Reply to reviewer 1 (Alessandro Samuel-Rosa):

We thank the reviewer for his positive comments about WoSIS (sub 1) as well as his specific comments (sub 2). The latter have been addressed below in the order they appear in the reviewer's comments.

The revised manuscript, which also incorporates our responses to Reviewer #2, will be uploaded separately.

RC1 – specific comments (listed per section, and paragraph)

' I have found a few typos in the abstract. There are unnecessary parenthesis around soil property names (lines 10 to 15). Also, there is no need to use capitals for soil property names. When you mention that the original soil classification is provided, please specify that it is only the international soil classification that you are considering. These typos and inconsistencies appear in other parts of the document and should be taken care of.'

AR1: line10-15: Unnecessary brackets have been removed.

Capitalisation of property names (Nitrogen and Phosphorus). Thanks remedied.

RC2: 'In the introduction (page 3, line 15), specify that you are discussing the changes since the preceding static snapshot. In line 20 of the same page, make it clear that you are referring to various spatial scale levels.'

AR2: p.13, line 15 (original manuscript): some text has been added with a reference to the preceding '2016 snapshot'. Snapshots inherently provide a static view of a database at a given moment in time; hence, the word static has not been added.

p 3, line 20, (original manuscript). The word spatial has been added, thanks.

RC3: 'In section WoSIS workflow you mention numbers of profiles several times. Add a percent estimate along these figures (between parenthesis) so that the user can more properly see them in relative terms. You should do the same in the other sections of the manuscript as well.'

AR3: Percentages have been added between brackets to express the specified figures in relative terms. This has also been done for other sections/tables in the revised manuscript

RC4: "In Consistency checks, be careful with generalizations such as 'in accord with current conventions' (page 6, line 11). I understand that the conventions that are used in WoSIS are those agree internationally but that not necessarily are used in various countries. Perhaps you should say something such as 'in accord with current internationally accepted conventions'. In the same page, explain how organic layers are flagged, e.g. do you create an auxiliary variable?"

AR4: Good point, we added 'with current internationally accepted standards …'.

Litter layers are flagged as an auxiliary variable in the dataset, please see App. B. Now indicated in revised manuscript.

RC5: 'In Flagging duplicate profiles, I think that readers and soil data users would like to know one or two examples of what is done when you perform 'additional visual checks. You also mention that this is a laborious process. Provide some time estimate so that readers can value your efforts.'

AR5: Additional detail on the screening procedure is provided in the revised manuscript, as copied below. Adding a time estimate for this effort, however, is considered beyond the scope of this paper.

'To avoid duplication in the WoSIS database, soil profiles located within 100 m of each other are flagged as possible duplicates. Upon additional, semi-automated checks concerning the first three layers (upper and lower depth), sand, silt and clay content, the duplicates with the least comprehensive component of attribute data are flagged and excluded from further processing. When still in doubt at this stage, additional visual checks are made with respect to other commonly reported soil properties such as pH_{water} and organic carbon content. This laborious, yet critical, screening process (see Ribeiro et al., 2018) led to the exclusion of some 50,000 additional profiles from the initial complement of soil profile data.'

RC6: 'In Ensuring naming consistency, first paragraph, you could use a sentence to say what happens when data do not pass a quality check.'

AR6: We added a sentence to describe what 'happens to data that do not meet a quality check':

Data that do not fulfil the requirements are flagged and not considered further in the workflow, unless the 'inconsistencies' can easily be fixed (e.g. blatant typos in pH).

RC7:' In Providing measures for geographic and attribute accuracy, add the respective units to measures of geographic accuracy.'

AR7: The units (metrics) are already indicated in Table 2, column 3 and mentioned in the text (e.g. less than 10 m); alternatively, column 1 shows the approximate accuracy in decimal degrees. However, we added a sentence about the accuracy of 'pre-GPS' observations in the revised manuscript:

The approximate accuracy of the point locations, as inferred from the original coordinates given in the source datasets, is less than 10 m (total= 196,498 profiles, see Section 4). Typically, geo-referencing of soil profiles described/sampled before the advent of GPS (Global Positioning Systems) in the 1970s is less accurate; sometimes we just do not know the 'true' accuracy. Digital soil mappers should duly consider the inferred geometric accuracy of the profile locations in their applications (Grimm and Behrens, 2010), since the soil observations and covariates may not actually correspond (Cressie and Kornak, 2003), both in space and time (see section 4, second paragraph).

RC8: 'In your second figure, you could highlight the new soil profiles included in the present snapshot so that one can have a better idea of the improvement. You could also have two images to show this improvement.'

AR8: We have looked at this issue yet found it impractical. Further, this information is already presented in the text in Section 4, third paragraph.

RC9: '(a) In the last paragraph of page 11, you mention soil depth, but you make no reference to the depth range spanned by the data. I think that this would be a valuable information. (b) Note that you're also not mentioning the period of time covered by the data, another interesting information for readers – especially if country or region specific.

AR9a: We can only provide data on the maximum profile depth sampled during the field work (see DSDS in Appendix A, under site data); few databases *explicitly* report the depth to bedrock. For clarification, we have added the following information in the revised manuscript:

The interquartile range for maximum soil depth sampled in the field is 56-152 cm, with a median value of 110 cm (mean 117 cm). In this respect, it should be noted that some specific purpose surveys only considered the topsoil (e.g. soil fertility surveys), while others systematically sampled soil layers up to depths exceeding 20 m.

AR9b: We have added a new paragraph (Section 4, par. 2) for this:

Being a compilation of national soil data, the profiles were sampled over a long period of time. The dates reported in the snapshot will reflect the year the respective data were sampled/analysed: 1397 (0.7%) profiles were sampled before 1920, 218 (0.1%) between 1921 and 1940, 7,657 (3.9%) between 1941 and 1960, 26,614 (13.5%) between 1961 and 1980, 62,691 (31.9%) between 1981 and 2000, and 31,084 (15.8%) between 2001 and 2020, while the date of sampling is unknown for 66,837 profiles (34.0%). This information should be taken into consideration when linking the point data with environmental covariates, such as land use, in digital soil mapping.

RC10: In Data availability you mention the CSV file format. However, in page 13, you state that the data is distributed in tab-separated values format, that is TSV. Please verify if this information is correct to avoid confusion among readers.

AR10: This has been changed to TSV in the manuscript (and in the ReadMe file that accompanies the dataset).

RC11: "In the conclusions, you state that 'important gaps in the geographic and feature space will be addressed in future releases'. I think that it would be useful – for soil data users – to present a calendar

for the planed releases of snapshots, at least for static snapshots. This way, soil data users could plan their activities with more certainty."

AR11: We would really love to be able to do this. However, we are depending on the willingness/ability of data providers worldwide to provide us with new soil profile data (as well as resources available). New additions, upon their standardization, are regularly ingested in 'wosis_latest', the dynamic version of the standardized database. Generating a new snapshot and a new data paper, however, would require a substantial increase in number of profiles served (say 75,000-100,000) or a larger range of soil properties standardized. Unfortunately, open-access for soil data is not that common yet (even though Arrouays et al., 2017, mention at least 800,000 profiles in stored in digital databases); hopefully, a wider stream of shared soil data will become available in the framework of the Global Soil Partnership and GLOSIS.