Interactive comment on "A new model for intra- and inter-institutional soil data sharing" *by* José Padarian and Alex B. McBratney

Anonymous Referee #3

Received and published: 26 November 2019

This paper descibes a new protocol for soil data exchange, based on the very popular blockchain technology. After introducing the technology itself, a specific application for a bottom-up global soil spectra library is presented as an illustration of the potentialities of blockchain for soil data sharing. The paper is well scoped and well written, and in particular makes the effort of introducing this complex technology to the soil scientist reader. It could however take a bit more time to explore some of the specific aspects laid down in section 4. I particular, the immutability is brushed off quickly as a great feature, but there need to be explanations of how one could, for example, implement data versioning in such as scheme (in the case of a transcription error that has been spotted, and needs correction, for example). The paper is written in a short and impactful way, which is good. The structure is good too, altough I'd argue that section 2

and 3 should be merged.

However, some parts of the paper need to either be corrected, amended, or added. The part about data ownership is very short and vague, and implementing a technological solution like blockchain does not preclude from having a reflection about data licensing, in my view: blockchain is a technological tool, but the license data is shared under should be acknowledged as the way the rules of engagement between data sharing parties are laid. I also have an issue with the brush statement in the "Data Governance" subsection, which states that "in practice, control and governance over the data are ceded to teh central node". This is simply untrue, if you consider eg federated data management. And one could argue that when the data governance gets decentralised, there is a risk that no governance is going on at all.

Which brings me to my most important problem with the paper as it stands: it ommits completely all the issues associated with blockchain, it does not show the other flip of the coin, so to speak. For example, while the integrity of an asset can be tracked, there is nothing in blockchain that can verify the original certification - in other words, when an actor signs an asset and puts it into the blockchain, there is no mechanism to check whether that asset is legitimate or not. In this case, one could think of a soil spectra that would be invlaid/noisy/faulty. There is also other issues with the technology, like its excessive energy use, its scalability as the database grows.

Don't get me wrong, I'm not trying to say that blockcahin should not be tried or even proposed, rather that a good paper introducing a new technology like bitcoin has to present the drawbacks too, otherwise it runs the risk of "overselling" that technology. There are hard questions to ask about the use of blockcahin in general, and for soil data in particular: is it not completely overkill? Do we want, as a community, to implement a solution we know has a consequent environmental cost, and one that increases with the number of transactions? Is it scalable enough? Integrating a section showing more discussion is, I think, a requirement in this paper to get away from the technological "buzz".

C2

Lastly, with all due respect, I think there are more established papers as a reference for the use of statistical modelling of spectral data than your 2019b paper.

Printer-friendly version

Discussion paper