HoliSoils
Working together for forest soils

D3.4 Open access harmonised soil database

Holistic management practices, modelling and monitoring for European forest soils, HoliSoils

Project duration: 1.5.2021–31.10.2025

H2020 Grant Agreement No 101000289

Coordinator: Natural Resources Institute Finland (Luke)

Deliverable D3.4: Open access harmonised soil database		
This report describes how forest soil datasets for Europe, as collated and shared in the framework of the EU HoliSoils project, were processed and standardised/harmonised with a view to making them freely available to the international community (where possible). Of the 7.2k point observations shared, only 2.8k are freely available. The remaining point data have a restrictive license, i.e. locations may not be shown. The latter can only be used for the soil property mapping component of the HoliSoils project. The shared point data themselves are queryable and accessible via various platforms, including the resources page of the HoliSoils Project. All data were processed and standardised using newly developed ETL (Extract, Transform, Load) procedures implemented within the framework of ISRIC's World Soil Information Service (WoSIS). The underpinning WoSIS database will be maintained, and augmented, by ISRIC WDC-Soils after finalisation of the EU project in 2025.		
Due date	10/2023	
Authors	N.H. Batjes; L. Calisto	
Date of publication	15.06.2023	
Dissemination level		
PU	Public, fully open e.g., web	x
CO	Confidential, only for members of the consortium (including the Commission Services)	
CI	Classified, information as referred to in Commission Decision 2001/844/EC	
Nature of the Deliverable		
R	Document, report	x
DEM	Demonstration, pilot, prototype, plan design	
DEC	Websites, patents filing, market studies, press & media actions, videos etc.	
OTHER	Software, technical diagram etc.	
Ethics	Ethics deliverables	

This project has received funding from the European Union's H2020 research and innovation programme under grant agreement No 101000289.

Contents

1. Introduction 3
2. Source of data and data processing 3
   2.1 Data sources 3
   2.2 Key soil properties 5
   2.3 Data processing 6
3. Results 7
   3.1 Characterisation of HoliSoils-provided data 7
   3.2 Augmenting the dataset for mapping purposes 9
4. Next steps 9
5. Concluding remarks 10
6. Acknowledgements 10
Appendix - Reference methods for chemical properties 11
References 12

Tables

Table 1. Main characteristics of datasets shared with HoliSoils 4

Figures

Figure 1. Schematic representation of the Extract, Transform and Load (ETL) workflow 5
Figure 2. Overview of HoliSoils-shared profile data under Corine 2018 forest cover 7
Figure 3. Dashboard for the HoliSoils-provided datasets 8
Figure 4. Descriptive statistics for key soil properties 8
Figure 5. HoliSoils forest soil dataset augmented with WoSIS forest soil holdings 9

This project has received funding from the European Union's H2020 research and innovation programme under grant agreement No 101000289.

1. Introduction

Serious problems can arise with the comparison of soil analytical data derived from different sources. Many countries use specific analytical methods for their soil analyses, be it for agricultural soils or forest soils, and this generally according to the prevailing national standards (Higgins *et al.* 2021). In some instances different standards are used within countries, for example for 'states' (Höhle *et al.* 2016; Hoffmann *et al.* 2019). These 'original' data often meet specific goals and are not necessarily intended to contribute to international transboundary studies. Standardisation and harmonisation of such data for wider use may imply a loss of appropriateness for originally intended purposes. However, once they are compiled under a common standard they importantly gain in appropriateness for use for cross border or international applications (Batjes *et al.* 2020).

The quality and possible extent of the research and policy advice that can be based on such standardised/harmonised data will strongly depend on the quality and comparability of the available (source) data (Van Egmond and Fantappiè 2021). Many projects and initiatives for Europe have underlined the difficulties that may arise when comparing and sharing data from national soil monitoring systems (Bispo et al. 2021; Batjes 2023).

This report discusses how datasets (T(ask) 3.3) collated for consideration in the HoliSoils mapping work (T3.5) have been processed and harmonised using an updated ETL (Extract, Transform, Load) procedure developed for WoSIS[1] (World Soil Information Service) (T3.4). The source of data and data processing procedure are outlined in Section 2. Subsequently, Section 3 provides a summary of the shared data and indicates how these data have been expanded using holdings held in ISRIC's WoSIS database. The expanded-dataset will be used to map key soil properties of forest soils in Europe at 100m resolutions, with measures for uncertainty (Section 4). Concluding remarks are made in Section 5.

2. Source of data and data processing

2.1 Data sources

Ultimately, the project provided four databases (Wellbrock et al. 2021). These include ICP Forest Level I and Level II (ICP Forests 2021), a compilation of forest soil data for Poland (Brozek *et al.* 2019)[2], and French soil monitoring system (RMQS 2021). Main characteristics of these datasets are provided in Table 1.

1 https://www.isric.org/explore/wosis/faq-wosis

[2] *Dataset compiled from 'Atlas siedlisk leśnych' by team of Prof. Jarosław Lasoto (Univ. of Agriculture, Krakow) in support of the HoliSoils project.*

This project has received funding from the European Union's H2020 research and innovation programme under grant agreement No 101000289.

Table 1. Main characteristics of datasets shared with HoliSoils

Abbreviation	Land cover	Main characteristics[a]	License	Source
ICPF L1-S	Forest	Transnational 16 x 16 km grid, Europe; ~5600 plots; monitored 2004-2014	restricted	http://icp-forests.net/
ICPF L2	Forest	Purposively selected plots across Europe (~730); monitored every 10 years	restricted	http://icp-forests.net/
RMQS	Forest and other lands	National 16 x 16 km square grid, ~2100 plots (of which ~550 plots next to ICPF Level 1 plots); first campaign 2000-2010, second campaign started in 2016.	open	www.gissol.fr
PO-Forests	Forest	Selected forest soil plots (177). Soil data include C, N, pH, bulk density and texture	open	Brozek et al. (2019)

[a] Approximate number of plots pre-ingestion and cleaning. Not all the plots (point data) have the desired soil data (see Section 2.2) and many data were shared with 'theoretical' coordinates.

The datasets, pre-screened for completeness (e.g., lineage, license), were imported 'as is' in the ISRIC Data Repository keeping the original data model, naming and coding conventions, abbreviations, domains respectively units of measurement, and so on. Subsequently, these diverse 'source' datasets were converted into PostgreSQL format using a newly developed, semi-automated ETL (extraction, transformation, load) procedure (Figure 1). During this process, the source data were 'mapped' to the WoSIS naming conventions, standard values and/or units of measurement. This corresponds with the first major step of standardisation: make the originally incongruent data queryable and usable. The second step of standardisation involves importing the standardised data into the WoSIS data model itself. This process includes automated checks on the plausibility of the reported values (e.g., pH H_2O should between 1 and 13).

This project has received funding from the European Union's H2020 research and innovation programme under grant agreement No 101000289.

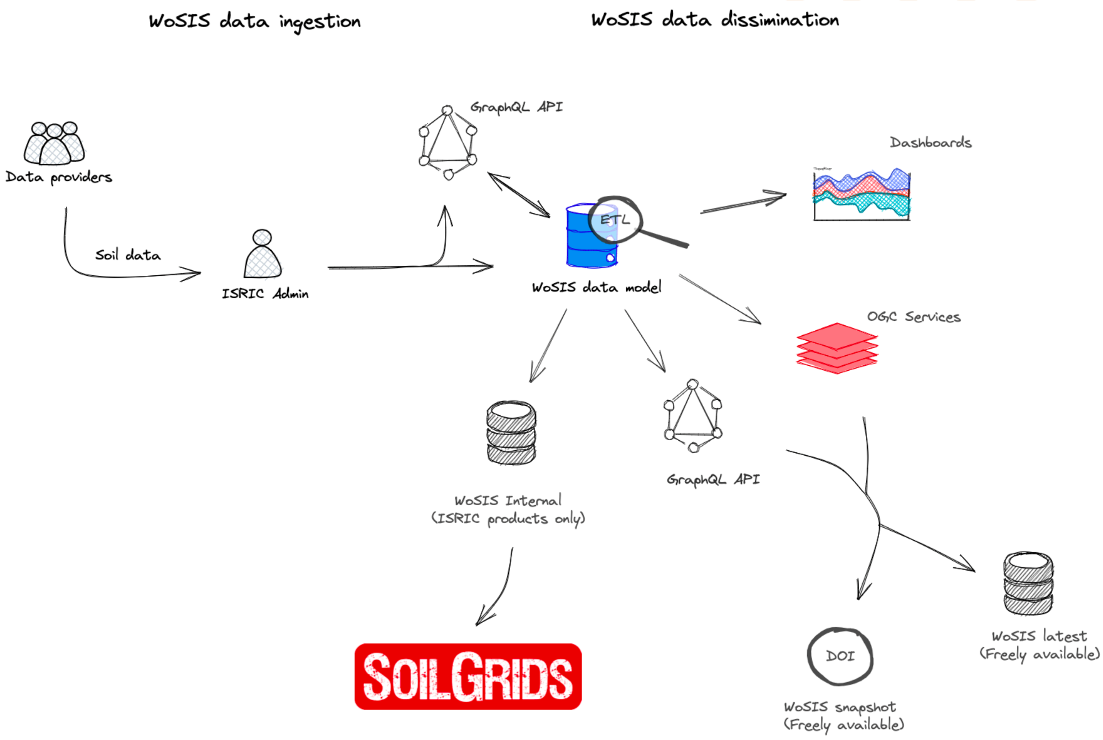

Figure 1. Schematic representation of the Extract, Transform and Load (ETL) workflow

A desired third step, as a follow up to the standardisation, full data harmonisation would involve making nominally similar soil chemical and physical data comparable, that is as 'if assessed by a commonly endorsed single reference method' (for pH, CEC, organic carbon, etc.). For Europe-wide soil monitoring the list of analytical methods (ISO) as implemented for the LUCAS topsoil monitoring programme (Orgiazzi et al. 2018) is recommended as reference (Bispo et al. 2021). Similarly, ISO standard operating procedures (SOP) are used by ICP Forest (2021) and RMQS (Jolivet *et al.* 2006). Nonetheless, many organisations in Europe use their own SOP's. The set of ISO SOP's adopted for HoliSoils is listed in the Appendix.

2.2 Key soil properties

During the initial phase of the HoliSoils project and 2nd General Assembly (Helsinki, June 2022), the list of soil properties required for the subsequent mapping/modelling work was defined as:

- bulk density
- Carbon (Total and Organic)
- Total Nitrogen
- Calcium carbonate
- Cation exchange capacity

This project has received funding from the European Union's H2020 research and innovation programme under grant agreement No 101000289.

- pH (H_2O)
- Texture (sand, silt, clay)
- water retention (at specified tensions).

2.3 Data processing

When confronted with soil analytical data derived from a wide range of laboratories that use different national standards, and considering the observations made in Batjes (2023, Section 3.3.1), a practical solution towards harmonisation is to cluster the shared soil data according to key elements of the chemical procedures under consideration according to 'operational definitions' (Soil Survey Staff 2022). In other words, analytical data obtained from different laboratories are grouped according to key criteria of the procedures (SOP's) themselves. This approach is illustrated in Table 2 using pH-KCl as an example. This information can then be used to group similarly analysed soil data (e.g., as like 'pH-KCl ISO SOP'). Of necessity, such an approach assumes interlaboratory differences for a given operationally-defined method (e.g., pH H20) are neglible, which is a simplification (van Leeuwen *et al.* 2022).

Table 2. Characterising soil analytical methods for pH-KCl according to key criteria

Key	ISO[a]	ISRIC[b]	USDA[c]	WEPAL[d]	GLOSOLAN[e]
Pretreatment	< 2 mm	< 2 mm	< 2 mm	< 2 mm	< 2 mm
Solution	KCl	KCl	KCl	KCl	KCl
Concentration	1 M	1 M	1 M	1 M	1 M
Ratio	1:5	1:2.5	1:1	1:5	1:5
Ratio base	v/v	w/v	w/v	v/v	w/v
Instrument	Electrode	Electrode	Electrode	Electrode	Electrode

Source: Ribeiro (2020). Footnotes:

a) ISO 10390:2021 specifies an instrumental method for the routine determination of pH using a glass electrode in a 1:5 (volume fraction) suspension of soil in water (pH in H_2O), in 1 mol/l potassium chloride solution (pH in KCl) or in 0.01 mol/l calcium chloride solution (pH in $CaCl_2$) (ISO-10390 2021); this coding example is for pH KCl.
b) ISRIC: Method 4-1 for pH-KCl (van Reeuwijk 2002).
c) USDA: Method 4C1a2a3 (Soil Survey Staff 2022).
d) WEPAL Wageningen Evaluating Programs for Analytical Laboratories – www.wepal.nl – WEPAL is an accredited world-leading organiser of proficiency testing programmes in the fields of plants, soil, sediments and organic waste. Participants in the International Soil-Analytical Exchange programme receive four times a year, four samples to be analysed for comparison of results. Participants describe the applied extraction/ digestion, and the method of detection of the particular element of their method applied.
e) GLOSOLAN/FAO (2021): The pH of the soil measured in either water, $CaCl_2$, or KCl systems.

This project has received funding from the European Union's H2020 research and innovation programme under grant agreement No 101000289.

3. Results

3.1 Characterisation of HoliSoils-provided data

Around 3.2k out of the shared 7.2k profiles occur under EU CORINE 2018[3] forest classes (Figure 1). This is a reflection of the fact that RMQS, the French national soil monitoring system, has national grid coverage thus also considers sites representing non-forest land.

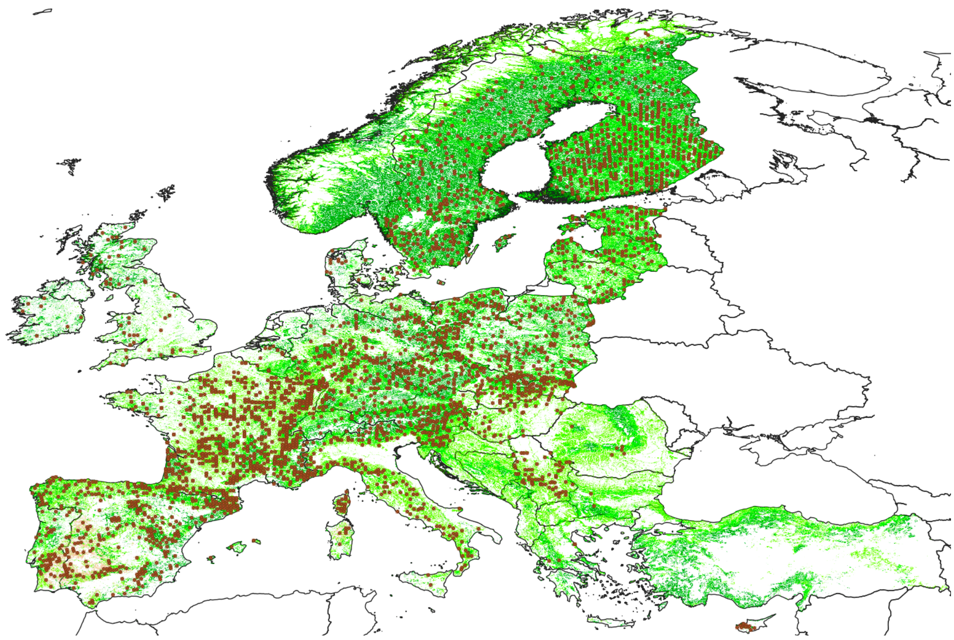

Figure 2. Overview of HoliSoils-shared profile data under Corine 2018 forest cover.

Key soil properties such as pH, organic carbon content, bulk density and soil texture of the shared soil profile data can be viewed and queried using a dashboard. The data can be filtered according to various criteria. Importantly, the dashboard also shows the licenses (i.e., open (33%) or restricted (67%)). Inherently, only data shared with an 'open' license can be downloaded through the HoliSoils dashboard.

[3] https://land.copernicus.eu/pan-european/corine-land-cover

This project has received funding from the European Union's H2020 research and innovation programme under grant agreement No 101000289.

Figure 3. Dashboard for the HoliSoils-provided datasets[4]

Descriptive statistics are provided for key soil properties. The data can be queried by dataset and country using the dashboard as shown in Figure 4.

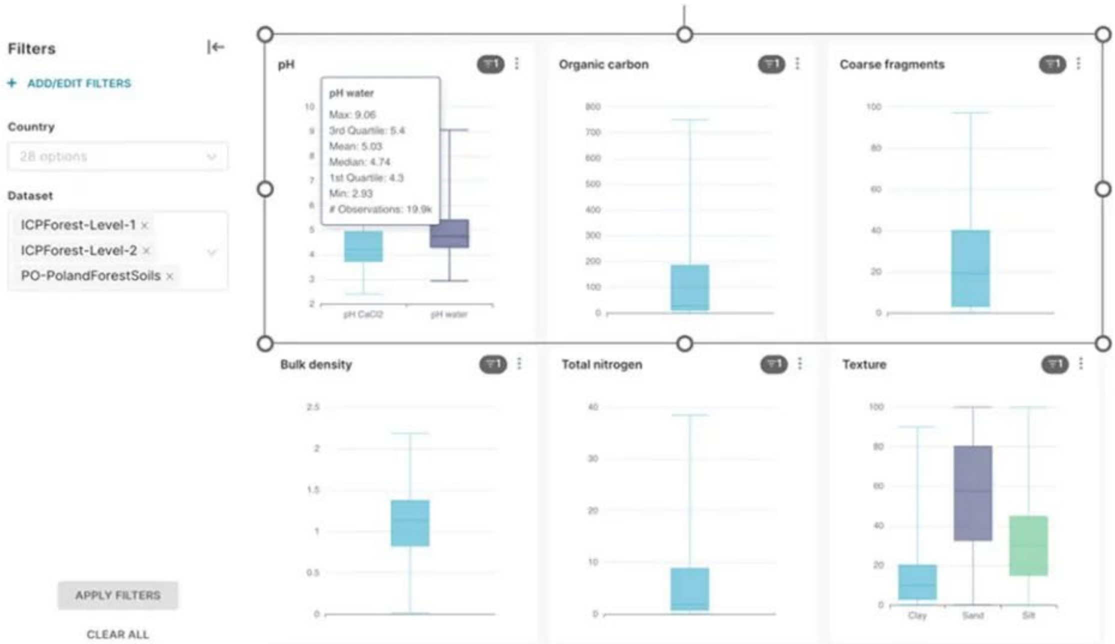

Figure 4. Descriptive statistics for key soil properties[4]

[4] *For higher solution please access the HoliSoils dashboard.*

This project has received funding from the European Union's H2020 research and innovation programme under grant agreement No 101000289.

3.2 Augmenting the dataset for mapping purposes

To augment the HoliSoils collection, which is to represent ~227 million ha forest soils in Europe, forest soil data held in WoSIS will also be used during the subsequent soil property mapping stage (~ 28k profiles, see Figure 2). Some of these point data are not freely available (i.e., plot coordinates may not be shown) yet can be used by ISRIC for mapping purposes based on license agreements ISRIC has with the various data providers.

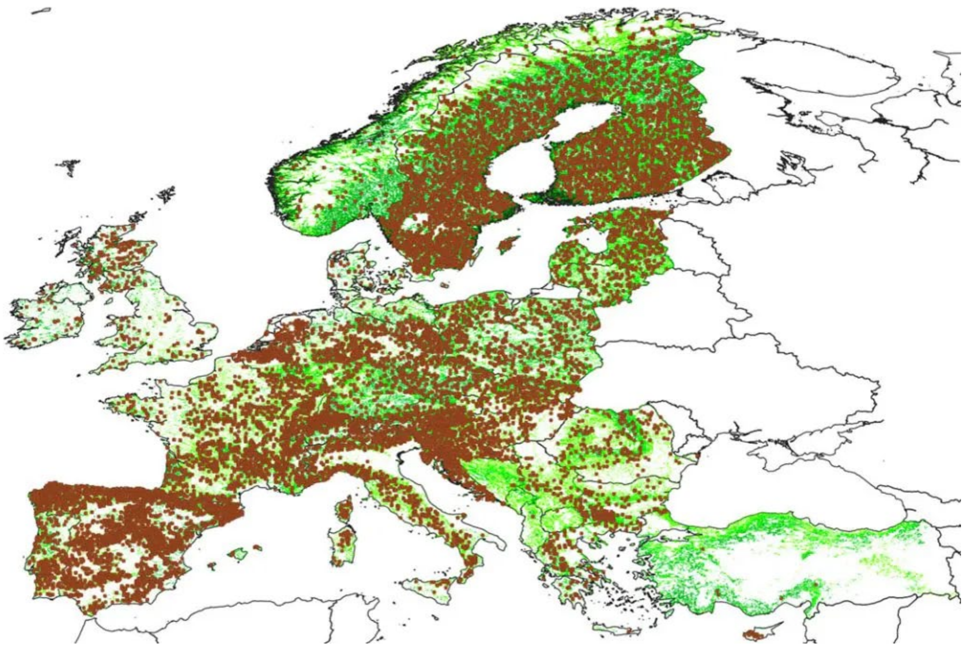

Figure 5. HoliSoils forest soil dataset augmented with WoSIS forest soil holdings

4. Next steps

For the mapping component itself we will use an approach featuring model tuning, covariates selection and regression random forest to map key soil properties at a resolution of 100 m and assess the spatial uncertainty for each pixel. This approach will build upon the work of Poggio (2021), and consider positional accuracy. The latter aspect is important here as many of the source data were shared with theoretical coordinates.

The resulting soil property maps, with associated uncertainty maps, will contribute to HoliSoils deliverable D3.6, 'Open access harmonised maps', web portal hosted at Thuenen Institute. The information will also be available through the ISRIC Soil Data Hub, with possible linkages to FAO's Global Soil Information System (GloSIS).

This project has received funding from the European Union's H2020 research and innovation programme under grant agreement No 101000289.

5. Concluding remarks

- Getting access to forest soil data for Europe for consideration in a continental scale, open-access database is problematic.
- The complement of shared datasets (4) was cleaned, standardised and ultimately harmonised using newly developed ETL procedures, implemented within WoSIS.
- Only one third (~2.32k) of the processed data can be shared freely with the international user community due to license restrictions. This subset can be queried and downoaded through the HoliSoils dashboard and Holisoils website.
- A larger complement of standardised forest soil data for Europe, including forest soil data extracted from WoSIS (~28k locations in total), can be used for the soil property mapping component of HoliSoils. Yet many of these data cannot be made available freely to the international community due to license constraints as defined by the data providers.
- Future submissions of forest soil data, collated and standardised through ISRIC's regular data acquisition programme, will be served freely to the international community through WoSIS (where allowed) after the project ends in 2025, and automatically updated in the HoliSoils-related dashboard.

6. Acknowledgements

This report was prepared in conjunction with the HoliSoils project, with funding from the European Union's Horizon 2020 research and innovation programme (grant agreement No. 101000289), and framework of ISRIC WDC-Soils long-term WoSIS (World Soil Information Service) programme.

This project has received funding from the European Union's H2020 research and innovation programme under grant agreement No 101000289.

Appendix - Reference methods for chemical properties

Parameter		Reference Analysis Method [1]					Unit [2]
		ISO	Extractant	Measurement method(s) [3]			
pH(CaCl2)		ISO 10390 (2005)	0.01 M CaCl2	pH-electrode			
pH(H2O)			H2O	pH-electrode			
Total nitrogen		ISO 13878 (1998)	-	Dry Combustion			g/kg
		ISO 11261 (1995)	-	Modified Kjeldahl			
Total organic carbon[4]		ISO 10694 (1995)	-	Dry Combustion at 900 °C			
Carbonates		ISO 10693 (1994)	HCl	Calcimeter			
P		ISO 11466 (1995)	Aqua Regia by reflux digestion	ICP	Colorimetry		mg/kg
K, Ca, Mg, Mn					AAS		
Heavy metals: Cu, Cd, Pb, Zn							
Other: Al, Fe, Cr, Ni, Na							
Hg				ICP	Cold vapour AAS		
S				ICP			
			CNS - analyser				
Free Acidity (or sum of AC) and free H$^+$		ISO 11254 (1994) Modified	0.1 M BaCl2[5]	titration to pH 7.8 or 'German' method			cmol(+) /kg
Exchangeable Cations	Al, Fe, Mn	ISO 11260 (1994) modified	0.1 M BaCl2	ICP	AAS	-	
	K, Ca, Mg, Na					FES	
Reactive Fe and Al Oxalate extractable P		ISRIC (2002)	Acid ammonium oxalate	AAS	ICP		mg/kg
Total Elements: Ca, Mg, Na, K, Al, Fe, Mn		ISO 14869-1 (2001)	HF or LiBO2	AAS	ICP		mg/kg

Source: ICP Forests (2021). Soil textural classes, pipette method, are defined as: clay ≤2 µm, silt = 2-63 µm and sand = 63-2000 µm. Only a selection of the listed properties, identified as 'key properties' in Section 2.2, were considered in this study.

This project has received funding from the European Union's H2020 research and innovation programme under grant agreement No 101000289.

References

Batjes NH 2023. *Options for harmonising soil data obtained from different sources [HoliSoils project - H2020 Grant Agreement No 101000289]*, ISRIC - World Soil Information, Wageningen, 21 p

Batjes NH, Ribeiro E and van Oostrum A 2020. 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

Bispo A, Arrouays D, Saby N, Boulonne L and Fantappiè M 2021. *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 p. https://ejpsoil.eu/fileadmin/projects/ejpsoil/WP6/EJP_SOIL_Deliverable_6.3_Dec_2021_final.pdf

Brozek S, Lasota J, Gruba P, Wanic T, Zwydak M, Blonska E and Pacanowki P 2019. *Atlas Siedlisk Lesnych Polski (Atlas of Polish forest habitats)*. Centrum Informacyjne Laszów Panstowych, Warsaw (Poland), 598 p. https://www.lasy.gov.pl/pl/publikacje/copy_of_gospodarka-lesna/urzadzanie/atlas-siedlisk-lesnych-4.pdf/@@download/file/Atlas%20siedlisk%20le%C5%9Bnych.pdf

GLOSOLAN/FAO 2021. *Standard operating procedure for soil pH determination*, GLOSOLAN / FAO, Rome, 23 p. https://www.fao.org/3/cb3637en/cb3637en.pdf

Higgins S, Kadziuliene Z, Paz A, Mason E, Vervuurt W, Astover A, Borchard N, Jacobs A, Laszio P, Wall D, Trinchera GA, Budai A, Mano R, Thorma S, Rok JM, Sanchez B, Hirte J and Madenogiu S 2021. *Stocktake study and recommendations for harmonizing methodologies for fertilization guidelines*, EJP Soil (European Joint Programme), 30 p. https://ejpsoil.eu/fileadmin/projects/ejpsoil/WP2/Deliverable_2.13_Stocktake_study_and_recommendations_for_harmonizing_methodologies_for_fertilization_guidelines.pdf

Hoffmann C, Schulz S, Eberhardt E, Grosse M, Stein S, Specka X, Svoboda N and Heinrich U 2019. *Data Standards for Soil- and Agricultural Research*, BonaRes Data Centre, Halle (Germany), 99 p. https://doi.org/10.20387/BonaRes-ARM4-66M2

Höhle J, Bielefeldt J, Hilbrig L, Wellbrock N, Eickenscheidt N, Kompa T, Grüneberg E, Hilbrig L and Wellbrock N 2016. *Bodenzustandserhebung im Wald. Dokumentation und Harmonisierung der Methoden*, Johann Heinrich von Thünen-Institut, 540 p. https://blumwald.thuenen.de/fileadmin/blumwald/BZE/Thuenen_workingpaper_97_Dokumentation_und_Harmonisierung_der_Methoden.pdf

ICP Forests 2021. *ICP Forests monitoring Manual. Part X: Sampling and analysis of soil.* https://storage.ning.com/topology/rest/1.0/file/get/9995584862?profile=original

ISO-10390 2021. *ISO 10390:2021: Soil, treated biowaste and sludge – Determination of pH*, ISO, 8 p. https://www.iso.org/standard/75243.html

Jolivet C, C., Boulonne L and Ratié C 2006. *Manuel du Réseau de Mesures de la Qualité des Sols (RMQS)*, 190 p. p. https://hal.inrae.fr/hal-02818011

Orgiazzi A, Ballabio C, Panagos P, Jones A and Fernandez-Ugalde O 2018. LUCAS Soil, the largest expandable soil dataset for Europe: a review. *EUROPEAN JOURNAL OF SOIL SCIENCE* 69, 140-153.

Poggio L, de Sousa LM, Batjes NH, Heuvelink GBM, Kempen B, Ribeiro E and Rossiter D 2021. SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty. *SOIL* 7, 217-240.

Ribeiro E, Batjes NH and Van Oostrum AJM 2020. *World Soil Information Service (WoSIS) - Towards the standardization and harmonization of world soil data. Procedures Manual 2020*. ISRIC Report 2020/01, ISRIC - World Soil Information, Wageningen, 153 p. http://dx.doi.org/10.17027/isric-wdc-2020-01

This project has received funding from the European Union's H2020 research and innovation programme under grant

RMQS 2021. Analyses physico-chimiques des sites du Réseau de Mesures de la Qualité des Sols (RMQS) du territoire métropolitain pour la 1ère campagne (2000-2009), avec coordonnées théoriques. In: INRAE (editor)https://dx.doi.org/10.15454/QSXKGA

Soil Survey Staff 2022. *Soil Survey Laboratory Methods Manual (Version 6.0., Part1: Curren methods)*, Soil Survey Investigations Report No. 42, v6. U.S. Department of Agriculture, Natural Resources Conservation Service, Lincoln (Nebraska), 1001 p. https://nrcs.app.box.com/s/fgrv9vdiiwrymtemw3ns8ocsdz7n5q3q/file/998358972771

Van Egmond F and Fantappiè M 2021. *Report on harmonized procedures for creation of databases and maps. Towards climate-smart sustainable management of agricultural soils (EU H2020-SFS-2018-2020 / H2020-SFS-2019)* EJP Soil, 391 p

van Leeuwen C, Mulder VL, Batjes NH and Heuvelink GBM 2022. Statistical modelling of measurement error in wet chemistry soil data. *European Journal of Soil Science* 73 https://doi.org/10.1111/ejss.13137

van Reeuwijk LP 2002. *Procedures for soil analysis (6th ed.)*. Technical Paper 9, ISRIC, Wageningen, 81 p. https://www.isric.org/sites/default/files/ISRIC_TechPap09.pdf

Wellbrock N, Lehtonen A, Mäkipää R, Salmivaara A and Nieminen TM 2021. *D3.1 List of confirmed data & data providers*, 3 p. https://holisoils.eu/deliverables/

This project has received funding from the European Union's H2020 research and innovation programme under grant agreement No 101000289.