ISRIC
World Soil Information

Tier 1 and Tier 2 data in the context of the federated Global Soil Information System (GLOSIS)

ISRIC Report 2019/01

Niels H. Batjes
Bas Kempen
Fenny van Egmond

ISRIC - World Soil Information
Droevendaalsesteeg 3 (building 101)
6708 PB Wageningen
The Netherlands
E-mail: soils@isric.org

Note: This report has been prepared in support of the development of GLOSIS (Global Soil Information System) by ISRIC – World Soil Information, as host institute of the Soil Data Facility (SDF) for the Global Soil Partnership. The document was endorsed by the International Network of Soil Information Institutions (INSII) in June 2020.

.

Citation

Batjes N.H., Bas Kempen and Fenny van Egmond 2019. Tier 1 and Tier 2 data in the context of the federated Global Soil Information System (GLOSIS). Report 2019/01, ISRIC – World Soil Information, Wageningen (doi: 10.17027/isric-dcmkd4-g431).

ISRIC
World Soil Information

ISRIC Report 2019/01

Tier 1 and Tier 2 data in the context of the federated Global Soil Information System (GLOSIS)

Niels H. Batjes, Bas Kempen and Fenny van Egmond

with contributions from the GSP Pillar 4 Working Group

and

the International Network of Soil Information Institutions

September 2019

CONTENTS

List of figures

List of tables

Acronyms and abbreviations

AfSP	Africa Soil Profiles database
countrySIS	Country Soil Information System (within GLOSIS framework)
DOI	Digital Object Identifier
eSOTER	Regional pilot platform as EU contribution to a Global Soil Observing System
FAO	Food and Agriculture Organization of the United Nations
GeoSciML	Geoscience Markup Language
GEOSS	Global Earth Observation System of Systems
GfSD	FAO guidelines for soil description
GML	Geography Markup Language
GLOSIS	Global Soil Information System (GSP)
GLOSOLAN	Global Soil Laboratory Network (established by the Global Soil Partnership)
GODAN	Global Open Data for Agriculture and Nutrition
GSP	Global Soil Partnership
ICSU	International Council for Science
INSII	International Network of Soil Information Institutions
ISO	International Organization for Standardization
ISRIC	ISRIC - World Soil Information (legally registered as International Soil Reference and Information Centre)
IUSS	International Union of Soil Sciences
OGC	Open Geospatial Consortium
SDI	Spatial Data Infrastructure
SDF	Soil Data Facility; technical backbone of GloSIS (ISRIC – World Soil Information)
SoilML	Soil Markup Language
SOILSTAT	A system for monitoring, forecasting and reporting periodically on the status of global soil resources. Part of the FAOSTAT family of global status databases and monitoring.
SOTER	Soil and Terrain database programme
WDC-Soils	World Data Center for Soils, regular member of the International Council for Science (ICS) World Data System (WDS)
WISE	World Inventory of Soil Emission potentials (harmonized soil profile data for the world)
WoSIS	World Soil Information Service (ISRIC's centralised server database)
XML	Extensible Markup Language

1 Introduction

1.1 Context

Soil is one of the most important natural resources, yet it is non-renewable on a human life scale (FAO-ITPS 2014). It plays a vital role in the Earth's ecosystem: foothold for plant roots, storage of water and nutrients for plants to grow, filtering of rainwater and regulating its discharge, storage of organic matter, buffering of pollutants, and biodiversity. Sustainable use of this resource requires adequate information on its spatial and temporal variation. In this context, there is a need for guidance and an infrastructure to collate, handle, standardise and make available the various sources of soil data as collated in different countries of the world, while recognising differences in technical capabilities of the various data holders. Pillar 4 of the Global Soil Partnership (GSP) will develop such a framework under the denomination of GLOSIS (Global Soil Information System).

The main components of GLOSIS are shown in Figure 1. They include: monitoring, forecasting and status reporting (SoilSTAT); soil profile (point) data; global soil polygon coverage; global grids of soil properties; and, a capacity development programme on soil information. For each of these, specific technical documents will be prepared building on earlier recommendations of the GSP (GSP and FAO 2016, 2017a).

The main purpose of GLOSIS is to enable exchange of consistent, standardised soil data within and between countries, thus facilitating their potential use due enhanced accessibility. The latter in accordance with the licences specified by the original data providers (GSP and FAO 2017b). Within such a federated system there is no single central database. Rather, a central discovery hub will provide a) a registry of national nodes (soil information systems, SIS) that are *compliant* with GLOSIS *exchange standards*, and b) provide a search engine through which users may query the standardised data shared by the various national GLOSIS nodes.

Figure 1. Schematic representation of GLOSIS (Global Soil Information System)

This document provides an overview of the range of soil profile data that may be assembled and exchanged within the structure of GLOSIS. Its purpose is to guide development of the GLOSIS data model itself, and its

subsequent implementation within the federated GLOSIS construct (see de Sousa *et al.* 2019). As such, this report is not meant to define standard abbreviations/codes for the various attributes nor to define the standard units of analysis and measurement.

GSP Pillar 4 identified the need to develop/support national level databases, based on common de facto standards, that contain a list of predefined, commonly required, soil properties for geo-referenced soil profiles (site and layers), and this within a federated structure. Originally, two types of soil profile *databases* (called Tier 1 and Tier 2) were envisaged for this in the implementation plan (GSP and FAO 2016). However, this would imply creation of a central GLOSIS database in contradiction to the federated, bottom-up structure envisaged by the GSP. Consequently, the foreseen GLOSIS spatial data infrastructure does not accommodate centralized data bases (de Sousa *et al.* 2019). Soil data providers will share (parts of) their data via web services that are registered by the discovery hub. Consequently, we propose to speak of Tier 1 and Tier 2 type data (hereafter referred to as T1 and T2) in the context of GLOSIS and not T1 and T2 *databases*. A T1 can accommodate all soil profile data that have been collated in a given country and standardised in accord with (emerging) GLOSIS protocols. Alternatively, T2 are subsets of the national T1's that can be used to address specific issues (i.e. soil functions and threats as required for SoilSTAT).

Summarising from de Sousa et al. (2019), there will be three different options for GSP members to join the GLOSIS federation. Countries with a 'SIS ready for connection' (stage A, top in Figure 2), those in 'the process of developing a national SIS' (stage B, middle position in Figure 2), and countries in the 'early stage of developing a SIS that will share their data through the support node' (Stage C, bottom position in Figure 2).

Online data exchange will require a dynamic connection between the national T1's using a standard data exchange model (de Sousa *et al.* 2019). Critical in this process will be the work towards 'global harmonisation and exchange of soil data' by ISO, OGC, GODAN, GSP pillar 5 and other groups aimed at developing a common *de facto* standard for soil data-interoperability (GSP and FAO 2017a; Wilson 2016).

In view of the bottom-up and federated approach adopted for GLOSIS, all national SIS's will be managed and updated at source level that is in the country of origin. General guidance on procedures for standardising the source data to a common GLOSIS standard will be provided by the GSP, in conjunction with a capacity-building component (mainly for 'case C' countries). However, type A and type B SIS's should also 'transform' their data for a T1 according to these procedures to permit consistent querying through the central discovery hub. It is considered good practice that all data providers maintain the original data in their own national soil database together with metadata on data lineage, 'conversion rules applied for categorical data', and a description of the analytical methods and units of measurement.

Some countries present part of their soil data in the form of so-called modal soil profile(s). These are then considered to be representative of a defined typological unit (soil polygon). Each component soil unit thereof is characterised by one or more representative soil profile(s). The properties thereof are given as the average or another statistic for the respective soil layers and soil properties under consideration. By its nature, this type of derived, polygon-referenced soil data are considered beyond the scope of this report that focuses on geo-referenced point data. However, in principle, they could be considered as specific 'thematic subsets' within the overall GLOSIS structure.

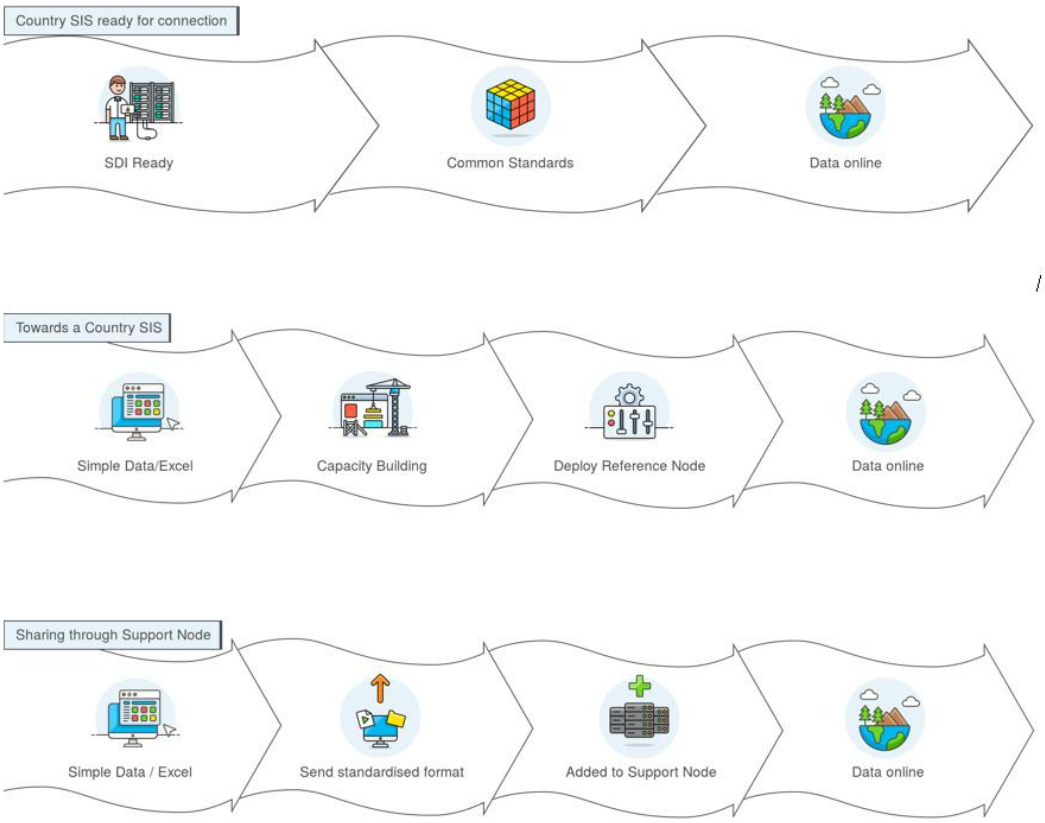

Figure 2. Three different levels of adhesion to the GloSIS federation (de Sousa *et al.* 2019). Top (A) – *tailored* implementation: data providers with an established soil spatial data infrastructure apply/implement the GloSIS data exchange, i.e. they bring their data to the common GloSIS standard and then publish their data through the GloSIS discovery hub. Middle (B) – reference implementation: data providers with soil data stored in simple databases or plain tables and with the ambition to establish a (national) soil information system will be trained in setting up a reference node. Once a reference node is deployed the data are served through a national web portal and are discoverable through the GloSIS data hub if allowed by the data provider. Bottom (C) – support implementation. Data providers with their soil data stored in simple databases or plain tables standardize their data according to the GloSIS data model and send their data to the GSP who will act as a custodian of the data by storing the data in the support node. The data will be discoverable through the GloSIS data hub when allowed by the data provider(s).

1.2 Scope of report

This report serves to give an indication of the range of soil properties that may be encountered in national T1, which in turn will be queried through the central discovery hub to prepare specific T2 subsets (e.g. for assessing soil salinity). This overview is meant to guide development of the GLOSIS data model (de Sousa *et al.* 2019) itself. Guidelines and technical specifications for developing a national SIS (or 'CountrySIS') for those countries that do yet have an operational system will be presented in a separate report (see GSP-SDF 2018a). The required vocabulary (service) will be prepared by GSP Pillar 5 (GSP and FAO 2017a).

As indicated by Pillar 4 (GSP and FAO 2016), T1 may include data for legacy profiles as well as newly-collected soil profiles (conventional wet chemistry and soil properties derived from spectral methods), both with date of collection and provenance (data lineage) documented.

General considerations about the diversity of national soil profile data(bases) are made in Chapter 2, to provide an insight concerning the range of soil properties that may be encountered for standardisation to the GSP conventions in a national T1. Chapter 3, gives an indication of specific purposes data that may be queried from the federated T1's, for example a specific T2 for assessing soil water retention. Concluding remarks on the way forward are made in Chapter 4. Supplementary information and examples are provided in the Appendices.

2 National soil profile holdings provide the basis for T1 data

2.1 General considerations

Many soil profiles have been collated worldwide. These profiles have been described, sampled, analysed and classified according to a wide range of procedures and standards, and stored according to various data models (AD-HOC Arbeitsgruppen Boden 2005; FAO 2006; Isbell 1996; Omuto *et al.* 2012; Ribeiro *et al.* 2018; Soil Science Division Staff 2017).

The *Guidelines for Soil Description*, now in its Fourth Edition (FAO 2006), and USDA Soil Survey Manual (Soil Science Division Staff 2017) provide a good consensus of the type and range of data that may be described and collected/sampled during a field survey. As a result, the range of properties that may occur in national soil databases is diverse. Importantly, as indicated, these disparate (source) data will need to be brought under a common (i.e., comparable) denominator to become queryable as a national T1 through the GLOSIS web portal.

The 'soil description status' provides a useful descriptor for the inferred quality of a soil profile description (FAO, 2006) in that it considers the quality of the soil description and the analytical data (Table 1). It serves to provide a qualitative measure of the degree to which the available field and laboratory data allow for a full characterisation of a soil profile according to the World Reference Base for Soil Resources (IUSS Working Group WRB 2015). Hence, the description status is allocated after completion of the analyses. The profile description status is indicative of the (inferred) reliability of soil profile information entered into a database (Table 1).

Table 1. Describing the inferred quality of soil data

No	Status[a]	Description
1	Reference profile description	No essential elements or details are missing from the description, sampling or analysis. The accuracy and reliability of the description and analytical results permit the full characterization of all soil horizons to a depth of 125 cm, or more if required for classification, or down to a C or R horizon or layer, which may be shallower.
2	Routine profile description	No essential elements are missing from the description, sampling or analysis. The number of samples collected is sufficient to characterize all major soil horizons, but may not allow precise definition of all subhorizons, especially in the deeper soil. The profile depth is 80 cm or more, or down to a C or R horizon or layer, which may be shallower. Additional augering and sampling may be required for lower level classification.
3	Incomplete description	Certain relevant elements are missing from the description, an insufficient number of samples was collected, or the reliability of the analytical data does not permit a complete characterization of the soil. However, the description is useful for specific purposes and provides

No	Status[a]	Description
		a satisfactory indication of the nature of the soil at high levels of soil taxonomic classification.
4	Soil augering description	Soil augerings do no permit a comprehensive soil profile description. Augerings are made for routine soil observation and identification in soil mapping, and for that purpose normally provide a satisfactory indication of the soil characteristics. Soil samples may be collected from augerings.
5	Other descriptions	Essential elements are missing from the description, preventing a satisfactory soil characterization and classification.

[a] See FAO (2006, p.6)

Observations and measurements for soil profiles as well as auger observations (by horizon or layer) may be considered noting though that the availability of well described, measured soil data is considered essential for GLOSIS. In so far as possible with the available data and available resources, all profiles should be correlated by national experts to the World Reference Base for Soil Resources (IUSS Working Group WRB 2015) system or the USDA Soil Taxonomy (Soil Survey Staff 2014b); both systems are endorsed by the International Union of Soil Sciences IUSS and used worldwide (Hempel 2014). It is also considered good practice to include the national classification system to allow for possible future correlation (see Appendix 2).

Laboratory (analytical) procedures required for consistent soil classification according to WRB are given as van Reeuwijk (2002) and Soil Survey Staff (2014a). Similarly, for consistent classification according to USDA Soil Taxonomy specific laboratory methods are required (Soil Survey Staff 2014a). In practice, however, these 'required' methods are not applied on a routine basis in many countries (Arrouays *et al.* 2017; Baritz *et al.* 2014; Batjes *et al.* 2017; FAO-Unesco 1981; Hannam *et al.* 2009), making strict adherence to the classification standards and correlation cumbersome.

Overall, the recommendation for a national database is that an 'as complete as possible set of attributes is provided for each soil profile, both for the site and horizon data' (FAO 2006; Soil Science Division Staff 2017). Each soil profile needs to be characterised by its site and location (x, y coordinates), year of sampling (t) as well as properties of the individual soil horizons (or layers) geo-referenced by their upper and lower boundaries (z), with supporting information on the lineage (e.g. data owner, license, data source, source laboratory, analytical methods used, vocabulary).

2.2 Possible soil properties

The range of soil properties that may occur in a T1 is presented in Appendix 1, as an example. This list is based on a worldwide data compilation of national soil profile data shared for possible standardisation in the World Soil Information Service (Batjes *et al.* 2017; Ribeiro *et al.* 2018), which unlike the GSP uses a top-down approach. The corresponding observations and measurements were collected and analysed according to a wide range of (inter)national standards, as documented in the lineage of the various soil data sets. In view of the wide range of properties and analytical procedures that may be encountered, the actual

(worldwide) standardisation thereof has generally been limited to a smaller number of so-called key properties (see Batjes 2009; Batjes *et al.* 2019; GlobalSoilMap 2015; van Engelen and Dijkshoorn 2013).

In the framework of the federated GLOSIS approach, all national soil data will have to be standardised to the GLOSIS conventions first, before they can be used to populate a T1 for a given country. Technical guidelines for this will be developed by GSP Pillar 5.

The complement of standardised T1 data (i.e., set/view as queried/accessed via the central discovery hub) may become large and its content diverse (though standardised). It is also likely that there will be numerous 'gaps' for those properties that are difficult or expensive to measure (e.g. bulk density and moisture retention). As such, the complement of national T1's will provide the basis for 'querying/extracting' the selection of soil properties required for a specific T2 (See Section 3).

3 T1 data provide the basis for thematic T2 data

3.1 General

As indicated, profiles/properties for consideration in a specific T2 set may be selected from the larger set of soil profiles queryable from the federated T1's. Examples include a T2 for assessing soil carbon stocks, soil salinity, soil fertility and and water retention (Section 3.2 to 3.5), respectively a T2 of representative soil profiles for the world (Section 3.6). Similarly, in principle, the T1's may be queried to extract time-series that focus on SOC stock changes, as required to support Land Degradation Neutrality projects (UNCCD 2015).

All T2's should reference the lineage (e.g. dataset_id, profile_id, laboratory_id, laboratory_methods), date of sampling, as well as the coordinates (x, y) and upper (Z_u) lower (Z_l) depth for each layer or horizon in a consistent format (to be defined during actual development of the overarching database model, see de Sousa (2019)).

3.2 Soil organic carbon

This data set may include:

- bulk density
- mass of organic carbon
- mass of total carbon (i.e. organic carbon + inorganic carbon)
- proportion of coarse fragments
- proportion of soil carbonates (i.e. inorganic carbon or total carbon equivalent, used for corrections)
- soil pH_{water} (used for corrections)

3.3 Soil salinity

This data set may include:

- pH measured in water
- pH measured in KCl
- pH measured in $CaCl_2$
- Electrical conductivity, measured in 1:x water solution
- Electrical conductivity, saturation extract
- Soluble Na^+ content of the saturated paste
- Soluble Ca^{2+} content of the saturated paste
- Soluble Mg^{2+} content of the saturated paste
- Soluble K^+ content of the saturated paste
- Soluble Cl^- content of the saturated paste
- Soluble SO_4^{2-} content of the saturated paste
- Soluble HCO_3^- content of the saturated paste
- Soluble CO_3^- content of the saturated paste

3.4 Soil fertility

This dataset may include:

- pH measured in water
- pH measured in KCl
- pH measured in $CaCl_2$
- Electrical conductivity of saturation extract
- Electrical conductivity of 1:X water extract
- Exchangeable Ca^{2+}
- Exchangeable Mg^{2+}
- Exchangeable K^{+}
- Exchangeable Na^{+}
- Exchangeable Al^{3+}
- Exchangeable acidity (determined in 1N KCl)
- Cation exchange capacity (CEC) of the soil at 'pH 7.0'
- Cation exchange capacity (CEC) of the soil at 'pH 8.0'
- bulk density
- mass of organic carbon
- mass of total carbon (i.e. organic carbon + inorganic carbon)
- proportion of soil carbonates (i.e. inorganic carbon or total carbon equivalent)
- Content of total N of the soil
- Gypsum content
- Available P-content of the soil (for defined operational methods, e.g. P-Olsen, B-Bray I, P-Mehlich 3)
- Total P-content of the soil
- Phosphate retention
- Proportion of coarse fragments

3.5 Water retention

This set may include:

- Coarse fragments, volume % of rock and/or coarse fragments in the soil matrix (FAO, 1990)
- Coarse fragments, mass % of rock and/or coarse fragments in the soil matrix
- Sand, very coarse fractions, mass % of particles 2.0 - 1.0 mm (very coarse sand) in fine earth fraction[1].
- Sand, coarse fraction, mass % of particles 0.5 - 0.25 mm (medium sand) in fine earth fraction
- Sand, medium fraction, mass % of particles 0.5 - 0.25 mm (medium sand) in fine earth fraction
- Sand, fine fraction, mass % of particles 0.25 - 0.1 mm (fine sand) in fine earth fraction
- Sand, very fine fraction, mass % of particles 0.1 - 0.05 mm (very fine sand) in fine earth fraction
- Silt, mass % of particles 0.002-0.05 mm (silt) in fine earth fraction
- Clay, mass % of particles < 0.002 mm (clay) in fine earth fraction
- Bulk density

[1] These are the 'default' particle size limits as given in the 2013 SOTER Procedures Manual. In practice, different limits are used for the clay, silt and sand-size fraction in various countries/laboratories. Hence, the exact size limits for the 'clay-size', 'silt-size' and 'sand-size' fraction should be explicitly mentioned in the database table that lists the laboratory methods, as a prerequisite for further harmonisation.

- Moisture content (volume %)[2] at -0.1 kPa tension
- Moisture content (v%) at -10 kPa tension
- Soil moisture (v%) at -20 kPa tension
- Soil moisture (v%) at -33 kPa tension (Field capacity according to USDA conventions)
- Soil moisture (v%) at -50 kPa tension
- Soil moisture (v%) at -100 kPa tension
- Soil moisture (v%) at -330 kPa tension
- Soil moisture (v%) at -1500 kPa tension (Permanent wilting point according to USDA conventions)

3.6 Soil reference profiles

To be considered in this T2, the underpinning T1 profiles should at least be classified according to an international system (IUSS Working Group WRB 2015; Soil Survey Staff 2014b), have general site characteristics described, as well as a 'minimum' set of measured physical and chemical soil attributes (see Appendix 1 to 3). Preferably, they should have a soil profile description status of 1 or 2 (see Table 1). This is a prerequisite if any realistic use/interpretation of the T2 data are to be made in the context of a possible update of the HWSD or world polygon map (GSP and FAO 2016). In the case of the SOTER programme (van Engelen and Dijkshoorn 2013), rather optimistically, a fairly diverse range of site and soil properties are termed 'highly desirable' or even 'mandatory'. In practice, however, all the 'desired' fields could seldom be filled in conventional SOTER databases (e.g. Dijkshoorn *et al.* 2016) as several 'desired' data may have been beyond the scope of the original field surveys.

For example, for the list for horizon data could include[3]:

- horizon or sampled layer depth (as defined by the upper and lower limit),
- horizon designation (FAO 2006) in case of a sampled horizon,
- matrix colour (moist and dry),
- structure (grade, size and type of aggregates),
- texture (clay, sand and silt-size fractions)
- coarse fragments,
- pH (H_2O)
- pH (KCl),
- electrical conductivity (EC_x and EC_e),
- cation exchange capacity (CEC, at a defined pH),
- exchangeable cation composition (exchangeable bases),
- $CaCO_3$ (total carbonate equivalent),
- organic Carbon,
- total Nitrogen,
- bulk density,
- moisture retention with special attention for the *mandatory* units of measurement (e.g. organic Carbon expressed in g kg^{-1} or pro mill), particle size fractions used (e.g. silt defined as '2-50 µm' or '2-63 µm'), size limit for the coarse fraction (e.g. >2 mm or >1 mm (Katchinsky scheme)), and

[2] The same tensions may be used when water retention is given on a gravimetric basis (mass %).
[3] The actual selection may be expanded or reduced as the number of national T1 set grows; technically, such simple 'changes' can readily be accommodated in the data model once developed.

analytical procedures used. Such information must be accommodated in specific tables in the data model (see e.g. Ribeiro *et al.* 2018).

4. Concluding remarks

This report describes soil properties that may be queried from the federated, national T1 data through the central GLOSIS user interface. It is anticipated that the actual list will evolve once countries start preparing their T1 data. Similarly, the smaller list of thematic T2 data is flexible; it may evolve subject to the recommendations of the GSP as well as the actual quantity/quality and type of data that will actually be shared for consideration in the federated T1 system.

Guidelines or *de facto* international standards as developed by FAO, USDA and ISRIC during the last decades may provide elements for the necessary initial standardisation (for soil measurements) and harmonisation (for observations) of the various national soil data to a common standard; developing consistent procedures for this resorts under the tasks of GSP Pillar 5 (GSP and FAO 2017a).

The actual compilation of T1's, in conformance with the upcoming specifications for 'CountrySIS' compilation, will be the responsibility of the data contributors themselves, with overall guidance and capacity building provided through the GSP.

References

AD-HOC Arbeitsgruppen Boden (editor) 2005. Bodenkundliche Kartieranleitung. Bundesanstalt für Geowissenschaften und Rohstoffe in Zussamenarbeit mit den Staatliche Geologischen Diensten der Bunder Republik Deutschanalnd, Hannover, 438 p.

Arrouays D, Leenaars JGB, Richer-de-Forges AC, 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 GM, Mulder VL, Minasny B, Montanarella L, Odeh I, Padarian J, Poggio L, Roudier P, Saby N, Savin I, Searle R, Solbovoy V, Thompson J, Smith S, Sulaeman Y, Vintila R, Rossel RV, Wilson P, Zhang G-L, Swerts M, Oorts K, Karklins A, Feng L, Ibelles Navarro AR, Levin A, Laktionova T, Dell'Acqua M, Suvannang N, Ruam W, Prasad J, Patil N, Husnjak S, Pásztor L, Okx J, Hallet S, Keay C, Farewell T, Lilja H, Juilleret J, Marx S, Takata Y, Kazuyuki Y, Mansuy N, Panagos P, Van Liedekerke M, Skalsky R, Sobocka J, Kobza J, Eftekhari K, Alavipanah SK, Moussadek R, Badraoui M, Da Silva M, Paterson G, Gonçalves MdC, Theocharopoulos S, Yemefack M, Tedou S, Vrscaj B, Grob U, Kozák J, Boruvka L, Dobos E, Taboada M, Moretti L and Rodriguez D 2017. Soil legacy data rescue via GlobalSoilMap and other international and national initiatives. *GeoResJ* 14, 1-19. https://doi.org/10.1016/j.grj.2017.06.001

Baize D and Girard M-C (editors) 2009. Référentiel pédologique 2008. Association française pour l'étude du sol (Afes)and Éditions Quae, Versailles (FR), 405 p. http://www.afes.fr/wp-content/uploads/2017/11/Referentiel_Pedologique_2008.pdf

Baritz R, Erdogan H, Fujii K, Takata Y, Nocita M, Bussian B, Batjes NH, Hempel J, Wilson P and Vargas R 2014. *Harmonization of methods, measurements and indicators for the sustainable management and protection of soil resources (Providing mechanisms for the collation, analysis and exchange of consistent and comparable global soil data and information)*, Global Soil Partnership, FAO, 44 p. http://www.fao.org/3/a-az922e.pdf

Batjes NH 2009. Harmonized soil profile data for applications at global and continental scales: updates to the WISE database. *Soil Use and Management* 25, 124-127 http://dx.doi.org/10.1111/j.1475-2743.2009.00202.x (supplemental information: https://www.isric.org/sites/default/files/isric_report_2008_02.pdf)

Batjes NH, Ribeiro E and van Oostrum A 2019. Standardised soil profile data to support global mapping and modelling (WoSIS snapshot 2019). *Earth Syst. Sci. Data Discuss.* https://doi.org/10.5194/essd-2019-164

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

CPCS 1967. *Classification des sols (Ëdition 1967)*, Grignon (FR), 87 p. http://horizon.documentation.ird.fr/exl-doc/pleins_textes/divers16-03/12186.pdf

de Sousa LM, Kempen B, Mendes de Jesus J, Yigini Y, Viatkin K, Medyckyj-Scott D, Richie A, Wilson P, van Egmond F and Baritz R 2019. *Conceptual desing of the Global Soil Information System infrastructure*, ISRIC, FAO, Manaaki Whenua (Landcare Research), CSIRO, Wageningen UR, European Environment Agencyr, 22 p

Dijkshoorn JA, Leenaars JGB, Huting J and Kempen B 2016. *Soil and terrain database for Malawi*, ISRIC - World Soil Information, Wageningen, 38 p. http://dx.doi.org/10.17027/isric-wdcsoils.20160001

FAO-ITPS 2014. *Report of the Second Session of the Intergovernmental Technical Panel on Soilsb (Rome, Italy, 7 - 11 April 2014)*. ITPS-II/14/Report, FAO, Rome, 15 p. http://www.fao.org/fileadmin/user_upload/GSP/docs/ITPS_Pillars/ITPS_II_14_FINAL.pdf

FAO-Unesco 1981. *Soil Map of the World, 1:5,000,000. Vol. 5 - Europe*. United Nations Educational, Scientific, and Cultural Organization, Paris

FAO 2006. *Guidelines for soil description (Fourth ed.)*, FAO, Rome, 97 p. http://www.fao.org/docrep/019/a0541e/a0541e.pdf

GlobalSoilMap 2015. *Specifications Tiered GlobalSoilMap products (Release 2.4)*, 52 p. https://www.isric.org/documents/document-type/globalsoilmap-specifications-v24-07122015

GSP and FAO 2016. *Pillar 4 Implementation Plan -- Towards a Global Soil Information System*, Rome, 36 p. http://www.fao.org/3/a-bl102e.pdf

GSP and FAO 2017a. *Implementation Plan for Pillar Five of the Global Soil Partnership - Harmonization of methods, measurements and indicators for the sustainable management and protection of soil resources (ver.1.1)*, Rome, 36 p. http://www.fao.org/3/a-bs756e.pdf

GSP and FAO 2017b. *GSP Soil Data Policy (June 2017)*, Food and Agriculture Organization of the Unted Nations (FAO) and Global Soil Parnership (GSP), 8 p. http://www.fao.org/3/a-bs975e.pdf

Hannam JA, Hollis JM, Jones RJA, Bellamy PH, Hayes SE, Holden A, Van Liedekerke MH and Montanarella L 2009 *SPADE-2: Soil Profile Analytical Database for Europe (Version 2.0)*. https://esdac.jrc.ec.europa.eu/Esdb_Archive/eusoils_docs/esb_rr/SPADE-2_Beta_Report.pdf

Hempel J 2014. Endorsement of USDA Soil Taxonomy by the International Union of Soil Sciences. *Soil Horizons* 55 http://dx.doi.org/10.2136/sh2014-55-5-gc

Isbell RF 1996. *The Australian soil classification*, Australian Soil and Land Survey Handbook. CSIRO Publishing, Collingwood, VIC, 143 p.

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

Lebedeva II, Tonkonogov VD and Gerasimova MI 2008. A New Classification System of Soils of Russia: Preliminary Results of Discussion. *Eurasian Soil Science* 41 93-99. https://link.springer.com/content/pdf/10.1134/S1064229308010110.pdf

Leenaars JGB, van Oostrum AJM and Ruiperez Gonzalez M 2014. *Africa Soil Profiles Database: A compilation of georeferenced and standardised legacy soil profile data for Sub Saharan Africa (version 1.2)*. Report 2014/01, Africa Soil Information Service (AfSIS) and ISRIC - World Soil Information, Wageningen, 160 p. http://www.isric.org/sites/default/files/isric_report_2014_01.pdf

Omuto C, Nachtergaele F and Vargas Rojas R 2012. *State of the Art Report on Global and Regional Soil Information: Where are we? Where to go?*, FAO, Italy, 69 p. http://www.fao.org/docrep/017/i3161e/i3161e.pdf

Ribeiro E, Batjes NH and Van Oostrum AJM 2018. *World Soil Information Service (WoSIS) - Towards the standardization and harmonization of world soil data. Procedures Manual 2018*. ISRIC Report 2018/01, ISRIC - World Soil Information, Wageningen, 166 p. http://dx.doi.org/10.17027/isric-wdcsoils.20180001

Soil Science Division Staff (editor) 2017. Soil survey manual. Government Printing Office, Washington, D.C., 639 p. https://www.nrcs.usda.gov/wps/PA_NRCSConsumption/download?cid=nrcseprd1333016&ext=pdf

Soil Survey Staff 2014a. *Kellogg Soil Survey Laboratory Methods Manual (Version 5.0. R. Burt and Soil Survey Staff (ed.))* Soil Survey Investigations Report No. 42, . U.S. Department of Agriculture, Natural Resources Conservation Service, Lincoln (Nebraska), 1001 p. https://www.nrcs.usda.gov/Internet/FSE_DOCUMENTS/stelprdb1253872.pdf

Soil Survey Staff 2014b. *Keys to Soil Taxonomy, 12th ed.* USDA-Natural Resources Conservation Service, Washington, DC.https://www.nrcs.usda.gov/wps/PA_NRCSConsumption/download?cid=stelprdb1252094&ext=pdf

UNCCD 2015. *Land Degradation Neutrality: Resilience at local, national and regional levels*, United Nations Convention to Combat Desertification, Bonn (DE), 24 p. http://catalogue.unccd.int/858_V2_UNCCD_BRO_.pdf

van Engelen VWP and Dijkshoorn JA 2013. *Global and National Soils and Terrain Digital Databases (SOTER) - Procedures manual (Ver. 2.0)*. ISRIC Report 2013/04, IUSS, ISRIC and FAO, Wageningen, 198 p. http://www.isric.org/sites/default/files/isric_report_2013_04.pdf

Wilson PL 2016. Global harmonisation and exchange of soil data, *SciDataCon 2016: Advancing the Frontiers of Data in Research*, Denver (11-13 September 2016), Colorado, USAhttp://www.scidatacon.org/2016/sessions/50/paper/174/

Appendices

These appendices serve to describe the type of soil properties that may be considered in a national T1; the list will likely grow as data providers start sharing/converting their data for consideration in GLOSIS. They include the properties considered in the FAO (2006) Guidelines for Soil Description, the SOTER Procedures Manual (van Engelen and Dijkshoorn 2013), and Africa Soil Profile Database (Leenaars *et al.* 2014) as well as properties derived from various 'specific purpose' surveys.

The properties shown here are presented in the format developed for the World Soil Information Service(WoSIS), a PostgreSQL server database (Ribeiro *et al.* 2018). It also provides the domains for categorical attributes (i.e., observations), and these are used as an example here. As indicated earlier, the actual data model, coding conventions/vocabulary, and interoperability standards for GLOSIS will be described in separate GSP reports.

The central domain table of WoSIS, which provides a description/list of the permissible data entries for categorical variables is far too lengthy to be included in this report. Table 2 provides an example for 'surface cover of rock outcrops'. Pragmatically, for categorical variables, Appendix 2 and 4 include a reference to the relevant the page and table in the Guidelines for Soil Description (FAO 2016). The actual domain values may be consulted on-line once the data model and associated viewers for the GLOSIS soil data infrastructure have been implemented.

For T1, possible *site* properties are listed in Appendix 1 and 2, while possible *horizon* (or layer) properties are listed in Appendix 3. Procedures for characterising soil laboratories and their analytical methods are described in Appendix 4. The critical aspect of registering data ownership, and possible access to the data, is described in Appendix 5. Appendix 6 provides an overview of the proposed set site and horizon properties that could be considered in the 'T2 with reference profiles'. Inherently, the information provided here is meant to inform the development of GLOSIS.

Table 2. Example of domain values for classification of rock outcrops

Code[a]	Class	% surface cover
N	None	0
V	Very few	0 - 2
F	Few	2 - 5
C	Common	5 - 15
M	Many	15 - 40
A	Abundant	40 - 80
D	Dominant	>80

[a] Source: FAO (2006), p. 21

Separate tables are required for the soil classification, which may be according to the (highly desired) IUSS Working Group WRB (2015), as well as the FAO Revised Legend (FAO 1988), and/or USDA Soil Taxonomy

(Soil Survey Staff 2014b). Where applicable, widely applied national systems such the French (CPCS 1967) system in Francophone Africa, the URSS classification system (Lebedeva *et al.* 2008) and the national system of classification should be documented also. In all cases, the year/version of the classification system has to be indicated to avoid confusion and permit international correlation. Ideally, a link to a scanned version of the corresponding document(s) should be provided for reference.

The list of attributes described in the Appendix is varied and long, being based on earlier collations of soil profile data worldwide. Rarely if ever, will all these attributes be available for incorporation in a national T1. As indicated, the complement of site and soil attributes collected in the field in a given area will vary with the original purpose of the survey (e.g., detail, semi-detailed, reconnaissance). Pragmatically, therefore, any input sheets/forms (for 'level C' partners, see Figure 2) should be customisable so that each user (data provider) may select those fields for which they have records (i.e., observations or measurements).

Appendix 1. List of possible site properties [4,5,6]

As indicated, for each profile site the coordinates (x, y) and upper and lower layer depths, and lineage are required. Further, the following may be characterised:

attribute / value	data type	unit	GfSD domain	GfSD page	GfSD table
Age of land surface - class	Text	unitless	Provisional coding for age of land surface	19	13
Altitude	Integer	m			
Author	Text	unitless			
Climate - Köppen class	Text	unitless	Köppen climate classification		
Coarse fragments - weathering	Text	unitless	Classification of weathering of coarse fragments	31	29
Coarse fragments - weathering class	Text	unitless	Classification of weathering of coarse fragments	31	29
Coarse surface fragments - cover class	Text	unitless	Classification of coarse surface fragments - surface cover	22	15
Coarse surface fragments - size class	Text	unitless	Classification of coarse surface fragments - size classes	22	15
Coarse surface fragments - surface cover	Text	unitless	Classification of coarse surface fragments - surface cover	22	15
Cracks - depth of surface cracks	Text	unitless	Classification of surface cracks - depth	24	21
Cracks - distance between surface cracks	Text	unitless	Classification of surface cracks - distance between cracks	24	21
Cracks - width of surface cracks	Text	unitless	Classification of surface cracks - width	24	21
Crop – class	Text	unitless	Crop codes	15	9
Crop - description	Text	unitless			
Depth of soil - observed	Integer	cm			
Depth of soil - rootable as class	Text	unitless			
Depth of soil - rootable (effective)	Integer	cm	Rootable depth		
Depth of soil - sampled	Integer	cm			
Depth of soil - to bedrock	Integer	cm			

[4] GfSD stands for FAO Guidelines for Soil Description (2006), with corresponding page number and number of Table that describes a given soil property (i.e. domain tables).

[5] List of soil properties considered in the Guidelines for Soil Description (GfSD, FAO 2006) . The classification of the soil profile should be documented in a set of separate tables, see for example Appendix 2. Classification according to IUSS Working Group WRB (2015) is mandatory for profiles to be considered in a T2 data set.

attribute / value	data type	unit	GfSD domain	GfSD page	GfSD table
Drainage - class	Text	unitless			
Erosion - area affected class	Text	unitless			
Erosion - category	Text	unitless	Classification of erosion, by category	22	16
Erosion - degree	Text	unitless	Classification of erosion, by degree	22	18
Erosion - period of activity	Text	unitless	Classification of erosion, by activity	23	19
Erosion - total area affected by erosion and deposition class	Text	unitless	Classification of total area affected by erosion and deposition	22	17
Flooding - duration	Text	unitless	Flooding duration	24	174
Flooding - duration class	Text	unitless			
Flooding - frequency class	Text	unitless	Flooding frequency	24	174
Geology	Text	unitless	Hierarchy of lithology	18	12
Groundwater - depth	Real	cm			
Human influence - class	Text	unitless	Recommended codes for human influence	15	10
Humus horizon	Text	unitless	Aeromorphic organic layers on forest floors	32	
Infiltration rate - class	Text	unitless	Infiltration rate		
Infiltration rate - numeric	Real	cm/h			
Landforms - subdivisions for complex forms	Text	unitless	Subdivisions for complex landforms	11	5
Land use - remarks	Text	unitless			
Land use - bare cover	Integer	%			
Land use - class	Text	unitless	Land-use classification	14	8
Land use - crop rotation	Text	unitless			
Land use - forest	Integer	%			
Land use - grass	Integer	%			
Land use - paved cover	Integer	%			
Land use - shrubs	Integer	%			
Lithology - class	Text	unitless	Hierarchy of lithology	18	12
Location - description	Text	unitless	Descriptive		
Major landforms - class	Text	unitless	Hierarchy of major landforms	11	4
Map sheet ID	Text	unitless			

attribute / value	data type	unit	GfSD domain	GfSD page	GfSD table
Parent material - class	Text	unitless	Hierarchy of lithology	18	12
Parent material - mode of deposition	Text	unitless	Classification of erosion, by category	22	16
Parent material - remarks	Text	unitless			
Parent material - texture class of unconsolidated material	Text	unitless	Texture of non-consolidated parent material		
Physiography - description	Text	unitless			
Physiography - physiographic unit	Text	unitless	Slope positions in undulating and mountainous terrain	11	2
Rock outcrops - cover class	Text	unitless	Recommended classification of rock outcrops - surface cover	21	14
Rock outcrops - distance between rock outcrops	Text	unitless	Recommended classification of rock outcrops - distance between rock outcrops	21	14
Salt - surface cover class	Text	unitless	Classification of salt characteristics - cover	24	22
Salt - surface presence	Boolean	unitless			
Salt - surface thickness class	Text	unitless			
Salt - thickness	Text	unitless	Classification of salt characteristics - cover	24	22
Sealing - consistence of surface sealing	Text	unitless	Classification of attributes of surface sealing - consistence	23	20
Sealing - surface thickness class	Text	unitless	Classification of attributes of surface sealing - thickness	23	20
Site - remarks	Text	unitless			
Slope - form class	Text	unitless	Classification of slope forms (categorical)	12	6
Slope - gradient class	Text	unitless	Slope gradient class	12	7
Slope - gradient numeric	Real	%			
Slope - orientation class	Text	unitless	Slope orientation	13	
Slope - orientation numeric	Integer	°			
Slope - surface pathways	Text	unitless	Slope forms and surface pathways	12	Fig. 3
Soil classification - FAO Legend	Text	unitless			
Soil classification - USDA Soil taxonomy	Text	unitless			
Soil classification - WRB	Text	unitless			
Soil profile description status	Text	unitless	Soil profile description status	6	1
Soil temperature and moisture regime - USDA moisture code	Text	unitless	Soil temperature and moisture regime codes - moisture	10	3

attribute / value	data type	unit	GfSD domain	GfSD page	GfSD table
Soil temperature and moisture regime - USDA temperature code	Text	unitless	Soil temperature and moisture regime codes - temperature	10	3
Surface bleached sand - cover class	Text	unitless			
Surface organic matter - class	Text	unitless	Surface organic matter		
Surface texture - class	Text	unitless			
Vegetation - class	Text	unitless	Vegetation classification	16	11
Vegetation - description	Text	unitless			
Vegetation - tree density	Integer	n/ha			
Weather - conditions at time of sampling	Text	unitless	Codes for weather conditions - present weather conditions	9	2
Weather – past weather conditions	Text	unitless	Codes for weather conditions - former weather conditions	9	2
Weathering - weathering status of solid rock	Text	unitless	Classification of weathering of coarse fragments	31	29

Appendix 2. Examples of soil classification systems that may represented in GloSIS

Codings below are according to WoSIS conventions, and only provided here as an example to support GLOSIS database model development.

Table name	Description
class_wrb	Soil name according to WRB (for given version)
class_wrb_horizon	Diagnostic horizons according to given WRB version
class_wrb_material	Diagnostic materials according to given WRB version
class_wrb_property	Diagnostic properties according to given WRB version
class_wrb_qualifier[7]	Qualifiers according to given WRB version
class_fao	Soil name according to given version of the FAO Legend
class_fao_horizon	Diagnostic horizons according to given FAO Legend version
class_fao_property	Diagnostic properties according to given FAO Legend version
class_soil_taxonomy	Soil name according to USDA Soil Taxonomy, for defined version (e.g. 1975 or 2014)
class_cpcs	Soil name according to French CPSS soil classification system (Commission Pédologique de Soils, 1967)
class_russia	Soil name according to the Russian, respectively former URSS, soil classification scheme, for defined version (with year of publication)
class_local	Soil name according to the national soil classification system, for example (AD-HOC Arbeitsgruppen Boden 2005) for Germany or Référentiel Pédologique for France (Baize and Girard 2009) .

[7] Tables with possible subdivisions for WRB and USDA Soil Taxonomy units may be found in the WoSIS Procedures Manual (Ribeiro *et al.* 2018).

Appendix 3. Example of horizon properties encountered in a global data compilation programme [8,9]

This list serves to provide an example of soil properties that are commonly described for horizons/layers; seldom will the full complement of attributes shown be available for each profile.

attribute	data type	unit	GfSD domain	GfSD page	GfSD table
Acidity - exchangeable	Real	cmol(c)/kg			
Acidity - extractable	Real	cmol(c)/kg			
Aluminium (Al) + 0.5 Fe oxalate	Real	%			
Aluminium (Al) - dithionite extractable	Real	g/100g			
Aluminium (Al+++) - exchangeable	Real	cmol(c)/kg			
Aluminium (Al) - oxalate extractable	Real	g/100g			
Aluminium (Al) - pyrophosphate extractable	Real	g/100g			
Aluminium (Al+++) - saturation (ESP)	Real	%			
Available water capacity - volumetric (FC to WP)	Real	$cm^3/100cm^3$			
Base saturation - calculated	Real	%			
Base saturation - sum of cations	Real	%			
Bases - exchangeable	Real	cmol(c)/kg			
Biological activity - abundance (class)	Text	unitless	Classification of the abundance of biological activity	60	81
Biological features - examples	Text	unitless	Examples of biological features	60	82
Boron (B) - extractable	Real	mg/kg			
Boron (B) - total	Real	mg/kg			

[8] This list serves to give an indication of the breadth of properties that may be accommodated in the Tier 1 database, based on soil profile data collated from over 100 countries worldwide.

attribute	data type	unit	GfSD domain	GfSD page	GfSD table
Boundary between mottle and matrix - class	Text	unitless	Classification of boundary between mottle and matrix	36	35
Bromite (Br-) - extractable	Real	mg/kg			
Bulk density - field estimation for mineral soils	Text	unitless	Field estimation of bulk density for mineral soils	51	58
Bulk density - field estimation of volume of solids and bulk density of peat soils	Text	unitless	Field estimation of volume of solids and bulk density of peat soils - bulk density	52	59
Bulk density - fine earth	Real	kg/dm^3			
Bulk density - whole soil	Real	kg/dm^3			
Calcium (Ca++) - exchangeable	Real	cmol(c)/kg			
Calcium (Ca++) - extractable	Real	mg/kg			
Calcium carbonate equivalent - fraction	Real	g/kg			
Calcium carbonate equivalent - total	Real	g/kg			
Calcium (Ca++) - soluble	Real	cmol(c)/l			
Calcium (Ca++) - total	Real	mg/kg			
Carbonate (CO3--) - soluble	Real	cmol(c)/l			
Carbonate reaction in the soil matrix - class	Text	unitless	Classification of carbonate reaction in the soil matrix	38	38
Carbon (C) - organic	Real	g/kg			
Carbon (C) - pyrophosphate extractable	Real	% w/w			
Carbon (C) - total	Real	g/kg			
Carbon - fulvic acid	Real	g/kg			
Carbon - humic acid	Real	g/kg			
Carbon/Nitrogen (C/N) ratio	Real	unitless			
Carbon - total humic	Real	g/kg			
Cation exchange capacity (CEC)	Real	cmol(c)/kg			
Cation exchange capacity effective (ECEC)	Real	cmol(c)/kg			
Cementation/compaction - continuity	Text	unitless	Classification of the continuity of cementation/compaction	56	69

attribute	data type	unit	GfSD domain	GfSD page	GfSD table
Cementation/compaction - degree of	Text	unitless	Classification of the degree of cementation/compaction	57	72
Cementation/compaction - nature	Text	unitless	Classification of the nature of cementation/compaction	57	71
Cemented/compacted layer - fabric	Text	unitless	Classification of the fabric of the cemented/compacted layer	56	70
Chloride (Cl-) - soluble	Real	cmol(c)/l			
Clay - carbonate	Real	%			
Clay - mineralogy	Text	unitless			
Clay - non-carbonate	Real	%			
Clay - size fraction 01[10]	Real	g/100g			
Clay - size fraction 02	Real	g/100g			
Clay - total	Real	g/100g			
Clay-water dispersible - total	Real	%			
Coarse fragments - field class	Text	unitless			
Coarse fragments - gravimetric fraction 01	Real	g/100g			
Coarse fragments - gravimetric fraction 02	Real	g/100g			
Coarse fragments - gravimetric fraction 03	Real	g/100g			
Coarse fragments - gravimetric fraction 04	Real	g/100g			
Coarse fragments - gravimetric total	Real	g/100g			
Coarse fragments - volumetric fraction 01	Real	$cm^3/100cm^3$			
Coarse fragments - volumetric fraction 02	Real	$cm^3/100cm^3$			
Coarse fragments - volumetric fraction 03	Real	$cm^3/100cm^3$			
Coarse fragments - volumetric total	Real	$cm^3/100cm^3$			
Coarse fragments - volumetric total, field estimated	Real	$cm^3/100cm^3$			
Coatings - abundance	Text	unitless	Classification of abundance of coatings	55	66

[10] The physical limits for each particle size fraction should be listed in a separate 'look up' or option table.

attribute	data type	unit	GfSD domain	GfSD page	GfSD table
Coatings - classification of their nature	Text	unitless	Classification of the nature of coatings	55	66
Coatings - contrast of coatings	Text	unitless	Classification of the contrast of coatings	55	65
Coatings - form	Text	unitless	Classification of the form of coatings	56	67
Coatings - location of coatings and clay accumulation	Text	unitless	Classification of the location of coatings and clay accumulation	56	68
Consistence - soil mass when dry	Text	unitless	Consistence of soil mass when dry	48	53
Consistence - soil mass when moist	Text	unitless	Consistence of soil mass when moist	49	54
Consistency - dry class	Text	unitless			
Consistency - moist class	Text	unitless			
Consistency - wet class, plasticity	Text	unitless			
Copper (Cu) - extractable	Real	mg/kg			
Copper (Cu) - total	Real	mg/kg			
Electrical conductivity	Real	dS/m			
Exchangeable cations	Real	cmol(c)/kg			
Field estimation and coding of the degree of decomposition and humification of peat	Text	unitless	Field estimation and coding of the degree of decomposition and humification of peat	32	31
Field estimation of volume of solids and bulk density of peat soils - drainage conditions	Text	unitless	Field estimation of volume of solids and bulk density of peat soils - drainage conditions	52	59
Field estimation of volume of solids and bulk density of peat soils - solid volume	Text	unitless	Field estimation of volume of solids and bulk density of peat soils - solid volume	52	59
Gypsum content - class	Text	unitless	Classification of gypsum content	39	40
Gypsum content - weight	Real	g/kg			
Horizon boundary - distinctness class	Text	unitless	Classification of horizon boundaries - distinctness	25	24
Horizon boundary - topography class	Text	unitless	Classification of horizon boundaries - topography	25	24
Hydraulic conductivity	Real	cm/h			

attribute	data type	unit	GfSD domain	GfSD page	GfSD table
Hydrocarbonate (HCO3-) - soluble	Real	cmol(c)/l			
Hydrogen (H+) - exchangeable	Real	cmol(c)/kg			
Iron (Fe) - dithionite extractable	Real	g/100g			
Iron (Fe) - extractable	Real	mg/kg			
Iron (Fe) - oxalate extractable	Real	g/100g			
Iron (Fe) - pyrophosphate extractable	Real	g/100g			
Iron (Fe) - total	Real	mg/kg			
Magnesium (Mg++) - exchangeable	Real	cmol(c)/kg			
Magnesium (Mg) - extractable	Real	mg/kg			
Magnesium (Mg++) - soluble	Real	cmol(c)/l			
Magnesium (Mg) - total	Real	mg/kg			
Manganese (Mn) - dithionite extractable	Real	% w/w			
Manganese (Mn) - extractable	Real	mg/kg			
Manganese (Mn) - KCl extractable	Real	mg/kg			
Manganese (Mn) - oxalate extractable	Real	mg/kg			
Manganese (Mn) - pyrophosphate extractable	Real	% w/w			
Manganese (Mn) - total	Real	mg/kg			
Mineral concentrations - abundance by volume	Text	unitless	Classification of the abundance of mineral concentrations, by volume	58	73
Mineral concentrations - colour	Text	unitless	Colour names of mineral concentrations	59	78
Mineral concentrations - hardness class	Text	unitless	Classification of the hardness of mineral concentrations	58	76
Mineral concentrations - kinds	Text	unitless	Classification of the kinds of mineral concentrations	58	74
Mineral concentrations - size and shape class	Text	unitless	Classification of the size and shape of mineral concentrations - size	58	75
Mineral concretions - abundance class	Text	unitless			
Mineral concretions - nature class	Text	unitless	Examples of the nature of mineral concentrations	59	77

attribute	data type	unit	GfSD domain	GfSD page	GfSD table
Mineral content	Real	%			
Mottles - abundance class	Text	unitless	Classification of the abundance of mottles	35	32
Mottles - colour	Text	unitless	Munsell colour codes	33	
Mottles - contrast of mottles	Text	unitless	Classification of the contrast of mottles	36	34
Mottles - presence boolean	Boolean	unitless			
Mottles - size class	Text	unitless	Classification of the size of mottles	35	33
Nitrate (NO3-) - soluble	Real	mmol(c)/l			
Nitrite (NO2-) - soluble	Real	mmol(c)/l			
Nitrogen (N) - total	Real	g/kg			
Optical density - oxalate extractable	Real	unitless			
Organic matter	Real	g/kg			
Particle size fractions - sum	Real	g/100g			
pH - CaCl2	Real	unitless			
pH - Field	Real	unitless			
pH - H2O	Real	unitless			
pH - KCl	Real	unitless			
pH - NaF	Real	unitless			
pH - NH4CL	Real	unitless			
Phosphate (PO4--) - soluble	Real	mmol(c)/l			
Phosphorus (P) - extractable	Real	mg/kg			
Phosphorus (P) - oxalate extractable	Real	mg/kg			
Phosphorus (P) - retention	Real	g/100g			
Phosphorus (P) - total	Real	mg/kg			
pH - unknown	Real	unitless			
Pores - abundance	Text	unitless	Classification of abundance of pores	53	63
Pores - size	Text	unitless	Classification of abundance of pores	53	63

attribute	data type	unit	GfSD domain	GfSD page	GfSD table
Porosity	Real	$cm^3/100cm^3$			
Porosity - class	Text	unitless	Classification of porosity	52	60
Potassium (K+) - exchangeable	Real	cmol(c)/kg			
Potassium (K) - extractable	Real	mg/kg			
Potassium (K+) - soluble	Real	cmol(c)/l			
Potassium (K) - total	Real	mg/kg			
Primary mineral fragments - class	Text	unitless	Codes for primary mineral fragments	31	30
Rock and artefacts - abundance	Text	unitless	Abundance of rock fragments and artefacts, by volume	29	26
Rock and artefacts - size	Text	unitless	Classification of rock fragments and artefacts - artefacts	30	27
Rock fragments - shape	Text	unitless	Classification of shape of rock fragments	31	28
Roots - abundance	Text	unitless	Classification of the abundance of roots	60	80
Roots - presence	Boolean	unitless			
Salt - content class	Text	unitless	Classification of salt content of soil	40	42
Sand - size fraction 01	Real	g/100g			
Sand - size fraction 02	Real	g/100g			
Sand - size fraction 03	Real	g/100g			
Sand - size fraction 04	Real	g/100g			
Sand - size fraction 05	Real	g/100g			
Sand - size fraction 06	Real	g/100g			
Sand - size fraction 07	Real	g/100g			
Sand - size fraction 08	Real	g/100g			
Sand - size fraction 09	Real	g/100g			
Sand - size fraction 10	Real	g/100g			
Sand - total	Real	g/100g			
Sand-water dispersible - fraction 01	Real	%			

attribute	data type	unit	GfSD domain	GfSD page	GfSD table
Sand-water dispersible - fraction 02	Real	%			
Sand-water dispersible - fraction 03	Real	%			
Sand-water dispersible - fraction 04	Real	%			
Sand-water dispersible - fraction 05	Real	%			
Sand-water dispersible - total	Real	%			
Secondary carbonates - forms class	Text	unitless	Classification of forms of secondary carbonates	38	39
Secondary gypsum - forms class	Text	unitless	Classification of forms of secondary gypsum	39	41
Selenium (Se) - extractable	Real	mg/kg			
Selenium (Se) - total	Real	mg/kg			
Silicon (Si) - oxalate extractable	Real	% w/w			
Silt - size fraction 01	Real	g/100g			
Silt - size fraction 02	Real	g/100g			
Silt - size fraction 03	Real	g/100g			
Silt - total	Real	g/100g			
Silt-water dispersible - fraction 01	Real	%			
Silt-water dispersible - fraction 02	Real	%			
Silt-water dispersible - total	Real	%			
Sodium (Na+) - exchangeable	Real	cmol(c)/kg			
Sodium (Na+) - exchangeable %	Real	%			
Sodium (Na) - extractable	Real	mg/kg			
Sodium (Na+) - soluble	Real	cmol(c)/l			
Sodium (Na) - total	Real	mg/kg			
Soil colour - dry (compounded)	Text	unitless	Munsell colour codes	33	
Soil colour (matrix) - dry, chroma	Text	unitless	Munsell colour codes	33	
Soil colour (matrix) - dry, hue	Text	unitless	Munsell colour codes	33	
Soil colour (matrix) - dry, value	Text	unitless	Munsell colour codes	33	

attribute	data type	unit	GfSD domain	GfSD page	GfSD table
Soil colour (matrix) - moist, chroma	Text	unitless	Munsell colour codes	33	
Soil colour (matrix) - moist, hue	Text	unitless	Munsell colour codes	33	
Soil colour (matrix) - moist, value	Text	unitless	Munsell colour codes	33	
Soil colour - moist (compounded)	Text	unitless	Munsell colour codes	33	
Soil moisture - status class	Text	unitless	Classification of moisture status of soil	50	57
Soil plasticity - classification	Text	unitless	Classification of soil plasticity	49	56
Soluble anions - total	Real	cmol(c)/l			
Soluble cations - total	Real	cmol(c)/l			
Stickiness - class	Text	unitless	Classification of soil stickiness	49	55
Structure - grade class	Text	unitless	Classification of structure of pedal soil materials	45	47
Structure - size class	Text	unitless	Size classes for soil structure types	47	50
Subdivisions of sandy textural classes	Text	unitless	Subdivisions of sandy textural classes	27	4
Sulfate (SO4--) - soluble	Real	cmol(c)/l			
Sulfur (S) - extractable	Real	mg/kg			
Sulfur (S) - total	Real	mg/kg			
Sum of cations	Real	cmol(c)/kg			
Texture field - class	Text	unitless			
Texture lab - class[d]	Text	unitless			
Total Salt - estimated	Real	% w/w			
Voids - classification	Text	unitless	Classification of voids	53	61
Voids - diameter as class	Text	unitless	Classification of diameter of voids	53	62
Water retention - gravimetric	Real	g/100g			
Water retention - gravimetric 100 kPa	Real	g/100g			
Water retention - gravimetric 10 kPa	Real	g/100g			
Water retention - gravimetric 1500 kPa	Real	g/100g			
Water retention - gravimetric 200 kPa	Real	g/100g			

attribute	data type	unit	GfSD domain	GfSD page	GfSD table
Water retention - gravimetric 33 kPa	Real	g/100g			
Water retention - gravimetric 500 kPa	Real	g/100g			
Water retention - gravimetric 6 kPa	Real	g/100g			
Water retention - volumetric	Real	$cm^3/100cm^3$			
Water retention - volumetric 100 kPa	Real	$cm^3/100cm^3$			
Water retention - volumetric 10 kPa	Real	$cm^3/100cm^3$			
Water retention - volumetric 1500 kPa	Real	$cm^3/100cm^3$			
Water retention - volumetric 200 kPa	Real	$cm^3/100cm^3$			
Water retention - volumetric 33 kPa	Real	$cm^3/100cm^3$			
Water retention - volumetric 500 kPa	Real	$cm^3/100cm^3$			
Water retention - volumetric 6 kPa	Real	$cm^3/100cm^3$			
Zinc (Zn) - extractable	Real	mg/kg			
Zinc (Zn) - total	Real	mg/kg			

a) List of categorical soil properties (GfSD domain) according to the Guidelines for Soil Description (FAO 2006).
b) Many of the horizon attributes listed above have to be considered in relation to the analytical method according to which they were analysed (see App. 4).
c) The above list of attributes will seldom be analysed/available for most soil surveys.
d) May require an attribute that stores the texture triangle used to classify the texture classes.

Appendix 4. Possible characterisation of laboratories and analytical method descriptions[11]

The following may serve as an example. It is based on procedures developed for WoSIS (Ribeiro *et al.* 2018). In this example, four tables are used for the characterisation.

Table name	Description
desc_laboratory	Listing of (unique) laboratories where the soil sample were analysed
desc_method_option	Criteria used to standardise disparate soil analytical method descriptions, for a given analytical method, according to a defined set of options
desc_method_source	Analytical method descriptions as defined in a given 'source' database. Should be comprehensive enough for Regional Soil Partnerships to standardize the analytical method descriptions.
desc_methods_standard	Result of the standardization of the soil analytical methods according to the evolving Pillar 5 standard for this.

In short, the following information will need to be recorded concerning the soil analytical methods. A detailed description is critical for any possible further standardisation and harmonisation, as this provides the basis for worldwide exchange of soil data using the federated GLOSIS approach.

Laboratory Information:

lab_id:	Unique code for the laboratory where the profile(s) were analysed. The code should be comprised of the country's ISO code followed by a number (e.g. BR01).
laboratory_name:	Name of the laboratory in full (with English translation when necessary).

Laboratory method:

lab_id:	Unique code for the laboratory the given analysis was performed (e.g. BR01).
lab_year:	Year during which the given analytical method was introduced in the given laboratory.
lab_attribute:	The analytical method under consideration (e.g. pHKCl).
lab_method_id:	Unique ID for the analytical method applied. For example, pHKCl006, where the first letters refer to the broad analytical method (pH measured in a KCl solution) and the number '006' to specific details of the analytical method as described below.

Analytical method:

[11] According to WoSIS procedures (Ribeiro et al. 2018).

lab_method_id:	Unique code for the analytical method applies to the given sample.
lab_method_description:	A concise, yet complete, description of main aspects of the analytical method under consideration. For example, for soil pH-KCl the following need to be described: the sample pre-treatment (e.g. for' < 2 mm' fraction), the solution (here KCl), the 'soil/solution' ratio, the ratio base (i.e. weight/volume or volume/volume), the concentration (e.g. 1M KCL), and the instrument that was used for the measurements. A worked out example is provided in the WoSIS Procedure Manual (Ribeiro *et al.* 2018, p. 80). For a bottom-up approach for GLOSIS, it is crucial that all data providers provide a concise description for their analytical methods; these will provide the basis for any subsequent standardisation/harmonisation to the T1 conventions. Further, globally unique identifiers should be provided for each soil laboratory.

Appendix 5. Registration of data ownership

The lineage of the source data and the associated licence (e.g. Creative Commons[12]) should be documented so that each data provider can be acknowledged/cited. Further, the licence will determine how a specific national dataset, once converted to T1 conventions, may be shared with the international community through GLOSIS. Options for this are documented in the GSP Data Policy (GSP and FAO 2017b).

Table name	Description
contact	Names of people(contact person(s)) that have contributed data for the T1 and T2 set.
contact_organization	As above, but for data-contributing organisations

Personal data should be managed in conformance with EU 'Data protection and privacy ethical guidelines'[13] or similar national legislation.

Appendix 6. Possible properties for T2 of reference profiles

As indicated in the report (Chapter 3), several types of T2 data are foreseen. Some are aimed at specific assessments of soil functioning, for example assessing changes in soil carbon stocks. Alternatively, this Appendix presents a list of 'desirable properties' for the 'T2 soil profile reference set' (see Section 3.6). The example below is based on the SOTER specifications (van Engelen and Dijkshoorn 2013) and should be fine-tuned to GSP's needs and overall data availability.

[12] https://creativecommons.org/licenses/
[13] http://ec.europa.eu/research/participants/data/ref/fp7/89827/privacy_en.pdf

Site data	Horizon data
Profile_id	*Profile_id + Horizon_id*
General:	*General:*
Soil profile description status	horizon / layer number
Description, year of	depth, top
Description, month of	depth, bottom
Depth of soil (described/sampled)	horizon designation
Number of horizons	matrix Munsell colour, moist
Lineage	matrix Munsell colour, dry
Soil classification:	*Chemical attributes*[a]*:*
Reference Soil Group WRB (mandatory)	organic carbon
Edition of WRB	total N
	extractable Phosphorus
USDA Soil Taxonomy (optional)	electrical conductivity
Edition (year) of USDA Soil Taxonomy	Free $CaCO_3$
Former USSR resp. Russian soil classification	$CaSO_4$
Edition (year) of above system	pH-H_2O
National soil classification (optional)	pH-KCl
Edition (year) of national system	pH-$CaCl_2$
	exchangeable Ca^{2+}
Location:	exchangeable Mg^{2+}
Country (ISO code)	exchangeable Na^{+}
Latitude (WGS 1984)	exchangeable K^{+}
Longitude (WGS 1984)	exch. Al^{3+} + H^{+} (exch. acidity)
Location (descriptive)	exch. Al^{3+} (exch. Aluminium)
Geometric accuracy	cation exchange capacity (CEC)
	base saturation (as percent of CEC)
General site data:	
Climate	*Physical attributes:*
Altitude	weight % sand
Major landform	weight % silt
Landscape position	weight % clay
Slope	volume % > 2 mm
Parent material	bulk density
Drainage class	Vol. per cent water held at -10 kPa (optional)
Land use	Vol. per cent water held at -33 kPa (opt.)
	Vol. per cent water held at -1500 kPa (opt.)

[a] Should always be given with details about the analytical methods used (i.e. as operational-definitions, see Soil Survey Staff 2014a). May also need to consider soluble salts.

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

www.isric.org