Response to Reviewer 2 (essd-2024-14-RC1):

Note: Our responses to Revier2 are given below in blue.

I have reviewed the ESSD paper entitled "Providing quality-assessed and standardised soil data to support global mapping and modelling (WoSIS snapshot 2023)".

The paper is well written and provides a comprehensive description and technical guidance on the available data. As it is a continuation of previous "snapshots", the documentation base is solid and it is nice to see the continuous improvement by Dr Batjes and his team in terms of quality assurance and expansion of the dataset. My comments are minor, also because reviewer 1 has already provided excellent feedback and the authors have responded to these issues (which I will not repeat unless my opinion on the comment or feedback differs). The large number of profiles added is impressive, but I think the paper could benefit from being clearer about some things and highlighting potential problems.

- Uncertainty assessment. I like the idea of the multi-stage uncertainty assessment in terms of location, time of sampling, methods. However, given that the intended data users here will often not be soil scientists (which is good, soil data should be used), I think it is important to explain more clearly why these limitations are important. In addition, clear warnings should be given to non-soil scientists about the incomplete coverage of certain land uses, and about soil data from the global south (certain sub-Saharan African or Arctic regions have not seen much improvement in data coverage since the last snapshot), or also information about which soil layers are covered (most data are probably still rich in topsoil data, but not in subsoil). If pedotransfer functions are used, it is important to know to what extent the profiles are genetically sound, or whether they are highly disturbed or not representative of a particular soil region. I know this is a lot to ask of the authors and you may disagree with what the purpose of this paper is, but I feel there is a risk that some soil profiles here may be interpreted as representative when we know they are clearly not, for the reasons and examples I have given above.

Thanks for these important comments. Keeping in mind the scope of the present paper, for clarification, we have added that the data served from WoSIS are based on the 'best available' data, openly shared with our centre. Many of these source data were collected using purposive sampling, hence not based on a probabilistic sampling design. We have added the following sentence to Section 4.3:

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

- Related to this point, the maps provided by the authors in response to a similar comment from reviewer 1 are a good first step towards more information about where we have soil data and

where we don't (at least in this database). However, I think more clarity could be provided by providing some sort of meta-analysis on such clear limitations as which regions can be mapped from, or whether soil profiles are reasonably representative of a region. In my opinion, the authors need to make it clear whether it makes sense to colour an area on these snapshot maps at all when we know that we only have a handful of profiles and no systematic soil data from these regions (anything yellow on the maps provided is essentially no data). Again, these limitations may be clear to soil scientists, but are often overlooked when the data are used by other research communities who will have a strong interest in soil data from these regions (which should be encouraged if these limitations are understood).

A sentence, and reference, has been added concerning the need to carefully consider the 'area of applicability' of the data:

'For example, large areas of the globe are still poorly represented in WoSIS (yellow areas in Fig. 3). As indicated, this issue can only be remedied when a larger selection of datasets is shared for consideration in WoSIS.'

- Similar to the tables for the total number and spatio-temporal variation of profiles in the database, I would find it useful to have more information available in terms of surface vs. subsurface data, land use, climate zone or soil type.

We will add a table showing the 'maximum' of soil depth sampled per continent for the 2023-snapshot, using three depth classes (0-30, 30-60 and > 60 cm Table 5) for illustration.

- Data inaccessibility: I find it quite shocking how much data is still not freely available, even from well-funded regions such as the EU, where essentially all soil data production is funded by taxpayers, no matter what the opinion of the individual data producer may be. This is not the fault of the authors, of course, but perhaps WoSIS needs to think about a mechanism to enforce true open access to all data (I have no idea what this would be, but if tens of thousands of profiles are not fully accessible, something is wrong with the system and against the spirit of open access).

We fully agree with your point of view and have been struggling with this issue for years. However, realistically, ISRIC itself is too small to resolve this important issue alone. As now indicated in the conclusions, in principle the Global Soil Partnership, through its affiliation with UN-FAO, would be a possible forum to tackle this challenge.

- This is more a question of interest or something to consider for the future: How should the dataset be viewed given the growing discrepancy between the time of sampling and assessment of soil parameters and the reliability of the values for a modern user? As we know that soil properties change over time, does this mean that we need to 'phase out'

certain parameters from profiles where we know that they may be significantly different today than they were decades ago? On a related note, Table 3 shows that more than a quarter of all profiles have no date. I think that's almost as bad as not knowing where these profiles are. Are these data points worth keeping at all, or will they cause confusion over time?

Certainly, we have been considering this. Hopefully, we can still trace the age of some of the older profiles. Older profiles remain relevant for the more stable soil properties such as soil texture. Alternatively, for soil carbon content and pH changes can be rapid. Data from different time periods remain useful, for example when using machine learning in space and time for modelling soil organic carbon change (Heuvelink et al., 2021).

In principle, as World Data Centre, we do not discard any legacy data. However, as clearly stressed in the manuscript, data users should 'filter' the available data according to their fitness for the intended use(s) they envision. That is, carefully assess the 'area of applicability of any' prediction.

References cited:

Brus, D. J., Kempen, B., and Heuvelink, G. B. M.: Sampling for validation of digital soil maps, European Journal of Soil Science, 62, 394-407, https://bsssjournals.onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2389.2011.01364.x, 2011.

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

Cramer, M. D., Wootton, L. M., van Mazijk, R., and Verboom, G. A.: New regionally modelled soil layers improve prediction of vegetation type relative to that based on global soil models, Diversity and Distributions, 25, 1736-1750, https://onlinelibrary.wiley.com/doi/abs/10.1111/ddi.12973, 2019.

FAO: Guidelines for soil description (4th ed.), FAO, Rome, 97 pp., http://www.fao.org/docrep/019/a0541e/a0541e.pdf, 2006.

Heuvelink, G. B. M., Brown, J. D., and van Loon, E. E.: A probabilistic framework for representing and simulating uncertain environmental variables, International Journal of Geographical Information Science, 21, 497-513, https://doi.org/10.1080/13658810601063951, 2007.

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

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

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