

## Author response to Referee #1 : ESSDD-2016-34-RC1

Dear Dr Hugelius,

Thank you for your constructive review and encouraging comments concerning the anticipated usefulness of present and future standardised datasets derived from WoSIS. These comments have helped us to improve the manuscript. In particular, we have followed up on your suggestion to improve the user-friendliness of the WoSIS-derived snapshots, hence the file-format for distributing the data.

Below we address your queries and suggestions in the order that they are raised in the referee report. Revised texts are indicated in **blue** in the author's comments and accompanying manuscript (which also includes our response to Reviewer 2).

In places, we have reworded some text for enhanced clarity. For example, the phrases concerning the total number of profiles managed in the central WoSIS server database, as of July 2016, versus the number of standardised profile data that we can provide in the 'July 2016' snapshot (see Section 3).

**Referee Comment 1 (RC1):** I find that data-users who want more details will likely need to consult the procedures manual of WoSIS (Ribeiro et al 2015). This applies e.g. to how standardization or harmonization of different protocols and methods was carried out (sections 2.6 and 2.7). Given the very wide scope of this work and the length of text required to explain it, I think this is sensible. However, I would suggest that it could be made more clear in the data "read me" file as well as this manuscript where the technical details can be found (as a side-note, the link to the Ribeiro et al (2015) report on <http://www.isric.org/data/wosis> does not work?).

**Author comment 1 (AC1):** Thank you for your observation. As recommended, we have expanded the 'ReadMe File.' We have added text to explain the structure of the files served in the snapshot (see details below, under RC3/AC3). We have also added a reference to the technical documentation (procedures manual) for the whole WoSIS database (Ribeiro et al. 2015), which can now be downloaded directly from the new 'ReadMe File' thus enhancing 'user-friendliness'.

The revised "Readme File" is included in the newly formatted version of the "July 2016 snapshot (on-line as of 13/12/2016).

For clarification, section 2.6 and 2.7 now include references to the relevant pages in the procedures manual, for example in Section 2.6: "**WoSIS criteria for coding standardised analytical method descriptions are listed in Ribeiro et al. (2015, p. 47-53).**"

The typo in the URL has been corrected.

**RC2:** The only details I would urge the authors to include a fuller presentation and description of in this manuscript is the flags of data quality (i.e. the flags referred to in the methods in section 2 or more specifically on page 6 line 27). How are included and expressed in the database?

**AC2:** These flags are processed in the central WoSIS server database, and gradually updated as the various steps of data processing/standardisation proceed (see Figure 2); this type of information is especially important for the database manager (see Ribeiro et al. 2015, p. 90 for table profile, p. 91 for profile\_attribute, and p. 93 for profile\_layer.attribute). Initially, all data entered/provided are assigned level of trust A for 'As entered, no validation (by ISRIC)'. Upon their standardisation, data are coded as having 'level of trust B'. Ultimately, but only upon full harmonisation, the trust level can be upgraded to level-C (see section 2.7, final paragraph).

For profile locations, we provide a measure for the positional accuracy:

"A measure for the positional accuracy is provided **for each profile (e.g., '0.01' when degrees, minutes and seconds are provided, see Ribeiro et al. 2015, p. 90).**"

**RC3:** I have found that an unexpectedly important variable to make sure a dataset finds the use it deserves from the scientific community is that it is user-friendly and easy to access. Researchers that are highly pressed for time may opt to use older data they have at hand if new data requires too much formatting before use. In light of this my only substantial comment regarding data presentation is that the authors should consider providing one file that contains all data in one large matrix/table. I realize that such a file would likely be too large for e.g. 32-bit MS Excel to handle smoothly, but many users will access the data using other softwares that would have no problem handling it. This would save many users the work of writing individual scripts that combine the different data layers for any individual profile. This will reduce the risk of unintentional processing errors and likely increase the usability of the dataset.

**AC3:** We greatly value this particular suggestion as user-friendliness is indeed a key factor. The WoSIS team has therefore re-considered the issue: various alternatives to present the standardised data (in addition to the 'dynamic' or latest WFS-set) were explored. The probably best option was then implemented as further described below and in the revised manuscript. Concerning the format we now use tab-delimited TXT-files; this is the most common and widely accepted format that can be opened by any application in any operating system (see ReadMe file in zipped data set).

In brief, we considered four options for grouping and exporting data from WoSIS in TXT format:

1. Flexible way (25 tables)
2. Easy way (23 tables)
3. Everything-in-one file (1 table)
4. New suggestion (3 tables)

Option 1, the 'flexible' way, would be to distribute the data using the minimal number of TXT-files corresponding with tables in a database. This option would give the following tables (files), one TXT-file for unique profiles (through profile\_id), another one for layers per profile (through unique profile\_id and layer\_id) and one TXT-file for each soil property, per soil layer and profile under consideration. Next, these data would have to be joined using the primary keys.

Option 2, the 'easy' way, corresponds to the approach that we have followed in the initial manuscript; there is no need to do any joins in order to see the full information about a certain soil property.

Option 3, 'everything-in-one' file, in our view, brings both disadvantages and some advantages. There will be one large file with many columns. With the anticipated increase in number and type of standard data provided from WoSIS in upcoming snapshots this would probably become 'incomprehensible', thus not user-friendly. Another disadvantage would be that for the less commonly measured soil properties there may be numerous empty records. A possible advantage, is that the columns gid, profile\_layer\_attribute\_id, profile\_layer\_id, profile\_id, dataset\_id, license\_type, profile\_code, country\_id, date, top, bottom, geom\_accuracy, geom would appear only once and need not be repeated in every TXT file (as is presently the case for the snapshot, see option 2). Then we would just need to add 7 columns (attribute, unit, method, licence, source\_database\_id, original profile code, and observation date) for each soil property. Such a large file would be "difficult to handle".

Option 4, our 'new suggestion', requires 3 tables. The first contains the profile ids 'plus information about the soil classification. The second is a 'long' file listing all soil properties (and descriptors) in sequence. The third file is essentially a compact 'look up' table file with all the codes, descriptions, and units of measurement. The 4-letter coding system is based on SOTER conventions (Van Engelen and Dijkshoorn, 2013), with some modifications necessary to accommodate for the level of detail/standardisation implemented in the WoSIS database.

Upon due deliberation, we consider option 4 to be the recommended solution for serving TXT-snapshots from WoSIS as it minimises repetition and should be readily understood/accessible by most data users. Also the file can readily be imported into SQL databases or accessed from R and similar.

The new format for serving WoSIS-snapshots is described in a new Appendix B, as well as in the revised 'Readme First' file.

**Specific comments:**

**RC4:** Title: I would consider changing “Serving” to “Providing” which may be a more appropriate term.

**AC4:** Agree, this has been done.

**RC5:** Figure 1. I find the layout counter-intuitive and would prefer to have it flow the other way around, from top to bottom.

**AC5:** Indeed, this has been done.

**RC6:** It would be beneficial to data users to get some more details regarding the procedures used to identify these duplicated profiles? Or a clear reference to another source that describes the procedures in detail.

**AC6:** The text was modified as follows: “Being derived from multiple data sources, some of which are compilations, there is a risk that the same profiles are imported several times into WoSIS *albeit* using different identifiers. Computerised procedures that consider lineage and geographical proximity checks were developed to screen for possible repeated profiles. [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. It 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 \(Ribeiro et al, p. 5-6\).](#)”

**RC7:** Please define what QA/QC is (Page 4, line 19). I also find the meaning of the subsequent sentence unclear. Does “(first) user” refer to the data provider? In that case, does this reasoning imply that data provided by “trusted sources” was screened differently?

**AC7:** All data sources are processed using the same, consistent procedures in WoSIS. The above query has been answered as follows:

“All data sources ~~data~~ are submitted to [the same routine QA/QC checks, building on procedures developed for the WISE \(Batjes, 1995, p. 52-53\) and AfSP \(Leenaars, 2012, p. 125-128\) databases. For example, this includes checks on referential integrity, data types, geo-location, units of expression, domain ranges, as well as possible ‘latitude-longitude inversions’ in profile coordinates. It is assumed that the quality requirements of the \(first\) ~~user~~ data provider are met and that basic quality checks and screening have taken place, this with due consideration for any soil-specific options in the laboratory procedures \(Ribeiro et al., 2015\).](#)”

**RC8:** Page 4 line 27. Here the term “pedon” appears. Is this used as a synonym to “soil profile” or does it have a different meaning to the authors?

**AC8:** Profile is meant here, this has been changed.

**RC9:** Sections 2.5 and 2.6 clearly state that: “... more details can be found in the procedures manual of the ISRIC World Information Service (Ribeiro et al., 2015).” This is already mentioned in section 2.7.

**AC9:** Indeed repetitive; this will be (has been) removed in the upcoming revised manuscript.

**RC10:** Page 5 line 19. “In first instance“ could be replaced by “In the first generations of WoSIS“ or similar, (provided that this is what you mean).

**AC10:** This has been reworded: “[As indicated, in the first version we limited ...](#)”.

**RC11:** Page 5 line 21. Here, the variables carbon and organic carbon content are not listed. This is different from organic matter content which is listed.

**AC11:** Changed to organic carbon.

**RC12:** I would suggest that you complement figure 4 with a table that summarizes the amount of profiles available from different key geographical regions (e.g. countries with more than a thousand profiles or continents).

**AC12:** This has been addressed as follows in the text: “The number of profiles per continent is as follows: Africa (17,153), Antarctica (9), Asia (3,089), Europe (1,908), North America (63,077), Oceania (235) and South America (8,970). Details by continent and country are provided in Appendix C.”

**RC13:** I found section 4 highly interesting and informative, and hopefully coming generations of interoperability may include more geographical regions (notably Asia and the Americas).

**AC13:** Thank you for your comment. The next WoSIS-derived snapshot will consider shared data sets for Australia, Cambodia and Canada. However, we have not yet been able to process this information as has now been indicated in the text at the end of Section 3. See also the acknowledgment list on our website under “2) *Data providers whose contributions still have to be processed in WoSIS*” <http://www.isric.org/content/wosis-cooperating-institutions-and-experts>.

# WoSIS: ~~Providing~~**Serving** standardised soil profile data for the world

Niels Batjes<sup>1</sup>, Eloi Ribeiro<sup>1</sup>, Ad van Oostrum<sup>1</sup>, Johan Leenaars<sup>1</sup>, Tom Hengl<sup>1</sup>, and Jorge Mendes de Jesus<sup>1</sup>

<sup>1</sup> ISRIC – World Soil Information, Wageningen, 6708 PB, The Netherlands

5 *Correspondence to:* Niels H. Batjes ([niels.batjes@isric.org](mailto:niels.batjes@isric.org))

**Abstract.** The aim of the World Soil Information Service (WoSIS) is to serve quality-assessed, georeferenced soil data (point, polygon, and grid) to the international community upon their standardisation and harmonisation. So far, the focus has been on developing procedures for legacy point data with special attention for the selection of soil analytical and physical properties considered in the *GlobalSoilMap* specifications (e.g. organic carbon, soil pH, soil texture (sand, silt, and clay), coarse fragments (< 2mm), cation exchange capacity, electrical conductivity, bulk density, and water holding capacity). Profile data managed in WoSIS were contributed by a wide range of soil data providers; the data have been described, sampled, and analysed according to methods and standards in use in the originating countries. Hence, special attention was paid to measures for soil data quality and the standardisation of soil property definitions, soil property values, and soil analytical method descriptions. ~~At the time of writing, The present version of the full~~ WoSIS database contained ~~s~~ some 15 118,400 unique “shared” soil profiles of which ~~some over~~ 96,000 are georeferenced within defined limits. In total, this corresponds with over 31 million soil records, of which some 20% have so far been quality-assessed and standardised using the ~~sequential~~ procedures discussed in this paper. The number of measured data for each property varies between profiles and with depth, generally depending on the purpose of the initial studies. Overall, the data lineage strongly determined which data could be standardised with acceptable confidence ~~in accord with WoSIS procedures (as flagged in the database),~~ 20 ~~corresponding with over 4 million records for 94,441 profiles.~~ The publically available data –WoSIS snapshot of July 2016– are persistently accessible from ISRIC WDC-Soils through <http://dx.doi.org/10.17027/isric-wdcsoils.20160003>.

## 1 Introduction

Soil is an important provider of ecosystem services (UNEP, 2012; MEA, 2005). Yet this natural resource, considered to be 25 non-renewable on a human life span, is being threatened (FAO and ITPS, 2015; UNEP, 2014). Worldwide, professionals, scientists, “decision makers and managers must have access to the information they need, when they need it, and in a format they can use” (GEO, 2010). Large numbers of consistent soil profile data of known provenance (lineage) are needed to accurately model and map the status of the world’s soil resources at increasingly detailed resolutions (Omuto et al., 2012; Hengl et al., 2014; Arrouays et al., 2014; FAO and ITPS, 2015).

This paper describes procedures for safeguarding, standardising/harmonising and subsequently serving of consistent world soil data to the international community as developed in the framework of the DataWoSIS (World Soil Information Service) project. In essence, the development of the WoSIS server database may be seen as a sequel to earlier collaborative, but still “stand-alone”, compilations of soil legacy data coordinated by ISRIC such as WISE (Batjes, 2009), SOTER (van Engelen and Dijkshoorn, 2013), and the Africa Soil Profiles database (Leenaars, 2013). Ultimately, WoSIS aims to serve consistent harmonised soil data (point, polygon and grids), derived both from a wide range of shared legacy holdings as well as from recently developed soil spectral libraries (e.g. Viscarra Rossel et al., 2016; Shepherd and Walsh, 2002), in an interoperable mode and this preferably in the setting of a federated global soil information system.

Harmonisation, as defined by the Global Soil Partnership (GSP, Baritz et al., 2014), involves “providing mechanisms for the collation, analysis and exchange of consistent and comparable global soil data and information”. Areas of harmonisation include those related to: a) soil description, classification and mapping, b) soil analyses, c) exchange of digital soil data, and d) interpretations. So far, seen the breadth and magnitude of the task, the focus in WoSIS has been on the standardisation of soil property definitions, soil analytical method descriptions and soil property values for those properties considered in the GlobalSoilMap specifications (GlobalSoilMap, 2013). Such a standardisation is a prerequisite for the development/testing of a soil information model that can underpin global soil data interoperability and modelling (Omuto et al., 2012). Quality assessed profile data served from WoSIS, and its predecessors as discussed above, may be used for various purposes such as conventional respectively digital mapping of soil properties and soil classes (Batjes, 2016; Hengl et al., 2015; Hengl et al., 2016). In turn, such derived products may be used in studies that address a range of global issues at various scale levels (e.g. Hendriks et al., 2016; Luo et al., 2016; Jones and Thornton, 2015; Maire et al., 2015).

20

## 2 Data and methods

### 2.1 Basic principles

Everyone may contribute data for inclusion in WoSIS. [Data may be submitted in various ways. Analogue data should be provided using a template with standardised variable names as described in the WoSIS Procedures Manual \(Ribeiro et al., 2015, p. 378-40\). Alternatively, large digital data sets comprising over thousand profiles can be provided to ISRIC as zip files containing both the database, documentation and metadata. Prior to any data processing at ISRIC, the data provider must agree in writing with the terms of the Data Policy \(ISRIC, 2016\).](#)

The access rights and data provenance (lineage), as documented in the metadata, will determine which quality assessed-data may later be served freely to the international community. Therefore, when processing the wealth of contributed data, priority is given to those datasets that have a “non-restrictive” Creative Commons licence, defined here as at least a CC BY (Attribution) or CC BY-NC (Attribution Non-Commercial). The corresponding source data will be gradually standardised

30

and harmonised to make them “comparable as if assessed by a given (reference) method” (Fig. 1). Ultimately, only the quality-assessed and standardised/harmonised “shared” data will be served to the international community.

<< Insert Figure 1 >>

## 5 2.2 Measures for data quality

As indicated, soil profile data submitted for consideration in WoSIS were collated according to various national or international standards. Therefore, proper documentation, in so far as possible, of the provenance and identification of each dataset, and ideally each observation or measurement, is essential to allow for efficient processing of the data; such aspects are discussed later in detail.

10 For soil observations and measurements, the following need to be specified: feature (x-y-z and time (t) referenced profiles and layers), attribute (class, site, layer-field, and layer-lab), method, and value, including units of expression (Leenaars, 2013; Leenaars et al., 2014; Ribeiro et al., 2015). As indicated by Chapman (2005), “too often, data are used uncritically without consideration of the error contained within, and this can lead to erroneous results, misleading information, unwise environmental decisions and increased costs”. WoSIS is being populated using data produced for different types of studies  
15 ranging from routine soil surveys to more specific assessments, each of these having their specific quality requirements (Landon, 1991; Soil Survey Division Staff, 1993). The corresponding samples were analysed in a range of laboratories or in the field according to a wide range of methods (e.g. wet chemistry or soil spectroscopy), each with their own uncertainty and costs. As indicated by Kroll (2008), issues of soil data quality are not restricted to uncertainty issues; they also include aspects like completeness, accessibility and verifiability (traceability) of data.

20 A review of quality aspects specifically related to soil data led to consideration of ~~three~~four quality indicators in WoSIS: (a) observation date (date of observation or measurement), (b) level of trust (a subjective measure inferred from soil expert knowledge), and (c) ~~data quality rating (based on expert judgement), and (d)~~ accuracy (an indicator for the laboratory and field-related uncertainty as well as the accuracy of georeferencing). These indicators provide measures that allow  
investigators the WoSIS database manager to recognise factors that may compromise the quality of certain data and hence  
25 their suitability for use. Consideration of such quality indicators ensures that objective methods are applied for evaluating data in the database, while at the same time the system enables soil expert knowledge to override these assessments when needed. In practice, however, the information provided with some source materials does not allow for a full characterisation of all ~~three~~four indicators. In particular, the accuracy of individual analytical measurements is seldom expressed in the source databases as such information is generally maintained in separate soil laboratory information systems (see van  
30 Reeuwijk, 1998; WEPAL, 2015) to which we have no access. Alternatively, a measure for the positional accuracy is provided for each profile (e.g., “0.01” when degrees, minutes and seconds are provided, see Ribeiro et al. 2015, p. 90).

### 2.3 Standard data model

Sometimes, the source data may be in paper (analogue) format in which case they must first be digitised following certain basic criteria (Ribeiro et al., 2015, p. 37-40; Leenaars et al., 2014, p. 52). Preferably, data entry is done by the data providers, as they best know their data. Basically, this first step amounts to “safeguarding soil data collections” at risk of being lost.

5 This is an important remit of ISRIC as World Data Centre for Soils (WDC-Soils) of the ICSU World Data System.

Any submitted digital data set is first assessed as regards its overall suitability for inclusion in WoSIS (e.g. licence and metadata). After this filtering, the data are imported “as is” into PostgreSQL, an open source database management system.

At this early stage of processing, the data are still organised according to numerous data models, data conventions and data contents. So, the next step (yet first step of standardisation) is to map this disparate data to the WoSIS *standard data* model

10 (Fig. 1); technical details are provided elsewhere (Ribeiro et al., 2015).

[<< Insert Figure 1 >>](#)

### 2.4 Identifying repeated profiles

15 Being derived from multiple data sources, some of which are compilations, there is a risk that the same profiles are imported several times into WoSIS *albeit* using different identifiers. Computerised procedures that consider lineage and geographical proximity checks were developed to screen for possible repeated profiles. [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](#)

20 [proximity check is based on the geographic coordinates. It 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 \(Ribeiro et al. p. 5-6\).](#) Figure 2 serves to show the results of this time-consuming process for four databases: ISIS (2016), the ISRIC Soil Information System (reference collection); WISE, World Inventory of Soil Emission potentials (compilation, Batjes, 2009); SOTER, Soil and Terrain databases (compilation, Van Engelen, 2011); and AfSP, the Africa Soil Profiles database (compilation, Leenaars, 2013). For example, 12,810 profiles are present only in

25 AfSP, 35 are shared among AfSP and ISIS (the original source), 164 are shared between AfSP, WISE and ISIS, and 10 profiles occur in the four databases. In case of duplicate profiles, [all the corresponding data will nonetheless be standardised as described below \(i.e. the “flagged” data are maintained in the WoSIS database\).](#) However, ultimately, [only the profile with the most complete data and detailed lineage will be distributed \(see Sect. 3\)](#) ~~is maintained for further processing.~~

30 << Insert Figure 2 >>



## 2.5 Basic data quality assessment and control

5 ~~For All data sources data~~ are submitted ~~subsequent~~ to ~~the same basic~~ QA/QC checks, building on procedures developed for the WISE (Batjes, 1995, p. 52-53) and AfSP (Leenaars, 2013, p. 125-128) database. For example, this includes checks on referential integrity, data types, geo-location, units of expression, domain ranges, as well as possible “latitude-longitude inversions” in profile coordinates. ~~I~~ it is assumed that the quality requirements of the ~~(first) user data provider~~ are met and that basic quality checks and screening have taken place, this with due consideration for any soil-specific options in the laboratory procedures (Ribeiro et al., 2015). This approach allows users of WoSIS ~~derived data sets~~ to make their own judgement on the quality of individual analytical data, for instance by the assumption that selected data have comparable quality characteristics or an acceptable (inferred) quality compared to their requirements.

10

## 2.6 Standardisation of soil analytical method descriptions

As indicated, there is often no detailed quantitative information on the quality and uniformity of the soil analytical data held in the diverse source databases. Full quality control, including verification of in-profile ~~edon~~ consistencies, requires the data to be harmonised according to an analytical reference method. The foreseen ultimate step of data harmonisation, converting property values assessed with analytical method X to values “as if” assessed by reference method Y, requires an unambiguous identification and definition of the various analytical methods. Therefore, it was first necessary to develop a qualitative procedure to describe the analytical methods, including their method features, in a flexible, yet comprehensive and consistent way.

15

The options selected for the analytical method features in WoSIS are assigned on basis of the descriptions in the respective (database) sources. This implies that information, as interpreted or distilled from the original report (source materials) by the data compilers, was used in WoSIS. In the future, some refinements may still prove possible or necessary should the original materials, such as laboratory manuals, be consulted again.

20

In essence, the WoSIS approach for the qualitative description of soil analytical methods can be seen as complementary to method descriptions used in reports from proficiency tests (NATP, 2015; van Reeuwijk, 1998; WEPAL, 2015). In such tests, results from participants are coded to provide details of the methods applied for a particular grouping (e.g. CEC, cation exchange capacity). As discussed in Ribeiro et al. (2015), the spread of these results may give an indication for the maximum spread in a compiled database.

25

In addition to the method description according to the standardised coding system developed for WoSIS, measures have been allocated for the inferred confidence in each “method conversion” (i.e. from low to high); of necessity, this qualitative assessment is based solely on the information embedded in the “summarised” method descriptions as provided in the various source databases. As indicated, such descriptions have often been generalised from a more detailed source, such as a

30

laboratory manual. Importantly, the provided confidence flags should not be seen as a measure for the quality of a particular laboratory.

## 2.7 Towards the harmonisation of world soil data

5 | Depending on the projected applications, user communities will require specific sets of data. [As indicated, in the first](#)  
| [version instance](#), we limited ourselves to the list of properties considered in the GlobalSoilMap specifications  
(GlobalSoilMap, 2013): soil pH, soil organic ~~matter~~-[carbon](#) content, effective cation exchange capacity, electrical  
conductivity, soil texture (sand, silt, and clay content), proportion of fragments > 2 mm, bulk density and water retention. In  
the respective source databases, these properties were determined using a range of analytical procedures, thus requiring  
10 | standardisation of the soil analytical method descriptions to make them “fit for use” and comparable (Leenaars et al., 2014).  
Key in the approach developed for WoSIS is that “a property is best described by key elements of the (laboratory) procedure  
applied” (Soil Survey Staff, 2011). Similarly, in WoSIS, major features of commonly used methods for determining a given  
soil property are characterised. For soil pH, for example, these are the solution, concentration, ratio (soil/solution), and  
instrument. As indicated, the key component features can be aggregated where considered as being comparable in the  
15 | context of global or regional level data analyses. For example, soil pH data measured in a KCl solution, 1M, at a soil/liquid  
ratio of 1:5, and using a conventional electrode can be aggregated within the group considered to meet the ISO 10390:2005  
criteria for pH-KCl (ISO, 2015). Similarly, the combination KCl solution, 1 M, 1:2.5 soil/liquid ratio, and conventional  
electrode broadly corresponds with ISRIC criteria (van Reeuwijk, 2002). Similar principles were applied for all soil  
| properties under consideration here; ~~methodological details are provided in~~ [as described in](#) Ribeiro et al. (2015, [p. 47-53](#)).  
20 | A next, desired step would be to make the data (e.g. pH, CEC or organic carbon) comparable, “as if” assessed by a single  
given (reference) method. That is, fully “harmonised” and unambiguously defined. However, there is generally no universal  
equation for converting property values from one method to another in all situations (GlobalSoilMap, 2013; Jankauskas et  
al., 2006; Lettens et al., 2007). Basically, this implies that within the framework of the Global Soil Partnership (GSP), for  
example, each regional or continental node will need to develop and apply node-specific conversion functions (towards the  
25 | yet to be defined GSP-adopted standard reference methods, see Baritz et al. 2014), building on comparative analyses using  
say archived soil samples and spectral libraries.

## 3 Serving consistent standardised data

| [The WoSIS server database itself](#) provides an important building block for the spatial data infrastructure (Fig. 3) through  
30 | which ISRIC WDC-Soils will be serving an increasing range of data (point, raster and polygon) to the international  
community (Batjes et al., 2013; Hengl et al., 2016). The most recent set of WoSIS-derived [point](#) data is served “[24/7](#)” via an

OGC-compliant WFS (Web Feature Service) provided by GeoServer instance. These data may be accessed freely via the following webpage: <http://www.isric.org/content/wosis-distribution-set>. By its nature, however, this dataset will be *dynamic* as it will grow when additional point data are processed, additional soil attributes are considered, and/or when possible corrections are required. Therefore, for consistent modelling and citation purposes, we provide static snapshots of the standardised data with clear time stamps, in tab separated values format. Each snapshot will have a unique name and Digital Object Identifier (DOI), for example file *WoSIS\_2016\_July.zip* with doi: [10.17027/isric-wdcsoils.20160003](https://doi.org/10.17027/isric-wdcsoils.20160003).

<< Insert Figure 3 >>

At the time of writing, the WoSIS server database contained some 118,400 unique “shared” soil profiles, out of which over 96,000 are georeferenced within defined limits, corresponding with some 31 million soil records in total. So far, some 20% thereof have been quality-assessed and standardised using the sequential procedure discussed in this paper. As indicated, the number of measured data for each property varies between profiles and with depth, generally depending on the purpose of the initial studies. Therefore, the data lineage strongly determined which specific data could be served with acceptable confidence (as flagged in the central WoSIS database, see Ribeiro et al. 2015, p. 92). By implication, the “July 2016” snapshot only includes standardised data for 94,441 geo-referenced profiles, representing some 455 thousand layers (or horizons). In total, this corresponds with over 4 million records that include both numeric (19, e.g. sand content or soil pH) as well as class (3, e.g. WRB soil classification) properties. The naming conventions, units of measurement, and file structure are described in Appendix A and B, respectively.

The number of profiles per continent is highest for North America (63,077) and Africa (17,153), followed by South America (8,970), Asia (3,089), Europe (1,908), Oceania (235), and Antarctica (9). These profiles come from 148 countries; the average density of observations is 0.7 profiles per 1000 km<sup>2</sup>. The actual density of observations varies greatly, both between countries (Appendix C) and within each country (Fig. 4). Such geographic gaps will be filled gradually in the future, this largely depending though on the willingness or ability of data providers to freely share (some of) their data for consideration in WoSIS. Alternatively, it should be noted here that some recently shared datasets are not yet included in the present snapshot (e.g. Australia, Canada and Cambodia).

<< Insert Figure 4 >>

#### 4 Towards global soil data interoperability

So far, all datasets managed in WoSIS were provided as “stand-alone” databases; as such their content is “static”. Steps are being made towards the development of a federated, and ultimately inter-operable, service or Spatial soil Data Infrastructure (SDI); through which source data are served and updated by the respective data providers and made queryable according to the agreed upon data standards. A first, possible step in this direction -though not yet interoperable- is the exchange of data in a PostgreSQL environment using a Foreign Data Wrapper (FWD). Subsequently, the “transferred” data can be mapped to the WoSIS data model for further standardisation and harmonisation as described earlier. A technically more challenging solution for the worldwide exchange of soil data was implemented during the OGC Soil Data Interoperability Experiment (soilIE).

- 10 SoilIE, undertaken in the second half of 2015, had the objective of developing and testing a soil standard that harmonised existing standards for data exchange defined in Europe and Oceania. During the SoilIE, partners from Europe and Oceania mapped their test data to the SoilML format. Multiple OGC Web Feature Services (WFS) providing data in soilML format were established, allowing for on-line derivation of new data (e.g. using pedotransfer functions). The SoilIE was successful in accessing data in multiple clients (servers) from several soil data providers, each using their own software configurations (Ritchie, 2016). Further collaboration will involve refinements to the SoilML schema, Resource Description Foundation (RDF) vocabularies, linked data, and other remaining issues.

Use of OGC web services and modelling data in XML is necessary for fulfilment of compliance requirements with regional interoperability initiatives (INSPIRE, 2015; GS Soil, 2008; Wilson, 2016). The output of the data can then be customised between different XML standards using Extensible Stylesheet Language (XSL) templates or using server schema mapping.

- 20 The above activities in support of a global soil SDI were initiated by the GlobalSoilMap consortium in Wageningen, 2009, and may be consolidated within the framework of the Global Soil Partnership (FAO-GSP, 2014b, a; IUSS WG-SIS, 2015) and related interoperability efforts in other domains (e.g. Porter et al., 2015; GEOSS, 2012; GODAN, 2015).

#### 5 Data availability

- 25 Version WoSIS\_2016\_July, as described in this paper, is archived for long-term storage at ISRIC – World Soil Information, the World Data Center (WDC) for Soils of the ICSU World Data System; it may be accessed freely through [doi: 10.17027/isric-wdcsoils.20160003](https://doi.org/10.17027/isric-wdcsoils.20160003). [The zip file \(32 Mb\) includes a detailed “readme first” file that describes key aspects of the dataset.](#)

#### 6 Conclusions

- 30 Bringing disparate soil databases from numerous sources under a common standard poses many and diverse challenges. So far, the focus in WoSIS has been on the standardisation of soil property definitions, soil analytical method descriptions and

soil property values in order to serve consistent, quality-assessed data that have been observed or measured according to analytical procedures (aggregates) that are functionally comparable.

5 Future releases of WoSIS-served data will consider a wider selection of soil site and layer properties, assessed by conventional soil analytical procedures as well as by soil spectroscopy. Further, grid and polygon maps will be gradually added to the server database. Each release (snapshot) will be given a unique time stamp and digital object identifier.

10 The WoSIS server database forms an important building block of ISRIC's evolving spatial data infrastructure. Instrumental to enhanced usability of the data served by WoSIS will be the actual harmonisation of soil property values as well as the further standardisation of identifiers and descriptions of soil analytical procedures. Development of corresponding interfaces will allow for the fulfilment of future exchange of, and demands, for global soil information and enable further processing of soil data shared by contributing parties.

WoSIS-related activities are already catalysing institutional collaboration with institutes in Africa, Europe and Latin America. Capacity building and cooperation among (inter)national soil institutes around the world is essential to create and share ownership of the soil information newly derived from the shared data. Also to strengthen the necessary expertise and capacity to further develop and test the world soil information service worldwide.

15

**Appendix A: Naming conventions and descriptions of variables provided in the “WOSIS July 2016” snapshot served from the ISRIC World Soil Information Service**

Code <sup>a</sup>	Attribute	Unit	Profiles	Layers	Description
BDFI	Bulk density fine earth	kg dm <sup>-3</sup>	20 727	105 848	Bulk density of the fine earth fraction < 2 mm (kg/dm <sup>3</sup> )
BDWS	Bulk density whole soil	kg dm <sup>-3</sup>	25 909	153 568	Bulk density of the whole soil including coarse fragments (kg/dm <sup>3</sup> )
TCEQ	Calcium carbonate equivalent total	g kg <sup>-1</sup>	27 809	115 448	The content of carbonate in a liming material or calcareous soil calculated as if all of the carbonate is in the form of CaCO <sub>3</sub> (g/kg in the fine earth fraction < 2 mm); also known as inorganic carbon
CECX	Cation exchange capacity (CEC)	cmol <sub>(c)</sub> kg <sup>-1</sup>	48 461	273 346	Capacity of the fine earth fraction < 2 mm to hold exchangeable cations, estimated by buffering the soil at specified pH (e.g. pH7 or pH8; cmol <sub>c</sub> /kg)
CLAY	Clay total	g kg <sup>-1</sup>	80 082	408 452	Gravimetric content of < 0.002 mm soil material in the fine earth fraction < 2 mm (g/100g)
CFGR	Coarse fragments gravimetric total	10 <sup>-2</sup> g g <sup>-1</sup>	27 050	159 206	Gravimetric content of coarse fragments > 2 mm in the whole soil (g/100g)
CFVO	Coarse fragments volumetric total	10 <sup>-2</sup> cm <sup>3</sup> cm <sup>-3</sup>	37 280	198 534	Volumetric content of the coarse fragments > 2 mm in the whole soil (cm <sup>3</sup> /100cm <sup>3</sup> )
ECEC	Effective cation exchange capacity (ECEC)	cmol <sub>(c)</sub> kg <sup>-1</sup>	23 189	102 665	Capacity of the fine earth fraction < 2 mm to hold exchangeable cations at the pH of the soil (ECEC, cmol <sub>c</sub> /kg). Conventionally approximated by summation of exchangeable bases (Ca <sup>2+</sup> , Mg <sup>2+</sup> , K <sup>+</sup> , and Na <sup>+</sup> ) plus 1 N KCl exchangeable acidity (Al <sup>3+</sup> and H <sup>+</sup> ) in acidic soils
ELCO	Electrical conductivity	dS m <sup>-1</sup>	28 266	120 039	Ability of a 1:x soil water extract to conduct electrical current (EC <sub>x</sub> , dS/m); EC <sub>e</sub> refers to values measured in a saturated soil extract
ORGC	Organic carbon	g kg <sup>-1</sup>	64 118	315 362	Gravimetric content of organic carbon in the fine earth fraction
PHCA	pH CaCl <sub>2</sub>	unitless	39 074	237 756	A measure of the acidity or alkalinity in soils, defined as the negative logarithm (base 10) of the activity of hydronium ions (H <sup>+</sup> ) in a CaCl <sub>2</sub> solution, as specified in the analytical method descriptions (dimensionless)
PHAQ	pH H <sub>2</sub> O	unitless	79 118	407 226	A measure of the acidity or alkalinity in soils, defined as the negative logarithm (base 10) of the activity of hydronium ions (H <sup>+</sup> ) in water (dimensionless)
PHKC	pH KCl	unitless	19 064	88 127	A measure of the acidity or alkalinity in soils, defined as the negative logarithm (base 10) of the activity of hydronium ions (H <sup>+</sup> ) in a KCl solution, as specified in the analytical method descriptions (dimensionless)
PHNF	pH NaF	unitless	4866	24 917	A measure of the acidity or alkalinity in soils, defined as the negative logarithm (base 10) of the activity of hydronium ions (H <sup>+</sup> ) in a NaF solution, as specified in the analytical method descriptions (dimensionless)
SAND	Sand total	10 <sup>-2</sup> g g <sup>-1</sup>	78 402	398 573	Larger than Y mm fraction of the < 2 mm soil material (g/100g); esd (equivalent spherical diameter), Y as specified in the analytical method descriptions
SILT	Silt total	10 <sup>-2</sup> g g <sup>-1</sup>	79 331	406 502	0.002 mm to Y mm fraction of the < 2 mm soil material (g/100g); esd (equivalent spherical diameter), X respectively Y as specified in the analytical method descriptions

Code <sup>a</sup>	Attribute	Unit	Profiles	Layers	Description
CFAO	Soil classification FAO	unitless	24 894	24 894	Classification of the soil profile according to specified edition (year) of the FAO-Unesco Legend, up to soil unit level when available
CSTX	Soil classification Soil taxonomy	unitless	21 614	21 614	Classification of the soil profile according to specified edition (year) of USDA Soil Taxonomy, up to subgroup level when available
CWRB	Soil classification WRB	unitless	24 628	24 628	Classification of the soil profile according to specified edition (year) of the World Reference Base for Soil Resources (WRB), up to qualifier level when available
TOTC	Total carbon	g kg <sup>-1</sup>	14 094	70 687	Gravimetric content of organic carbon and inorganic carbon in the fine earth fraction < 2 mm (g/kg)
WRGR	Water retention gravimetric	10 <sup>-2</sup> g g <sup>-1</sup>	28 701	173 972	Soil moisture content by weight, at the tension specified in the analytical method descriptions (g/100g)
WRVO	Water retention volumetric	10 <sup>-2</sup> cm <sup>3</sup> cm <sup>-3</sup>	17 124	82 695	Soil moisture content by volume, at the tension specified in the analytical method descriptions (cm <sup>3</sup> /100cm <sup>3</sup> )

<sup>a</sup> A full complement of all these attributes is generally not available for many profiles (see text).

## Appendix B: Structure of the “July 2016” WoSIS snapshot

This Appendix describes the structure of the data files presented in the “July 2016” WoSIS snapshot:

- [wosis\\_201607\\_attributes.txt](#)
- [wosis\\_201607\\_profiles.txt](#), and
- [wosis\\_201607\\_layers.txt](#).

The first file lists the four letter code for each attribute, a short explanation, and the units of measurement, respectively (Appendix A). This file also gives the number of profiles and layers in the present snapshot.

The second file lists the unique profile ID (i.e. primary key), country name and ISO country code, geometric accuracy, latitude and longitude (WGS 1984) as well as information on the soil classification system and edition. Depending on the soil classification system used, the number of fields will vary. For example, for US Soil Taxonomy, coded here as “cstx”, these are: order, suborder, great group and subgroup as indicated in the column headings.

The third, largest file, lists all the soil properties by layer and profile to enhance user-friendliness. It starts with:

<u>profile_id</u>	<u>identifier for profile, links to file <a href="#">wosis_201607_profiles</a>.</u>
<u>profile_layer_id</u>	<u>unique identifier for layer for given profile (primary key)</u>
<u>top</u>	<u>upper depth of layer (or horizon)</u>
<u>bottom</u>	<u>lower depth of layer</u>

Subsequently, the following items are listed sequentially per attribute (“xxxx”) as defined under “code” in file [wosis\\_201607\\_attributes.txt](#):

<u>xxxx_value</u>	<u>array listing all values for soil property “xxxx” for the given layer; thus, more than one observation can be reported when available, for example 3 values for ORGC: {1:0.55, 2:1.01, 3:0.85}</u>
-------------------	---

<u>xxxx_value_avg</u>	<u>average, for above (use this value for “routine” modelling)</u>
-----------------------	--

<u>xxxx_method</u>	<u>array listing the method descriptions for each value</u>
<u>xxxx_date</u>	<u>array listing the date of observation for each value</u>
<u>xxxx_dataset_id</u>	<u>abbreviation for source data set (e.g. WD-ISIS).</u>
<u>xxxx_profile_code</u>	<u>code for given profile</u>
<u>xxxx_licenset</u>	<u>licence for given data (CC-BY-NC or CC-BY)</u>



(...) as above, but for the next attribute

All fields in the above files are tab-delimited, while double quotation marks serve as text delimiters; file coding is according to the UTF-8 unicode transformation format. As such, the files can be easily imported into an SQL database or statistical software such as R, after which they may be joined using the unique profile id.

5

10

## Appendix C: Number of profiles by country and continent.

Continent	Country_name	ISO code	N of profiles	Area (km <sup>2</sup> )	Profile density (per 1000 km <sup>2</sup> )
Africa	Algeria	DZ	4	2 308 647	0.002
	Angola	AO	1035	1 246 690	0.830
	Benin	BJ	738	115 247	6.404
	Botswana	BW	898	578 247	1.553
	Burkina Faso	BF	887	273 281	3.246
	Burundi	BI	36	26 857	1.340
	Cameroon	CM	455	465 363	0.978
	Central African Republic	CF	87	619 591	0.140
	Chad	TD	5	1 265 392	0.004
	Congo	CG	70	340 599	0.206
	Côte d'Ivoire	CI	254	321 762	0.789
	Dem. Rep. of the Congo	CD	374	2 329 162	0.161
	Egypt	EG	22	98 2161	0.022
	Ethiopia	ET	1583	1 129 314	1.402
	Gabon	GA	46	264 022	0.174
	Ghana	GH	163	238 842	0.682
	Guinea	GN	62	243 023	0.255
	Guinea-Bissau	GW	18	30 740	0.586
	Kenya	KE	504	582 342	0.865
	Lesotho	LS	33	30 453	1.084
	Liberia	LR	48	96 103	0.499
	Libya	LY	14	1 620 583	0.009
	Madagascar	MG	52	588 834	0.088
	Malawi	MW	2985	118 715	25.144
	Mali	ML	756	1 251 471	0.604
	Mauritania	MR	11	1 038 527	0.011
	Morocco	MA	27	414 030	0.065
	Mozambique	MZ	275	787 305	0.349
	Namibia	NA	62	823 989	0.075
	Niger	NE	488	1 182 602	0.413
	Nigeria	NG	1203	908 978	1.323
	Rwanda	RW	92	25 388	3.624
	Senegal	SN	311	196 200	1.585
	Sierra Leone	SL	12	72 281	0.166
	Somalia	SO	245	632 562	0.387
	South Africa	ZA	649	1 220 127	0.532
	Sudan	SD	116	1 843 196	0.063
	Swaziland	SZ	14	17 290	0.810
	Togo	TG	9	56 767	0.159
	Tunisia	TN	60	155 148	0.387
	Uganda	UG	12	241 495	0.050

Continent	Country_name	ISO code	N of profiles	Area (km <sup>2</sup> )	Profile density (per 1000 km <sup>2</sup> )
	United Republic of Tanzania	TZ	1647	939 588	1.753
	Zambia	ZM	472	751 063	0.628
	Zimbabwe	ZW	319	390 648	0.817
Antarctica	Antarctica	AQ	9	12 537 967	0.001
Asia	Afghanistan	AF	19	641 827	0.030
	Armenia	AM	6	29 624	0.203
	Azerbaijan	AZ	4	164 780	0.024
	Bahrain	BH	2	673	2.970
	Bangladesh	BD	16	139 825	0.114
	Bhutan	BT	80	37 674	2.123
	China	CN	1490	9 345 214	0.159
	Georgia	GE	9	69 785	0.129
	India	IN	139	2 961 118	0.047
	Indonesia	ID	108	1 888 620	0.057
	Iran (Islamic Republic of)	IR	2	1 677 319	0.001
	Iraq	IQ	14	435 864	0.032
	Israel	IL	16	20 720	0.772
	Japan	JP	39	373 651	0.104
	Jordan	JO	40	89 063	0.449
	Lebanon	LB	6	10 136	0.592
	Malaysia	MY	46	329 775	0.139
	Mongolia	MN	7	1 564 529	0.004
	Nepal	NP	141	147 437	0.956
	Oman	OM	9	308 335	0.029
	Pakistan	PK	43	788 439	0.055
	Philippines	PH	68	296 031	0.230
	Republic of Korea	KR	17	99 124	0.172
	Sri Lanka	LK	13	66 173	0.196
	Syrian Arab Republic	SY	66	188 128	0.351
	Taiwan	TW	33	36 127	0.913
	Tajikistan	TJ	5	142 004	0.035
	Thailand	TH	285	515 417	0.553
	Turkey	TR	68	781 229	0.087
	United Arab Emirates	AE	6	71 079	0.084
	Uzbekistan	UZ	8	449 620	0.018
	Yemen	YE	284	453 596	0.626
Europe	Albania	AL	63	28 682	2.197
	Belarus	BY	94	207 581	0.453
	Belgium	BE	190	30 669	6.195
	Bulgaria	BG	45	111 300	0.404
	Czech Republic	CZ	38	78 845	0.482

Continent	Country_name	ISO code	N of profiles	Area (km <sup>2</sup> )	Profile density (per 1000 km <sup>2</sup> )
	Denmark	DK	20	44 458	0.450
	Estonia	EE	123	45 441	2.707
	Finland	FI	24	336 892	0.071
	France	FR	53	548 785	0.097
	Germany	DE	51	357 227	0.143
	Greece	GR	11	132 549	0.083
	Hungary	HU	61	93 119	0.655
	Iceland	IS	11	102 566	0.107
	Ireland	IE	36	69 809	0.516
	Italy	IT	86	301 651	0.285
	Latvia	LV	10	64 563	0.155
	Lithuania	LT	18	64 943	0.277
	Luxembourg	LU	128	2621	48.842
	Netherlands	NL	192	35 203	5.454
	Norway	NO	10	324 257	0.031
	Poland	PL	128	311 961	0.410
	Portugal	PT	35	91 876	0.381
	Republic of Moldova	MD	32	33 798	0.947
	Romania	RO	43	238 118	0.181
	Russian Federation	RU	156	16 998 830	0.009
	Slovakia	SK	40	49 072	0.815
	Spain	ES	42	505 752	0.083
	Sweden	SE	26	449 212	0.058
	Switzerland	CH	10	41 257	0.242
	Ukraine	UA	79	600 526	0.132
	United Kingdom	GB	53	244 308	0.217
North America	Barbados	BB	3	433	6.928
	Belize	BZ	21	21 764	0.965
	Canada	CA	148	9 875 646	0.015
	Costa Rica	CR	55	51 042	1.078
	Cuba	CU	52	110 863	0.469
	Dominican Republic	DO	1	48 099	0.021
	El Salvador	SV	9	20 732	0.434
	Greenland	GL	5	2 165 159	0.002
	Guatemala	GT	22	109 062	0.202
	Honduras	HN	11	112 124	0.098
	Jamaica	JM	76	10 965	6.931
	Mexico	MX	12 223	1 949 527	6.270
	Netherlands Antilles	AN	4	790	5.066
	Nicaragua	NI	26	128 376	0.203
	Panama	PA	25	74 850	0.334
	Puerto Rico	PR	30	8937	3.357
	Trinidad and Tobago	TT	2	5144	0.389

Continent	Country_name	ISO code	N of profiles	Area (km <sup>2</sup> )	Profile density (per 1000 km <sup>2</sup> )
	United States of America	US	50 361	9 315 946	5.406
	United States Virgin Islands	VI	3	352	8.514
Oceania	Australia	AU	142	7 687 634	0.018
	Cook Islands	CK	1	241	4.142
	Fiji	FJ	9	18 293	0.492
	Micronesia (Feder. States of)	FM	14	740	18.917
	New Zealand	NZ	20	270 415	0.074
	Papua New Guinea	PG	31	462 230	0.067
	Samoa	WS	17	2835	5.996
	Solomon Islands	SB	1	28 264	0.035
South America	Argentina	AR	238	2 780 175	0.086
	Bolivia (Plurinational State of)	BO	77	1 084 491	0.071
	Brazil	BR	7842	8 485 946	0.924
	Chile	CL	45	753 355	0.060
	Colombia	CO	166	1 137 939	0.146
	Ecuador	EC	77	256 249	0.300
	French Guiana	GF	7	83 295	0.084
	Guyana	GY	43	211 722	0.203
	Peru	PE	147	1 290 640	0.114
	Suriname	SR	27	145 100	0.186
	Uruguay	UY	131	177 811	0.737
	Venezuela (Bolivarian Rep. of)	VE	170	912 025	0.186
World	World (total)	WD	94 441	137 770 610	0.685

Note: Country names and areas are based on the Global Administrative Layers (GAUL) database, see:

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

## Acknowledgements

The development of WoSIS has been made possible thanks to the contributions and shared knowledge of a steadily growing number of data providers, including soil survey organisations, research institutes and individual experts, whose contributions are gratefully acknowledged. A detailed list of data providers is available at [http://www.isric.org/content/wosis-cooperating-](http://www.isric.org/content/wosis-cooperating-institutions-and-experts)

5 [institutions-and-experts](http://www.isric.org/content/wosis-cooperating-institutions-and-experts) for details.

## References

- 10 Arrouays, D., Grundy, M. G., Hartemink, A. E., Hempel, J. W., Heuvelink, G. B. M., Hong, S. Y., Lagacherie, P., Lelyk, G., McBratney, A. B., McKenzie, N. J., Mendonca-Santos, M. d. L., Minasny, B., Montanarella, L., Odeh, I. O. A., Sanchez, P. A., Thompson, J. A., Zhang, G.-L., and Donald, L. S.: GlobalSoilMap: Toward a Fine-Resolution Global Grid of Soil Properties, *Advances in Agronomy*, 125, 93-134, <http://dx.doi.org/10.1016/B978-0-12-800137-0.00003-0>, 2014.
- 15 Baritz, R., Erdogan, H., Fujii, K., Takata, Y., Nocita, M., Bussian, B., Batjes, N. H., Hempel, J., Wilson, P., and Vargas, R.: Harmonization of methods, measurements and indicators for the sustainable management and protection of soil resources (Providing mechanisms for the collation, analysis and exchange of con-sistent and comparable global soil data and information), *Global Soil Partnership, FAO*. 44 pp., 2014.
- Batjes, N. H.: *World Inventory of Soil Emission Potentials: WISE 2.1 - Profile database user manual and coding protocols*, ISRIC, Wageningen. Tech. Pap. 26, 56 pp., 1995.
- 20 Batjes, N. H.: Harmonised soil profile data for applications at global and continental scales: updates to the WISE database, *Soil Use and Management*, 25, 124-127 <http://dx.doi.org/10.1111/j.1475-2743.2009.00202.x>, 2009.
- Batjes, N. H.: Harmonised soil property values for broad-scale modelling (WISE30sec) with estimates of global soil carbon stocks, *Geoderma*, 269, 61-68, <http://dx.doi.org/10.1016/j.geoderma.2016.01.034>, 2016.
- 25 Chapman, A. D.: *Principles of Data Quality*, version 1.0, Global Biodiversity Information Facility, Copenhagen., Copenhagen. 58 pp., 2005.
- FAO-GSP: *Plan of action for Pillar Four of the Global Soil Partnership : Building an enduring and authoritative global system to monitor and forecast the condition of the Earth's soil resources* FAO Global Soil Partnership, Rome. 33 pp., 2014a.
- FAO-GSP: *Plan of Action for Pillar Five of the Global Soil Partnership: Harmonization of methods, measurements and indicators for the sustainable management and protection of soil resources*, FAO Global Soil Partnership, Rome. 37 pp., 30 2014b.
- FAO and ITPS: *Status of the world's soil resources (SWSR) - Main report*, Food and Agriculture Organization of the United Nations and Intergovernmental Technical Panel on Soils, Rome, 650 pp., 2015.
- GEO: *GEOSS data sharing action plan*, Group on Earth Observations, Geneva. 8 pp., 2010.
- Global Earth Observation System of Systems (GEOSS): <http://www.earthobservations.org/geoss.shtml> access: 14 October 35 2016, 2012.
- GlobalSoilMap: *Specifications Version 1 GlobalSoilMap.net products (Release 2.1)*, 50 pp., 2013.
- GODAN: *How can we improve agriculture, food and nutrition with open data?*, Open Data Institute London. Report UK20150528, 33 pp., 2015.

- GS Soil: Discover INSPIRE compliant harmonised soil data and services, GS Soil Consortium, Paris-Lodon University Salzburg GS Soil 06/2009 -05-2012, 60 pp., 2008.
- Hendriks, C. M. J., Stoorvogel, J. J., and Claessens, L.: Exploring the challenges with soil data in regional land use analysis, *Agricultural Systems*, 144, 9-21, <http://dx.doi.org/10.1016/j.agsy.2016.01.007>, 2016.
- 5 Hengl, T., Mendes de Jesus, J., MacMillan, R. A., Batjes, N. H., Heuvelink, G. B. M., Ribeiro, E. C., Samuel-Rosa, A., Kempen, B., Leenaars, J. G. B., Walsh, M. G., and Gonzalez, M. R.: SoilGrids1km— global soil information based on automated mapping, *PLoS ONE*, 9, e105992, <http://dx.doi.org/10.1371/journal.pone.0105992>, 2014.
- Hengl, T., Heuvelink, G. B. M., Kempen, B., Leenaars, J. G. B., Walsh, M. G., Shepherd, K. D., Sila, A., MacMillan, R. A.,  
10 Mendes de Jesus, J., Tamene, L., and Tondoh, J. E.: Mapping Soil Properties of Africa at 250 m Resolution: Random Forests Significantly Improve Current Predictions, *PLoS ONE*, 10, e0125814, <http://dx.doi.org/10.1371/journal.pone.0125814>, 2015.
- Hengl, T., Mendes de Jesus, J., Heuvelink, G. B. M., Ruiperez Gonzalez, M., Kilibarda, M., Blagotic, A., Shangguan, W., Wright, M. N., Geng, X., Bauer-Marschallinger, B., Guevara, M. A., Vargas, R., MacMillan, R. A., Batjes, N. H., Leenaars, J. G. B., Wheeler, I., Mantel, S., and Kempen, B.: SoilGrids250m: Global gridded soil information based on machine  
15 learning, *PLoS ONE*, ([http://gsif.isric.org/lib/exe/fetch.php?media=wiki:soilgrids250m\\_global\\_gridded\\_preprint.pdf](http://gsif.isric.org/lib/exe/fetch.php?media=wiki:soilgrids250m_global_gridded_preprint.pdf)), 2016.
- INSPIRE Data specifications - Infrastructure for spatial information in the European Community: <http://inspire.ec.europa.eu/index.cfm/pageid/2>, access: 25.04.2016, 2015.
- ISRIC Soil Information System (ISIS): <http://isis.isric.org/>, access: 23 June 2016, 2016.
- ISO 10390:2005: Soil quality -- Determination of pH: [http://www.iso.org/iso/catalogue\\_detail.htm?csnumber=40879](http://www.iso.org/iso/catalogue_detail.htm?csnumber=40879), access:  
20 10 October 2016, 2015.
- ISRIC: Data and Software Policy, ISRIC - World Soil Information (WDC - Soils) Wageningen. 6 pp., 2016.
- IUSS Working Group on Soil Information Standards <http://www.soilinformationstandards.org/>, access: 12 March 2016, 2015.
- Jankauskas, B., Jankauskiene, G., Slepetiene, A., Fullen, M. A., and Booth, C. A.: International comparison of analytical methods of determining the soil organic matter content of Lithuanian eutric Albeluvisols, *Communications in Soil Science and Plant Analysis*, 37, 707-720, <http://dx.doi.org/10.1080/00103620600563499>, 2006.
- 25 Jones, P. G., and Thornton, P. K.: Representative soil profiles for the Harmonized World Soil Database at different spatial resolutions for agricultural modelling applications, *Agricultural Systems*, 139, 93-99, <http://dx.doi.org/10.1016/j.agsy.2015.07.003>, 2015.
- Landon, J. R.: Booker Tropical Soil Manual, Longman Scientific & Technical, New York, 474 pp., 1991.
- 30 Leenaars, J. G. B.: Africa Soil Profiles Database -- A compilation of georeferenced and standardised legacy soil profile data for Sub Saharan Africa (version 1.1), Africa Soil Information Service (AfSIS) and ISRIC - World Soil Information, Wageningen. Report 2013/03, 160 pp., 2013.
- Leenaars, J. G. B., Kempen, B., van Oostrum, A. J. M., and Batjes, N. H.: Africa Soil Profiles Database: a compilation of georeferenced and standardised legacy soil profile data for Sub-Saharan Africa GlobalSoilMap - Basis of the global spatial  
35 information system, Orléans 2014, 51-57,
- Letten, S., Vos, B. D., Quataert, P., van Wesemael, B., Muys, B., and van Orshoven, J.: Variable carbon recovery of Walkley-Black analysis and implications for national soil organic carbon accounting, *European Journal of Soil Science*, 58, 1244-1253, <http://dx.doi.org/10.1111/j.1365-2389.2007.00916.x> 2007.
- Luo, Y., Ahlström, A., Allison, S. D., Batjes, N. H., Brovkin, V., Carvalhais, N., Chappell, A., Ciais, P., Davidson, E. A., Finzi, A., and al., e.: Towards more realistic projections of soil carbon dynamics by Earth System Models, *Global Biogeochem. Cycles*, 30, 40-56, <http://dx.doi.org/10.1002/2015GB005239>, 2016.
- 40 Maire, V., Wright, I. J., Prentice, I. C., Batjes, N. H., Bhaskar, R., van Bodegom, P. M., Cornwell, W. K., Ellsworth, D., Niinemets, Ü., Ordonez, A., Reich, P. B., and Santiago, L. S.: Global effects of soil and climate on leaf photosynthetic traits and rates, *Global Ecology and Biogeography*, 24, 706-715, <http://dx.doi.org/10.1111/geb.12296>, 2015.
- 45 MEA: Ecosystems and Human Well-Being: Synthesis, Millennium Ecosystem Assessment Island Press, Washington, 2005.

- North American Proficiency Testing (NAPT) Program: <http://www.naptprogram.org/>, access: 27.04.2016, 2015.
- Omuto, C., Nachtergaele, F., and Vargas Rojas, R.: State of the Art Report on Global and Regional Soil Information: Where are we? Where to go?, FAO, Italy. 69 pp., 2012.
- Porter, C. H., Villalobos, C., Holzworth, D., Nelson, R., White, J. W., Athanasiadis, I. N., Zhang, M., Janssen, S., Knapen, R., Jones, J. W., Boote, K. J., Hargreaves, J., and Antle, J. M.: Data Interoperability Tools for Regional Integrated Assessments, in: Handbook of Climate Change and Agroecosystems, 147-171, 2015.
- Ribeiro, E., Batjes, N. H., Leenaars, J. G. B., Van Oostrum, A. J. M., and Mendes de Jesus, J.: Towards the standardization and harmonization of world soil data: Procedures Manual ISRIC World Soil Information Service (WoSIS version 2.0) ISRIC - World Soil Information, Wageningen. Report 2015/03, 110 pp., 2015.
- Ritchie, A.: OGC Soil Data Interoperability Experiment, Open Geospatial Consortium, 74 pp., 2016.
- Shepherd, K. D., and Walsh, M. G.: Development of reflectance spectral libraries for characterization of soil properties, Soil Science Society of America Journal, 66, 988–998, 2002.
- Soil Survey Division Staff: Soil survey manual, Soil Conservation Service, U.S. Department of Agriculture, Washington, 503 pp., 1993.
- Soil Survey Staff: Soil Survey Laboratory Information Manual (Ver. 2.0), National Soil Survey Center, Soil Survey Laboratory, USDA-NRCS, Lincoln (NE). Soil Survey Investigation Report No. 45, 506 pp., 2011.
- UNEP: The benefits of soil carbon - managing soils for multiple, economic, societal and environmental benefits, in: UNEP Yearbook - Emerging issues in our global environment 2012, United Nations Environmental Programme, Nairobi, 19-33, 2012.
- UNEP: UNEP Yearbook - Emerging issues in our global environment 2014, UNEP Yearbooks, United Nations Environmental Programme, Nairobi, 68 pp., 2014.
- Van Engelen, V. W. P.: Standardizing soil data (*e-SOTER* regional pilot platform as EU contribution to a Global Soil Information System), International Innovation, June, 48-49, 2011.
- van Engelen, V. W. P., and Dijkshoorn, J. A.: Global and National Soils and Terrain Digital Databases (SOTER) - Procedures manual (Ver. 2.0), IUSS, ISRIC and FAO, Wageningen. ISRIC Report 2013/04, 191 pp., 2013.
- van Reeuwijk, L. P.: Guidelines for quality management in soil and plant laboratories, FAO, Rome. 143 pp., 1998.
- van Reeuwijk, L. P.: Procedures for soil analysis (6th ed.), ISRIC, Wageningen. Technical Paper 9, 81 pp., 2002.
- Viscarra Rossel, R. A., Behrens, T., Ben-Dor, E., Brown, D. J., Demattê, J. A. M., Shepherd, K. D., Shi, Z., Stenberg, B., Stevens, A., Adamchuk, V., Aichi, H., Barthès, B. G., Bartholomeus, H. M., Bayer, A. D., Bernoux, M., Böttcher, K., Brodský, L., Du, C. W., Chappell, A., Fouad, Y., Genot, V., Gomez, C., Grunwald, S., Gubler, A., Guerrero, C., Hedley, C. B., Knadel, M., Morrás, H. J. M., Nocita, M., Ramirez-Lopez, L., Roudier, P., Campos, E. M. R., Sanborn, P., Sellitto, V. M., Sudduth, K. A., Rawlins, B. G., Walter, C., Winowiecki, L. A., Hong, S. Y., and Ji, W.: A global spectral library to characterize the world's soil, Earth-Science Reviews, doi: 10.1016/j.earscirev.2016.1001.1012, <http://dx.doi.org/10.1016/j.earscirev.2016.01.012>, 2016.
- WEPAL. Wageningen Evaluating Programmes for Analytical Laboratories: [http://www.wepal.nl/website/about\\_wepal/History.htm](http://www.wepal.nl/website/about_wepal/History.htm), access: 28.05.2016, 2015.
- Wilson, P. L.: Global harmonisation and exchange of soil data, SciDataCon 2016: Advancing the Frontiers of Data in Research, Denver (11-13 September 2016), Colorado, USA, 2016.



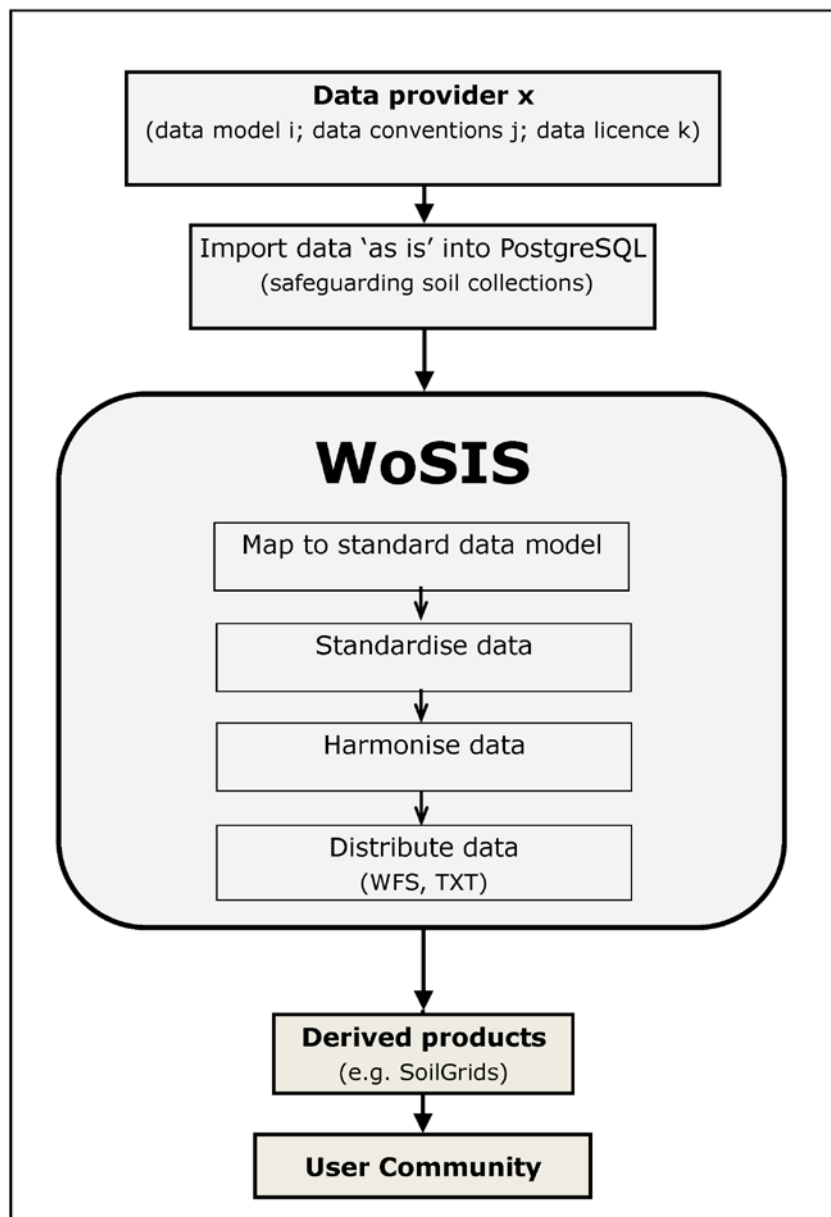
## Figure captions

**Figure 1.** General procedure for processing data in WoSIS

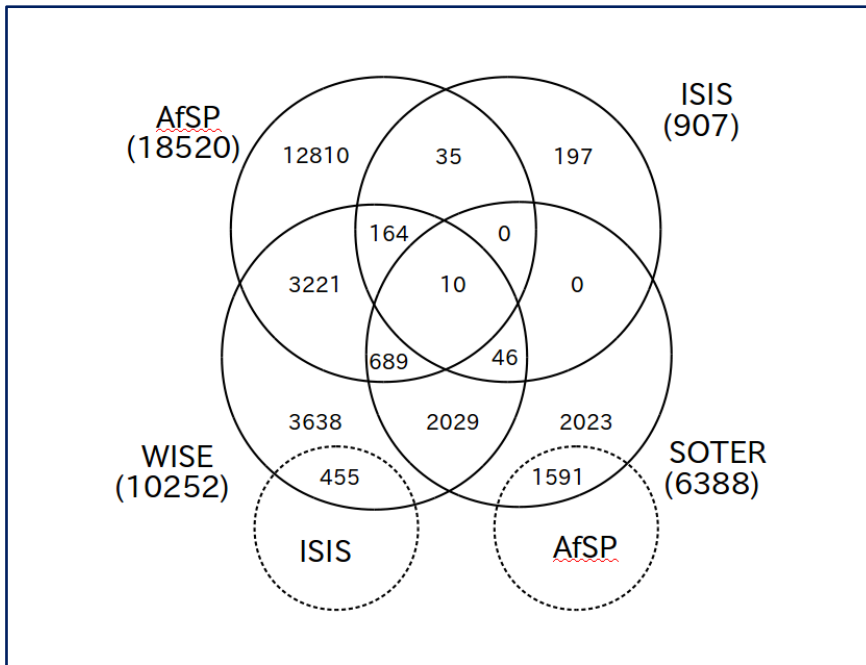
- 5 **Figure 2.** Flagging of repeated profiles between the AfSP, ISIS, WISE and SOTER databases (see text for explanation of abbreviations)

**Figure 3.** Serving consistent soil layers from WoSIS to the user community through ISRIC's evolving spatial data infrastructure

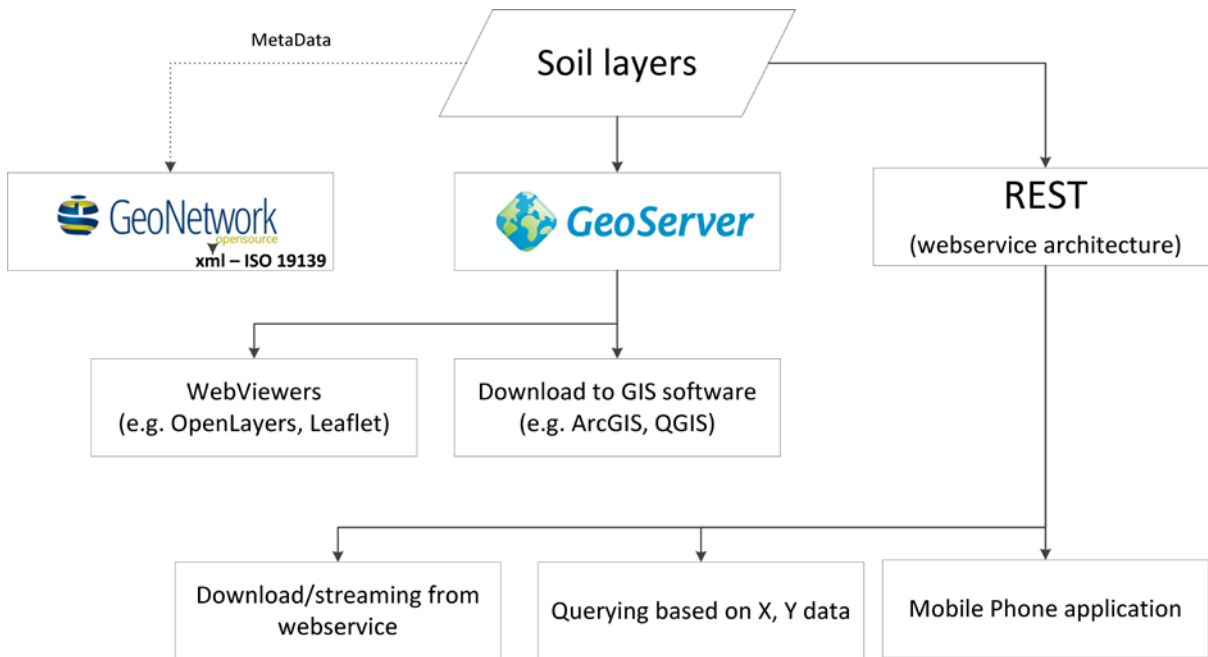
- 10 **Figure 4.** Location of [fp](#) soil profiles [provided](#) in the "July 2016" snapshot of WoSIS.  
[\(See Appendix C for the number of profiles by country\)](#)



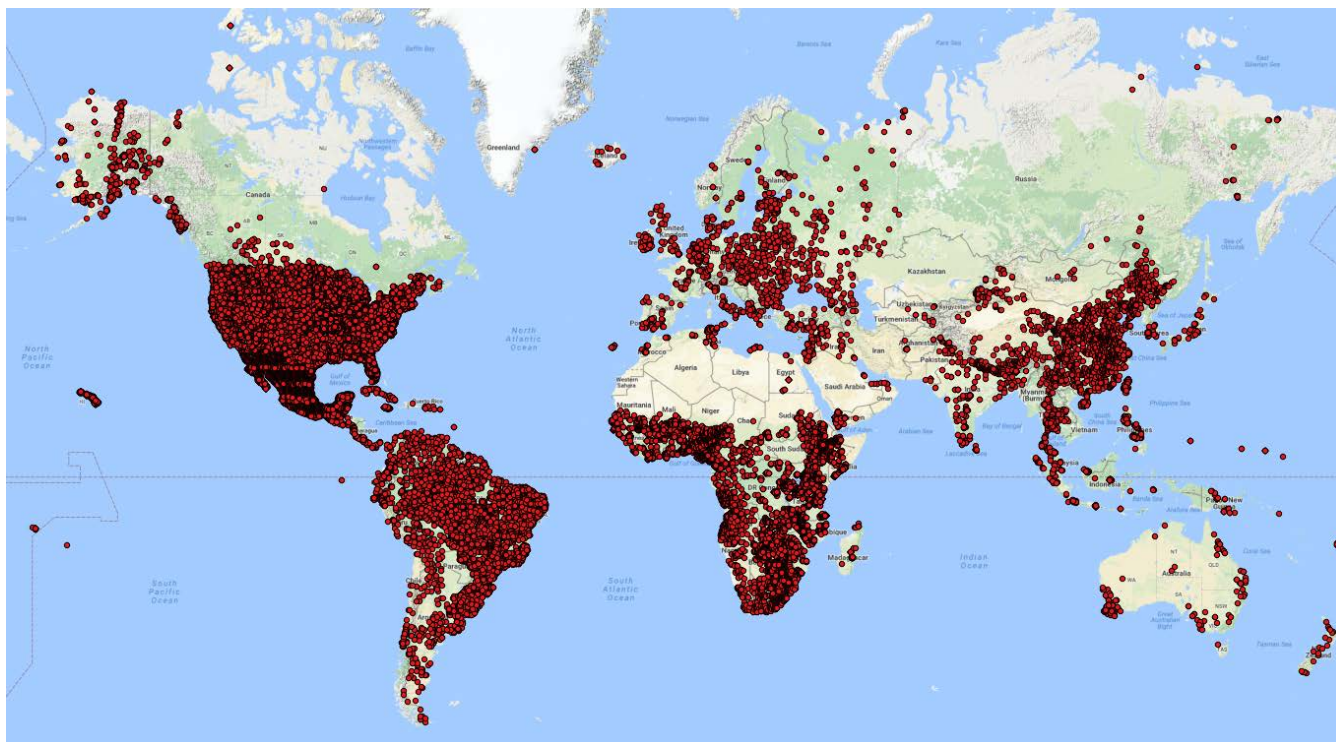
**Figure 1.** General procedure for processing data in WoSIS



**Figure 2.** Flagging of repeated profiles between the AfSP, ISIS, WISE and SOTER databases (see text for explanation of abbreviations)



**Figure 3.** Serving consistent soil layers from WoSIS to the user community through ISRIC's evolving spatial data infrastructure



5

**Figure 4.** Location of  $p$  soil profiles [provided](#) in the “July 2016” snapshot of WoSIS.  
[\(See Appendix C for the number of profiles by country\)](#)