Earth Syst. Sci. Data Discuss.,
doi:10.5194/essd-2016-34-RC1, 2016

Open Access Earth System Science Data Discussions

Interactive comment on "WoSIS: Serving standardised soil profile data for the world" *by* Niels H. Batjes et al.

G. Hugelius (Referee)

gustaf.hugelius@natgeo.su.se

Received and published: 20 November 2016

The WoSIS "Standardized soil profiles for the world" constitutes an immensely valuable dataset that will undoubtedly be used widely and frequently by the scientific and decision-making communities. The dataset as it is presented here builds on many decades of hard work and experience. It is also clear that the data presentation, harmonization and formatting described in this paper follows from a long-process of planning and thinking about how data should be effectively provided to the community (see e.g. Baritz et al., 2014, Leenars et al., 2014 and Ribeiro et al., 2015). As such, I commend the authors for this work and find little reason to suggest major/significant changes or amendments to the present structure of the data. I do have a few suggestion and/or request for clarity (see comments below).

The manuscript is well written, clear and logical in its presentation of the data and the process by which data has been compiled. However, I find that data-users who want more details will likely need to consult the procedures manual of WoSIS (Ribeiro et al 2015). This applies e.g. to how standardization or harmonization of different protocols and methods was carried out (sections 2.6 and 2.7). Given the very wide scope of this work and the length of text required to explain it, I think this is sensible. However, I would suggest that it could be made more clear in the data "read me" file as well as this manuscript where the technical details can be found (as a side-note, the link to the Ribeiro et al (2015) report on http://www.isric.org/data/wosis does not work?).

The only details I would urge the authors to include a fuller presentation and description of in this manuscript is the flags of data quality (i.e. the flags referred to in the methods in section 2 or more specifically on page 6 line 27). How are included and expressed in the database?

I have found that an unexpectedly important variable to make sure a dataset finds the use it deserves from the scientific community is that it is user-friendly and easy to access. Researchers that are highly pressed for time may opt to use older data they have at hand if new data requires too much formatting before use. In light of this my only substantial comment regarding data presentation is that the authors should consider providing one file that contains all data in one large matrix/table. I realize that such a file would likely be too large for e.g. 32-bit MS Excel to handle smoothly, but many users will access the data using other softwares that would have no problem handling it. This would save many users the work of writing individual scripts that combine the different data layers for any individual profile. This will reduce the risk of unintentional processing errors and likely increase the usability of the dataset.

Specific comments

Title: I would consider changing "Serving" to "Providing" which may be a more appropriate term.

Figure1. I find the layout counter-intuitive and would prefer to have it flow the other way around, from top to bottom

I find section 2.4 and figure 2 very assuring with regards to the data quality. I believe that duplication of profile data is a big problem for many databases. With that said, it would be beneficial to data users to get some more details regarding the procedures used to identify these duplicated profiles? Or a clear reference to another source that describes the procedures in detail.

This same concern also applies to section 2.5 which is very brief and would benefit from some more detail. Please define what QA/QC is (Page 4, line 19). I also find the meaning of the subsequent sentence unclear. Does "(first) user" refer to the data provider? In that case, does this reasoning imply that data provided by "trusted sources" was screened differently?

Page 4 line 27. Here the term "pedon" appears. Is this used as a synonym to "soil profile" or does it have a different meaning to the authors?

Sections 2.5 and 2.6 clearly state that ". . .more details can be found in the procedures manual of the ISRIC World Information Service (Ribeiro et al., 2015)." This is already mentioned in section 2.7.

Page 5 line 19. "In first instance" could be replaced by "In the first generations of WoSIS" or similar, (provided that this is what you mean).

Page 5 line 21. Here, the variables carbon and organic carbon content are not listed. This is different from organic matter content which is listed.

Figure 4 is informative, but it is difficult to judge the quantity of profiles in different regions since there are so many symbols. I believe readers will be interested to get an idea of the total amount of profiles from different regions. I would suggest that you complement figure 4 with a table that summarizes the amount of profiles available from different key geographical regions (e.g. countries with more than a thousand profiles

or continents).

I found section 4 highly interesting and informative, and hopefully coming generations of interoperability may include more geographical regions (notably Asia and the Americas).

Interactive comment on Earth Syst. Sci. Data Discuss., doi:10.5194/essd-2016-34, 2016.