Earth Syst. Sci. Data, 16, 4735–4765, 2024
https://doi.org/10.5194/essd-16-4735-2024

Open Access Earth System Science Data

# Providing quality-assessed and standardised soil data to support global mapping and modelling (WoSIS snapshot 2023)

**Niels H. Batjes, Luis Calisto, and Luis M. de Sousa**

ISRIC – World Soil Information, Droevendaalsesteeg 3, 6708 PB Wageningen, the Netherlands

**Correspondence:** Niels H. Batjes (niels.batjes@isric.org)

Received: 18 January 2024 – Discussion started: 27 March 2024
Revised: 20 August 2024 – Accepted: 30 August 2024 – Published: 22 October 2024

**Abstract.** Snapshots derived from the World Soil Information Service (WoSIS) are served freely to the international community. These static datasets provide quality-assessed and standardised soil profile data that can be used to support digital soil mapping and environmental applications at broad scale levels. Since the release of the preceding snapshot in 2019, refactored ETL (extract, transform and load) procedures for screening, ingesting and standardising disparate source data have been developed. In conjunction with this, the WoSIS data model was overhauled, making it compatible with the ISO 28258 and Observations and Measurements (O&M) domain models. Additional procedures for querying, serving and downloading the publicly available standardised data have been implemented using open software (e.g. GraphQL API). Following up on a short discussion of these methodological developments we discuss the structure and content of the "WoSIS 2023 snapshot". A range of new soil datasets was shared with us, registered in the ISRIC World Data Centre for Soils (WDC-Soils) data repository and subsequently processed in accordance with the licences specified by the data providers. An important effort has been the processing of forest soil data collated in the framework of the EU-HoliSoils project. We paid special attention to the standardisation of soil property definitions, description of the soil analytical procedures and standardisation of the units of measurement. The 2023 snapshot considers soil chemical properties (total carbon, organic carbon, inorganic carbon (total carbonate equivalent), total nitrogen, phosphorus (extractable P, total P and P retention), soil pH, cation exchange capacity and electrical conductivity) and physical properties (soil texture (sand, silt and clay), bulk density, coarse fragments and water retention), grouped according to analytical procedures that are operationally comparable. Method options are defined for each analytical procedure (e.g. pH measured in water, KCl or $CaCl_2$ solution, molarity of the solution, and soil / solution ratio). For each profile we also provide the original soil classification (i.e. FAO, WRB and USDA system with their version) and pedological horizon designations as far as these have been specified in the source databases. Three measures for "fitness for intended use" are provided to facilitate informed data use: (a) positional uncertainty of the profile's site location, (b) possible uncertainty associated with the operationally defined analytical procedures and (c) date of sampling. The most recent (i.e. *dynamic*) dataset, called *wosis_latest*, is freely accessible via various web services. To permit consistent referencing and citation, we also provide a *static* snapshot (in this case, December 2023). This snapshot comprises quality-assessed and standardised data for 228 000 georeferenced profiles. The data come from 174 countries and represent more than 900 000 soil layers (or horizons) and over 6 million records. The number of measurements for each soil property varies (greatly) between profiles and with depth, this generally depending on the objectives of the initial soil sampling programmes. In the coming years, we aim to gradually fill gaps in the geographic distribution of the profiles, as well as in the soil observations themselves, this subject to the sharing of a wider selection of "public" soil data by prospective data contributors; possible solutions for this are discussed. The WoSIS 2023 snapshot is archived and freely available at https://doi.org/10.17027/isric-wdcsoils-20231130 (Calisto et al., 2023).

## 1 Introduction

The World Soil Information Service (WoSIS) draws on a large complement of soil profile data that have been shared by numerous data providers. Nonetheless, a large proportion of the 800 000 "so-called" freely available soil profiles (see Arrouays et al., 2017), in practice, still remain "inaccessible" due to various licence constraints (e.g. Cornu et al., 2023). Soil data submitted for consideration in WoSIS come from a wide range of legacy holdings (e.g. traditional soil surveys) and increasingly include data derived from proximal sensing (e.g. Shepherd et al., 2022; Viscarra Rossel et al., 2016). The source data come in various formats and were determined according to a range of field sampling and soil analytical procedures, requiring standardisation and harmonisation during their ingestion/processing into WoSIS.

Prior to discussing the "2023 snapshot", we provide a short retrospective of activities that led to the development of WoSIS. In the early days of desktop computers, ISRIC with its partners compiled a range of project-specific databases such as ISIS (van de Ven and Tempel, 1994), created to manage data for the ISRIC World Soil Reference Collection; several national- and continental-scale SOil and TERrain (SOTER) databases (e.g. FAO and ISRIC, 2003; FAO et al., 2007, 1998); the WISE database (Batjes, 1997; Batjes and Bridges, 1994); and the Africa Soil Profiles Database (AfSP) (Leenaars et al., 2014). While these different databases were structured along the general principles and criteria of the FAO Guidelines for Soil Description (FAO, 1977, 2006) and USDA Soil Survey Manual (Soil Survey Division Staff, 1993), the ISIS, SOTER, WISE and AfSP databases each had their own data models and conventions. Further, out of necessity at the time, the databases were developed and implemented on stand-alone computers using a range of commercial software packages. In 2009, ISRIC management decided to bring the above stand-alone products together in a centralised enterprise database, known as WoSIS (World Soil Information Service), developed using PostgreSQL with the PostGIS extension for handling spatial data. After the initial ingestion and standardisation of the above "ISRIC holdings", the service was to be expanded with datasets shared by a diverse range of soil data providers.

The original aim of WoSIS was to accommodate any type of soil data (profile, vector and grid) (Ribeiro et al., 2015; Tempel et al., 2013). However, from 2015 onwards, in view of technical considerations and institutional developments, the scope of WoSIS was changed to "safeguarding, processing, standardising and serving geo-referenced soil profile (point) data for the world" (Ribeiro et al., 2020). Alternatively, vector and grid maps derived from traditional soil mapping (e.g. Batjes, 2016; Dijkshoorn et al., 2005; FAO et al., 2012; van Engelen et al., 2006) and digital soil mapping (e.g. Hengl et al., 2017; Poggio et al., 2021; Turek et al., 2023) would be managed and served through other components of our spatial data infrastructure, such as the ISRIC Data Hub (https://data.isric.org, last access: 24 April 2024) and the SoilGrids/WoSIS portal (https://soilgrids.org, last access: 24 April 2024). All these web services were developed using free and open-source software (FOSS).

The ultimate goal of WoSIS, like for related global data compilation activities (Baritz et al., 2017; de Sousa et al., 2019), is full data harmonisation (Batjes et al., 2020; Ribeiro et al., 2015, 2020). According to the Global Soil Partnership (GSP; Baritz et al., 2014), harmonisation involves "providing mechanisms for the collation, analysis and exchange of consistent and comparable global soil data and information" and considers the following domains: (a) soil description, classification and mapping; (b) soil analyses; (c) exchange of digital soil data; and (d) interpretations. In view of the breadth and magnitude of the task, as well as the limited availability of comparative "multiple analytical procedures" datasets as required for full harmonisation (Batjes, 2023; Bispo et al., 2021; van Leeuwen et al., 2022), we have limited ourselves to the standardisation of soil property definitions, soil analytical procedure descriptions, plausibility checks for soil observation values and the standardisation of measurement units for commonly required soil properties (see Appendix A). Importantly, users should always keep in mind that the source datasets themselves (e.g. Armas et al., 2023; NPDB, 2023; USDA-NCSS, 2021) will provide more detailed information than WoSIS, albeit not in a consistent, globally standardised format.

This paper discusses methodological changes to the WoSIS workflow and new data additions since the release of the previous snapshot (Batjes et al., 2020). First, we describe the new data model and the refactored data screening/ingestion process and indicate how the "shared" data are being served to the user community upon their standardisation. Thereafter, we describe the actual data screening, quality control and standardisation process. Subsequently, we describe the spatial distribution of soil profile sites and list the number of soil observations represented in the "WoSIS 2023 snapshot" (hereafter referred to as the 2023 snapshot). In conjunction with this, we provide three measures for "fitness for intended use" of the standardised data and discuss possible limitations of the snapshot. Finally, following up on a discussion concerning the scope for "full data harmonisation" in WoSIS, future developments and possible constraints arising are outlined.

The naming conventions and standard units of measurement are listed in Appendix A, while the structure of the snapshot files is described in Appendix B. In Appendix C we list the number of sites by country/area and continent (Table C1) as well as their distribution by World Terrestrial Ecosystems (Table C2) and biomes (Table C3).

Soils are important providers of ecosystem services (FAO and ITPS, 2015). WoSIS-served data have been used for a range of applications, such as predictive soil property mapping (Guevara et al., 2018; Moulatlet et al., 2017; Nenkam et al., 2022; Poggio et al., 2021; Turek et al., 2023), space and time modelling of soil organic carbon stock change (Heuvelink et al., 2021), and a diverse range of environmental assessments (e.g. Hassani et al., 2024; Huang et al., 2024; Luo et al., 2021; Lutz et al., 2019; Maire et al., 2015; Sanderman et al., 2017; Sothe et al., 2022). For example, based on the “2016 snapshot” and “2019 snapshot” respectively, Ivushkin et al. (2019) mapped global soil salinity change, while Wang et al. (2024) analysed responses of soil organic carbon under warming across global biomes. Ultimately, such information can help to inform the global conventions such as the UNCCD (United Nations Convention to Combat Desertification) and UNFCCC (United Nations Framework Convention on Climate Change) so that policymakers and business leaders can make informed decisions about the environment, biodiversity and human well-being at an appropriate scale level.

## 2 WoSIS data model and workflow

### 2.1 Workflow

The data model and workflow for acquiring, ingesting, processing and serving data as described in Ribeiro et al. (2020) were overhauled. This proved necessary as this procedure was essentially designed as a series of dataset-specific Python and SQL scripts, which was adequate as long as WoSIS was still relatively small. However, in view of the rapidly growing population of shared soil data and overall complexity of the data model itself, it proved necessary to implement a new, state-of-the-art ISO domain model (de Sousa, 2023; de Sousa et al., 2023), with refactored ETL (extract, transform and load) procedures, to ultimately better serve our diverse user community in our capacity as the World Data Centre for Soils (WDC-Soils).

The main stages of the new workflow are visualised in Fig. 1: (a) data providers share their data with ISRIC WDC-Soils; (b) the submitted datasets with associated metadata are screened for “completeness of information provided” (e.g. the licence defining access rights and description of terms and units) and, once considered adequate, subsequently stored “as is” in the WDC-Soils data repository (see “ISRIC Admin” in Fig. 1); and (c) the source datasets are imported into the new WoSIS PostgreSQL relational database (see Sect. 2.2), using refactored ETL procedures (see Sect. 2.3). Step (c) includes (c1) basic data quality assessment and control, (c2) standardising descriptions for the soil analytical procedures and units of measurement, and (c3) automated checks against plausibility limits for each soil observation; see Sect. 3 for details. Subsequently, (d) the quality-assessed and standardised data are distributed via various services such as dashboards and WFS (OpenGIS web feature service) as well as a metadata catalogue service.

### 2.2 Data model

As indicated earlier, a new data model for WoSIS was developed, aligned where possible with the ISO 28258 domain model (de Sousa et al., 2023) and the GloSIS web ontology (Palma et al., 2024), both stemming from O&M (Observations and Measurement; Cox and David, 2011), all the while preserving legacy data. Main features of interest are the dataset (describes source of data), site (geo-spatial location where a soil investigation took place) and profile (sequence of pedogenetic horizons along the depth of the profile). The key modification vis à vis the previous data model (Ribeiro et al., 2020) is the conditioning of analytical methods to the observation (see https://git.wur.nl/isric/databases/wosis-docs, last access: 26 April 2024). Changes made to the database schema and data over time are tracked using a migration tool (https://github.com/graphile/migrate, last access: 24 April 2024). It maintains a record of the history; state; and dependencies of the database, including the conversion to the new data model.

Special attention was paid to the succinct description of the analytical procedures (see c2 above) using seven database tables, as summarised below:

- `thes_method_value`. The thesaurus of values that match the keys is used to define an analytical method, for example, “natural clod” for the “sample type” key for the “bulk density” method.

- `thes_method_key`. This is the thesaurus of keys that is used to define an analytical method, for example, “reported pH”, “exchange solution” and “index cation”.

- `thes_method_option`. This encodes the possible combinations of key–value pairs for each numerical observation on layers. Note that only a small subset of observations can be associated with particular method options.

- `method_source`. Analytical method descriptions are as defined in the respective source databases (i.e. prior to standardisation). This table was imported as is from the old data model (Ribeiro et al., 2020) with the addition of a synthetic primary key. The records in this table remain essential to identify the method referred by each result.

- `method_standard`. This distinguishes each source method by the particular observation to which it applies. It can be regarded as a standardised description of the source method. Each record corresponds to a collection of key–value pairs in the `method_option` table for a single observation. Results for numerical observations

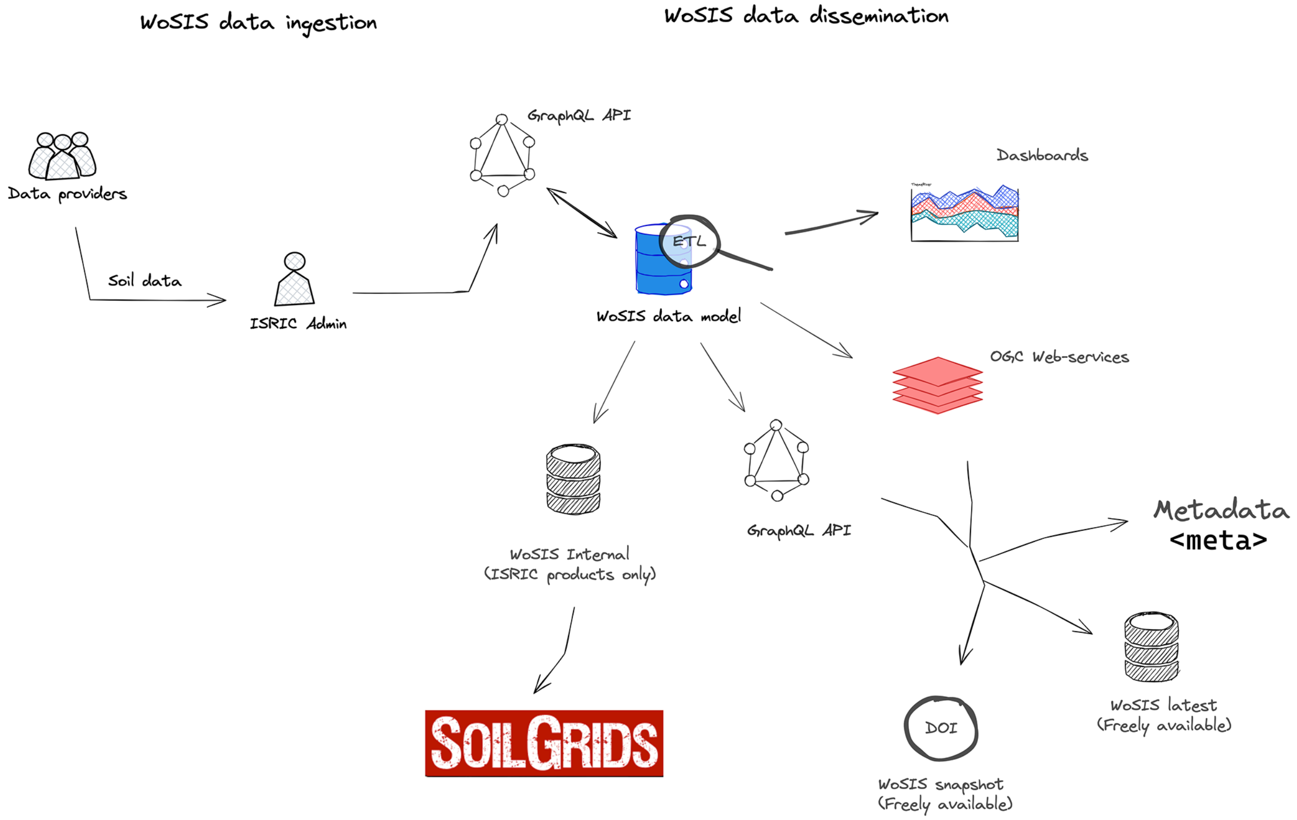

**Figure 1.** Schematic WoSIS workflow for ingesting, standardising and distributing soil profile data.

reference this table to identify the corresponding analytical method.

- `method_option_standard`. This defines a many-to-many relationship between the `method_standard` and `method_option` tables. It determines the exact collection of key–value option pairs that constitute a standard method. The standard method is a specialisation of the source method for a specific observation.

## 2.3 ETL procedure

Extract, transform and load (ETL) is a standardised, semi-automatic process that guides the data processor during the ingestion of new datasets. Refactored ETL procedures were developed to align with the structure of the ISO data model. During the initial phase, newly shared datasets are submitted to a quick consistency check (i.e. format, data model, metadata and licence) after which they are uploaded as is to a staging area in the WoSIS system. Subsequently, during the transform stage the uploaded datasets are parsed by the system. During this process, validation and standardisation occur (see Sect. 3.3 for details). In the case of (possible) unconformities the system will generate descriptive messages that guide the data processor towards possible actions that may be needed to resolve the flagged unconformities. The data processor then needs to correct these issues in conformity with the requirements of the WoSIS procedure manual in steps guided by the system; in some cases the original data providers may need to be consulted. At the end of this phase, the cleaned and standardised data remain in the staging area for final verification by a soil expert. After this verification, the final stage of the ETL process, “load”, can start. This is a fully automated process during which the cleaned and standardised data are copied into the WoSIS database and subsequently removed from the staging area (note that the source data themselves are permanently preserved in the ISRIC data repository). The newly ingested data can now be used to create a range of WoSIS-derived products (e.g. *wosis_latest*, *wosis_internal* and dashboards; see Fig. 1) in accordance with the licences and possible restrictions specified by the data providers.

## 2.4 Operational definitions

Soil characteristics, such as texture, bulk density and organic carbon content, are collated according to a wide range of procedures in different countries. For such incongruent data to be interpreted correctly during the ETL process, the procedures for their collection, analysis and reporting need to be well documented and understood. Results can differ when different analytical procedures are used, even though these procedures may carry the same name (e.g. clay, silt and sand size fraction) or concept (see Soil Survey Staff, 2011). This makes the inter-comparison of different datasets difficult if it is not known how these data were collected/analysed. Therefore we use "operational definitions", as defined by USDA Soil Survey Staff (2011), for soil properties that are linked to specific analytical procedures. To properly characterise the "pH of a soil", for example, we need information on sample pre-treatment, soil / solution ratio and description of solution (e.g. $H_2O$, 1 M KCl, 0.02 M $CaCl_2$ or 1 M NaF). Soil pH measured in sodium fluoride (pH NaF), for example, provides a measure for the phosphorus (P) retention of a soil, whereas pH measured in water (pH $H_2O$) is an indicator for soil nutrient status. Consequently, in WoSIS, soil properties are named according to and defined by the analytical procedures and corresponding "method options", based on common practice in soil science (e.g. BDFIOD for "bulk density (BD), fine earth fraction (FI), oven dry (OD)"). The current list of soil properties standardised in WoSIS is described in Sect. 3.3.

## 2.5 Data provisioning

Upon completion of the semi-automated ETL process, the quality-assessed and standardised data are distributed freely through various channels (see Fig. 1), this in accordance with the license agreements (see Sect. 2.6):

- Through *wosis_latest* (*dynamic*), the data are served via WFS, the respective endpoints are catalogued at the ISRIC Data Hub (https://data.isric.org/geonetwork/srv/eng/catalog.search#/search?any=wosis_latest, last access: 26 April 2024).
- The data are also served as "fixed" snapshots (in TSV format) with a unique digital object identifier (DOI) to permit consistent citation (https://data.isric.org/geonetwork/srv/eng/catalog.search#/search?any=wosis_snapshot, last access: 26 April 2024).
- The contents of *wosis_latest* can also be visualised using a dashboard with some querying and zooming facilities (https://dashboards.isric.org/superset/dashboard/wosis_latest, last access: 26 April 2024).
- Profile data from *wosis_latest* can also be queried through the "SoilGrids web platform" (https://soilgrids.org/, last access: 26 April 2024), which also provides access to a range of soil property maps derived from the WoSIS-served profile data and a set of environmental covariates using digital soil mapping (Poggio et al., 2021; Turek et al., 2023).
- The *wosis_latest* holdings can also be queried using a GraphQL interface (https://graphql.isric.org/, last access: 26 April 2024) that facilitates exploration of the data (e.g. select data for organic carbon, bulk density and proportion of coarse fragments per layer (horizon) for profiles located in a given geography). Results of such tailor-made queries can then be exported as input in scripting languages such as Python or R (R Core Team, 2021), for example to calculate regional carbon stocks.

## 2.6 Licence agreements

It is not a simple task to find potential providers of "open" soil data (Arrouays et al., 2017; Batjes, 2009; Cornu et al., 2023). This may be due to technical issues, access arrangements, reasons for sharing (e.g. "Why share the data and for what purpose? What is in it for us?"), and legal requirements (Bispo et al., 2021; Robinson et al., 2019). All datasets that are shared with our centre are first registered in the ISRIC Data Repository together with their metadata; data sharing agreements should align with the ISRIC Data Policy (ISRIC, 2016). During the subsequent WoSIS standardisation workflow, we are faced with three different types of datasets: first, those with a non-restrictive Creative Commons (CC-BY) licence, defined here as at least a CC-BY (Attribution) or CC-BY-NC (Attribution Non-Commercial) licence (these are later served as *wosis_latest*); second, datasets with a more "restrictive" licence in the sense that they can exclusively be used for "visualisations", such as SoilGrids™ (i.e. *wosis_internal*; see Fig. 1), by ISRIC itself (the latter generally because the coordinates cannot be disclosed as stipulated by certain data providers; for details see https://www.isric.org/explore/wosis/wosis-contributing-institutions-and-experts, last access: 26 April 2024); and finally, several datasets with licences that stipulate that they should only be safeguarded in the ISRIC repository and cannot be used for any data processing (i.e. permanent embargo).

The number of profiles in WoSIS per licence category, i.e. "public" and "restricted", can be viewed and filtered using a dashboard (https://dashboards.isric.org/superset/dashboard/wosis_licenses/, last access: 26 April 2024). As shown in Table 1, the number of "public access" profiles served from WoSIS as snapshots increased from 96 000 in 2016 to 228 000 in 2023. Conversely, it should be noted here that a large proportion of the forest soil data shared in the framework of the EU-HoliSoils project, for instance, could not be included in the "2023 snapshot" due to licence restrictions specified by the data providers. As a result, only 34 000 out of the total of 107 000 profiles shared with ISRIC between 2019 and 2023 could actually be included in the 2023 snapshot (i.e. *wosis_latest*).

**Table 1.** Number of soil profiles and properties served in successive WoSIS snapshots.

| Snapshot | No. of profiles | No. of properties* |
|---|---|---|
| Jul 2016 | 96 000 | 22 |
| Sep 2019 | 196 000 | 45 |
| Dec 2023 | 228 000 | 45 |

* Property names are based on "operational definitions", i.e. a combination of a property and procedure in the terminology of the WoSIS data model (see Sects. 2.4 and 3.3).

## 3 Data screening, quality control and standardisation

### 3.1 Consistency checks

Soil profile data shared for possible consideration in WoSIS were sampled and analysed according to various national or international standards and presented in various formats (from paper to digital). They are of varying degrees of completeness as discussed below. To be considered in the WoSIS standardisation workflow (Fig. 1), each soil profile must meet several criteria, as described earlier in Batjes et al. (2020, p. 301). In summary, they must be associated with a site correctly geo-referenced, have consistently defined upper and lower depths for each layer (or pedogenetic horizon), and have observations for at least some of the soil properties that are being served (e.g. sand, silt, clay and pH) as well as a succinct description of the analytical procedures and units of measurement. A soil (taxonomic) classification is considered desirable though not mandatory. Profiles associated with a valid site, for which only the classification is specified in the source data, can still be useful for mapping of soil taxonomic classes.

Consistency in layer depth (i.e. sequential increase in the upper and lower depth reported for each layer down the profile) is checked using automated procedures (see Sect. 3.2). In line with current internationally accepted conventions, such depth increments are given as "measured from the soil surface, including organic layers and mineral covers" (FAO, 2006; IUSS Working Group WRB, 2022; Schoeneberger et al., 2012; Soil Survey Staff, 2022b). Until 1993, however, the beginning (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, 1990). Organic horizons were recorded as above and mineral horizons recorded as below, relative to the mineral surface (Schoeneberger et al., 2012, p. 2–6). As far as possible, such "organic_surface" layers are flagged in the snapshot (see Appendix B) so that they may be filtered out during auxiliary computations of soil organic carbon stocks, for example.

### 3.2 Screening for duplicate profiles

In the early stage of WoSIS, many source databases were compilations of shared soil profile data necessitating intricate procedures for identifying and flagging possibly repeated profiles (see Batjes et al., 2017; Ribeiro et al., 2020). Soil profiles located within 100 m of each other are flagged as possible duplicates, provided the year of sampling is identical (this criterion allows for reporting results of soil monitoring campaigns at the same site). Upon additional automated checks concerning the thickness of the first three soil layers (i.e. upper and lower depth), sand, silt and clay content, the duplicate profiles with the least-comprehensive component of observations are flagged and excluded from further processing (i.e. distribution). When still in doubt after these rigorous tests, a final visual "similarity check" is made with respect to other commonly reported soil properties such as $\mathrm{pH_{water}}$ and organic carbon content, possibly leading to the flagging (exclusion) of some additional profiles.

### 3.3 Standardisation of property names, analytical procedure descriptions and units of measurement

A crucial step during data ingestion is the standardisation of the, regularly non-English, soil property names used in the source databases to the WoSIS conventions, as well as the standardisation of the soil analytical procedures according to consistent "operational definitions" (see Appendix A). Subsequently, the units of measurement are standardised and the reported measurement values assessed according to soil-observation-specific plausibility ranges for the respective soil properties (i.e. likely minimum and maximum). Some of these plausibility limits may change when more data become available for soil observations that are so far under-represented, similar to ICP Forests (2020, p. 25), and appropriate PostgreSQL "trigger mechanisms" have been implemented for this. Data that do not meet these conditions are flagged and not processed further in the ETL workflow (see above), unless the observed "inconsistencies" can easily be solved (e.g. blatant typos in pH values). Alternatively, the data provider(s) may be contacted to resolve the observed errors.

Similar to the 2019 snapshot, the following soil properties are considered in the 2023 snapshot:

- *Chemical*. Properties include total carbon (i.e. organic plus inorganic carbon), organic carbon, inorganic carbon (i.e. total carbonate equivalent), total nitrogen, soil pH, cation exchange capacity, electrical conductivity and phosphorus (extractable P, total P and P retention).
- *Physical*. Properties include soil texture (clay, silt, sand), coarse fragments, bulk density and water retention.

All measurement values are served as recorded in the source data, after the above consistency checks and standardisation of the units of measurement to the target units (see Appendix A). As such, we *do not* apply any "gap-filling" procedures during ETL, nor do we apply any pedotransfer functions (PTFs) to derive missing bulk density data or soil hydrological properties or harmonise particle class size limits to a common standard, for example. This follow-up stage of data processing is seen as the task of the data users (modellers) themselves. In practice, the required PTFs or ways for depth-aggregating the layer data will be determined by the projected use(s) of the standardised data (see Finke, 2006; Heuvelink et al., 2021; Poggio et al., 2021; Turek et al., 2023; van Leeuwen et al., 2024; Van Looy et al., 2017). It should be noted, however, that inadvertently some PTF-derived values (e.g. for bulk density) could have slipped through the above consistency checks in situations where procedures were miscoded in the metadata of a source dataset; critical modellers should exclude such values during their analyses.

### 3.4 Providing measures for fitness for intended use

As indicated earlier, data served from WoSIS are used for a wide range of environmental applications (e.g. Guevara et al., 2018; Heuvelink et al., 2021; Luo et al., 2021; Maire et al., 2015; Moulatlet et al., 2017; Poggio et al., 2021; Sanderman et al., 2017; Sothe et al., 2022; Turek et al., 2023), but many of these assessments do not explicitly consider the uncertainties that are associated with the data. However, it is well known that "soil observations used for calibration and interpolation are themselves not error-free" (e.g. Baroni et al., 2017; Cressie and Kornak, 2003; Folberth et al., 2016; Grimm and Behrens, 2010; Guevara et al., 2018; Heuvelink, 2014; van Leeuwen et al., 2022). Therefore, since 2019, we have provided three measures for fitness for intended use in *wosis_latest*, namely (a) positional uncertainty of the profiles (i.e. site location), (b) inferred accuracy of the laboratory measurements and (c) date of sampling. These three measures, although approximative, should be duly considered in digital soil mapping and subsequent earth system modelling as they can affect the prediction uncertainty and "area of applicability" of the resulting derived products (Dai et al., 2019; Meyer and Pebesma, 2021; Shi et al., 2023). For example, large areas of the globe are still poorly represented in WoSIS (basically the yellow areas in Fig. 3). As indicated earlier, this issue can only be remedied when a larger selection of datasets is shared by the international soil community for consideration in WoSIS.

Importantly, prospective data users should also realise that the point/profile data shared for consideration in WoSIS are largely based on purposive sampling. During such "traditional" surveys, soil surveyors identify sample locations based on their knowledge of the survey area, desired level of detail (scale) and objective of the survey, for example detailed or exploratory surveys (FAO, 2006; IUSS Working Group WRB, 2022; Soil Survey Staff, 2017). Hence, such "legacy" data are not based on a probabilistic sampling scheme, as recommended for digital soil mapping (Brus et al., 2011; Brus, 2022; Cramer et al., 2019; Heuvelink et al., 2007).

**Table 2.** Positional uncertainty of profile site locations.

| Positional uncertainty | Number of profiles | |
|---|---|---|
| | $n$ | % |
| $\sim$ 100 m | 195 554 | 86 |
| 100 m–1 km | 21 653 | 9 |
| 1–10 km | 3846 | 2 |
| Over 10 km | 7037 | 3 |

#### 3.4.1 Positional uncertainty

Profiles in WoSIS are geo-referenced through the site in which they were sampled in accordance with ISO 28258 standards (de Sousa et al., 2023). The coordinates themselves are presented according to the World Geodetic System datum ensemble (i.e. WGS84, EPSG code 4326) upon their conversion from a diverse range of national projections. For most profiles (86 %; see Table 2) the approximate positional uncertainty of the profile locations, as inferred from the coordinates given in the source datasets, is $\sim$ 100 m. Typically, georeferencing before the advent of GPS (Global Positioning Systems) in the 1970s is less accurate; often we just do not know the "true" accuracy. Nonetheless, digital soil mappers should be aware of this issue (Grimm and Behrens, 2010) because the soil observations and environmental covariates may not actually overlap (Cressie and Kornak, 2003), both in space and time.

#### 3.4.2 Measurement uncertainty

Soil data managed in WoSIS have been analysed according to a diverse range of analytical procedures in multiple laboratories. A measure for measurement uncertainty is thus desired. Soil-laboratory-specific quality management systems and laboratory proficiency testing (PT) can provide this type of information (GLOSOLAN, 2023; Magnusson and Örnemark, 2014; Munzert et al., 2007; NATP, 2015; WEPAL, 2019). Calculation of laboratory-specific measurement uncertainty for a single procedure, as well as multiple analytical procedures, will require several measurement rounds (years of observation) and solid statistical analyses (van Leeuwen et al., 2022). Generally, however, this type of information is not provided with the source datasets submitted to the ISRIC data repository. Therefore, pragmatically, we have distilled the required information from the PT literature (Al-Shammary et al., 2018; ICP Forests, 2021a; Kalra and Maynard, 1991; Rayment and Lyons, 2011; Rossel and McBratney, 1998; van Reeuwijk, 1983; WEPAL, 2019), as far as technically feasible. In the case of organic carbon content, for example, the mean variability was 17 % (with a range of 12 % to 42 %) and for "CEC buffered at pH 7" it was 18 % (range 13 % to 25 %) when multiple laboratories analyse a standard set of reference materials using similar operational procedures (WEPAL, 2019).

The figures for measurement accuracy presented in Appendix A represent first approximations. They are derived from the inter-laboratory comparison of analyses on well-homogenised reference samples for a still relatively small range of soil types. These indicatory figures should be refined, for example, using probability distribution functions (Heuvelink et al., 2007; van Leeuwen et al., 2022), once sufficient laboratory and procedure-related accuracy (i.e. systematic and random error) information is provided with the shared soil data (Magnusson and Örnemark, 2014). Alternatively, this type of information may be collated in the context of international laboratory PT networks such as GLOSOLAN and WEPAL and in the framework of the ongoing LUCAS topsoil monitoring round (Bispo et al., 2021; Cornu et al., 2023). Meanwhile, the present first estimates can already be considered when calculating the uncertainty of predictive digital soil maps and of any interpretations derived from them (e.g. studies of soil organic carbon stock change).

Realistically, full harmonisation of analytical data derived from disparate sources, the ultimate ambition in WoSIS, will first become feasible once results of a representative set of multi-procedure, inter-laboratory comparison datasets become (freely) available, as discussed by Baritz et al. (2014), Bispo et al. (2021) and Batjes (2023), and a common set of reference standard operating procedures (SOPs) has been accepted as a global standard.

### 3.4.3 Year of sampling

For each profile site, the date of sampling has been recorded as far as documented in the source data. This information is important to consider when superimposing the profile data with environmental covariates, such as land cover, for example, in the context of space and time analyses (Giller et al., 2006; Heuvelink et al., 2021). Most (54 %) profiles represented in the snapshot were described/sampled between 1980 and 2020 (Table 3) and less than 4 % before 1960. Alternatively, the date of site description and sampling is not known for almost 27 % of the profiles as the information was not provided in the source materials.

**Table 3.** Period of sampling/analysis.

| Period | *N* of profiles | Percentage |
|---|---:|---:|
| < 1920 | 37 | 0 |
| 1920–1940 | 253 | 0.1 |
| 1940–1960 | 8632 | 3.8 |
| 1960–1980 | 35 358 | 15.5 |
| 1980–2000 | 75 686 | 33.2 |
| 2000–2020 | 47 768 | 20.9 |
| Not specified | 60 356 | 26.5 |

## 4 Spatial distribution of soil profiles and number of observations

### 4.1 Spatial distribution

The 2023 snapshot includes standardised data for 228 000 profiles, sampled at 217 000 different sites (Fig. 2). The greatest number of profiles comes from North America (35 %), followed by Oceania (19 %) and Europe (17 %), while there are still few profiles for Asia (3 %) and Antarctica (Table 4). The profiles come from sites in 174 countries. The average density of observations varies greatly both between countries (Table C1) and within each country/area.

**Table 4.** Number of soil profiles per continent.

| Continent | Number of profiles | | |
|---|---:|---:|---:|
| | 2023 snapshot | 2019 snapshot | 2016 snapshot |
| Africa | 32 198 | 27 688 | 17 153 |
| Antarctica | 35 | 9 | 0 |
| Asia | 7763 | 6704 | 3089 |
| Europe | 39 728 | 35 311 | 1908 |
| North America | 78 996 | 73 604 | 63 066 |
| Oceania | 43 013 | 42 918 | 235 |
| South America | 26 457 | 10 218 | 8790 |

Changes in the spatial distribution and density of profiles (per 1000 km$^2$) in the successive WoSIS snapshots (Fig. 3) reflect the degree to which our data acquisition efforts were successful, as further discussed in Sect. 6. Overall, the density of soil observations is still low for central Asia, southeast Asia, central and eastern Europe, Russia, and the northern circumpolar region in the 2023 snapshot.

The number of profiles by biome (R. J. Olson et al., 2001) and broad climatic region (Sayre et al., 2014), as derived from GIS overlays, is listed in Tables C2 and C3.

### 4.2 Number and depth of observations

In total, the profiles considered in the 2023 snapshot are described by 0.9 million soil layers (or horizons). This corresponds with over 6.1 million records that include both numeric (e.g. silt content, soil pH and cation exchange capacity) and class (e.g. WRB soil classification and horizon designation) properties. There are more observations for the chemical properties than the physical properties (see Table A1). Further, the number of observations generally decreases with depth, largely depending on the objectives of the original soil surveys. The interquartile range (Q1–Q3) for maximum depth of soil sampled in the field is 33–150 cm, with a me-

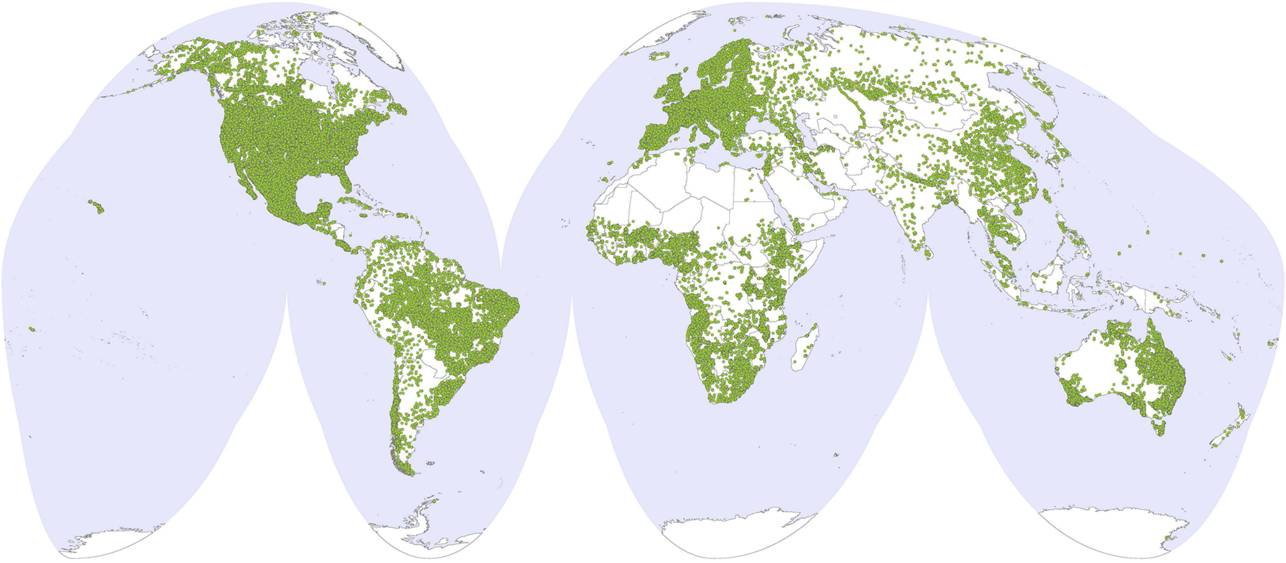

**Figure 2.** Distribution of sites represented in the 2023 snapshot of WoSIS (Goode homolosine equal-area projection).

**Table 5.** Maximum depth of soil sampled per continent.

| Continent | Maximum depth sampled (cm) | | |
|---|---|---|---|
| | $\leq 30$ | 30–60 | $> 60$ |
| Africa | 5307 | 1779 | 22 458 |
| Antarctica | 6 | 7 | 17 |
| Asia | 635 | 505 | 4310 |
| Europe | 9411 | 2848 | 23 195 |
| North America | 6190 | 6728 | 60 698 |
| Oceania | 9216 | 3792 | 27 839 |
| South America | 1730 | 810 | 8477 |

dian (Q2) of 100 cm (mean = 107 cm). It should be noted here that most specific purpose surveys only consider the topsoil (e.g. soil fertility surveys), while others systematically sample soil layers up to depths exceeding 20 m (with a maximum of 32 m). When data from such "specific purpose surveys" (defined here as $< 30$ cm and $> 300$ cm) are excluded, the figures for maximum depth sampled become Q1 = 90 cm, Q2 = 122 cm and Q3 = 155 cm, with a mean of 126 cm.

Table 5 provides an overview of the maximum depth of soil sampled during the various surveys that underpin WoSIS, by continent. Unfortunately, we are not able to show the "depth to bedrock" as this information is seldom made explicit in the source databases.

## 5 Distributing the standardised data

The standardised data are distributed through ISRIC's Spatial Data Infrastructure (SDI). The SDI is based on open-source technologies and open web services (WFS, WMS, WCS, CSW) following Open Geospatial Consortium (OGC) standards and aimed specifically at handling soil data. Our metadata are organised following standards of the International Organization for Standardization (ISO-19139, 2019) using GeoNetwork (see https://data.isric.org, last access: 26 April 2024). The WoSIS database is hosted in a PostgreSQL database, with the spatial extension PostGIS. The PostgreSQL database itself is connected to MapServer to permit data download from GeoNetwork. These processes are aimed at facilitating global data interoperability and citation in compliance with FAIR principles. The data should be "findable, accessible, interoperable and reusable" (Wilkinson et al., 2016).

*Static* snapshots are given a unique DOI (digital object identifier) to permit consistent citation. The 2023 snapshot is distributed in tab-separated values format (see Appendix B for file structure) and as a GeoPackage (https://doi.org/10.17027/isric-wdcsoils-20231130; Calisto et al., 2023). An online Readme file, which includes links to two short tutorials, provides additional technical information (https://www.isric.org/sites/default/files/Readme_WoSIS_202312_v2.pdf, last access: 26 April 2024). Alternatively, the evolving dynamic version of the standardised data (i.e. *wosis_latest*) can be accessed/queried through the ISRIC Data Hub (https://data.isric.org, last access: 26 April 2024) and the SoilGrids platform (https://soilgrids.org, last access: 26 April 2024). Tutorials describing how to access *wosis_latest* from QGIS using WFS and with GraphQL (Calisto, 2023) can be found on the ISRIC website (see https://www.isric.org/explore/wosis/faq-wosis, last access: 26 April 2024).

By its nature, the dynamic version will grow when new profile data are shared and processed, additional soil properties are considered in the WoSIS workflow,

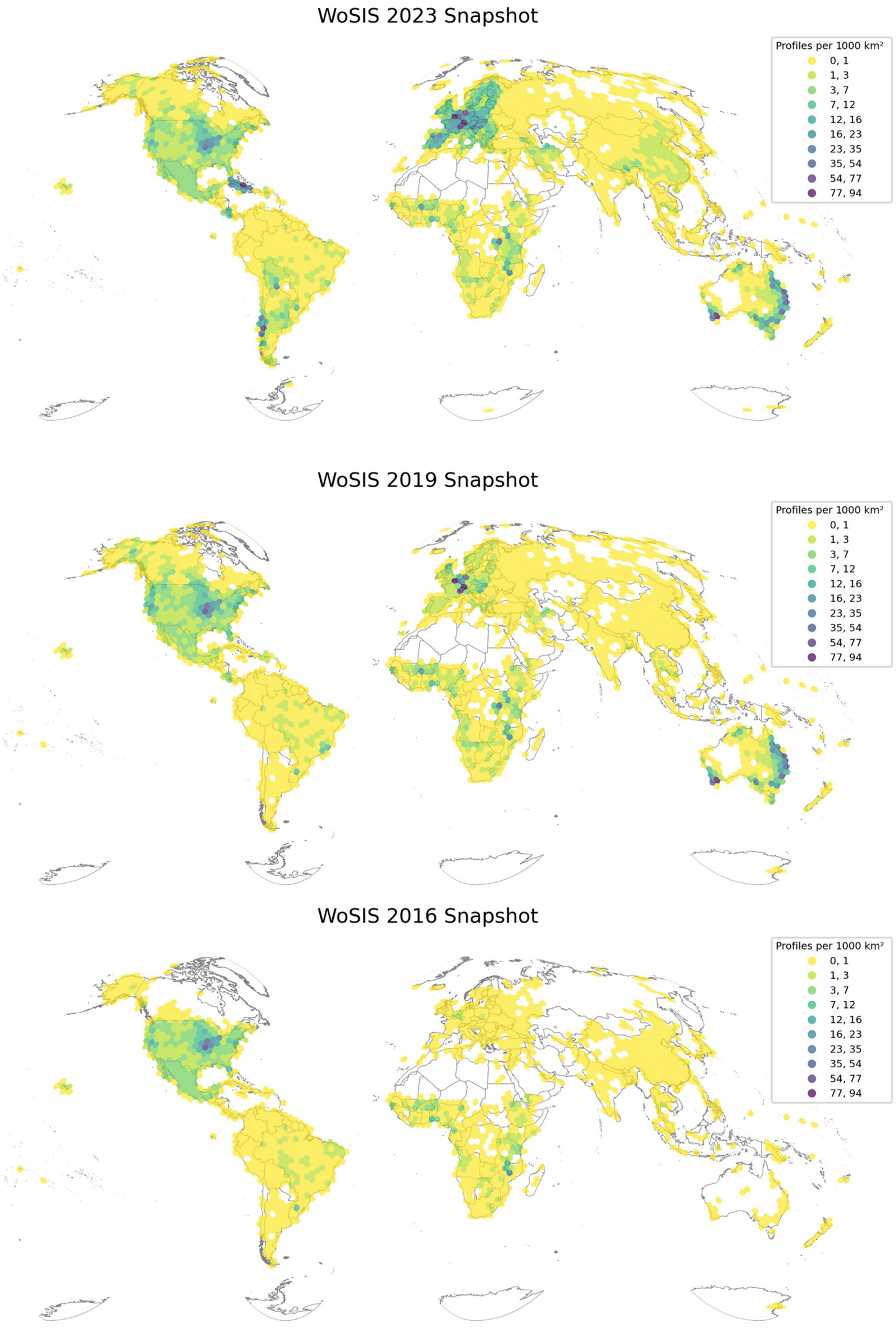

**Figure 3.** Density and spatial distribution of profiles served with the 2016, 2019 and 2023 WoSIS snapshots.

and/or when possible corrections are required. Potential errors can be reported via a "Google group" (https://groups.google.com/forum/#!forum/isric-world-soil-information, last access: 26 April 2024) so that these may be addressed in the dynamic version.

## 6 Discussion

We describe new procedures for handling and standardising disparate world soil profile data in WoSIS. The data model was fully harmonised to ISO 25828 and O&M requirements, with minor adjustments, and refactored ETL procedures were implemented. Alternatively, it should be stressed that the ultimate, desired full harmonisation of observations to an agreed reference analytical procedure $Y$, for example, "pH $H_2O$, 1 : 2.5 soil/water solution" for, say, all "pH 1 : $x$ $H_2O$" measurements, will first become feasible once the target procedure ($Y$) for analysing each property has been defined and subsequently accepted as a "global standard" by the international soil community. A next step would be to collate/develop "comparative" datasets for each soil property (i.e. sets with samples analysed according to a given reference procedure ($Y_i$) and the corresponding national procedures ($X_j$)) for pedotransfer function development. These relationships, however, will often be soil-type- and region-specific (GlobalSoilMap, 2015) and difficult to develop (i.e. calibrate and validate) when datasets for the comparisons do not yet exist or are simply not freely shared/available (Batjes, 2023; Bispo et al., 2021; Cornu et al., 2023; van Leeuwen et al., 2024). Hence, regional laboratory inter-comparison programmes, such as those undertaken in the framework of, for example, ANSIS (2023), GLOSOLAN (2023), ICP Forests (2021a) and LUCAS (Bispo et al., 2021), which aim to develop consistent, context-specific (e.g. by country or land use/soil type) pedotransfer functions towards an agreed set of SOPs, are important. However, it should be noted that the standard type of SOPs specified by these various programmes need not be comparable. In this context, Suvannang et al. (2018) observed that "comparable and useful soil information (at the global level) will only be attainable once laboratories agree to follow common standards and norms". Over the years, however, many organisations/countries have implemented analytical procedures and quality assurance systems that are well suited for their specific purposes (e.g. ANSIS, 2023; Cornu et al., 2023; Orgiazzi et al., 2018; Soil Survey Staff, 2022a). Consequently, they may not be inclined to harmonise their data to a (still to be decided) set of global "reference" SOPs. However, agreed-upon procedures for such a full-scale harmonisation will be required when developing a globally federated, and ultimately interoperable, spatial soil data infrastructure (GLOSIS, de Sousa et al., 2021) through which (pre-harmonised) source data are served and updated by the respective data providers and made queryable according to a common standard (de Sousa et al., 2023; OGC, 2019).

It is our intention to gradually fill gaps in the geographic distribution (Fig. 3) and range of soil properties (Appendix A) in the coming years. This work is part of ISRIC's remit as a regular member of the World Data System (https://worlddatasystem.org, last access: 26 April 2024). The degree to which this will be feasible, however, will largely depend on the willingness and ability of data providers to share (some of) their data for consideration in WoSIS. For the northern boreal and Arctic region, for example, ISRIC can draw on new profiles collated by the International Soil Carbon Network (ISCN; see Malhotra et al., 2019). Alternatively, it should be reiterated that several datasets in our repository (e.g. ICP Forests, 2021a) can *only* be standardised and used for SoilGrids™ applications due to existing licence restrictions. Conversely, some countries such as Aotearoa / New Zealand distribute their national soil profile dataset with a CC-BY-ND 4.0 licence, which implicitly precludes making any derivatives; hence they cannot be considered in WoSIS (see https://viewer-nsdr.landcareresearch.co.nz/datasets/downloads/1042-2, last access: 10 June 2024).

Concerning the actual scope for expanding *wosis_latest* in the coming years, we noted that getting positive responses to our requests for sharing soil data is becoming increasingly cumbersome; the overall success rate during the "2019–2023" acquisition effort was around 25 %. However, many of these datasets are being shared with ISRIC with the provision that the profile coordinates themselves may not be shown; hence, the corresponding soil data cannot be "openly" served to our user community through *wosis_latest*. Further, the site and profile coordinates are then regularly shared as "theoretical coordinates" only (e.g. ICP Forests, 2021b; Poeplau et al., 2020), highlighting the need for considering positional uncertainty in digital soil mapping and other applications. Another source of concern is that major soil monitoring programmes, such as LUCAS (e.g. Ballabio et al., 2016; Orgiazzi et al., 2018), only consider the top 20 or 30 cm of the soil. That is, they do not consider the actual soil profile depth as required for more comprehensive soil assessments such as computing changes in global carbon stocks or mapping plant-available water-holding capacity in the root zone (e.g. Batlle-Bayer et al., 2010; Leenaars et al., 2018; von Haden et al., 2020; Wang et al., 2022).

## 7 Data availability

The 2023 snapshot is archived for long-term storage at ISRIC – World Soil Information, the World Data Centre for Soils (WDC-Soils) of the ISC (International Council for Science) World Data System (WDS). It is freely accessible at https://doi.org/10.17027/isric-wdcsoils-20231130 (Calisto et al., 2023). The zip file (446 Gb) includes a copy of the Readme file and the data in TSV format (see Appendix B) and OGC GeoPackage format.

## 8 Conclusions

Bringing disparate soil profile data from different sources under a common global standard poses many and diverse challenges. A major improvement has been the harmonisation of the WoSIS data model to ISO 28258 and O&M domain specifications. In conjunction with this, refactored ETL procedures greatly improved the data ingestion and standardisation process, and new ways for visualising, querying and serving the data were developed to better serve our user community.

There are still numerous gaps in terms of geographic distribution as well as the range of soil taxonomic units and/or soil properties represented. We aspire to address such gaps in future updates of *wosis_latest*. However, as the World Data Centre for Soils, we are largely dependent on the ability of soil data owners to share some of their data freely for the greater benefit of the international community. To facilitate and stimulate this process, we are developing a web-based facility (front-end) to permit data providers to directly upload their soil data to WoSIS in a consistent format based on the refactored ETL procedures. As an incentive, upon their standardisation, we aim to provide each data provider with a tailor-made dashboard for viewing and querying the datasets they shared, possibly with a DOI to facilitate citation.

Various sources of uncertainty are associated with the data. Therefore, we provide three measures for fitness for intended use of the standardised data. This information, although coarse, should be duly considered by prospective users of the snapshot.

Unfortunately, numerous soil datasets worldwide are not freely accessible for various reasons. Standardised procedures, mechanisms, policies and incentives aimed at encouraging soil data sharing by different categories of data owners/providers are needed (e.g. Fantappie et al., 2021; Gobezie and Biswas, 2023; Padarian and McBratney, 2020; Robinson et al., 2019). At a transnational level, these pressing and complex issues are being addressed by the Global Soil Partnership, hosted by UN-FAO, in the context of the evolving federated Global Soil Information System.

## Appendix A: Coding conventions

**Table A1.** Coding conventions for observations (i.e. a combination of property, procedure and unit of measurement), number of profiles and layers provided in the WoSIS 2023 snapshot and inferred accuracy of measurements (codes are listed in alphabetical order).

| Code | Property | Procedure[a] | Unit | Profiles | Layers | Accuracy (± %)[b] |
|---|---|---|---|---|---|---|
| BDFI33 | Bulk density fine earth[c] | Bulk density of a soil sample that has been desorbed to 33 kPa (1/3 bar) | $kg\,dm^{-3}$ | 14 886 | 78 007 | 25.0 |
| BDFIAD | Bulk density fine earth | Bulk density of a soil sample that has been air-dried | $kg\,dm^{-3}$ | 4238 | 14 485 | 25.0 |
| BDFIFM | Bulk density fine earth | Bulk density of a soil sample at field-soil water content at time of sampling | $kg\,dm^{-3}$ | 5265 | 14 075 | 25.0 |
| BDFIOD | Bulk density fine earth | Bulk density of a soil sample that has been dried in an oven at 110 °C | $kg\,dm^{-3}$ | 26 064 | 131 623 | 25.0 |
| BDWSAD | Bulk density whole soil[e] | Bulk density of a soil sample that has been air-dried | $kg\,dm^{-3}$ | 0 | 0 | 25.0 |
| BDWSOD | Bulk density whole soil | Bulk density of a soil sample that has been dried in an oven at 110 °C | $kg\,dm^{-3}$ | 14 596 | 75 397 | 25.0 |
| CECPH7 | Cation exchange capacity | CEC estimated by buffering the soil at pH 7 (e.g. $NH_4Oac$) | $cmol(c)\,kg^{-1}$ | 60 339 | 320 532 | 20.0 |
| CECPH8 | Cation exchange capacity | CEC estimated by buffering the soil at pH 8 (e.g. $BaCl_2$) | $cmol(c)\,kg^{-1}$ | 6838 | 25 100 | 20.0 |
| CFGR | Coarse fragments | Gravimetric content of soil material larger than 2 mm[c] | g per 100 g | 39 481 | 202 414 | 20.0 |
| CFVO | Coarse fragments | Volumetric content of soil material larger than 2 mm[c] | $cm^3$ per 100 $cm^3$ | 48 891 | 246 580 | 30.0 |
| CLAY | Clay[d] | Determination of total gravimetric content of clay size fraction (for class size limits and analytical methods, see "method_options") | g per 100g | 153 319 | 652 347 | 15.0 |
| ECEC | Cation exchange capacity | Effective CEC conventionally approximated by summation of exchangeable bases ($Ca^{2+}$, $Mg^{2+}$, $K^+$ and $Na^+$) and 1 M KCl exchangeable acidity ($Al^{3+}$ and $H^+$) in acidic soils | $cmol(c)\,kg^{-1}$ | 35 123 | 143 693 | 25.0 |
| ELCO20 | Electrical conductivity | Electrical conductivity assessed on a 1 : 2 soil water extract. Used for saline soils. | $dS\,m^{-1}$ | 7971 | 44 350 | 10.0 |

**Table A1.** Continued.

| Code | Property | Procedure[a] | Unit | Profiles | Layers | Accuracy (± %)[b] |
|---|---|---|---|---|---|---|
| ELCO25 | Electrical conductivity | Electrical conductivity assessed on a 1 : 2.5 soil water extract. Used for saline soils. | $dS\,m^{-1}$ | 4395 | 17 825 | 10.0 |
| ELCO50 | Electrical conductivity | Electrical conductivity assessed on a 1 : 5 soil water extract. Used for saline soils. | $dS\,m^{-1}$ | 23 121 | 90 959 | 10.0 |
| ELCOSP | Electrical conductivity | Electrical conductivity assessed on water-saturated soil paste. Used for saline soils. | $dS\,m^{-1}$ | 22 052 | 85 020 | 10.0 |
| NITKJD | Total nitrogen (N) | Kjeldahl wet-oxidation digestion procedure | $g\,kg^{-1}$ | 72 905 | 240 433 | 10.0 |
| ORGC | Organic carbon (C) | Amount of organic carbon determined according to method specified under "method_options" | $g\,kg^{-1}$ | 135 655 | 526 953 | 15.0 |
| ORGM | Organic matter | Determination of organic compounds that accompany soil particles through a 2 mm sieve using loss on ignition (LOI) at about 400 °C. | $g\,kg^{-1}$ | 3871 | 16 282 | 15.0 |
| PHAQ | pH | A measure of the acidity or alkalinity in soils, defined as the negative logarithm (base 10) of the activity of hydronium ions ($H^+$) in water[a]. | unitless | 140 326 | 655 336 | 0.3 |
| PHCA | pH | A measure of the acidity or alkalinity in soils, defined as the negative logarithm (base 10) of the activity of hydronium ions ($H^+$), in the specified $CaCl_2$ solution. | unitless | 69 437 | 325 153 | 0.3 |
| PHKC | pH | A measure of the acidity or alkalinity in soils, defined as the negative logarithm (base 10) of the activity of hydronium ions ($H^+$), in the specified KCl solution. | unitless | 38 022 | 173 464 | 0.3 |
| PHNF | pH | A measure of the acidity or alkalinity in soils, defined as the negative logarithm (base 10) of the activity of hydronium ions ($H^+$), in the specified NaF solution. | unitless | 4965 | 25 409 | 0.3 |
| PHETB1 | Phosphorus (P) | Phosphorus determined according to the Bray-I method, a combination of HCl and $NH_4$-F to remove easily acid soluble P forms, largely Al and Fe phosphates (mainly applicable for acid soils) | $mg\,kg^{-1}$ | 10 719 | 40 379 | 40.0 |

**Table A1.** Continued.

| Code | Property | Procedure[a] | Unit | Profiles | Layers | Accuracy (± %)[b] |
|---|---|---|---|---|---|---|
| PHETM3 | Phosphorus (P) | Determined according to the Mehlich 3 method, a weak acid soil extraction procedure that is considered suitable for removing P and other elements in acid and neutral soil. The extract is composed of 0.2 M glacial acetic acid, 0.25 M ammonium nitrate, 0.015 M ammonium fluoride, 0.013 M nitric acid and 0.001 M ethylenediaminetetraacetic acid (EDTA). | $\mathrm{mg\,kg^{-1}}$ | 1444 | 7230 | 25.0 |
| PHETOL | Phosphorus (P) | Phosphorus determined according to the Olsen method (0.5 M sodium bicarbonate ($NaHCO_3$) solution at a pH of 8.5); used to extract P from calcareous, alkaline and neutral soils. | $\mathrm{mg\,kg^{-1}}$ | 4266 | 12 291 | 25.0 |
| PHPRTN | Phosphorus (P) | Phosphorus retention measured according to the New Zealand method (Blakemore et al., 1981). | g per 100 g | 5599 | 26 569 | 20.0 |
| PHPTOT | Phosphorus (P) | Phosphorus determined with a "harsh" digest procedure to liberate and measure all forms of element. | $\mathrm{mg\,kg^{-1}}$ | 7561 | 19 310 | 15.0 |
| PHPWSL | Phosphorus (P) | Phosphorus soluble in soluble in water | $\mathrm{mg\,kg^{-1}}$ | 282 | 1241 | 15.0 |
| SAND | Sand | Determination of total gravimetric content of sand size fraction (for class size limits and analytical methods, see "method_options"). | g per 100 g | 119 127 | 542 463 | 15.0 |
| SILT | Silt[f] | Determination of total gravimetric content of silt size fraction (for class size limits and analytical methods, see "method_options"). | g per 100 g | 145 906 | 620 790 | 15.0 |
| TCEQ | Calcium carbonate equivalent (TCEQ) | Determination of the gravimetric loss of carbonates as carbon dioxide in the presence of excess hydrochloric acid. The quantity of carbonate ($CO_3$) in the soil is expressed as $CaCO_3$ and as a weight percentage of the size fraction that is less than 2 mm. | $\mathrm{g\,kg^{-1}}$ | 59 294 | 247 368 | 10.0 |

**Table A1.** Continued.

| Code | Property | Procedure[a] | Unit | Profiles | Layers | Accuracy (± %)[b] |
|---|---|---|---|---|---|---|
| TOTC | Total carbon (C) | Total C is quantified by two basic methods: wet or dry combustion (see "method options"). In total C determinations, all forms of C in a soil are converted to $CO_2$ followed by a quantification of the evolved $CO_2$. Total C can be used to estimate the organic C content of a soil. The difference between total and inorganic C is an estimate of the organic C. | $g\,kg^{-1}$ | 33 527 | 112 787 | 10.0 |
| WG0006 | Water retention gravimetric | Water retention assessed at tension 6 kPa (see method options') | g per 100 g | 827 | 3828 | 20.0 |
| WG0010 | Water retention gravimetric | Water retention assessed at tension 10 kPa (see "method_options"). | g per 100 g | 2970 | 12 517 | 20.0 |
| WG0033 | Water retention gravimetric | Water retention assessed at tension 33 kPa (see "method_options"). | g per 100 g | 20 994 | 94 707 | 20.0 |
| WG0100 | Water retention gravimetric | Water retention assessed at tension 100 kPa (see "method_options"). | g per 100 g | 687 | 3360 | 20.0 |
| WG0200 | Water retention gravimetric | Water retention assessed at tension 200 kPa (see "method_options"). | g per 100 g | 4391 | 27 773 | 20.0 |
| WG0500 | Water retention gravimetric | Water retention assessed at tension 500 kPa (see "method_options"). | g per 100 g | 326 | 1414 | 20.0 |
| WG1500 | Water retention gravimetric | Water retention assessed at tension 1500 kPa (see "method_options"). | g per 100 g | 33 782 | 181 999 | 20.0 |
| WV0010 | Water retention volumetric | Water retention assessed at tension 10 kPa (see "method_options"). | $cm^3$ per 100 $cm^3$ | 1914 | 6883 | 20.0 |
| WV0033 | Water retention volumetric | Water retention assessed at tension 33 kPa (see "method_options"). | $cm^3$ per 100 $cm^3$ | 7444 | 22 291 | 20.0 |
| WV0100 | Water retention volumetric | Water retention assessed at tension 100 kPa (see "method_options"). | $cm^3$ per 100 $cm^3$ | 747 | 2553 | 20.0 |

**Table A1.** Continued.

| Code | Property | Procedure[a] | Unit | Profiles | Layers | Accuracy (± %)[b] |
|---|---|---|---|---|---|---|
| WV0500 | Water retention volumetric | Water retention assessed at tension 500 kPa (see "method_options"). | $cm^3$ per 100 $cm^3$ | 702 | 1758 | 20.0 |
| WV1500 | Water retention volumetric | Water retention assessed at tension 1500 kPa (see "method_options"). | $cm^3$ per 100 $cm^3$ | 7904 | 23 331 | 20.0 |

[a] Method options for each analytical procedure are described in Batjes and van Oostrum (2023) and provided in the *wosis_202312_xxxx.tsv* file; see Appendix C.

[b] Inferred accuracy (or uncertainty), rounded to the nearest 5 %, unless otherwise indicated (i.e. units for soil pH), as derived from various sources (Al-Shammary et al., 2018; Kalra and Maynard, 1991; Rayment and Lyons, 2011; Rossel and McBratney, 1998; van Reeuwijk, 1983; WEPAL, 2019). These figures are first approximations that should be fine-tuned once more specific results of laboratory proficiency tests, i.e. national Soil Quality Management systems, become freely available (e.g. from the GLOSOLAN laboratory proficiency programme).

[c] Generally, the fine earth fraction is defined as being $< 2$ mm. Alternatively, an upper limit of 1 mm was used in the former Soviet Union and its satellite states (Katchynsky scheme). The actual size limits are specified under "method_options" (see Appendix C).

[d] Provided only when the sum of clay, silt and sand fraction is $\geq 90$ and $\leq 100$ % (note that users should normalise the totals to 100 % before using them for mapping or modelling purposes; further, more stringent limits (e.g. $\geq 98$ and $\leq 102$) may be considered).

[e] No data are being served for this property because the associated licences are flagged as "restricted" by the data providers.

[f] The lower and upper limits for the "silt" size fraction can vary markedly between countries; hence these limits have been specified explicitly in WoSIS under "method_options" (see Appendix B). Development and application of conversion procedures to one common "silt" fraction (e.g. 0.002–0.05 mm) are beyond the remit of the WoSIS project itself. The necessary pedotransfer functions should be developed (and tested) prior to generating particle-size-class-related soil property maps for a given geography. Research in this direction is being undertaken by the SoilGrids team, based on the "best available" comparative datasets for calibration.

**Table A2.** Coding conventions and brief descriptions for soil classification, horizon designations and number of occurrences in the WoSIS 2023 snapshot.

| Code | Description | Count |
|---|---|---|
| CSTX | Classification of the soil profile according to specified edition (year) of USDA Soil Taxonomy, at least at soil order level | 31 400 |
| CWRB | Classification of the soil profile according to specified edition (year) of the World Reference Base for Soil Resources (WRB), at least at reference soil group level | 39 649 |
| CFAO | Classification of the soil profile according to specified edition (year) of the FAO-UNESCO legend, at least at major group level | 38 792 |
| HODS[a] | Horizon designations as provided in the source databases | 80 849/396 522[b] |

[a] Where available, the "cleaned" (original) layer/horizon designation is provided for general information; these codes have not been standardised as they vary widely between different classification systems (Bridges, 1993; Gerasimova et al., 2013). When no horizon designations are provided in the source data bases, we have flagged all layers with an upper depth given as being negative (e.g. −10 to 0 cm, that is using pre-1993 conventions (see Sect. 3.1) in the source databases as likely being a shallow "organic surface" layer above a mineral soil layer.

[b] Number of profiles with horizon descriptions as well as the total number of layers with horizon designations.

## Appendix B: Structure of the WoSIS 2023 snapshot

This appendix describes the structure of the data files served with the WoSIS 2023 snapshot, namely *wosis_202312_observations.tsv*, *wosis_20312_site.tsv, wosis_202312_profiles.tsv*, *wosis_202312_layers.tsv* and *wosis_202312_xxxx.tsv* (where "*xxxx"* is the name of the observation). The data files are also distributed in OGC GeoPackage format, which stores the files within an SQLite database. Technical details are provided in a Readme file (https://www.isric.org/sites/default/files/Readme_WoSIS_202312_v2.pdf, last access: 26 April 2024).

*wosis_202312_observations.tsv*. This file lists the four- to six-letter codes for each observation, whether the observation is for a site/profile or layer (horizon), the unit of measurement, and the number of profiles and layers represented in the snapshot. It also provides the inferred accuracy for the laboratory measurements (see Appendix A).

| | |
|---|---|
| code | Code for the observation |
| property | Description of soil property |
| procedure | Description of analytical procedure |
| unit | Standard unit of measurement |
| profiles | Number of profiles that have at least one measurement for the observation |
| layers | Number of layers that have measurements for the observation |
| accuracy | Inferred accuracy of the laboratory measurements (first approximation; see Sect. 3.4.2) |

*wosis_202312_site.tsv*. This file characterises the site location where profiles were sampled. The following field names are used.

| | |
|---|---|
| site_id | Primary key |
| longitude | Longitude in degrees (WGS84) |
| latitude | Latitude in degrees (WGS84) |
| positional_uncertainty | Positional uncertainty of the profile's site location, expressed in four classes (see Table 2) |
| country_name | Name of country/area where site is located |
| region | Region in which site is located |
| continent | Continent in which site is located |

*wosis_202312_profiles.tsv*. Presents the unique profile ID (i.e. primary key), site_id, source of the data, country ISO code and name, positional uncertainty, latitude and longitude (WGS84), and maximum depth of soil described and sampled, 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 the World Soil Reference Base (WRB) system, the options are publication year (i.e. version), reference_soil_group_code, reference_soil_group_name, and the name(s) of the prefix (primary) qualifier(s) and suffix (supplementary) qualifier(s). The terms principal qualifier and supplementary qualifier have been used since 2015 (IUSS Working Group WRB, 2015, 2022); earlier WRB versions used prefix and suffix for this (e.g. IUSS Working Group WRB, 2006). Alternatively, for USDA Soil Taxonomy, the version (year), order, suborder, great group and subgroup can be accommodated (Soil Survey Staff, 2014). The following field names are used.

| | |
|---|---|
| profile_id | Primary key |
| profile_code | Code for the profile |
| dataset_code | Identifier for source dataset |
| site_id | Identifier for site where profile is located |
| positional_uncertainty | Positional uncertainty of the profile's site location, expressed in four classes (see Table 2) |
| country_name | Name of country/area where site is located |
| longitude | Longitude in degrees (WGS84) |
| latitude | Latitude in degrees (WGS84) |
| wrb_reference_soil_group_code | Code for WRB group (in given version of WRB) |
| wrb_reference_soil_group | Full name for reference soil group |
| wrb_prefix_qualifiers | Name for prefix (i.e. for WRB1988) |
| wrb_suffix_qualifiers | Name for suffix (i.e. for WRB1988) |
| wrb_principal_qualifiers | Name for principal qualifiers (i.e. for WRB 2015 and WRB 2022) |
| wrb_supplementary qualifiers | Name for supplementary qualifiers (i.e. for WRB 2015 and WRB 2022) |
| wrb_publication _year | Version of World Reference Base for Soil Resources |
| fao_major_group_code | Code for major group (in given version of the legend), |
| fao_major_group | Name of major group |
| fao_soil_unit_code | Code for soil unit |
| fao_soil_unit | Name of soil unit |
| fao_publication _year | Version of FAO legend (e.g. 1974 or 1988) |
| usda_order_name | Name of USDA Soil Taxonomy order |
| usda_suborder | Name of USDA Soil Taxonomy suborder |
| usda_great_group | Name of USDA Soil Taxonomy greatgroup |
| usda_subgroup | Name of USDA Soil Taxonomy subgroup |
| usda_publication_year | Version of USDA Soil Taxonomy |

*wosis_202312_layers.tsv*. This file characterises the layers (or horizons) per profile.

| | |
|---|---|
| profile_id | Primary key |
| layer_id | Sequential number for the layer (or horizon) |
| profile_code | Code for the profile |
| site_id | Identifier for site where profile is located |
| layer_name | Name of pedogenetic horizon (as is) |
| upper_depth | Upper depth of layer |
| lower_depth | Lower depth of layer |
| layer_number | Sequential number for the layer (or horizon) |
| organic_surface | Flag for the presence of an organic layer above the mineral soil |
| dataset_id | Abbreviation for source dataset (e.g. WD-ISCN) |
| licence | Licence for observation as indicated by the data provider (e.g. CC BY) |

*wosis_202312_xxxx.tsv*. For each observation (e.g. “xxxx” = “BDFIOD”), as defined under “code” in file *wosis_202312_observation.tsv*, the following are listed.

| | |
|---|---|
| profile_id | Primary key |
| layer_id | Primary key (number, sequential from top to bottom) |
| profile_code | Code for given profile |
| layer_name | Name of pedogenetic horizon (as is) |
| upper_depth | Upper depth of layer |
| lower_depth | Lower depth of layer |
| organic_surface | Indicates if there is an organic layer above the mineral surface |
| value | Array listing all measurement values for observation “*xxxx*” for the given layer. (In some cases, more than one observation is reported for a given horizon (layer) in the source, for example four values for TOTC: [1 : 5.4, 2 : 8.2, 3 : 6.3, 4 : 7.7] (see value_avg below).) |
| method_options | Array listing the method options for each analytical procedure as distilled from the source data. (The content of this array varies with the soil observation under consideration as described in the method option table for each analytical procedure. For example, in the case of electrical conductivity (ELCO), the method options include sample pretreatment (e.g. sieved over 2 mm size, solution (e.g. water), ratio (e.g. 1 : 5) and ratio base (e.g. weight / volume)). For details, see Batjes and van Oostrum (2023).) |
| value_avg | Average, for above (it is recommended to use this value for “routine” modelling) |
| dataset_id | Abbreviation for source dataset (e.g. WD-ISCN) |
| country_name | Name of country/area where site is located |
| longitude | Longitude in degrees (WGS84) |
| latitude | Latitude in degrees (WGS84) |
| positional_uncertainty | Positional uncertainty of the profile’s site location (see Table 2) |
| region | Region in which site is located |
| continent | Continent where the profile’s site is located |
| date | Date the profile was described/sampled |
| licence | Licence for given data, as indicated by the data provider (i.e. CC BY or CC BY-NC) |

*Format*. All fields in the above files are tab-delimited, with double quotation marks as text delimiters. File coding is according to the UTF-8 Unicode transformation format.

*Using the data*. Tutorials for downloading and querying the data, using various platforms, are provided on the WoSIS FAQ web page (https://www.isric.org/explore/wosis/faq-wosis, last access: 24 April 2024).

## Appendix C: Distribution of sites

**Table C1.** Number of sites per continent and country/area.

| Continent | Country/area[a] | Code | No. of sites | Area (km$^2$) | Site density (per 1000 km$^2$) |
|---|---|---|---|---|---|
| Africa | Abyei[b] | 4 | 0 | 9943 | 0 |
| | Algeria | DZ | 10 | 230 8647 | 0.004 |
| | Angola | AO | 1168 | 1 246 690 | 0.937 |
| | Benin | BJ | 743 | 115 247 | 6.447 |
| | Botswana | BW | 994 | 578 247 | 1.719 |
| | British Indian Ocean Territory | IO | 0 | 49 | 0 |
| | Burkina Faso | BF | 2023 | 273 281 | 7.403 |
| | Burundi | BI | 36 | 26 857 | 1.34 |
| | Cameroon | CM | 1417 | 465 363 | 3.045 |
| | Cabo Verde | CV | 0 | 4056 | 0 |
| | Central African Republic | CF | 88 | 619 591 | 0.142 |
| | Chad | TD | 7 | 1 265 392 | 0.006 |
| | Comoros | KM | 0 | 1652 | 0 |
| | Republic of Congo | CG | 70 | 340 599 | 0.206 |
| | Côte d’Ivoire | CI | 255 | 321 762 | 0.793 |
| | Democratic Republic of the Congo | CD | 378 | 2 329 162 | 0.162 |
| | Djibouti | DJ | 0 | 21 670 | 0 |
| | Egypt | EG | 26 | 982 161 | 0.026 |
| | Equatorial Guinea | GQ | 0 | 27 000 | 0 |
| | Eritrea | ER | 0 | 120 763 | 0 |
| | Ethiopia | ET | 1712 | 1 129 314 | 1.516 |
| | Gabon | GA | 47 | 264 022 | 0.178 |
| | Gambia | GM | 0 | 11 203 | 0 |
| | Ghana | GH | 432 | 238 842 | 1.809 |
| | Guinea | GN | 128 | 243 023 | 0.527 |
| | Guinea-Bissau[b] | GW | 15 | 30 740 | 0.488 |
| | Hala’ib Triangle[b] | 10 | 0 | 17 684 | 0 |
| | Ilemi Triangle[b] | 13 | 0 | 3179 | 0 |
| | Kenya | KE | 1603 | 582 342 | 2.753 |
| | Lesotho | LS | 33 | 30 453 | 1.084 |
| | Liberia | LR | 50 | 96 103 | 0.52 |
| | Libya | LY | 14 | 1620 583 | 0.009 |
| | Madagascar | MG | 130 | 588 834 | 0.221 |
| | Malawi | MW | 3050 | 118 715 | 25.692 |
| | Mali | ML | 885 | 1 251 471 | 0.707 |
| | Ma’tan al-Sarra[b] | 11 | 0 | 1993 | 0 |
| | Mauritania | MR | 13 | 1 038 527 | 0.013 |
| | Mauritius | MU | 0 | 2014 | 0 |
| | Mayotte | YT | 0 | 378 | 0 |
| | Morocco | MA | 113 | 414 030 | 0.273 |
| | Mozambique | MZ | 565 | 787 305 | 0.718 |
| | Namibia | NA | 1569 | 823 989 | 1.904 |
| | Niger | NE | 520 | 1 182 602 | 0.44 |
| | Nigeria | NG | 1402 | 908 978 | 1.542 |
| | Réunion | RE | 0 | 2504 | 0 |
| | Rwanda | RW | 1016 | 25 388 | 40.018 |
| | Saint Helena, Ascension and Tristan da Cunha | SH | 0 | 399 | 0 |
| | São Tomé and Príncipe | ST | 0 | 991 | 0 |
| | Senegal | SN | 312 | 196 200 | 1.59 |
| | Seychelles | SC | 0 | 499 | 0 |
| | Sierra Leone | SL | 12 | 72 281 | 0.166 |

**Table C1.** Continued.

| Continent | Country/area | Code | No. of sites | Area ($km^2$) | Site density (per 1000 $km^2$) |
|---|---|---|---|---|---|
| | Somalia | SO | 245 | 632 562 | 0.387 |
| | South Africa | ZA | 879 | 1 220 127 | 0.72 |
| | South Sudan | SS | 82 | 629 821 | 0.13 |
| | Sudan | SD | 130 | 1 843 196 | 0.071 |
| | Eswatini (formerly Swaziland) | SZ | 14 | 17 290 | 0.81 |
| | Togo | TG | 9 | 56 767 | 0.159 |
| | Tunisia | TN | 60 | 155 148 | 0.387 |
| | Uganda | UG | 84 | 241 495 | 0.348 |
| | United Republic of Tanzania | TZ | 1910 | 939 588 | 2.033 |
| | Western Sahara[b] | EH | 0 | 268 617 | 0 |
| | Zambia | ZM | 603 | 751 063 | 0.803 |
| | Zimbabwe | ZW | 413 | 390 648 | 1.057 |
| Antarctica | Antarctica | AQ | 30 | 12 537 967 | 0.002 |
| | Bouvet Island | BV | 0 | 45 | 0 |
| | French Southern and Antarctic Territories | TF | 0 | 7738 | 0 |
| | Heard Island and McDonald Islands | HM | 0 | 412 | 0 |
| | South Georgia and the South Sandwich Islands | GS | 0 | 3870 | 0 |
| Asia | Afghanistan | AF | 19 | 641 827 | 0.03 |
| | Aksai Chin[b] | 1 | 0 | 30 666 | 0 |
| | Armenia | AM | 509 | 29 624 | 17.182 |
| | Arunachal Pradesh[b] | 2 | 2 | 67 965 | 0.029 |
| | Azerbaijan | AZ | 28 | 164 780 | 0.17 |
| | Bahrain | BH | 2 | 673 | 2.97 |
| | Bangladesh | BD | 207 | 139 825 | 1.48 |
| | Bhutan | BT | 85 | 37 674 | 2.256 |
| | Brunei Darussalam | BN | 0 | 5899 | 0 |
| | Cambodia | KH | 424 | 181 424 | 2.337 |
| | China | CN | 1644 | 9 345 214 | 0.176 |
| | China/India disputed territory[b] | 3 | 0 | 3526 | 0 |
| | Christmas Island | CX | 0 | 136 | 0 |
| | Cocos (Keeling) Islands | CC | 0 | 16 | 0 |
| | Cyprus | CY | 12 | 9249 | 1.297 |
| | Democratic People’s Republic of Korea | KP | 0 | 122 465 | 0 |
| | Georgia | GE | 18 | 69 785 | 0.258 |
| | Hong Kong SAR | HK | 2 | 1081 | 1.851 |
| | India | IN | 199 | 2 961 118 | 0.067 |
| | Indonesia | ID | 179 | 1 888 620 | 0.095 |
| | Iran (Islamic Republic of) | IR | 2010 | 1 677 319 | 1.198 |
| | Iraq | IQ | 14 | 435 864 | 0.032 |
| | Israel | IL | 17 | 20 720 | 0.82 |
| | Jammu and Kashmir[b] | 12 | 4 | 186 035 | 0.022 |
| | Japan | JP | 197 | 373 651 | 0.527 |
| | Jordan | JO | 47 | 89 063 | 0.528 |
| | Kazakhstan | KZ | 52 | 2 841 103 | 0.018 |
| | Kuril Islands[b] | 5 | 0 | 4996 | 0 |
| | Kuwait | KW | 1 | 17 392 | 0.057 |
| | Kyrgyzstan | KG | 1 | 199 188 | 0.005 |
| | Lao People’s Democratic Republic | LA | 20 | 230 380 | 0.087 |
| | Lebanon | LB | 10 | 10 136 | 0.987 |
| | Macau SAR | MO | 0 | 17 | 0 |
| | Malaysia | MY | 155 | 329 775 | 0.47 |
| | Maldives | MV | 0 | 223 | 0 |
| | Mongolia | MN | 9 | 1 564 529 | 0.006 |

**Table C1.** Continued.

| Continent | Country/area | Code | No. of sites | Area ($km^2$) | Site density (per 1000 $km^2$) |
|---|---|---|---|---|---|
| | Myanmar | MM | 0 | 667 085 | 0 |
| | Nepal | NP | 142 | 147 437 | 0.963 |
| | State of Palestine[b] | PS | 18 | 6225 | 2.892 |
| | Oman | OM | 11 | 308 335 | 0.036 |
| | Pakistan | PK | 45 | 788 439 | 0.057 |
| | Paracel Islands[b] | 6 | 0 | 8 | 0 |
| | Philippines | PH | 78 | 296 031 | 0.263 |
| | Qatar | QA | 0 | 11 549 | 0 |
| | Republic of Korea | KR | 23 | 99 124 | 0.232 |
| | Saudi Arabia | SA | 7 | 1 925 621 | 0.004 |
| | Scarborough Reef[b] | 7 | 0 | 44 | 0 |
| | Senkaku Islands[b] | 8 | 0 | 5 | 0 |
| | Singapore | SG | 1 | 594 | 1.683 |
| | Spratly Islands[b] | 9 | 0 | 1 | 0 |
| | Sri Lanka | LK | 73 | 66 173 | 1.103 |
| | Syrian Arab Republic | SY | 69 | 188 128 | 0.367 |
| | Taiwan | TW | 35 | 36 127 | 0.969 |
| | Tajikistan | TJ | 5 | 142 004 | 0.035 |
| | Thailand | TH | 479 | 515 417 | 0.929 |
| | Timor-Leste | TL | 0 | 14 892 | 0 |
| | Türkiye | TR | 69 | 781 229 | 0.088 |
| | Turkmenistan | TM | 0 | 555 052 | 0 |
| | United Arab Emirates | AE | 12 | 71 079 | 0.169 |
| | Uzbekistan | UZ | 9 | 449 620 | 0.02 |
| | Vietnam | VN | 29 | 327 575 | 0.089 |
| | Yemen | YE | 284 | 453 596 | 0.626 |
| Europe | Albania | AL | 97 | 28 682 | 3.382 |
| | Andorra | AD | 0 | 475 | 0 |
| | Austria | AT | 128 | 83 964 | 1.524 |
| | Belarus | BY | 96 | 207 581 | 0.462 |
| | Belgium | BE | 7013 | 30 669 | 228.667 |
| | Bosnia and Herzegovina | BA | 32 | 51 145 | 0.626 |
| | Bulgaria | BG | 134 | 111 300 | 1.204 |
| | Croatia | HR | 78 | 56 589 | 1.378 |
| | Czech Republic | CZ | 666 | 78 845 | 8.447 |
| | Denmark | DK | 72 | 44 458 | 1.619 |
| | Estonia | EE | 241 | 45 441 | 5.304 |
| | Faroe Islands | FO | 0 | 1400 | 0 |
| | Finland | FI | 442 | 336 892 | 1.312 |
| | France | FR | 3183 | 548 785 | 5.8 |
| | Germany | DE | 4362 | 357 227 | 12.211 |
| | Gibraltar | GI | 0 | 6 | 0 |
| | Greece | GR | 374 | 132 549 | 2.822 |
| | Guernsey | GG | 0 | 79 | 0 |
| | Holy See | VA | 0 | 0 | 0 |
| | Hungary | HU | 1421 | 93 119 | 15.26 |
| | Iceland | IS | 17 | 102 566 | 0.166 |
| | Ireland | IE | 124 | 69 809 | 1.776 |
| | Isle of Man | IM | 0 | 573 | 0 |
| | Italy | IT | 576 | 301 651 | 1.909 |
| | Jersey | JE | 0 | 120 | 0 |
| | Latvia | LV | 102 | 64 563 | 1.58 |
| | Liechtenstein | LI | 0 | 151 | 0 |
| | Lithuania | LT | 127 | 64 943 | 1.956 |

**Table C1.** Continued.

| Continent | Country/area | Code | No. of sites | Area (km$^2$) | Site density (per 1000 km$^2$) |
|---|---|---|---|---|---|
| | Luxembourg | LU | 142 | 2621 | 54.184 |
| | Malta | MT | 0 | 316 | 0 |
| | Monaco | MC | 0 | 8 | 0 |
| | Montenegro | ME | 12 | 13 776 | 0.871 |
| | Netherlands | NL | 958 | 35 203 | 27.214 |
| | Norway | NO | 507 | 324 257 | 1.564 |
| | Poland | PL | 796 | 311 961 | 2.552 |
| | Portugal | PT | 455 | 91 876 | 4.952 |
| | Republic of Moldova | MD | 35 | 33 798 | 1.036 |
| | Romania | RO | 113 | 238 118 | 0.475 |
| | Russian Federation | RU | 1464 | 16 998 830 | 0.086 |
| | San Marino | SM | 0 | 60 | 0 |
| | Serbia | RS | 69 | 88 478 | 0.78 |
| | Slovakia | SK | 161 | 49 072 | 3.281 |
| | Slovenia | SI | 67 | 20 320 | 3.297 |
| | Spain | ES | 907 | 505 752 | 1.793 |
| | Svalbard and Jan Mayen Islands | SJ | 4 | 63 464 | 0.063 |
| | Sweden | SE | 594 | 449 212 | 1.322 |
| | Switzerland | CH | 10 928 | 41 257 | 264.874 |
| | Republic of North Macedonia | MK | 20 | 25 424 | 0.787 |
| | Ukraine | UA | 462 | 600 526 | 0.769 |
| | United Kingdom | GB | 1727 | 244 308 | 7.069 |
| North America | Anguilla | AI | 0 | 79 | 0 |
| | Antigua and Barbuda | AG | 0 | 452 | 0 |
| | Aruba | AW | 0 | 180 | 0 |
| | Bahamas | BS | 0 | 11 904 | 0 |
| | Barbados | BB | 3 | 433 | 6.928 |
| | Belize | BZ | 26 | 21 764 | 1.195 |
| | Bermuda | BM | 0 | 63 | 0 |
| | British Virgin Islands | VG | 0 | 154 | 0 |
| | Canada | CA | 8778 | 9 875 646 | 0.889 |
| | Cayman Islands | KY | 0 | 269 | 0 |
| | Clipperton Island | CP | 0 | 9 | 0 |
| | Costa Rica | CR | 560 | 51 042 | 10.971 |
| | Cuba | CU | 53 | 110 863 | 0.478 |
| | Dominica | DM | 0 | 751 | 0 |
| | Dominican Republic | DO | 10 | 48 099 | 0.208 |
| | El Salvador | SV | 38 | 20 732 | 1.833 |
| | Greenland | GL | 2 | 2 165 159 | 0.001 |
| | Grenada | GD | 0 | 318 | 0 |
| | Guadeloupe | GP | 5 | 1697 | 2.947 |
| | Guatemala | GT | 28 | 109 062 | 0.257 |
| | Haiti | HT | 0 | 27 022 | 0 |
| | Honduras | HN | 38 | 112 124 | 0.339 |
| | Jamaica | JM | 74 | 10 965 | 6.749 |
| | Martinique | MQ | 0 | 1104 | 0 |
| | Mexico | MX | 12 599 | 1 949 527 | 6.463 |
| | Montserrat | MS | 0 | 101 | 0 |
| | Netherlands Antilles | AN | 4 | 790 | 5.066 |
| | Nicaragua | NI | 21 | 128 376 | 0.164 |
| | Panama | PA | 50 | 74 850 | 0.668 |
| | Puerto Rico | PR | 280 | 8937 | 31.329 |
| | Saint Kitts and Nevis | KN | 0 | 262 | 0 |
| | Saint Lucia | LC | 0 | 603 | 0 |

**Table C1.** Continued.

| Continent | Country/area | Code | No. of sites | Area (km$^2$) | Site density (per 1000 km$^2$) |
|---|---|---|---|---|---|
| | Saint Pierre and Miquelon | PM | 0 | 233 | 0 |
| | Saint Vincent and the Grenadines | VC | 0 | 427 | 0 |
| | Trinidad and Tobago | TT | 2 | 5144 | 0.389 |
| | Turks and Caicos Islands | TC | 0 | 530 | 0 |
| | United States Minor Outlying Islands | UM | 0 | 348 | 0 |
| | United States of America | US | 56 322 | 9 315 946 | 6.046 |
| | United States Virgin Islands | VI | 46 | 352 | 130.555 |
| Oceania | American Samoa | AS | 0 | 200 | 0 |
| | Australia | AU | 42 767 | 7 687 634 | 5.563 |
| | Cook Islands | CK | 0 | 241 | 0 |
| | Fiji | FJ | 6 | 18 293 | 0.328 |
| | French Polynesia | PF | 0 | 3967 | 0 |
| | Guam | GU | 15 | 544 | 27.579 |
| | Kiribati | KI | 0 | 1020 | 0 |
| | Marshall Islands | MH | 0 | 268 | 0 |
| | Micronesia (Federated States of) | FM | 75 | 740 | 101.343 |
| | Nauru | NR | 0 | 22 | 0 |
| | New Caledonia | NC | 2 | 18 574 | 0.108 |
| | Aotearoa / New Zealand | NZ | 52 | 270 415 | 0.192 |
| | Niue | NU | 0 | 263 | 0 |
| | Norfolk Island | NF | 0 | 38 | 0 |
| | Northern Mariana Islands | MP | 0 | 476 | 0 |
| | Palau | PW | 18 | 451 | 39.924 |
| | Papua New Guinea | PG | 24 | 462 230 | 0.052 |
| | Pitcairn Islands | PN | 0 | 49 | 0 |
| | Samoa | WS | 18 | 2835 | 6.349 |
| | Solomon Islands | SB | 1 | 28 264 | 0.035 |
| | Tokelau | TK | 0 | 15 | 0 |
| | Tonga | TO | 0 | 700 | 0 |
| | Tuvalu | TV | 0 | 48 | 0 |
| | Vanuatu | VU | 1 | 12 236 | 0.082 |
| | Wake Island[b] | WK | 0 | 7 | 0 |
| | Wallis and Futuna Islands | WF | 0 | 142 | 0 |
| South America | Argentina | AR | 253 | 2 780 175 | 0.091 |
| | Bolivia (Plurinational State of) | BO | 87 | 1 084 491 | 0.08 |
| | Brazil | BR | 9262 | 8 485 946 | 1.091 |
| | Chile | CL | 13 662 | 753 355 | 18.135 |
| | Colombia | CO | 236 | 1 137 939 | 0.207 |
| | Ecuador | EC | 94 | 256 249 | 0.367 |
| | Falkland Islands (Malvinas)[b] | FK | 0 | 12 084 | 0 |
| | French Guiana | GF | 30 | 83 295 | 0.36 |
| | Guyana | GY | 43 | 211 722 | 0.203 |
| | Paraguay | PY | 2 | 399 349 | 0.005 |
| | Peru | PE | 158 | 1 290 640 | 0.122 |
| | Suriname | SR | 31 | 145 100 | 0.214 |
| | Uruguay | UY | 136 | 177 811 | 0.765 |
| | Venezuela (Bolivarian Republic of) | VE | 204 | 912 025 | 0.224 |

[a] Country names and areas are based on the Global Administrative Unit Layers (GAUL) database; see https://data.apps.fao.org/map/catalogsrv/eng/catalog.search?id=12691#/metadata/9c35ba10-5649-41c8-bdfc-eb78e9e65654 (last access: 26 April 2024). [b] Disputed territory.

**Table C2.** Number of sites by World Terrestrial Ecosystems (WTE)*.

| Temperature zone | Moisture zone | No. of sites | Percent (%) |
|---|---|---:|---:|
| Polar | Dry | 224 | 0.1 |
| Polar | Moist | 532 | 0.2 |
| Boreal | Dry | 1789 | 0.8 |
| Boreal | Moist | 3398 | 1.6 |
| Cool temperate | Desert | 25 | 0 |
| Cool temperate | Dry | 10 968 | 5 |
| Cool temperate | Moist | 53 245 | 24.5 |
| Warm temperate | Desert | 238 | 0.1 |
| Warm temperate | Dry | 29 209 | 13.4 |
| Warm temperate | Moist | 46 533 | 21.4 |
| Subtropical | Desert | 296 | 0.1 |
| Subtropical | Dry | 25 748 | 11.8 |
| Subtropical | Moist | 17 906 | 8.2 |
| Tropical | Desert | 178 | 0.1 |
| Tropical | Dry | 11 315 | 5.2 |
| Tropical | Moist | 11 095 | 5.1 |
| No data | – | 4674 | 2.2 |

* World Terrestrial Ecosystems (WTE) as defined by Sayre (2022). Total may differ from 100 % due to rounding.

**Table C3.** Number of sites by WWF biome*.

| WWF biome | No. of sites | Percent (%) |
|---|---:|---:|
| Boreal forests/taiga | 5519 | 2.5 |
| Deserts and xeric shrublands | 13 410 | 6.2 |
| Flooded grasslands and savannas | 792 | 0.4 |
| Lakes | 85 | 0 |
| Mangroves | 765 | 0.4 |
| Mediterranean forests, woodlands and scrub | 24 459 | 11.3 |
| Montane grasslands and shrublands | 2796 | 1.3 |
| Rock and ice | 20 | 0 |
| Temperate broadleaf and mixed forests | 74 068 | 34.1 |
| Temperate coniferous forests | 14 436 | 6.6 |
| Temperate grasslands, savannas and shrublands | 23 890 | 11 |
| Tropical and subtropical coniferous forests | 2363 | 1.1 |
| Tropical and subtropical dry broadleaf forests | 4120 | 1.9 |
| Tropical and subtropical grasslands, savannas and shrublands | 31 376 | 14.4 |
| Tropical and subtropical moist broadleaf forests | 16 478 | 7.6 |
| Tundra | 2072 | 1 |
| No data | 724 | 0.3 |

* Biomes defined according to *Terrestrial Ecoregions of the World* (WWF) (D. M. Olson et al., 2001). Total may differ from 100 % due to rounding.

**Author contributions.** NHB is the scientific lead of the WoSIS project and wrote the first draft. LC developed the ETL and GraphQL procedures, while LMdS developed the new data model. All authors performed quality checks and data analyses and contributed to the writing and editing of the final paper.

**Competing interests.** The contact author has declared that none of the authors has any competing interests.

**Disclaimer.** ISRIC – World Soil Information remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**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, for which we are grateful. Regrettably, we can impossibly acknowledge all contributors (e.g. field surveyors, laboratory personnel, soil experts and database experts) individually. Therefore, we do this largely in a generic way (see https://www.isric.org/explore/wosis/wosis-contributing-institutions-and-experts, last access: 26 April 2024).

Our special thanks go to Eloi Ribeiro, former WoSIS database management expert at ISRIC, for his sustained support and advice during the refactoring of the ETL procedure. We also thank our colleague Laura Poggio for useful methodological discussions concerning methodological linkages between WoSIS and SoilGrids.

We gratefully acknowledge Alessandro Samuel-Rosa, Sebastian Doetterl and the topic editor Sibylle K. Hassler for their comments which greatly improved the scope of the initial submission.

The ETL procedures and new data model were co-developed in the framework of the European Union's Horizon 2020 HoliSoils project (grant agreement 101000289) and ISRIC's "global soil information and standards" workstream.

ISRIC – World Soil Information, legally registered as International Soil Reference and Information Centre, receives core funding from the Dutch Government.

**Financial support.** ISRIC receives core funding from the Dutch Government. This research has been supported by the EU Horizon 2020 (grant no. 101000289).

**Review statement.** This paper was edited by Sibylle K. Hassler and reviewed by Alessandro Samuel-Rosa and Sebastian Doetterl.

## References

Al-Shammary, A. A. G., Kouzani, A. Z., Kaynak, A., Khoo, S. Y., Norton, M., and Gates, W.: Soil Bulk Density Estimation Methods: A Review, Pedosphere, 28, 581–596, https://doi.org/10.1016/S1002-0160(18)60034-7, 2018.

ANSIS: Australian National Soil Information System, Australian National Soil Information System, Canberra (AU), https://ansis.net/ (last access: 26 April 2024), 2023.

Armas, D., Guevara, M., Bezares, F., Vargas, R., Durante, P., Osorio, V., Jiménez, W., and Oyonarte, C.: Harmonized Soil Database of Ecuador (HESD): data from 2009 to 2015, Earth Syst. Sci. Data, 15, 431–445, https://doi.org/10.5194/essd-15-431-2023, 2023.

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

Ballabio, C., Panagos, P., and Monatanarella, L.: Mapping topsoil physical properties at European scale using the LUCAS database, Geoderma, 261, 110–123, https://doi.org/10.1016/j.geoderma.2015.07.006, 2016.

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 consistent and comparable global soil data and information), Global Soil Partnership, FAO, 44 pp., http://www.fao.org/3/a-az922e.pdf (last access: 26 April 2024), 2014.

Baritz, R., Erdogan, H., Ahmadov, H., Ghanma, I., Lalljee, V. B., Wongmaneeroj, A., Collins, A., Monger, C., Ribeiro, J. L., Bertsch, F., Lalljee, V. B., with, Montanarella, L., Comerma, J., Khan, A., VandenBygaart, B., Gaistardo, C. C., Constantini, E., Galbraith, J. M., Schad, P., Lame, F., Suvannang, N., Hartmann, C., Medyckyj-Scott, D., Batjes, N. H., van Liedekerke, M., and Ziadat, F.: Implementation Plan for Pillar Five of the Global Soil Partnership: Providing mechanisms for the collation, analysis and exchange of consistent and comparable global soil data and information, ITPS, Rome, 48 pp., http://www.fao.org/3/a-bs756e.pdf (last access: 26 April 2024), 2017.

Baroni, G., Zink, M., Kumar, R., Samaniego, L., and Attinger, S.: Effects of uncertainty in soil properties on simulated hydrological states and fluxes at different spatio-temporal scales, Hydrol. Earth Syst. Sci., 21, 2301–2320, https://doi.org/10.5194/hess-21-2301-2017, 2017.

Batjes, N. H.: A world dataset of derived soil properties by FAO–UNESCO soil unit for global modelling, Soil Use Manage., 13, 9–16, https://doi.org/10.1111/j.1475-2743.1997.tb00550.x, 1997.

Batjes, N. H.: Harmonized soil profile data for applications at global and continental scales: updates to the WISE database, Soil Use Manage., 25, 124–127, https://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, https://doi.org/10.1016/j.geoderma.2016.01.034 2016.

Batjes, N. H.: Options for harmonising soil data obtained from different sources ISRIC – World Soil Information, Wageningen, 21 pp., https://doi.org/10.17027/isric-wdc-6ztd-eb19 2023.

Batjes, N. H. and Bridges, E. M.: Potential emissions of radiatively active gases from soil to atmosphere with special reference to methane: development of a global database (WISE), J. Geophys. Res., 99, 16479–16489, https://doi.org/10.1029/93JD03278, 1994.

Batjes, N. H., Ribeiro, E., van Oostrum, A., Leenaars, J., Hengl, T., and Mendes de Jesus, J.: WoSIS: providing standardised soil profile data for the world, Earth Syst. Sci. Data, 9, 1–14, https://doi.org/10.5194/essd-9-1-2017, 2017.

Batjes, N. H., Ribeiro, E., and van Oostrum, A.: Standardised soil profile data to support global mapping and modelling (WoSIS snapshot 2019), Earth Syst. Sci. Data, 12, 299–320, https://doi.org/10.5194/essd-12-299-2020, 2020.

Batjes, N. H. and van Oostrum, A. J. M.: WoSIS Procedures for standardizing soil analytical method descriptions, ISRIC – World Soil Information, Wageningen, 46 pp., https://doi.org/10.17027/isric-1dq0-1m83, 2023.

Batlle-Bayer, L., Batjes, N. H., and Bindraban, P. S.: Changes in organic carbon stocks upon land use conversion in the Brazilian Cerrado: A review, Agr. Ecosyst. Environ., 137, 47–58, https://doi.org/10.1016/j.agee.2010.02.003, 2010.

Bispo, A., Arrouays, D., Saby, N., Boulonne, L., and Fantappiè, M.: Proposal of methodological development for the LUCAS programme in accordance with national monitoring programmes. Towards climate-smart sustainable management of agricultural soils (EU H2020-SFS-2018-2020/H2020-SFS-2019) EJP Soil, 135 pp., https://ejpsoil.eu/fileadmin/projects/ejpsoil/WP6/EJP_SOIL_Deliverable_6.3_Dec_2021_final.pdf (last access: 26 April 2024), 2021.

Blakemore, L. C., Searle, P. L., and Daly, B. K.: Methods for chemical analysis of soils, Department of Scientific and Industrial Research, Lower Hutt, NZ, https://cdm20022.contentdm.oclc.org/digital/collection/p20022coll2/id/139/ (last access: 17 October 2024), 1981.

Bridges, E. M.: Soil horizon designations: past use and future prospects, CATENA, 20, 363–373, https://doi.org/10.1016/S0341-8162(05)80002-5, 1993.

Brus, D. J., Kempen, B., and Heuvelink, G. B. M.: Sampling for validation of digital soil maps, Eur. J. Soil Sci., 62, 394–407, https://doi.org/10.1111/j.1365-2389.2011.01364.x, 2011.

Brus, J.: Spatial sampling with R, Chapman and Hall R/C, New York, 2022.

Calisto, L.: ISRIC GraphQL web services for WoSIS and ISIS data access, ISRIC – World Soil Information, Wageningen, https://graphql.isric.org/ (last access: 26 April 2024), 2023.

Calisto, L., de Souza, L. M., and Batjes, N. H.: Standardised soil profile data for the world (WoSIS, December snapshot), ISRIC – World Soil Information, Wageningen [data set], https://doi.org/10.17027/isric-wdcsoils-20231130, 2023.

Cornu, S., Keesstra, S., Bispo, A., Fantappie, M., van Egmond, F., Smreczak, B., Wawer, R., Pavlù, L., Sobocká, J., Bakacsi, Z., Farkas-Iványi, K., Molnár, S., Møller, A. B., Madenoglu, S., Feiziene, D., Oorts, K., Schneider, F., Gonçalves, M. d. C., Mano, R., Garland, G., Skalský, R., O'Sullivan, L., Kasparinskis, R., and Chenu, C.: National soil data in EU countries, where do we stand?, Eur. J. Soil Sci., e13398, https://doi.org/10.1111/ejss.13398, 2023.

Cox, S. and David, J.: ISO 19156:2011 Geographic information – Observations and measurements International Organization for Standardization, https://www.iso.org/standard/32574.html (last access: 26 April 2024), 2011.

Cramer, M. D., Wootton, L. M., van Mazijk, R., and Verboom, G. A.: New regionally modelled soil layers improve prediction of vegetation type relative to that based on global soil models, Divers. Distrib., 25, 1736–1750, https://doi.org/10.1111/ddi.12973, 2019.

Cressie, N. and Kornak, J.: Spatial statistics in the presence of location error with an application to remote sensing of the environment, Stat. Sci., 18, 436–456, https://doi.org/10.1214/ss/1081443228, 2003.

Dai, Y., Shangguan, W., Wei, N., Xin, Q., Yuan, H., Zhang, S., Liu, S., Lu, X., Wang, D., and Yan, F.: A review of the global soil property maps for Earth system models, SOIL, 5, 137–158, https://doi.org/10.5194/soil-5-137-2019, 2019.

de Sousa, L., Kempen, B., Mendes de Jesus, J., Yigini, Y., Viatkin, K., Medyckyj-Scott, D., Richie, D. A., Wilson, P., van Egmond, F., and Baritz, R.: Conceptual design of the Global Soil Information System infrastructure, Rome, FAO and ISRIC, Wageningen, Netherlands, 30 pp., http://www.fao.org/3/cb4355en/cb4355en.pdf (last access: 26 April 2024), 2021.

de Sousa, L. M.: WoSIS data model 2023. Procedures Manual – Technical documentation, ISRIC – World Soil Information, Wageningen, https://git.wur.nl/isric/databases/wosis-docs (last access: 26 April 2024), 2023.

de Sousa, L. M., Kempen, B., Mendes de Jesus, J., Yigini, Y., Viatkin, K., Medyckyj-Scott, D., Richie, A., Wilson, P., van Egmond, F., and Baritz, R.: Conceptual desing of the Global Soil Information System infrastructure, ISRIC, FAO, Manaaki Whenua (Landcare Research), CSIRO, Wageningen UR, European Environment Agency, 30 pp., http://www.fao.org/3/cb4355en/cb4355en.pdf (last access: 26 April 2024), 2019.

de Sousa, L. M., Calisto, L., van Genuchten, P., Turdukulov, U., and Kempen, B.: Data model for the ISO 28258 domain model, ISRIC – World Soil Informatiom, https://iso28258.isric.org/ (last access: 26 April 2024), 2023.

Dijkshoorn, J. A., Huting, J. R. M., and Tempel, P.: Update of the 1:5 million Soil and Terrain Database for Latin America and the Caribbean (SOTERLAC, ver. 2.0), ISRIC – World Soil Information, Wageningen, Report 2005/01, https://www.isric.org/documents/document-type/isric-report-200501-update-15-million-soil-and-terrain-database-latin (last access: 24 April 2024), 2005.

Fantappie, M., Peruginelli, G., Conti, S., Rennes, S., van Egmond, F. M., and Le Bas, C.: Towards climate-smart sustainable management of agricultural soils: Deliverable 6.2 Report on the national and EU regulations on agricultural soil data sharing and national monitoring activities, 202 pp., https://edepot.wur.nl/642353 (last access: 24 April 2024), 2021.

FAO: Guidelines for the description of soils, 2nd edn., FAO, Rome, 66 pp., 1977.

FAO: Guidelines for soil description, 3rd rev. edn., FAO, Rome, 45 pp., https://edepot.wur.nl/570291 (last access: 24 April 2024), 1990.

FAO: Guidelines for soil description, 4th edn., FAO, Rome, 97 pp., http://www.fao.org/docrep/019/a0541e/a0541e.pdf (last access: 24 April 2024), 2006.

FAO and ISRIC: Soil and Terrain database for Southern Africa (1:2 million scale), ISRIC and FAO, Rome, FAO Land and Water Digital Media Series 25, 2003.

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., http://www.fao.org/3/a-i5199e.pdf (last access: 24 April 2024), 2015.

FAO, ISRIC, UNEP, and CIP: Soil and terrain digital database for Latin America and the Caribbean at 1 : 5 million scale, Food and Agriculture Organization of the United Nations, Rome, Land and Water Digital Media Series No. 5, 1998.

FAO, ISRIC, and UG: Soil and terrain database for central Africa (Burundi and Rwanda 1 : 1 million scale; Democratic Republic of the Congo 1 : 2 million scale), Food and Agricultural Organization of the United Nations, ISRIC – World Soil Information and Universiteit Gent, Rome, Land and Water Digital Media Series 33, https://www.isric.org/sites/default/files/isric_report_2006_07.pdf (last access: 24 April 2024), 2007.

FAO, IIASA, ISRIC, ISSCAS, and JRC: Harmonized World Soil Database (version 1.2), prepared by: Nachtergaele, F. O., van Velthuizen, H., Verelst, L., Wiberg, D., Batjes, N. H., Dijkshoorn, J. A., van Engelen, V. W. P., Fischer, G., Jones, A., Montanarella, L., Petri, M., Prieler, S., Teixeira, E., and Xuezheng, S., Food and Agriculture Organization of the United Nations (FAO), International Institute for Applied Systems Analysis (IIASA), ISRIC – World Soil Information, Institute of Soil Science – Chinese Academy of Sciences (ISSCAS), Joint Research Centre of the European Commission (JRC), Laxenburg, Austria, http://webarchive.iiasa.ac.at/Research/LUC/External-World-soil-database/HWSD_Documentation.pdf (last access: 24 April 2024), 2012.

Finke, P.: Quality assessment of digital soil maps: producers and users perspectives, in: Digital soil mapping: An introductory perspective, edited by: Lagacherie, P., McBratney, A., and Voltz, M., Elsevier, Amsterdam, 523–541, 2006.

Folberth, C., Skalsky, R., Moltchanova, E., Balkovic, J., Azevedo, L. B., Obersteiner, M., and van der Velde, M.: Uncertainty in soil data can outweigh climate impact signals in global crop yield simulations, Nat. Commun., 7, 11872, https://doi.org/10.1038/ncomms11872, 2016.

Gerasimova, M. I., Lebedeva, I. I., and Khitrov, N. B.: Soil horizon designation: State of the art, problems, and proposals, Eurasian Soil Sci., 46, 599–609, https://doi.org/10.1134/S1064229313050037, 2013.

Giller, K. E., Rowe, E. C., de Ridder, N., and van Keulen, H.: Resource use dynamics and interactions in the tropics: Scaling up in space and time, Agr. Syst., 88, 8–27, https://doi.org/10.1016/j.agsy.2005.06.016, 2006.

GlobalSoilMap: Specifications Tiered GlobalSoilMap products (Release 2.4), 52 pp., https://www.isric.org/documents/document-type/globalsoilmap-specifications-v24-07122015 (last access: 24 April 2024), 2015.

GLOSOLAN: GLOSOLAN best practice manual (on-line), FAO, GSP, Rome, https://www.fao.org/global-soil-partnership/glosolan-old/soil-analysis/standard-operating-procedures/en/#c763834 (last access: 24 April 2024), 2023.

Gobezie, T. B. and Biswas, A.: Break barriers in soil data stewardship by rewarding data generators, Nat. Rev. Earth Environ., 4, 353–354, https://doi.org/10.1038/s43017-023-00439-4, 2023.

Grimm, R. and Behrens, T.: Uncertainty analysis of sample locations within digital soil mapping approaches, GEODERMA, 155, 154–163, https://doi.org/10.1016/j.geoderma.2009.05.006, 2010.

Guevara, M., Olmedo, G. F., Stell, E., Yigini, Y., Aguilar Duarte, Y., Arellano Hernández, C., Arévalo, G. E., Arroyo-Cruz, C. E., Bolivar, A., Bunning, S., Bustamante Cañas, N., Cruz-Gaistardo, C. O., Davila, F., Dell Acqua, M., Encina, A., Figueredo Tacona, H., Fontes, F., Hernández Herrera, J. A., Ibelles Navarro, A. R., Loayza, V., Manueles, A. M., Mendoza Jara, F., Olivera, C., Osorio Hermosilla, R., Pereira, G., Prieto, P., Ramos, I. A., Rey Brina, J. C., Rivera, R., Rodríguez-Rodríguez, J., Roopnarine, R., Rosales Ibarra, A., Rosales Riveiro, K. A., Schulz, G. A., Spence, A., Vasques, G. M., Vargas, R. R., and Vargas, R.: No silver bullet for digital soil mapping: country-specific soil organic carbon estimates across Latin America, SOIL, 4, 173–193, https://doi.org/10.5194/soil-4-173-2018, 2018.

Hassani, A., Smith, P., and Shokri, N.: Negative correlation between soil salinity and soil organic carbon variability, P. Natl. Acad. Sci. USA, 121, e2317332121, https://doi.org/10.1073/pnas.2317332121, 2024.

Hengl, T., de Jesus, J. M., Heuvelink, G. B. M., Gonzalez, M. R., Kilibarda, M., Blagotic, A., Shangguan, W., Wright, M. N., Geng, X. Y., Bauer-Marschallinger, B., Guevara, M. A., Vargas, R., MacMillan, R. A., Batjes, N. H., Leenaars, J. G. B., Ribeiro, E., Wheeler, I., Mantel, S., and Kempen, B.: SoilGrids250m: Global gridded soil information based on machine learning, PLoS ONE, 12, e0169748, https://doi.org/10.1371/journal.pone.0169748, 2017.

Heuvelink, G. B. M.: Uncertainty quantification of GlobalSoilMap products in: GlobalSoilMap. Basis of the Global Spatial Soil Information System, edited by: Arrouays, D., McKenzie, N., Hempel, J., Forges, A. R. D., and McBratney, A., Taylor & Francis Group, London, UK, 335–240, 2014.

Heuvelink, G. B. M., Brown, J. D., and van Loon, E. E.: A probabilistic framework for representing and simulating uncertain environmental variables, Int. J. Geogr. Inf. Sci., 21, 497–513, https://doi.org/10.1080/13658810601063951, 2007.

Heuvelink, G. B. M., Angelini, M. E., Poggio, L., Bai, Z. G., Batjes, N. H., van den Bosch, R., Bossio, D., Estella, S., Lehmann, J., Olmedo, G. F., and Sanderman, J.: Machine learning in space and time for modelling soil organic carbon change, Eur. J. Soil Sci., 72, 1607–1623, https://doi.org/10.1111/ejss.12998, 2021.

Huang, Y., Song, X., Wang, Y.-P., Canadell, J. G., Luo, Y., Ciais, P., Chen, A., Hong, S., Wang, Y., Tao, F., Li, W., Xu, Y., Mirzaeitalarposhti, R., Elbasiouny, H., Savin, I., Shchepashchenko, D., Rossel, R. A. V., Goll, D. S., Chang, J., Houlton, B. Z., Wu, H., Yang, F., Feng, X., Chen, Y., Liu, Y., Niu, S., and Zhang, G.-L.: Size, distribution, and vulnerability of the global soil inorganic carbon, Science, 384, 233–239, https://doi.org/10.1126/science.adi7918, 2024.

ICP Forests: ICP Forests monitoring Manual. Part XVI: Quality assurance and control in laboratories (ver 2020-1), Eberswalde, Germany, 46 pp., https://www.icp-forests.org/pdf/manual/2020/ICP_Manual_part16_2020_QAQC_Labs_version_2020-1.pdf (last access: 26 April 2024), 2020.

ICP Forests: ICP Forests monitoring Manual. Part X: Sampling and analysis of soil, Eberswalde, Germany, https://storage.ning.com/topology/rest/1.0/file/get/9995584862?profile=original (last access: 26 April 2024), 2021a.

ICP Forests: ICP Forests monitoring Manual Eberswalde (Germany), http://icp-forests.net/page/icp-forests-manual (last access: 26 April 2024), 2021b.

ISO-19139: Geographic information XML schema implementation Part 1: Encoding rules, https://www.iso.org/standard/67253.html (last access: 26 April 2024), 2019.

ISRIC: Data and Software Policy, ISRIC – World Soil Information (WDC – Soils) Wageningen, 6 pp., https://www.isric.org/sites/default/files/user/ISRIC_Data_Policy_2016jun21doi.pdf (last access: 26 April 2024), 2016.

IUSS Working Group WRB: World Reference Base for Soil Resources, 2nd edn., FAO, Rome, World Soil Resources Report 103, 145 pp., http://www.fao.org/ag/agl/agll/wrb/doc/wrb2006final.pdf (last access: 26 April 2024), 2006.

IUSS Working Group WRB: World Reference Base for soil resources 2014 – International soil classification system for naming soils and creating legends for soil maps (update 2015), Global Soil Partnership, International Union of Soil Sciences, and Food and Agriculture Organization of the United Nations, Rome, World Soil Resources Reports 106, 182 pp., http://www.fao.org/3/i3794en/I3794en.pdf (last access: 26 April 2024), 2015.

IUSS Working Group WRB: World Reference Base for soil resources 2022 – International soil classification system for naming soils and creating legends for soil maps, International Union of Soil Sciences, Vienna (Austria), 284 pp., https://www.isric.org/sites/default/files/WRB_fourth_edition_2022-12-18.pdf (last access: 26 April 2024), 2022.

Ivushkin, K., Bartholomeus, H., Bregt, A. K., Pulatov, A., Kempen, B., and de Sousa, L.: Global mapping of soil salinity change, Remote Sens. Environ., 231, 111260, https://doi.org/10.1016/j.rse.2019.111260, 2019.

Kalra, Y. P. and Maynard, D. G.: Methods manual for forest soil and plant analysis, Forestry Canada, Edmonton (Alberta), 116 pp., https://cfs.nrcan.gc.ca/publications/download-pdf/11845 (last access: 26 April 2024), 1991.

Leenaars, J. G. B., van Oostrum, A. J. M., and Ruiperez Gonzalez, M.: Africa Soil Profiles Database: A compilation of georeferenced and standardised legacy soil profile data for Sub Saharan Africa (version 1.2), Africa Soil Information Service (AfSIS) and ISRIC – World Soil Information, Wageningen, Report 2014/01, 160 pp., http://www.isric.org/sites/default/files/isric_report_2014_01.pdf (last access: 26 April 2024), 2014.

Leenaars, J. G. B., Claessens, L., Heuvelink, G. B. M., Hengl, T., Ruiperez González, M., van Bussel, L. G. J., Guilpart, N., Yang, H., and Cassman, K. G.: Mapping rootable depth and root zone plant-available water holding capacity of the soil of sub-Saharan Africa, Geoderma, 324, 18–36, https://doi.org/10.1016/j.geoderma.2018.02.046, 2018.

Luo, Z., Viscarra-Rossel, R. A., and Qian, T.: Similar importance of edaphic and climatic factors for controlling soil organic carbon stocks of the world, Biogeosciences, 18, 2063–2073, https://doi.org/10.5194/bg-18-2063-2021, 2021.

Lutz, F., Stoorvogel, J. J., and Müller, C.: Options to model the effects of tillage on $N_2O$ emissions at the global scale, Ecol. Model., 392, 212–225, https://doi.org/10.1016/j.ecolmodel.2018.11.015, 2019.

Magnusson, B. and Örnemark, U.: The Fitness for Purpose of Analytical Methods – A Laboratory Guide to Method Validation and Related Topics, 2nd edn., Eurachem, https://www.eurachem.org/images/stories/Guides/pdf/MV_guide_2nd_ed_EN.pdf (last access: 26 April 2024), 2014.

Maire, V., Wright, I. J., Prentice, I. C., Batjes, N. H., Bhaskar, R., van Bodegom, P. M., Cornwell, W. K., Ellsworth, D., Niinemets, U., Ordonez, A., Reich, P. B., and Santiago, L. S.: Global effects of soil and climate on leaf photosynthetic traits and rates, Glob. Ecol. Biogeogr., 24, 706–717, https://doi.org/10.1111/geb.12296, 2015.

Malhotra, A., Todd-Brown, K., Nave, L. E., Batjes, N. H., Holmquist, J. R., Hoyt, A. M., Iversen, C. M., Jackson, R. B., Lajtha, K., Lawrence, C., Vinduskova, O., Wieder, W., Williams, M., Hugelius, G., and Harden, J.: The landscape of soil carbon data: emerging questions, synergies and databases, Prog. Phys. Geogr.-Earth and Environment, 43, 707–719, https://doi.org/10.1177/0309133319873309, 2019.

Meyer, H. and Pebesma, E.: Predicting into unknown space? Estimating the area of applicability of spatial prediction models, Methods Ecol. Evol., 12, 1620–1633, https://doi.org/10.1111/2041-210X.13650, 2021.

Moulatlet, G. M., Zuquim, G., Figueiredo, F. O. G., Lehtonen, S., Emilio, T., Ruokolainen, K., and Tuomisto, H.: Using digital soil maps to infer edaphic affinities of plant species in Amazonia: Problems and prospects, Ecol. Evol., 7, 8463–8477, https://doi.org/10.1002/ece3.3242, 2017.

Munzert, M., Kießling, G., Übelhör, W., Nätscher, L., and Neubert, K.-H.: Expanded measurement uncertainty of soil parameters derived from proficiency-testing data, J. Plant Nutr. Soil Sci., 170, 722–728, https://doi.org/10.1002/jpln.200620701, 2007.

NATP: North American Proficiency Testing (NAPT) Program, http://www.naptprogram.org/ (last access: 26 April 2024), 2015.

Nenkam, A. M., Wadoux, A. M. J. C., Minasny, B., McBratney, A. B., Traore, P. C. S., Falconier, G. N., and Whitbread, A. M.: Using homosoils for quantitative extrapolation of soil mapping models, Eur. J. Soil Sci., 73, e13285, https://doi.org/10.1111/ejss.13285, 2022.

NPDB: National Pedon Database Canada, Agriculture and Agri-food Canada, https://sis.agr.gc.ca/cansis/nsdb/npdb/index.html (last access: 26 April 2024), 2023.

OGC: Soil Data IE (Interoperability Experiment), Open Geospatial Consortium (OGC), https://www.opengeospatial.org/projects/initiatives/soildataie (last access: 26 April 2024), 2019.

Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V. N., Underwood, E. C., D'amico, J. A., Itoua, I., Strand, H. E., Morrison, J. C., Loucks, C. J., Allnutt, T. F., Ricketts, T. H., Kura, Y., Lamoreux, J. F., Wettengel, W. W., Hedao, P., and Kassem, K. R.: Terrestrial Ecoregions of the World: A New Map of Life on Earth: A new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity, BioScience, 51, 933–938, https://doi.org/10.1641/0006-3568(2001)051[0933:TEOTWA]2.0.CO;2, 2001.

Olson, R. J., Johnson, K. R., Zheng, D. L., and Scurlock, J. M. O.: Global and regional ecosystem modelling: databases of model drivers and validation measurements, Oak Ridge National Laboratory, Oak Ridge, ORNL/TM-2001/196, 95 pp., http://www-eosdis.ornl.gov/npp/GPPDI/comp/NPP_TM196.pdf (last access: 26 April 2024), 2001.

Orgiazzi, A., Ballabio, C., Panagos, P., Jones, A., and Fernandez-Ugalde, O.: LUCAS Soil, the largest expandable soil dataset for Europe: a review, Eur. J. Soil Sci., 69, 140–153, https://doi.org/10.1111/ejss.12499, 2018.

Padarian, J. and McBratney, A. B.: A new model for intra- and inter-institutional soil data sharing, SOIL, 6, 89–94, https://doi.org/10.5194/soil-6-89-2020, 2020.

Palma, R., Janiak, B., Sousa, L. M. D., Schleidt, K., Tomáš Rezník, Egmond, F. v., Leenaars, J., Moshou, D., Mouazen, A., Peter Wilson, Medyckyj-Scott, D., Ritchie, A., Yigini, Y., and Vargas, R.: GloSIS: The Global Soil Information System Web Ontology, arXiv [preprint], 2403.16778, https://doi.org/10.48550/arXiv.2403.16778, 2024.

Poeplau, C., Don, A., Flessa, H., Heidkamp, A., Jacobs, A., and Prietz, R.: Erste Bodenzustandserhebung Landwirtschaft – Kerndatensatz, Thünen-Institut, I. f. A., Göttingen, 2020.

Poggio, L., de Sousa, L. M., Batjes, N. H., Heuvelink, G. B. M., Kempen, B., Ribeiro, E., and Rossiter, D.: SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty, SOIL, 7, 217–240, https://doi.org/10.5194/soil-7-217-2021, 2021.

Rayment, E. R. and Lyons, D. J.: Soil chemical methods – Australasia, CSIRO Publishing, 495 pp., 2011.

R Core Team: R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, https://www.R-project.org (last access: 26 April 2024), 2021.

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., http://www.isric.org/sites/default/files/isric_report_2015_03.pdf (last access: 26 April 2024), 2015.

Ribeiro, E., Batjes, N. H., and Van Oostrum, A. J. M.: World Soil Information Service (WoSIS) – Towards the standardization and harmonization of world soil data. Procedures Manual 2020, ISRIC – World Soil Information, Wageningen, ISRIC Report 2020/01, 153 pp., https://doi.org/10.17027/isric-wdc-2020-01, 2020.

Robinson, N. J., Dahlhaus, P. G., Wong, M., MacLeod, A., Jones, D., and Nicholson, C.: Testing the public–private soil data and information sharing model for sustainable soil management outcomes, Soil Use Manage., 35, 94–104, https://doi.org/10.1111/sum.12472, 2019.

Rossel, R. A. V. and McBratney, A. B.: Soil chemical analytical accuracy and costs: implications from precision agriculture, Australian J. Exp. Agric., 38, 765–775, 1998.

Sanderman, J., Hengl, T., and Fiske, G. J.: Soil carbon debt of 12,000 years of human land use, P. Natl. Acad. Sci. USA, 114, 9575–9580, https://doi.org/10.1073/pnas.1706103114, 2017.

Sayre, R.: World Terrestrial Ecosystems (WTE) 2020, U.S. Geological Survey data release [data set], https://doi.org/10.5066/P9DO61LP, 2022.

Sayre, R., Dangermond, J., Frye, C., Vaughan, R., Aniello, P., Breyer, S., Cribbs, D., Hopkins, D., Nauman, R., Derrenbacher, W., Burton, D., Grosse, A., True, D., Metzger, M., Hartmann, J., Moosdorf, N., Dürr, H., Paganini, M., DeFourny, P., Arino, O., and Maynard, S.: A New Map of Global Ecological Land Units – An Ecophysiographic Stratification Approach, Association of American Geographers, Washington DC, 46 pp., https://www.aag.org/wp-content/uploads/2021/12/AAG_Global_Ecosyst_bklt72.pdf (last access: 26 April 2024), 2014.

Schoeneberger, P. J., Wysocki, D. A., Benham, E. C., and Soil Survey Staff: Field book for describing and sampling soils (ver. 3.0, Reprint 2021), National Soil Survey Center Natural Resources Conservation Service, U.S. Department of Agriculture, Lincoln (NE), 2012.

Shepherd, K. D., Ferguson, R., Hoover, D., van Egmond, F., Sanderman, J., and Ge, Y.: A global soil spectral calibration library and estimation service, Soil Security, 7, 100061, https://doi.org/10.1016/j.soisec.2022.100061, 2022.

Shi, G., Shangguan, W., Zhang, Y., Li, Q., Wang, C., and Li, L.-J.: Reducing Location Error of Legacy Soil Profiles Leads to Significant Improvement in Digital Soil Mapping, SSRN, https://doi.org/10.2139/ssrn.4643055, 2023.

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., http://www.nrcs.usda.gov/Internet/FSE_DOCUMENTS/nrcs142p2_052226.pdf (last access: 26 April 2024), 2011.

Soil Survey Staff: Keys to Soil Taxonomy, 12th ed., USDA-Natural Resources Conservation Service, Washington, DC, 2014.

Soil Survey Staff: Soil Survey Manual (rev. ed.), edited by: Ditzler, C., Scheffe, K., and Monger, H. C., United States Agriculture Handbook 18, USDA, Washington, 2017.

Soil Survey Staff: Soil Survey Laboratory Methods Manual (Version 6.0., Part1: Curren methods), U.S. Department of Agriculture, Natural Resources Conservation Service, Lincoln (Nebraska), 1001 pp., 2022a.

Soil Survey Staff: Keys to Soil Taxonomy, 13th edn., USDA-Natural Resources Conservation Service, Washington, DC., 2022b.

Sothe, C., Gonsamo, A., Arabian, J., and Snider, J.: Large scale mapping of soil organic carbon concentration with 3D machine learning and satellite observations, Geoderma, 405, 115402, https://doi.org/10.1016/j.geoderma.2021.115402, 2022.

Suvannang, N., Hartmann, C., Yakimenko, O., Solokha, M., Bertsch, F., and Moody, P.: Evaluation of the First Global Soil Laboratory Network (GLOSOLAN) online survey for assessing soil laboratory capacities, Global Soil Partnership (GSP)/Food and Agriculture Organization of the United Nations (FAO), Rome, GLOSOLAN/18/Survey Report, 54 pp., http://www.fao.org/3/CA2852EN/ca2852en.pdf (last access: 26 April 2024), 2018.

Tempel, P., van Kraalingen, D., Mendes de Jesus, J., and Reuter, H. I.: Towards an ISRIC World Soil Information Service (WOSIS ver. 1.0), ISRIC – World Soil Information, Wageningen, ISRIC Report 2013/02, 188 pp., https://www.isric.org/sites/default/files/isric_report_2013_02.pdf (last access: 26 April 2024), 2013.

Turek, M. E., Poggio, L., Batjes, N. H., Armindo, R. A., de Jong van Lier, Q., de Sousa, L., and Heuvelink, G. B. M.: Global mapping of volumetric water retention at 100, 330 and 15 000 cm suction using the WoSIS database, Int. Soil Water Conserv. Res., 11, 225–239, https://doi.org/10.1016/j.iswcr.2022.08.001, 2023.

USDA-NCSS: National Cooperative Soil Survey (NCSS) Soil Characterization Database, United States Department of Agriculture, Natural Resources Conservation Service, Lincoln, https://ncsslabdatamart.sc.egov.usda.gov/database_download.aspx (last access: 26 April 2024), 2021.

van de Ven, T. and Tempel, P.: ISIS 4.0 – ISRIC Soil Information System: User Manual, International Soil Reference and Information Centre, Wageningen, Technical Paper 15 (rev. ed.), https://www.isric.org/sites/default/files/ISRIC_TechPap15b.pdf (last access: 26 April 2024), 1994.

van Engelen, V. W. P., Verdoodt, A., Dijkshoorn, K., and van Ranst, E.: SOTER database for Central Africa – DR Congo, Burundi and Rwanda (SOTERCAF; ver. 1.0), Laboratory of Soil Science (University of Ghent), FAO and ISRIC - World Soil Information, Wageningen, ISRIC REport 2006/07, 28 pp., http://www.isric.org/Isric/Webdocs/Docs/ISRIC_Report_2006_07.pdf (last access: 15 August 2007), 2006.

van Leeuwen, C., Mulder, V. L., Batjes, N. H., and Heuvelink, G. B. M.: Statistical modelling of measurement error in wet chemistry soil data, Eur. J. Soil Sci., 73, 13137, https://doi.org/10.1111/ejss.13137, 2022.

van Leeuwen, C. C. E., Mulder, V. L., Batjes, N. H., and Heuvelink, G. B. M.: Effect of measurement error in wet chemistry soil data on the calibration and model performance of pedotransfer functions, Geoderma, 442, 116762, https://doi.org/10.1016/j.geoderma.2023.116762, 2024.

Van Looy, K., Bouma, J., Herbst, M., Koestel, J., Minasny, B., Mishra, U., Montzka, C., Nemes, A., Pachepsky, Y., Padarian, J., Schaap, M., Tóth, B., Verhoef, A., Vanderborght, J., van der Ploeg, M., Weihermüller, L., Zacharias, S., Zhang, Y., and Vereecken, H. C. R. G.: Pedotransfer functions in Earth system science: challenges and perspectives, Rev. Geophys., 55, 1199–1256, https://doi.org/10.1002/2017RG000581, 2017.

van Reeuwijk, L. P.: On the way to improve international soil classification and correlation: the variability of soil analytical data, ISRIC, Wageningen, Annual Report 1983, 7–13 pp., https://www.isric.org/sites/default/files/isric_annual_report_1983.pdf (last access: 26 April 2024), 1983.

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., Aïchi, 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-Sci. Rev., 155, 198–230, https://doi.org/10.1016/j.earscirev.2016.01.012, 2016.

von Haden, A. C., Yang, W. H., and DeLucia, E. H.: Soils' dirty little secret: Depth-based comparisons can be inadequate for quantifying changes in soil organic carbon and other mineral soil properties, Global Change Biol., 26, 3759–3770, https://doi.org/10.1111/gcb.15124, 2020.

Wang, M., Guo, X., Zhang, S., Xiao, L., Mishra, U., Yang, Y., Zhu, B., Wang, G., Mao, X., Qian, T., Jiang, T., Shi, Z., and Luo, Z.: Global soil profiles indicate depth-dependent soil carbon losses under a warmer climate, Nat. Commun., 13, 5514, https://doi.org/10.1038/s41467-022-33278-w, 2022.

Wang, M., Zhang, S., Guo, X., Xiao, L., Yang, Y., Luo, Y., Mishra, U., and Luo, Z.: Responses of soil organic carbon to climate extremes under warming across global biomes, Nat. Clim. Change, 14, 98–105, https://doi.org/10.1038/s41558-023-01874-3, 2024.

WEPAL: ISE Reference Material – A list with all available ISE reference material samples, WEPAL (Wageningen Evaluating Programmes for Analytical Laboratories), Wageningen, 110 pp., http://www.wepal.nl/website/products/RefMatISE.htm (last access: 26 April 2024), 2019.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., Gonzalez-Beltran, A., Gray, A. J. G., Groth, P., Goble, C., Grethe, J. S., Heringa, J., 't Hoen, P. A. C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S. J., Martone, M. E., Mons, A., Packer, A. L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M. A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., and Mons, B.: The FAIR Guiding Principles for scientific data management and stewardship, Sci. Data, 3, 160018, https://doi.org/10.1038/sdata.2016.18, 2016.