and Terrain Digital Databases (SOTER) - Procedures manual (Ver. 2.0).* ISRIC Report 2013/04, IUSS, ISRIC and FAO, Wageningen, 191 p. http://www.isric.org/sites/default/files/ISRIC_Report_2012-04_TOTAL_LR-SEC.pdf

van Reeuwijk LP, 1998. Guidelines for quality management in soil and plant laboratories. FAO Soil Bulletin 74, ISRIC and FAO, Rome, 143 p. http://www.fao.org/docrep/W7295E/W7295E00.htm

van Reeuwijk LP and Houba VJG, 1998. *External quality control of data*. In: Van Reeuwijk LP (editor), Guidelines for quality management in soil and plant laboratories. ISRIC and FAO, Rome, pp 121-127. http://www.fao.org/docrep/W7295E/W7295E00.htm

van Reeuwijk LP, 2002. *Procedures for soil analysis (6th ed.)*. Technical Paper 9, ISRIC, Wageningen, 1.1 to 19.5 p. http://www.isric.org/Isric/Webdocs/Docs/ISRIC_TechPap09_2002.pdf

Walker WE, Harremoes P, Rotmans J, Sluijs JP van der, Asselt MBA van, Janssen P and Krayer von Krauss MP, 2005. *Defining Uncertainty: A Conceptual Basis for Uncertainty Management in Model-Based Decision Support.* Integrated Assessment 4,5-17.

WEPAL, 2015. Wageningen Evaluating Programmes for Analytical Laboratories. http://www.wepal.nl/website/about_wepal/History.htm

WHO, 2011. Laboratory quality management system: handbook. World Health Organization Lyon (FR), 247 p. http://whqlibdoc.who.int/publications/2011/9789241548274_eng.pdf?ua=1

Appendix A

Basic principles for compiling a soil profile dataset

To be considered in WoSIS 2, a soil dataset should include data for commonly required soil properties (FAO, 2006; Soil Survey Staff, 2012; Van Engelen and Dijkshoorn, 2013), but no minimum dataset is prescribed. However, sufficient information (metadata) should be provided to assess the source and quality of the data as well as the access category (licence). A dictionary table describing the meaning of all (often abbreviated) column headings used in the dataset tables should be provided with the metadata. Similarly, the use of dictionary tables is recommended for describing all coded data entries, like ID's as well as any abbreviated descriptive soil property value (e.g. 'W' means 'well drained').

Soil records are considered complete and thus processable into WoSIS when: 1) the lineage¹ of the soil record is well specified and 2) the soil data are consistently expressed as the result of observations and measurements (O&M). A single soil profile record can be considered as a collection of observations and measurements, with similar lineage, applied to the soil (to the soil profile site, to the profile itself, and to the profile depth intervals (e.g. horizons or layers)), by using (defined) field and laboratory methods to assess values for the soil properties or attributes concerned. Those values, either numeric, categorical or descriptive, are expressed according to the associated value domain as dictated by the references used for defining units of expression or pick lists. (Note that the profile itself is a depth interval spanning the layer or horizon depth intervals).

The following should be specified for each soil record:

- Soil profile record ID. The identifier for each soil profile record should be unique for the given dataset.
- 2. Lineage ID. The lineage ID relates to a corresponding dictionary which gives full reference to the dataset itself and to all underlying original data sources (e.g. papers, reports), if any, preferably also including (an URL to) a copy of such original data source so that these may be added to the ISRIC library and metadata catalogue.
- 3. Soil profile data consistently given as the results of observations and measurements (O&M). Here, an observation is the outcome of an "act of measuring or otherwise determining the value of a property", while a measurement is the outcome of a "set of operations having the object of determining the value of a quantity" (OGC, 2013b):
 - (a) What piece of soil is observed or measured? This is the soil feature. Define the position of the soil in 4D space:
 - x-y location (lon-lat coordinates),
 - z depth interval (top and bottom, in cm) of either the whole soil profile² or an associated soil profile layer, and

https://en.wikipedia.org/wiki/Data_lineage

²If the bedrock or an impenetrable layer is observed, this deepest layer (R horizon) should be specified in the dataset to make the observation explicit.

- t moment in time (year and possibly month, day).
- (b) What soil property is observed or measured? Specify the soil properties or soil property IDs P1, P2, P3, etc. Relate the ID's to a corresponding dictionary giving property descriptions.
- (c) What method is used to observe or measure the soil-property? Specify the methods or method IDs M1, M2, M3, etc. Relate the ID's to a corresponding dictionary giving method descriptions. Note: When possible also give the laboratory name and the laboratory manual, or field manual (as PDF or URL).
- (d) What value is observed or measured by the method applied to the soil? Specify the values (numeric) or value IDs V1, V2, V3, etc. (descriptive). Relate the values to the corresponding value domains, as described under e).
- (e) What value domain applies to the given property values? Specify the units of expression (for numeric property-values). Specify the value domain IDs VD1, VD2, VD3, etc. (for descriptive values). Relate the ID's to a corresponding dictionary giving domain descriptions (references) and relate the value ID / value domain ID combinations to a (same) dictionary giving the value descriptions (value pick list).

In addition to the above, soil profile data need to be prepared towards standardization. For this, the original soil profile data given following O&M conventions and including lineage and value domain (according to step 3) are, together with metadata, presented in the form of dictionaries where necessary, adequate for being converted to standardized soil profile data meeting the (emerging) WoSIS standard data conventions. This standardization will apply to all O&M data, thus including:

- lineage ID,
- feature/record ID,
- · attribute ID.
- · method ID,
- value,
- · value ID and,
- value domain ID.

Such work will take into consideration ongoing developments within Pillar 5 of the Global Soil Partnership (Baritz et al, 2014) and related activities of the IUSS Working Group on Soil Information Standards (IUSS WG-SIS, 2015).

As indicated, sharing soil data for distribution through WoSIS does not require the use of a specific data entry template with a priori standards nor is there a minimum dataset size. A soil dataset is considered complete and thus processable when compiled according to a few basic principles, as described above, which will permit compilation of the source data under the common WoSIS standard. Basically, these guiding principles imply sharing of a simple dataset conventionally build up or otherwise.

A suggested template for small dataset submission, which may also serve as a potential template for web automated dataset upload, is presented below. It consists of 1 spreadsheet with 4 sheets (**dataset**, **profile**, **horizon** and **metadata**)³. As a general rule, sheet and column names should not contain:

- · diacritical marks
- symbols
- spaces
- upper-case characters, and
- not start with a number.

Sheet dataset, has 17 rows to fill in the following data:

• dataset_id - Dataset code, this will be the dataset_id in the WoSIS database.

 $^{{\}it ^3} Templates for this may be downloaded from: LibreOffice - http://www.isric.org/sites/default/files/template.ods; \\ MicrosofOffice http://www.isric.org/sites/default/files/template.xlsx$

- dataset_title Dataset title, project or thesis title.
- dataset_description Description of the dataset.
- publication_date Publication date.
- dataset_version Dataset version.
- dataset_license Access and use constraints of the dataset.
- organization_name Organization name.
- organization_country Organization country.
- organization_city Organization city.
- organization_postal_code Organization postal code.
- organization_street_name Organization street name.
- organization_street_number Organization street number.
- author_first_name Author first name.
- author_last_name Author last name.
- author_email Author email.
- dataset_reference_1 Reference to a document or scientific paper. A DOI is preferred, but URLs may also be used.
- dataset_reference_x Another reference, in so far as needed just increment the last number of the column name and provide the DOI or URL.

Any further dataset description can follow after these rows.

Sheet **profile** should start with the following columns:

- profile_id Unique identifier of the profile as used in the source dataset.
- *observation_date* Date of the observation in format (yyyy-mm-dd).
- coordinate_system Coordinate system used. Please indicate the correspondent EPSG code (e.g. WGS 84, EPSG: 4326).
- *x*_coord X coordinate, if in geographic coordinates (degrees), the same as Longitude.
- *y_coord* Y coordinate, if in geographic coordinates (degrees), the same as Latitude.
- gps_measurement_date Date of the GPS measurement.
- gps_measurement_time Time of the GPS measurement.
- gps_accuracy GPS error in meters.
- accuracy_type 'GPS' if coordinates come from GPS or 'MAP' if coordinates come from a map.
- country_id ISO 3166-1 alpha-2 country code.
- sample_type Either 'single' or 'composite' sample.
- sample_number Number of samples. One when the previous is 'single' or more for 'composite'. Express the value with integers.
- sample_area Area sampled (m^2) .
- soil_classification_system_name The soil classification system used to classify the profile.
- soil_classification_system_year The publication year of the soil classification system used.
- profile_soil_classification_name The name of the profile classification.
- profile_soil_classification_code The code of the profile classification.

Any other profile description attribute can follow after these columns. Each row should contain a different profile description.

Sheet horizon should start with following columns:

- profile_id Unique identifier of the profile. Refers to sheet 'profile'.
- horizon_number Consecutive layer number rated from top to bottom.
- horizon_name Horizon name. A, B, etc.
- sample_code Laboratory code of the sample.
- upper_depth Depth of upper horizon boundary (cm).
- lower_depth Depth of lower horizon boundary (cm).

Any other horizon description attribute can follow after these columns (e.g. ph_h2o, ph_kcl, clay, silt, sand, ...). Each row should contain a different combination of profile and horizon descriptions.

Sheet **metadata** serves to describe all the attributes (except for the default columns) in profile.csv and horizon.csv. It should start with the following columns:

- sheet_name Either 'profile' or 'horizon'.
- column_name The exact name of the column added after the default ones.
- attribute_name Name of the attribute.
- attribute_description Description of the attribute.
- attribute_unit Units used, if not used, 'unitless' should be indicated.
- attribute_data_type Data type, use one out of ('Array', 'Binary', 'Boolean', 'Date', 'Integer', 'Real', 'Text', 'Time').
- analytical_method Analytical method used in the laboratory, if none is used enter 'Not applicable'.
- laboratory_name Laboratory name.
- laboratory_country Laboratory country.
- laboratory_city Laboratory city.
- laboratory_postal_code Laboratory postal code.
- *laboratory_street_name* Laboratory street name.
- laboratory_street_number Laboratory street number.

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

Appendix B

Quality aspects related to laboratory data

B.1 Context

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

ISRIC, for example, published reference procedures for soil analysis as a step towards standardization of analytical methods in soil laboratories (van Reeuwijk, 2002). These procedures cover the range of analytical methods required for soil characterization according to the Revised FAO Legend (FAO, 1988) and the World Reference Base (IUSS Working Group WRB, 2014). The Natural Resources Conservation Service of the United States Department of Agriculture published a Soil Survey Laboratory Methods Manual (Soil Survey Staff, 2011), which is the reference source for the National Cooperative Soil Survey Soil Characterization Database (NCSS, 2010). 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.

B.2 Laboratory error

Important quality characteristics for any measured data are the random and systematic error. 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 different 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 set-ups. Laboratories participating in inter-laboratory studies such as ring tests or round robin tests receive feedback on their quality performance with the particular methods by comparing their results with those from other participants. Examples are WEPAL (2015), the Wageningen Evaluating Programmes for Analytical Laboratories, and NATP (2015), the North American Proficiency Testing Program. These programs

often are certified according to ISO/IEC 17043. These programs, however, do not consider the influence of differences in sampling procedure and pre-treatment at individual laboratories as these programs 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 at 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 disparate 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 lab error is not necessarily the major component of the total error in derived interpretations; spatial variability can contribute even more (e.g. 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, 2006). While cost-efficient and cost-effective procedures for field sampling procedures 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 Staff, 2011) and may apply for specific user groups only. 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 soil/solution ratio and description of solution (e.g. water, KCI 1M) to be fully understood. In WoSIS, soil properties also are defined by the analytical methods and terminology used is 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 important for selection of alternative methods are Cation Exchange Capacity and available Phosphate. The capacity of a soil to adsorb and exchange cations from exchange sites depends importantly on the actual pH and the ionic strength of the solution. However, the need for a sufficiently detailed description of analytical procedures is particularly reflected in the case of so-called plant 'available phosphate', where the choice of the appropriate laboratory methods is largely determined by soil pH as a proxy for soil mineralogy and soil type (Elrashidi, 2010). Hence, 'vague' descriptions for available-P methods are basically useless, unless used in a specific context such as a (local) fertilizer recommendation scheme. For example, correlation studies have shown that only in specific cases (i.e. soils and intended use) region-specific conversions can be made for available-P values determined according to different analytical procedures (e.g. P-Olsen and P-Bray, see Mallarino, 1995; Modified P-Morgan and Mehlich III, see Ketterings, 2002), making international harmonization of the results of such methods cumbersome or possibly at best 'broad brush'. An example of such an harmonization effort for Tunisia is discussed in Ciampalini et al. (2013). According to GlobalSoilMap (GSM, 2013 p. 25) there is generally no universal equation for converting from one 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 will need to develop and apply node-specific conversions (towards the adopted standard methods and soils), building on comparative analyses of say archived samples.

B.3 Standardization of soil analytical method descriptions

Lacking detailed quantitative information on the quality of the soil analytical data held in the diverse source databases shared for use in WoSIS, it was necessary to develop a qualitative procedure to describe the analytical methods in a flexible, yet comprehensive and consistent way. For all source data, 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. 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 method features in WoSIS are assigned on basis of the descriptions 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 are consulted again.

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. As explained above, the spread of these results may be an indication for the maximum spread in a compiled database.

For soil property 'pH KCl', for example, the selected features within WoSIS are the soil/solution ratio, the molarity of the KCl solution, and the measurement technique (see Appendix B.4). It is assumed here 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 may differ per soil type and (pair of) electrode(s) used, but are considered of minor importance for differentiating methods in the WoSIS database (Table B.1). Once a feature is identified, based on the available (source) information, the appropriate option / value is added (i.e. 0.1, 0.5, 1 Molar). This allows users of the database to select those data that are analysed according to defined (and comparable) methods and may be judged as having equal quality as well as those that are suited for specific use. When new data are entered, the table is used for describing (coding) the added data. If required, values / options not yet tabled can be added. Additional soil properties and features for methods will be added gradually in future versions of the WoSIS document.

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 from a more detailed source, such as a laboratory manual. Importantly, the present confidence flags should not be seen as a measure for the quality of a particular laboratory. Procedures for coding 'standardized analytical methods' in WoSIS 2, developed so far, are presented in Appendix D.

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

Feature name	Feature option
solution	Calcium chloride [CaCl2]
solution	Potassium chloride [KCI]
solution	Sodium fluoride [NaF]
solution	unknown
solution	water [H2O]
ratio	1:1
ratio	1:10
ratio	1:2
ratio	1:2.5
ratio	1:5
ratio	1:50
ratio	saturated paste
ratio	unknown
concentration	0.01 M
concentration	0.02 M

Feature name	Feature option
concentration	0.2 M
concentration	1 M
concentration	not applied
concentration	unknown
instrument	electrode
instrument	electrode (field measured)
instrument	indicator paper (field measured)
instrument	unknown

B.4 Worked out example (soil pH)

As indicated, when selecting (alternative) lab methods for specific uses or data for further use, it should be remembered that numerous soil properties are based on 'operational definitions' (Soil Survey 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 2; the procedure is illustrated below using pH as an example.

Major characteristics (features) of commonly used methods for determining a given soil property are identified first. For soil property pH, for example, these are the extractant solution (water or salt solution), and in case of salt solutions the salt concentration (molarity), 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 features per method, the options (specifications) that are used in data descriptions or known from reference laboratory manuals are tabled. For soil property (i.e. field *attribute_agg_name*) 'pH' and *feature_name* soil/solution 'ratio', the available options range from 'unknown' to 'saturated paste' (Table B.1).

The above system for the description of laboratory methods (data) in WoSIS will allow for flexible and straightforward database queries.

Appendix C

Background information soil analytical procedures

C.1 General

Information similar to that presented below (Table C.2) for soil pH (as example) will gradually be added to future releases of the WoSIS report as it becomes available. In first instance, this will relate to the soil properties listed in Table C.1. As indicated, these correspond with the list presented in the Global Soil Map Specifications (GSM, 2013). All measurement values in WoSIS are expressed using SI units or non-SI units accepted for use with the International System of Units.

Table C.1: Initial list of soil properties for standardization (i.e. for which standardized descriptions are being developed)

Soil property	Status	Standard units ¹	Decimals
рН	OK	unitless	_
Organic carbon	OK	g/kg	1
CEC	In prep.	cmol(c)/kg	1
Bulk Density	OK	kg/dm^3	2
Water Holding Capacity ²	OK	cm^3 / 100 cm^3	2
Calcium carbonate equivalent	OK	g/kg	1
Sand, silt and clay fractions	OK	g/100g	1
Coarse fragments	OK	cm^3 / $100\ cm^3$	0
Electrical conductivity	In prep.	dS/m	1

C.2 Soil pH

To determine the pH of a soil sample a range of solutions is used to bring H^+ ions into solution. Distilled water and solutions with low ionic strength are used to estimate the soil solution. Measurements in saturated pastes are aimed to reproduce the conditions of the natural environment as closely as possible (Pansu and Gautheyrou, 2006).

 $^{^{1}}$ 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 $meq\ /\ 100\ g^{-1}$. Layer depth (top resp. bottom) expressed in cm, measured from the surface, including organic layers and mineral covers (see Section 3.2.5).

²Water holding capacity is calculated here as the amount of water held between 1/3 bar and 15 bar (USDA conventions, Soil Survey Staff, 2014). At a later stage, in case of missing measured data, this may be done using a range of pedotransfer functions (e.g. Botula et al., 2014).

Methods differ also in the soil / solution ratio, which may be expressed as weight / volume, or volume / volume. With each ratio the pH measurement in the resulting supernatant solutions differs. More ions dissolve in a larger volume up to maximum solubility is reached. Agitation time and method of shaking, as well as measurement 'in rest' or opposite 'actively stirred' have to be standardized in a laboratory to have consistent measurement conditions adapted to their range of soils. To obtain reliable measurements for pH water in soils with high organic matter content usually a higher water: soil ratio is used (Pansu and Gautheyrou, 2006). Measurements at soil / solution ratios especially those where the electrodes are in contact with the sediment, may save a 'suspension' effect. This effect can modify the results by +/- 1 pH unit.

With high molar salt solutions exchange processes are enforced. For instance 1 M KCl solution is used to release the hydrogen ions and Al^{3+} ions from the exchange complex. The difference with the pH measured in water is a measure for the potential acidity (pH delta). pH delta should be measured with equal conditions. With 1 M NaF, OH^- is released. The increase in pH is an indication for 'active aluminium' (Van Reeuwijk, 2002).

Conditions for measurement of pH in the field may differ too much from usual lab conditions to obtain comparable results. They are differentiated by separate feature options.

pH methods within WoSIS are grouped by extracting solution, the soil / solution ratio and electrodes used, only. Although for detailed studies other details of the particular methods are important also, they are not used in this grouping. As indicated, that information is hard to deduct from the resources which just refer to titles of lab manuals and condensed descriptions of particular methods.

In Table C.2 the options from the WoSIS system for method descriptions are assigned to routine methods for pH KCl, as described in ISO 10390:2005, the Soil Survey Laboratory Methods Manual from the Natural Resources Conservation Service of the United States Department of Agriculture (Soil Survey Staff, 2011) as procedure '4C1a2a3', and ISRIC procedure '4-1' pH H2O and pH- KCl (Van Reeuwijk 2002).

The flexibility of the system for coding soil analytical method descriptions is illustrated in Table C.2 where it has been applied to code three different reference methods for pH KCl as used by 'ISO' (ISO 2015), 'USDA' (Soil Survey Staff, 2014) and ISRIC (van Reeuwijk, 2002).

Table C.2: Unique option codes for describing 'pH KCl' as determined according to ISO, ISRIC, and USDA laboratory protocols

	Analytical method		
Feature name	ISO ³	USDA ⁴	ISRIC ⁵
Solution	KCI	KCI	KCI
Concentration	1 M	1 M	1 M
Ratio	1:5	1:1	1:2.5
Instrument	Electrode	Electrode	Electrode

³ISO 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, 2015); this coding example is for pH KCl.

⁴USDA: Method 4C1a2a3 (Soil Survey Staff, 2014).

⁵ISRIC: Method 4-1 pH-H2O and pH-KCI (van Reeuwijk 2002).

Appendix D

Procedures for coding standardized analytical methods

This appendix lists the criteria used for standardizing disparate analytical method descriptions to the WoSIS standard.

To facilitate data entry, that is the standardization of soil analytical method descriptions by third parties, the recommended sequence (1,2, *) for describing attribute-specific features is listed under *feature_order_option* in table **desc_method_feature**.

Table D.1: Procedure for coding Bulk density

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

Table D.2: Procedure for coding Calcium carbonate equivalent

Feature name	Feature option
reaction	dissolution of Carbonates by Hydrochloric acid [HCI], or
	Perchloric acid $[HClO_4]$
reaction	dissolution of Carbonates by Phosphoric acid $[H_3PO_4]$

Feature name	Feature option
reaction	dissolution of Carbonates by Sulfuric acid $[H_2SO_4]$
reaction	unknown
temperature	external heat, elevated temperature; ignition \leq 400 deg. Celcius
temperature	extrenal heat, combustion (element analyzer)
temperature	no external heat
temperature	unknown
detection	gravimetric - weight increase (from trapped Carbon dioxide $[CO_2]$ evolved)
detection	gravimetric - weight loss (from Carbon dioxide $[CO_2]$ evolved)
detection	pressure (i.e. pressure build bij Carbon dioxide $[CO_2]$ evolved, manometric)
detection	sensoric (as in element analyzer)
detection	titrimetric (i.e. excess acid, absorption Carbon dioxide $[CO_2]$)
detection	unknown
detection	volumetric (i.e. volume of Carbon dioxide $[CO_2]$ evolved) (1 Pa, room temperature)
calculation	(in)direct estimates of Carbonates $[XXCO_3.xxH_2O]$ or Inorganic Carbon, expressed as Calcium carbonate equivalent
calculation	subtraction; (Total C - Organic C) expressed as Calcium carbonate equivalent
calculation	unknown

Table D.3: Procedure for coding Carbon

Feature name	Feature option
pretreatment	inorganic carbon removed; Hydrochloric acid [HCI]
pretreatment	inorganic carbon removed; Phosphoric acid $[H_3PO_4]$
pretreatment	not applied
pretreatment	unknown
reaction	dry oxidation (i.e loss on ignition)
reaction	dry oxidation (such as element analyzer)
reaction	unknown
reaction	wet oxidation - other methods
reaction	wet oxidation with Sulphuric acid $[H_2SO_4]$ -
	Potassiumbichromate $[K_2Cr_2O_7]$ (and Phosphoric acid
	$[H_3PO_4]$) mixture
temperature	controlled, at 960 deg Celsius and higher (assumed:
	element analyzer)
temperature	controlled, at elevated temperature (wet oxidation,
	temperature (not) specified)
temperature	controlled, lower temperature range (assumed; loss on
	ignition, muffle furnace)
temperature	no external heat
temperature	unknown
detection	colorimetry (i.e. by graphing a standard curve)
detection	gravimetric; increase weight by trapping evolved Carbon
	dioxide $[CO_2]$
detection	sensoric (in element analyzer)
detection	titrimetric
detection	unknown
detection	volumetric
detection	weight loss (i.e. "loss on ignition" method)
calculation	complete recovery (assumed)

Feature name	Feature option
calculation	conversion factor "organic matter to total carbon" = 1/1.7
	(1.7 = Van Bemmelen factor)
calculation	correction factor = 1.03
calculation	correction factor = 1.15
calculation	correction factor = 1.18
calculation	correction factor for recovery not specified
calculation	default correction factor for recovery of 1.3 - assumed
calculation	default correction factor recovery 1.3 - assumed
calculation	default (Walkley and Black) correction factor for recovery
	of 1.3 applied
calculation	not applied
calculation	Total Carbon minus Total inorganic Carbon
calculation	unknown

Table D.4: Procedure for coding Clay

Feature name	Feature option
size	0 - 0.001 mm
size	0 - 0.002 mm
size	0 - 0.005 mm
size	unknown
pretreatment	Hydrogen peroxide $[H_2O_2]$ plus Hydrochloric acid [HCI] or
	Acetic acid $[CH_3COOH]$ (if pH- H_2O >6.5)
pretreatment	Hydrogen peroxide $[H_2O_2]$ plus mild Acetic acid
	$[CH_3COOH]$ / Sodium acetate $[CH_3COONa]$ buffer
	treatments (if pH- H_2O >6.5)
pretreatment	no pretreatment
pretreatment	pretreatment, deferration included
pretreatment	unknown
dispersion	Ammonium $[NH_4]$
dispersion	no dispersion
dispersion	Sodium hexametaphosphate $[(NaPO_3)6]$ - Calgon type
	(ultrasonic treatment might be included)
dispersion	Sodium hydroxide [NaOH]
dispersion	unknown
instrument	analyzer
instrument	field hand estimate
instrument	hydrometer
instrument	pipette
instrument	unknown

Table D.5: Procedure for coding Coarse fragments

Feature name	Feature option
size	>2 mm
size	unknown
pretreatment	no pretreatment
pretreatment	rock fragments, course concretions, roots and adhering
	finer particles removed from field sample
pretreatment	unknown
instrument	field hand estimate
instrument	sieve
instrument	unknown

Table D.6: Procedure for coding pH

Feature name	Feature option
solution	Calcium chloride $[CaCl_2]$
solution	Potassium chloride [KCI]
solution	Sodium fluoride [NaF]
solution	unknown
solution	water $[H_2O]$
ratio	1:1
ratio	1:10
ratio	1:2
ratio	1:2.5
ratio	1:5
ratio	1:50
ratio	saturated paste
ratio	unknown
concentration	0.01 M
concentration	0.02 M
concentration	0.2 M
concentration	1 M
concentration	not applied
concentration	unknown
instrument	electrode
instrument	electrode (field measured)
instrument	indicator paper (field measured)
instrument	unknown

Table D.7: Procedure for coding Sand

Feature name	Feature option
size	0.02 - 2 mm
size	0.05 - 0.1 mm
size	0.05 - 1.7 mm
size	0.05 - 1 mm
size	0.05 - 2 mm
size	0.06 - 2 mm
size	0.063 - 2 mm
size	0.10 - 0.25 mm
size	0.2 - 2 mm
size	0.25 - 0.5 mm
size	1 - 2 mm
size	unknown
pretreatment	Hydrogen peroxide $[H_2O_2]$ plus Hydrochloric acid [HCI] or
	Acetic acid $[CH_3COOH]$ (if pH- $H_2O > 6.5$)
pretreatment	Hydrogen peroxide $[H_2O_2]$ plus mild Acetic acid
	$[CH_3COOH]$ / Sodium acetate $[CH_3COONa]$ buffer
	treatments (if pH- H_2O >6.5)
pretreatment	no pretreatment
pretreatment	pretreatment, deferration included
pretreatment	unknown
dispersion	Ammonium $[NH_4]$
dispersion	no dispersion
dispersion	Sodium hexametaphosphate $[(NaPO_3)6]$ - Calgon type
	(ultrasonic treatment might be included)
dispersion	Sodium hydroxide [NaOH]
dispersion	unknown
instrument	analyzer
instrument	field hand estimate

Feature name	Feature option
instrument	hydrometer
instrument	sieve
instrument	unknown

Table D.8: Procedure for coding Silt

Feature name	Feature option
size	0.001 - 0.05 mm
size	0.002 - 0.02 mm
size	0.002 - 0.05 mm
size	0.002 - 0.06 mm
size	0.005 - 0.05 mm
size	0.02 - 0.05 mm
size	0.02 - 0.063 mm
size	unknown
pretreatment	Hydrogen peroxide $[H_2O_2]$ plus Hydrochloric acid [HCl] or
	Acetic acid $[CH_3COOH]$ (if pH- H_2O >6.5)
pretreatment	Hydrogen peroxide $[H_2O_2]$ plus mild Acetic acid
	$[CH_3COOH]$ / Sodium acetate $[CH_3COONa]$ buffer
	treatments (if pH- H_2O >6.5)
pretreatment	no pretreatment
pretreatment	pretreatment, deferration included
pretreatment	unknown
dispersion	Ammonium $[NH_4]$
dispersion	no dispersion
dispersion	Sodium hexametaphosphate $[(NaPO_3)6]$ - Calgon type
	(ultrasonic treatment might be included)
dispersion	Sodium hydroxide [NaOH]
dispersion	unknown
instrument	analyzer
instrument	field hand estimate
instrument	hydrometer
instrument	pipette
instrument	unknown

Table D.9: Procedure for coding Water retention

Feature name	Feature option
soil water tension applied	kPa=0.1, cm water head=1.0, bar=0.001, pF=0.0
soil water tension applied	kPa=0.3, cm water head=3.2, bar=0.003, pF=0.5
soil water tension applied	kPa=0.5, cm water head=5.0, bar=0.005, pF=0.7
soil water tension applied	kPa=100, cm water head=1021.6, bar=1.00, pF=3.0
soil water tension applied	kPa=10, cm water head=102.2, bar=0.10, pF=2.0
soil water tension applied	kPa=12, cm water head=125.0, bar=0.12, pF=2.1
soil water tension applied	kPa=1500, cm water head=15324.0, bar=15.00, pF=4.2
soil water tension applied	kPa=15, cm water head=150.0, bar=0.15, pF=2.2
soil water tension applied	kPa=1, cm water head=10.2, bar=0.01, pF=1.0
soil water tension applied	kPa=200, cm water head=2043.2, bar=2.00, pF=3.3
soil water tension applied	kPa=20, cm water head=204.3, bar=0.20, pF=2.3
soil water tension applied	kPa=24, cm water head=250.0, bar=0.24, pF=2.4
soil water tension applied	kPa=250, cm water head=2554.0, bar=2.50, pF=3.4
soil water tension applied	kPa=33, cm water head=337.1, bar=0.33, pF=2.5
soil water tension applied	kPa=3, cm water head=30.6, bar=0.03, pF=1.5
soil water tension applied	kPa=400, cm water head=4086.4, bar=40.90, pF=3.6
soil water tension applied	kPa=40, cm water head=408.6, bar=0.40, pF=2.6
soil water tension applied	kPa=500, cm water head=5085.0, bar=5.00, pF=3.7

Feature name	Feature option
soil water tension applied	kPa=500, cm water head=5108.0, bar=51.10, pF=3.7
soil water tension applied	kPa=50, cm water head=510.8, bar=0.50, pF=2.7
• •	kPa=580, cm water head=510.6, bar=5.80, pF=3.8
soil water tension applied	·
soil water tension applied	kPa=5, cm water head=51.1, bar=0.05, pF=1.7
soil water tension applied	kPa=60, cm water head=613.0, bar=0.60, pF=2.8
soil water tension applied	kPa=6, cm water head=61.3, bar=0.06, pF=1.8
soil water tension applied	kPa=70, cm water head=715.1, bar=0.70, pF=2.9
soil water tension applied	kPa=7, cm water head=75.0, bar=0.07, pF=1.9
soil water tension applied	kPa=80, cm water head=817.3, bar=0.80, pF=2.9
soil water tension applied	kPa=90, cm water head=919.4, bar=0.90, pF=3.0
soil water tension applied	not applied
soil water tension applied	unknown
sample type	<2 mm (sieved) disturbed samples
sample type	clod, reconstituted / disturbed
sample type	natural clod
sample type	undisturbed soil in metal/PVC-ring (soil core)
sample type	unknown
sample pretreatment	air dry, then saturated
sample pretreatment	field moist condition, then saturated
sample pretreatment	not applied
sample pretreatment	oven dried, no saturation applied (i.e.: absorption curve)
sample pretreatment	oven dry, then saturated
sample pretreatment	saturated, desorbed, rewetted and desorbed again
sample pretreatment	unknown
method	absorption into oven dry sample (curve, dry to wet)
method	desorption, evaporation
method	desorption, oven drying
method	desorption, pressure
method	desorption, suction (hanging water column + Hg
	manometer)
method	desorption, suction (hanging water column, water
	manometer)
method	saturation (pF 0)
method	unknown
device	balans, tensiometers (wind evaporation method)
device	fine textured medium; (pressumed) kaolin box
device	fine textured medium; (presumed) sandbox
device	not applied
device	porous plate and burette
device	pressure plate extractor
device	tension table
device	unknown
result expression	dry mass basis; mass water per unit mass of soil solids
	(w/w, gravimetric water content)
result expression	unknown
result expression	volume base; volume of water per unit volume of moist
	soil (v/v, volumetric water content)
result expression	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
result expression	volume base; volume of water per unit volume of moist
	soil (v/v, volumetric water content). w/w % converted by
manufa armana este e	bulk density oven dry
result expression	volume base; volume of water per unit volume of moist
	soil (v/v, volumetric water content). w/w % converted by
	bulk density pKa 33

Feature name	Feature option
result expression	volume base; volume of water per unit volume of moist soil (v/v, volumetric water content). w/w % converted by bulk density rewet
result expression	volume base; volume of water per unit volume of moist soil (v/v, volumetric water content). w/w % converted by Unknown bulk density
result expression	wet mass basis; mass of water per unit mass of wet soil (w/w)

Appendix E

Database design

This appendix explains the structure of all tables considered in WoSIS 2.

Table E.1: class_cpcs - Soil name according to the French soil classification system (Commission de pedologie et de cartographie des sols, CPCS 1967)

Column	Туре	Modifiers Comment	Comment
class_cpcs_id	integer	sednence	Primary key
dataset_id	character varying(20)		Foreign key that together with "profile_id" refers to table
			"dataset_profile"
profile_id	integer		Foreign key that together with "dataset_id" refers to table
			"profile_profile"
publication_year	smallint		Year of publication of the version used for the characterization
group_code	character varying(5)		Soil group code
group_name	character varying(50)		Soil group name
unit_code	character varying(5)		Soil unit code
unit_name	character varying(100)		Soil unit name
note	text		Comments field
trust	character(1)	'A'::bpchar	Level of trust: 'A' as entered, no validation; 'B' standardized;
			'C' harmonized; 'D' the highest level, data validated by a soil
			scientist

Table E.2: class_fao - Soil name according to the Legend of the 1:5M scale Soil Map of the World (FAO-Unesco 1974) resp. Revised Legend (FAO 1990)

Column	Туре	Modifiers	Comment
class_fao_id	integer	sedneuce	Primary key
dataset₋id	character varying(20)		Foreign key that together with "profile_id" refers to table "dataset_profile"
profile_id	integer		Foreign key that together with "dataset_id" refers to table "dataset_profile"
publication_year	smallint		Year of publication of the version used for the characterization
major_group_code	character varying(2)		Major soil group code
major_group	character varying(30)		Major soil group name
formative_element_code	character varying(1)		Formative element code
formative_element	character varying(30)		Formative element name
subunit_code	character varying(5)		Soil unit code
subunit	character varying(30)		Soil unit name
phase_code	character varying(2)		Phase code - limiting factor related to surface or subsurface
			features of the land
phase	character varying(30)		Phase name - limiting factor related to surface or sub surface
			features of the land
verified	boolean		Was the classification verified (True/False)
verified_user	character varying(50)		Who made the verification?
verified_date	date		Date when was the classification verified?
note	text		Comments field
trust	character(1)	'A'::bpchar	Level of trust: 'A' as entered, no validation; 'B' standardized;
			'C' harmonized; 'D' the highest level, data validated by a soil
			scientist

Table E.3: class_fao_horizon - Diagnostic horizon according to the Soil Map of the World Legend (FAO-Unesco 1974, resp. FAO 1990)

Column		Modifiers Comment	Comment
class_fao_horizon_id integer	integer	sedneuce	sequence Primary key
class_fao_id	integer		Foreign key that refers to table "class_fao"
diagnostic_horizon	character varying(50)		Name of the diagnostic horizon
horizon	character varying(20)		Soil horizon
upper_depth	smallint		Depth of upper horizon boundary (cm)
lower_depth	smallint		Depth of lower horizon boundary (cm)

Table E.4: class_fao_property - Diagnostic property according to the Soil Map of the World Legend (FAO-Unesco 1974, resp. FAO 1990)

Column	Туре	Modifiers	Modifiers Comment
class_fao_property_id	integer	sedneuce	sequence Primary key
class_fao_id	integer		Foreign key that refers to table "class_fao"
diagnostic_property	character varying(80)		Name of the diagnostic property
upper_depth	smallint		Depth of upper horizon boundary (cm)
lower_depth	smallint		Depth of lower horizon boundary (cm)

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

Comment	Primary key	Foreign key that together with "profile_id" refers to table	″dataset_profile″ Foreign key that together with "dataset⊥id" refers to table	"dataset_profile"	Name of local soil classification system	Year of publication of the version used for the characterization	Taxon name	Taxon common name	Comments field	Level of trust: 'A' as entered, no validation; 'B' standardized;	'C' harmonized; 'D' the highest level, data validated by a soil	scientist
Modifiers	sedneuce									'A'::bpchar		
Туре	integer	character varying(20)	integer		character varying(100)	smallint	text	character varying(255)	text	character(1)		
Column	class_local_id	dataset_id	profile_id		system_name	publication_year	classification_name	common_name	note	trust		

Table E.6: class_soil_taxonomy - Soil name according to USDA Soil Taxonomy (with defined version)

Column	Туре	Modifiers	Comment
class_soil_taxonomy_id	integer	sedneuce	Primary key
dataset_id	character varying(20)		Foreign key that together with "profile_id" to table
			"dataset_profile"
profile_id	integer		Foreign key that together with "dataset_id" refers to table
			"dataset_profile"
publication_year	smallint		Year of publication of the version used for the characterization
order_name	character varying(50)		Order name
suborder	character varying(50)		Suborder name
great_group	character varying(50)		Great group name
subgroup	character varying(50)		Subgroup name
temperature_regime	text		Temperature regime
moisture_regime	text		Moisture regime
mineralogy	text		Mineralogy
textural_class	text		Textural class
other	text		Other information
note	text		Comments field
trust	character(1)	'A'::bpchar	Level of trust: 'A' as entered, no validation; 'B' standardized; 'C' harmonized: 'D' the highest lovel data validated by a goil
			scientist

Table E.7: class_wrb - Soil name according to World Reference Base for Soil Resources with defined version (e.g. WRB 2006, WRB 2014)

<u>jq</u>			
7: +000+07	integer	sedneuce	Primary key
מממאפורום	character varying(20)		Foreign key that together with "profile_id" refers to table
			"dataset_profile"
profile_id ir	integer		Foreign key that together with "dataset_id" refers to table
			"dataset_profile"
publication_year s	smallint		Year of publication of version used for the characterization
code	text[]		Prefix qualifier code
	text[]		Prefix qualifier name
reference_soil_group_code c	character varying(4)		Reference soil group code
reference_soil_group	character varying(30)		Reference soil group name
code	text[]		Sufix qualifier code
sufix_qualifier te	text[]		Sufix qualifier name
	boolean		Is the classification verified (True/False)
verified_user c	character varying(50)		Who made the verification?
verified_date c	date		Date when the classification was verified
note	text		Comments field
trust	character(1)		Level of trust: 'A' as entered, no validation; 'B' standardized;
			'C' harmonized; 'D' the highest level, data validated by a soil
			scientist

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

Column	Туре	Modifiers Comment	Comment
class_wrb_horizon_id	integer	sedneuce	sequence Primary key
class_wrb_id	integer		Foreign key that refers to table "class_wrb"
diagnostic_horizon	character varying(50)		Name of the diagnostic horizon
upper_depth	smallint		Depth of upper horizon boundary (cm)
lower_depth	smallint		Depth to lower horizon boundary (cm)

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

Column	Туре	Modifiers Comment	Comment
class_wrb_material_id	integer	sedneuce	sequence Primary key
class_wrb_id	integer		Foreign key that refers to table "class_wrb"
diagnostic_material	character varying(50)		Name of the diagnostic material
upper_depth	smallint		Depth of upper horizon boundary (cm)
lower_depth	smallint		Depth of lower horizon boundary (cm)

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

7-1	F	-: -: j; V	
Column	lype	Modifiers Comment	Comment
class_wrb_property_id	integer	sedneuce	sequence Primary key
class_wrb_id	integer		Foreign key that refers to table "class_wrb"
diagnostic_property	character varying(50)		Name of diagnostic property
upper_depth	smallint		Depth of upper horizon boundary (cm)
lower_depth	smallint		Depth of lower horizon boundary (cm)

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

Column	Туре	Modifiers	Modifiers Comment
class_wrb_qualifier_id	integer	sedneuce	sequence Primary key
class_wrb_id	integer		Foreign key that refers to table "class_wrb"
qualifier_position character	character varying(6)		Qualifier position (prefix/suffix)
qualifier	character varying(50)		Qualifier name
specifier	character varying(50)		Specifier name
order	smallint		Qualifier ordering number

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

	Column		odifiers	Modifiers Comment
l	country_id	character(2)		Primary key and ISO 3166-1 alpha-2, two-letter country code
	iso3_code	character(3)		ISO 3166-1 alpha-2, three-letter country code
	gaul_code	integer		Global Administrative Unit Layers (GAUL) country code
	color_code	character(3)		Country map colour by GAUL
	ar	text		Country name in Arabic
	en	text		Country name in English
	es	text		Country name in Spanish
	fr	text		Country name in French
	pt	text		Country name in Portuguese
	2	text		Country name in Russian
	zh	text		Country name in Chinese
	status	text		Status of the country
	disp_area	character varying(3)		Unsettled Territory (True/False)
	capital	text		Country capital name
	continent	text		Continent name
	un_reg	text		UN region name
	unreg_note	text		Note about UN region
	geom	geometry (Multi Polygon, 4326)		Country polygon geometry
	geom_label	geom_label geometry(Point,4326)		Point geometry in order to place a label in a map
ı				

Table E.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	Туре	Modifiers	Comment
dataset_id	character varying(20)		Primary key
dataset_title	character varying(220)		Dataset title
summary	text		Dataset summary
access_constraints	text		Access constraints to the dataset
use_constraints	text		Use constraints to the dataset
dataset_progress	character varying(50)		Dataset import progress (Planned/In progress/Complete/Not
70000000000000000000000000000000000000	; ; ;		Didinieu/
dataset_organization_id	Integer		Foreign key that refers to table dataset_organization The defect on he chard coording to the license provided
pilvate	DOOLEGAL		itte dataset cari be strated according to tite licerice provided with the data source (True/False)
dataset_type	character varying(15)		Type of dataset (Source/Compilation/Citation)
dataset_rank	smallint		Ranking of dataset based on expert knowledge; a measure for
			the inferred degree of confidence in given dataset
author	character varying(150)		Dataset author
publication_date	character varying(10)		Publication date
dataset_version	character varying(5)		Dataset version
nnid	character varying(250)		Foreign key that refers to table "geonetwork.metadata" for
			further detailed metadata about the dataset
n_profiles	integer		Number of profiles in the dataset
n_layers	integer		Number of layers in the dataset
n_profile_attr	integer		Number of different attributes in the dataset describing the site
			where the profile was taken
n_layer_attr	integer		Number of different attributes in the dataset describing layers
n_profile_rows_inserted	integer		Number of inserted records from the dataset describing the site
n_profile_rows_standard	integer		Number of standerdized records from the dataset describing the
n_layer_rows_inserted	integer		Number of inserted records from the dataset describing the
			layers
n_layer_rows_standard	integer		Number of standerdized records from the dataset describing the
			layers

Table E.14: dataset_organization - Holds contact information about organizations that in some capacity have played a role in the creation, gathering, management, or dissemination of data held in WoSIS

Column	Туре	Modifiers Comment	Comment
dataset_organization_id integer	integer	sedneuce	Primary key
	character varying(255)		Acronym or abbreviation
	character varying(255)		Organization main name
	character varying(255)		Organization sub name
	character varying(5)		Telephone international prefix
	character varying(20)		Telephone number
	character varying(255)		E-mail
	text		Web site
country_id	character(2)		Foreign key that refers to table "country"
	character varying(255)		City
	character varying(255)		Street name
	character varying(255)		Street number
	character varying(20)		Postal (ZIP) code
note	text		Comments field

Table E.15: dataset_organization_contact - Holds contact information about people that in some capacity have played a role in the creation, gathering, management, or dissemination of data held in WoSIS

Column	Туре	Modifiers Comment	Comment
dataset_organization_contact_id	integer	sednence	sequence Primary key
	character varying(25)		Contact title
first_name	character varying(25)		First name
	character varying(10)		Prename (e.g. van)
	character varying(50)		Last name
job_title	character varying(100)		Job title
dataset_organization_id	integer		Foreign key that refers to table "dataset_organization"
department	character varying(100)		Department
telephone	character varying(50)		Work telephone
email	character varying(65)		Work E-mail
note	text		Comments field

Table E.16: 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	Туре	Modifiers	Modifiers Comment
dataset_id	character varying(20)		Primary key that together with "profile_id" and Foreign key refers
			to table "dataset"
profile_id	integer		Primary key that together with "dataset_id" and Foreign key
			refers to table "profile"
profile_code	character varying(80)		Code for soil profile as used in the source database
note	text		Comments field

Table E.17: desc_attribute - Description of all the soil properties for each dataset that has been imported in WoSIS

Column	Туре	Modifiers	Comment
desc_attribute_id	integer	sedneuce	Primary key
dataset_id	character varying(20)		Foreign key that refers to table "dataset"
schema_name	character varying(25)		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	character varying(50)		Column name for the source where de soil property comes from
source_attribute_name	character varying(255)		Source soil property name
source_attribute_description	text		Source soil property description
source_attribute_unit	character varying(15)		Source soil property unit
source_attribute_type	character varying(10)		Source soil property type
source_attribute_domain	smallint		Source soil property domain
attribute_type	character varying(7)		Standard soil property name
desc_attribute_standard_id	character varying(100)		Standard soil property unit
conversion	character varying(10)		Standard soil property type
sql_insert	text		SQL code for moving the data from the source into WoSIS
sql_standard	text		SQL code used to standardize the data
number_rows_source	integer		Number of records in the source database
number_rows_inserted	integer		Number of records inserted into WoSIS
number_rows_standard	integer		Number of records that have been standardized
conflict_source_min_value	text[]		Array of conflict values from source compared with the allowed
			minimum value
conflict_source_max_value	text[]		Array of conflict values from source compared with the allowed
	1		liaximum value
conflict_source_domain	text[]		Array of conflict values from source compared with the allowed
conflict_standard_min_value	text[]		Array of conflict values from standard compared with the
	1		allowed minimum value
conflict_standard_max_value	text[]		Array of conflict values from standard compared with the
			allowed maximum value
conflict_standard_domain	text[]		Array of conflict values from standard compared with the
			allowed dolliall values
user_name start date	criaraciei varyirig(zb) date		wito iniported the soil property Date when this soil property was imported
start_time	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

Aodifiers Comment	IP address from where the soil property was imported	IP address to where the soil property was exported	Comments field
Modifiers			
Туре	character varying(25)	character varying(25)	text
Column	client_addr	server_addr	note

Table E.18: desc_attribute_standard - Discription of standard attributes and standardisation progress

Column	Туре	Modifiers	Comment
desc_attribute_standard_id	character varying(100)		Standard attribute name
	character varying(7)		Attribute type (Site/Horizon)
attribute_agg_name	character varying(100)		Standard attribute aggregation name
	character varying(15)		Standard attribute unit
data_type	character varying(10)		Standard attribute type
decimals	smallint		Number of decimals
	numeric(6,2)		Minimum value
	numeric(6,2)		Maximum value
desc_domain_id	smallint		Standard attribute domain
progress_name	boolean		If the attribute name has been standardised
progress_method	boolean		If the attribute analytical methods have been standardised
progress_value	boolean		If the attribute measured values have been standardised
priority	smallint		Standardisation priority
distribute	boolean		If distribution is intended
dsm	boolean		If the ttribute is from Global Soil Map
gfsd	boolean		If the ttribute is from Guide lines for soil description FAO (2006)
note	text		Comments field
observed_minimum	numeric(6,2)		Minimum observed value
ximum	numeric(6,2)		Maximum observed value
	integer		Number attributes that have been standardized
	integer		Number of profiles
n_rows_inserted	integer		Number of rows inserted
	integer		Number of rows that have been standardized
	integer		Number of rows that have been sta

Table E.19: 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	Column Type	Modifiers Comment	Comment
desc_domain_id	bigint	sedneuce	sequence Primary key
reference_id	bigint		Foreign key that refers to table "reference"
domain_name	character varying(240)		Domain name
description	text		Domain description
note	text		Comments field
page_num	smallint		Page number, from the referenced document
object_num	smallint		Figure or table number from the referenced document

Table E.20: 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	Туре	Modifiers	Modifiers Comment
alue_id	bigint	esdneuce	sequence Primary key
desc_domain_id			Foreign key that refers to table "desc_domain"
domain_value	character varying(240)		Domain class code
description	text		Domain class description
explanation	text		Domain class explanation

Table E.21: desc_laboratory - Listing of laboratories where soil samples have been analysed

Column	Туре	Modifiers Comment	Comment
desc_laboratory_id integer	integer	sedneuce	Primary key
acronym	character varying(25)		Laboratory acronym or abbreviation
lab_name_1	character varying(250)		Laboratory name, level 1
lab_name_2	character varying(250)		Laboratory name, level 2
lab_name_3	character varying(250)		Laboratory name, level 3
country_id	character(2)		Foreign key that refers to table "country"
city	character varying(30)		The city where the laboratory is located
postal_code	character varying(20)		Laboratory postal code
street_name	character varying(100)		Laboratory street name
street_number	character varying(10)		Laboratory street number
note	character varying(250)		Comments field
lab_name_original	lab_name_original character varying(250)		Laboratory source description

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

Column	Туре	Modifiers	Modifiers Comment
attribute_agg_name	character varying(100)		Standard attribute aggregation name
feature_name	character varying(50)		Standard analytical method feature name. Combined primary
			key ("standard_attribute_name","feature_name","feature_option")
feature_option	text		Standard analytical method feature option. Combined primary
			key ("standard_attribute_name", "feature_name", "feature_option")
feature_option_code	smallint		Standard analytical method feature option code
feature_option_order	smallint		Value to order the features

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

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

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

Column	Tvne	difiers	Modifiers Comment
	296.		
descriptor_id	smallint		Foreign key that refers to table "descriptor"
standard_attribute_name	text		Standard soil property name. Combined foreign key
			("standard_attribute_name","feature_name","feature_option")
feature_name	text		Standard analytical method feature name. Combined foreign
			key ("standard_attribute_name","feature_name","feature_option")
feature_option	text		Standard analytical method feature option. Combined foreign
			key ("standard_attribute_name", "feature_name", "feature_option")
confidence	character varying(10)		Confidence in the standardization procedure
reference_id	smallint		Foreign key that refers to table "reference"

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

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

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

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

Table E.27: image - Images that ilustrate soil profiles or the site of their provenance

Colimb	1,00°	Modifion	Commont
Column	lype	MODIFIES	Colliner
image_id	integer	esdneuce	Primary key
country_id	character(2)		Foreign key that refers to table "country"
identification	character varying(50)		Image internal code
image_year	smallint		Year in which the image was produced
image_timestamp	timestamp without time zone		Timestamp of the image
author	character varying(50)		Image author
image_medium_type	text		Image medium type
extent_height	integer		Image height extent
extent_width	integer		Image width extent
extent_unit	character varying(10)		Image height and width units
file_path	text		Path to image file
file_path_description	text		File path description
description	text		Image description
description_location	text		Image location description
note	text		Comments field
internal_note	text		Image internal note
coordinates	character varying(250)		Image coordinates (WGS 84)
copyright	boolean		The image is copyrighted (True/False)
qualified_for_web_publication	boolean		Image resolution is suited for web publication (True/False)
geom	geometry(Point, 4326)		Point geometry of the location of the picture

Table E.28: image_profile - Links between images and the corresponding profiles

s Comment	Primary key	Foreign key that refers to table "image"	Comments field
Modifiers			
Туре	integer	integer	text
Column	profile_id	image_id	note

Table E.29: image_subject - Images listed by subject (meant to store pictures tags)

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

Table E.30: map_attribute - Holds values of any characteristic associated with "map_unit", "map_unit_component" and "map_unit_soil_component"

Column	Туре	Modifiers Comment	Comment
map_attribute_id	integer	sedneuce	Primary key
map_unit_id	integer		Foreign key that refers to table "map_unit"
map_unit_component_id	smallint		Foreign key that refers to table "map_unit_component"
t id	smallint		Foreign key that refers to table "map_unit_soil_component"
	integer		Foreign key that refers to table "descriptor"
observation_date	time with time zone		Date of observation or measurement
	character varying(150)		Attribute source value
standard_value	character varying(150)		Attribute standardised value
accuracy	character(50)		Accuracy of an observation or measurement precision
trust	character(1)	'A'∷bpchar	Level of trust: 'A' as entered, no validation; 'B' standardized;
			'C' harmonized; 'D' the highest level, data validated by a soil
			scientist
quality	character(1)	'0'::bpchar	'0'::bpchar Quality indicator - 0 lowest, 4 highest

Table E.31: map_unit - Polygon geometry of homogeneous map unit features

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

Table E.32: map_unit_component - Information about mapping unit component

Column	Туре	Modifiers Comment	Comment
map_unit_component_id	integer	integer sequence F	Primary key
map_unit_id	integer		Foreign key that refers to table "map_unit"
unit_number	smallint		Map unit component number
proportion	smallint		Map unit component proportion

Table E.33: map_unit_soil_component - Coding of soil components

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

Table E.34: map_unit_soil_component_x_profile - Links soil components, listed in table "map_unit_soil_component", to one or more reference profiles (table still void in this version of WoSIS)

Column	Туре	Modifiers Comment	Comment
map_unit_soil_component_x_profile_id integer	integer	sedneuce	sequence Primary key
map_unit_soil_component_id	integer		Foreign key that refers to table "map_unit_soil_component"
dataset₋id	character varying(20)		Foreign key that together with "profile_id" refers to table
profile_id	integer		dataset_prome Foreign key that together with "dataset.id" refers to table
			"dataset_profile"

Table E.35: profile - list of soil profiles, with their location (geometry)

Column	Туре	Modifiers	Comment
profile_id	integer	sedneuce	Primary key
country_id	character(2)		Foreign key that refers to table "country"
geom_accuracy	real		Accuracy of the geometry in degrees. Example: If degree,
			minutes and seconds are provided then geom_accuracy is
			asigned with the value 0.01, if seconds are missing then 0.1,
			if seconds and minutes are missing then 1
nuid	nuid		Universally unique identifier profile code
geom	geometry(Point,4326)		Point geometry of the location of the profile
trust	character(1)	'A'::bpchar	Level of trust: 'A' as entered, no validation; 'B' standardized;
			'C' harmonized; 'D' the highest level, data validated by a soil
			scientist
date_time	timestamp without time zone		Timestamp of the GPS reading. May be used for differential
			correction purposes
geom_accuracy_type	character varying(3)	'UNK'::bpchar	GPS - Coordinates come from a GPS; MAP - Coordinates come
			from a map; UNK - Unknown coordinates source
sample_type	character varying(10)	'single'::character varying	Either single or composite sample
sample_number	integer	-	Number of samples
sample_area	integer	-	Area sampled (m2)
country_geom_closest	character(2)		Closest country to profile, based on geometry
country_geom_distance	integer		Distance of the closest country to profile, based on geometry

Table E.36: profile_attribute - Values of characteristics that are associated with the profiles site

Column	Туре	Modifiers Comment	Comment
profile_attribute_id integer	integer	sedneuce	Primary key
profile_id	integer		Foreign key that together with "profile_id" refers to table
			"dataset_profile"
dataset_id	character varying(20)		Foreign key that together with "dataset_id" refers to table
			"dataset_profile"
descriptor_id	integer		Foreign key that refers to table "descriptor"
accuracy	character(50)		Accuracy of an observation or measurement precision
quality	character(1)	'0'∷bpchar	Quality indicator - 0 lowest, 4 highest
trust	character(1)	'A'::bpchar	Level of trust: 'A' as entered, no validation; 'B' standardized;
			'C' harmonized; 'D' the highest level, data validated by a soil
			scientist
observation_date	date		Date of observation or measurement
source_value	text		Profile attribute, source value
standard_value	text		Profile attribute, standard value

Table E.37: profile_layer - Lists of Layers depths and samples definition per profile and dataset

Column	Туре	Modifiers	Modifiers Comment
profile_layer_id	integer	sedneuce	Primary key
dataset_id	character varying(20)		Foreign key that together with "profile_id" refers to table
			"dataset_profile"
profile_id	integer		Foreign key that together with "dataset id" refers to table
			"dataset_profile"
layer_number	smallint		Consecutive layer number rated from top to bottom
layer_name	character varying(50)		Layer name. Horizon A, B,
sample_type	character varying(50)		Type of sample taken
sample_code	character varying(255)		Sample code
sample_composition	character varying(50)		Sample composition
sample_available	character varying(50)		Sample availability
upper_depth	smallint		Depth of upper horizon boundary (cm)
lower_depth	smallint		Depth of lower horizon boundary (cm)
upper_depth_source	real		Depth of upper horizon boundary (cm) from source
lower_depth_source	real		Depth of lower horizon boundary (cm) from source
note	text		Comments field

Table E.38: profile_layer_attribute - Lists values of any characteristic associated with a profile layer

Column	Туре	Modifiers	Comment
profile_layer_attribute_id	integer	sedneuce	Primary key
profile_layer_id	integer		Foreign key that refers to table "profile_layer"
descriptor_id	integer		Foreign key that refers to table "descriptor"
accuracy	character(50)		Accuracy of an observation or measurement precision
quality	character(1)	'0'::bpchar	Quality indicator - 0 lowest, 4 highest
trust	character(1)	'A'∷bpchar	Level of trust: 'A' as entered, no validation; 'B' standardized;
			'C' harmonized; 'D' the highest level, data validated by a soil
			scientist
observation_date	date		Date of observation or measurement
source_value	text		Layer attribute, source value
standard_value	text		Layer attribute, standard value

Table E.39: profile_layer_thinsection - Lists information related with thin sections

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

Table E.40: raster - Registed raster layers

Column	Туре	Type Modifiers Comment	Comment
rid	integer	sednence P	Primary key
descriptor_id	smallint		Foreign key that refers to table "descriptor"
rast	raster		Raster binary data
filename	text		Path to raster file

Table E.41: reference - List of references to source materials managed in WoSIS

		n/Map/URL/Digital media)																				
Comment	Primary key	Type of reference (Publication/Map/URL/Digital media)	ISBN number	ISRIC library code	Title	Subtitle	Issue	Series	Page	Publication year	Publisher	URL	Map type	Map sheet	Map scale	Min latitude	Max latitude	Min longitude	Max longitude	Digital format	Content description	Comments field
Modifiers	eouenbes																					
Туре	integer	character varying(13)	character varying(20)	character varying(20)	text	text	character varying(255)	character varying(255)	character varying(50)	smallint	character varying(255)	text	character varying(50)	character varying(255)	integer	numeric(7,5)	numeric(7,5)	numeric(8,5)	numeric(8,5)	character varying(50)	text	text
Column	reference_id	reference_type	isbn	isn	title	subtitle	issue	serie	page	publication_year	publisher	url	map_type	map_sheet	map_scale	min_latitude	max_latitude	min_longitude	max_longitude	digital_format	content	note

Table E.42: reference_author - List of author names considered in table references

Column	Туре	Modifiers	Modifiers Comment
reference_author_id	integer	sedneuce	Primary key
reference_id	smallint		Foreign key that refers to table "reference"
author	character varying(255)		Publication author

Table E.43: reference_dataset - List of references by imported dataset

		able "dataset"
Comment	Primary key	Foreign key that refers to table "dataset"
Modifiers		ər varying(20)
Туре	integer	characte
Column	reference_id	dataset_id

Table E.44: reference_dataset_profile - List of references by imported dataset and their profiles

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

Table E.45: reference_file - List of references that have a file as support

Column	Туре М	Modifiers Comment	Comment
reference_file_id	bigint	sedneuce	sequence Primary key
reference_id	integer		Foreign key that refers to table "reference"
file_type	character varying(50)		File type
file_name	character varying(254)		File name
file_description	text		File description
file_path	text		File path



ISRIC – World Soil Information has a mission to serve the international community with information about the world's soil resources to help addressing major global issues.

More information: www.isric.org

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