

## ***Interactive comment on “P<sup>3</sup> – PetroPhysical Property Database – a global compilation of lab measured rock properties” by Kristian Bär et al.***

### **Anonymous Referee #2**

Received and published: 2 April 2020

In this paper the authors describe the P3 petrophysical database they developed thanks to the support of a EU funded project i.e., IMAGE. How the data were collected and organised and the content of the database are well described in the manuscript. This topic is of interest for the scientific community aimed to a quick search for rocks petrophysical properties for different purposes. Moreover, the authors report about pre-upload data selection criteria and provide an interesting way to classify the content data in term of quality.

However, there are some general questions and others more related to the text and consequently in the dataset that have to be clarified:

- Did the authors consider the possibility to use the GeoSciML model of geological features to build the database? GeoSciML is a recognised international standard frame-

C1

work aligned on INSPIRE. GeoSciML is useful for basic data exchange and it allows to easily extend the model to address more complex scenarios. Did the author consider a possible INSPIRE compliance of the database? If yes how? If no why?

- The dataset presented by the authors, follows almost all the FAIR (Findable Accessible Interoperable Reusable) requirements (i.e., I checked the dataset with this online tool: <https://www.andis-nectar-rds.org.au/fair-tool>). It is worthwhile that this important fact is mentioned in the text.

- Regarding the datasets provided together with the manuscript, I would put more emphasis on the interoperable file format. Although a txt file is provided (which is readable by different computer machines), it would be formally better to have the database in a comma separated format (.csv) as specified in the file “Reeadme\_P\_\_Petrophysical\_Property\_Database\_V1\_Release\_2019.pdf” available in the repository. The pdf version can be avoided because it is a too large file and in my case it wasn't visible.

- In the appended file named: “Reeadme\_P\_\_Petrophysical\_Property\_Database\_V1\_Releas downloadable from the repository reported in the manuscript (<ftp://datapub.gfz-potsdam.de/download/10.5880.GFZ.4.8.2019.P3.s/>) in the section 4. ‘File format’ is mentioned that the dataset is published in comma separated ASCII file (csv, MS-DOS) with columns delimited by “;”. However, there is no csv file in the linked repository (cfr. point above). It exists a txt file where column are separated by TABs. In the same txt file (but also in the xls spreadsheet) the decimal marker seems to be a comma “,” and not the dot “.” as stated. Please check it out and in adjust accordingly.

- Page 4 – Line 24 to 35 – The authors explain the choice to use a flat file instead of a relational database with the pros and cons. They built the database following a relational database system (Line 27), but this statement is not clear. I recommend the rewording in order to clarify that point. What does it mean the sentence at Line 27? Did the author create an Entity-Relation (ER) model? The ER model is highly recom-

C2

mended even if a flat file database is developed, so it would be good to add the ER model into the paper describing at least the cardinalities among the different entities. At Line 30 the authors describe the positive aspects to have a relational database. Among them there are: i) data is uniquely stored just once and ii) eliminating data duplication that they can be referred to the important property named “referential integrity constraint”. How the authors guarantee this property? Did they perform any check on the xls file to guarantee the database consistency?

- Page 7 – Line 7 – It is mentioned that the latitude/longitude coordinates are UTM based. The UTM system is a projected system that implies planar coordinates (i.e., easting and northing). In the database the field related to coordinates report geographical coordinates as decimal degree. It is not clear which is finally the used coordinate system or why there is this difference between the text and the database. This has to be sort out. Beside that it would be good to explain also how the coordinates retrieved from literature were treated if samples had different coordinate typology.

Some more minor suggestions and technical changes:

- Page 2 - Line 15-16 - a closing bracket is missing to the citation

- Page 2 - Line 18 – substitute ‘hosted by the United States Geological Survey’ with ‘hosted by a federate infrastructure including national organization and academia (e.g., the United States Geological Survey, Southern Methodist University, Association of American State Geologists, U.S. Department of Energy’s Geothermal Data Repository, ...)’

- Page 3 – Figure 1 – the text inside the pictures are not well readable (mainly those where the text has a white border) – suggestion: if the text goes inside the picture, put text in filled white, black if text is located outside the picture.

- Page 6 - Figure 2 (structure) – in the table Magnetic susceptibility the blue is to dark and it hard to read the content of the box

C3

- Page 15 – Line 6 – probably a comma is missing after the word ‘rocks’

In the conclusions and perspectives the plan to develop a public accessible web-based interface is highly recommended for the future and should be prioritized, because the high number of rows (more than 75000 ) and columns (around 300) doesn’t make the database so easily query-able and browsable, even in the excel software package.

---

Interactive comment on Earth Syst. Sci. Data Discuss., <https://doi.org/10.5194/essd-2020-15>, 2020.

C4