D6.1: Executive Summary

Many research and policy advice depends on the presence and quality of adequate information. For soil related research and policies, this is existing or newly collected data about soil, soil properties, use, functions, quality and threats. Often the quality and the possible extent of the research and policy advice strongly depends on the available data. The proper collection, organisation, management and analysis of soil data towards useful soil information is therefore crucial in any project. This includes performing an inventory of existing data and knowledge, organising and annotating this data such that it is findable, accessible, reusable and if possible interoperable for various projects and purposes, choosing the best strategy for new data collection through sampling or other techniques and then analysing the data towards adequate and understandable spatial soil information products such as maps.

There are however many ways to approach each step of this soil data workflow, it requires a range of different expertise and the best options to choose depend on the aim and scale of the project. This report aims to provide a common basis, a synthesis and a reference for available knowledge on the best practices in this soil data workflow at EJP SOIL partners, aimed at soil institutes.

Data sources

Various recent and ongoing EU and global projects and initiatives have addressed one or more of the topics in the soil data workflow and are therefore presented and discussed in the first chapter and for each topic separately. To understand which soil and soil related data is available in Europe a stocktake on soil data sources was performed. The main conclusions of this stocktake are:

1. Basic soil property (such as soil organic carbon, pH, particle size) databases are available in each country, but some use different measurement methods. If not harmonised before mapping soil properties across Europe, this will result in sudden value changes at national or survey borders of (transboundary issues). A third of the EJP SOIL partners also collect spectral data collection on soil properties.
2. In less than half of the countries data on soil threats are available for soil pollution, compaction, water erosion, and organic matter decline.
3. Many data on soil properties are freely available, but their spatio-temporal resolution varies a lot and often uncertainty quantification is missing. The launch of the European Soil Observatory in 2020 can accelerate a comparison between national databases and LUCAS.
4. There are only a few databases available on measures to control soil erosion. Most partners reported limited access to national soil management databases. This is regulated by the national ministries of agriculture.

In addition to the stocktake on data sources the GSP CountrySIS survey was updated for EJP SOIL partner countries. The main conclusions are:

1. Basic soil data are stored in databases of very different formats, with very different data standards. Their accessibility is variable among countries, among different data owners inside the countries, and for different types of soil data. General soil properties and plant nutrients are always recorded. This is not the case for data on soil salinity, pollution, and contamination.
2. Not all countries that have soil databases, also have a soil information system (SIS) and/or a soil monitoring system. The maintenance of a SIS needs skilled staff, which is not always available.

Some countries reported to have a SIS in 2018, which is no more accessible in 2020. There is a general complaint about the lack of skilled staff, lack of financial resources, lack of time.

3. Other complaints are on the lack of communication/coordination between organizations, which makes it difficult to organize and maintain a national soil database, a connected SIS and a soil monitoring system. Another complaint is on the lack of common standards needed to integrate different soil data sources.
4. Some countries reported the lack of specific legislation for the legal implementation of soil surveys and a soil information system, specifically for soil data protection and data ownership.

Data organisation

In response to the results of the GSP CountrySIS survey, a chapter is dedicated to the ongoing initiatives, background, basic principles and choices for setting up a soil information system while using available standards for data storage, exchange and harmonisation of soil data. Following these structures, it will be easier to organise, store, use and exchange soil data for research, policy and other applications. When setting up a SIS, there is not one single best way to do it because every situation has its own requirements and therefore appropriate choices in architecture, data standards etc. Best practices that apply in all cases include a good documentation of metadata, adherence to existing standards such as INSPIRE, OGC, ISO, Dublin Core etc., and making data findable, accessible, interoperable and reusable (FAIR).

Within the EJP SOIL programme, we aim to set up a distributed soil information system that adheres to and uses the INSPIRE Directive specifications for metadata and soil (Annex III). This means that we choose that data remain at and is curated and updated by its owner (institute) and can be exchanged in a common infrastructure using the INSPIRE soil domain model and appropriate technology. This can be independent from the way partners choose to organise their data. The examples and the overview of harmonisation possibilities show that there is still quite some work to be done before harmonised soil data can be exchanged effortlessly by partners and or member states and the EC/ESDAC/EU Soil Observatory following the INSPIRE model. Currently ongoing activities are aimed at resolving as much as possible the present impediments for full and easy implementation of INSPIRE Soil by partner institutions and member states. This is geared towards at some point in time arriving at a full-fledged standardised decentralised soil information system for Europe that allows harmonisation of soil data for many different applications.

Sampling

Often the existing soil data is not sufficient to answer new questions and new soil data needs to be collected. As becomes clear in the chapter on statistical sampling methods, there is not one sampling method that fits all possible aims and campaigns. Depending on the purpose(s) of the sampling campaign (estimating a mean, mapping, monitoring, gap filling/additional sampling into an existing scheme) a choice needs to be made on the most appropriate design. In general, we can conclude that to estimate global quantities, such as means and totals, probability sampling approaches with design-based inference are more suitable than model-based methods. For sampling for mapping, model based designs are considered more appropriate. In the designs for monitoring, not only spatial variation is a factor, but temporal variation must be taken into account as well.

When choosing a design, a general rule of thumb is to keep a sampling design as simple as possible. The primary concern when designing a monitoring scheme should be to develop an adequate design that

makes good use of the available resources and not to construct the perfect, optimal design. Practical convenience and simplicity cannot be sacrificed to achieve optimal statistical efficiency. On the other hand, practical convenience and simplicity should not be the reasons for cumbersome and complicated statistical inferences.

When the aim is to combine data from two designs by far the most important aspect is to know which designs have been used including the details of the construction of the design, such as which strata were used for instance. When the design and, for probability sampling, the inclusion probabilities of the sampling units are known, this can be used to obtain an estimate of a mean or total for the area of interest. How to combine national and European monitoring schemes and other aspects of sampling such as metadata storage and sampling protocol will be elaborated more in deliverable D6.3. An overview of soil monitoring networks in Europe has been published along with suggested options for harmonizing these networks.

Mapping

When the research or policy question requires a map of a soil property, function or threat and the input data is collected it is important to choose the most suitable mapping method. Different intended uses for a map and the availability of existing (in situ point or covariate) data will result in different preferred approaches, there is not one best method. At the same time, there are a few general best practices that we advise to adhere to and a stepwise procedure is proposed to select a suitable method. This starts with defining the purpose of the map and inventorying the existing data. Thereby using knowledge of soil forming factors and the SCORPAN model to make effective choices and verifying the quality of the input data and eliminating possible errors, i.e. 'garbage in is garbage out'. During the entire data collection and mapping process a good documentation of methods, metadata, sampling design strategy and protocol, used data, chosen method, resulting uncertainty metrics and maps, validation of the result and a continuous effort to decrease possible sources of uncertainty are very important and result in a better quality map that can be validated using described methods and a repeatable mapping process. Within EJP SOIL we will adhere to the INSPIRE grid specifications for European and national maps and will aim to reduce transboundary inconsistencies and/ or propose possibilities to address these such as (lab) method harmonisation, combining different sample designs and protocols, using GPS etc.

A good soil data workflow is centered around the defined aims and (research or policy) questions. It uses existing knowledge, experience (from projects, initiatives, literature), and data, a good data organisation, a well-chosen sampling strategy, and the most suitable mapping method followed by validation of the result. This allows to adequately address research and policy questions based on relevant and sufficient quality data, thereby enabling reliable information-based decision making.

EJP SOIL has received funding from the European Union's Horizon 2020 research and innovation programme: Grant agreement No 862695