

Comment on **essd-2021-312**

Anonymous Referee #2

Referee comment on "The Active Faults of Eurasia Database (AFEAD): the ontology and design behind the continental-scale dataset" by Egor Zelenin et al., Earth Syst. Sci. Data Discuss., <https://doi.org/10.5194/essd-2021-312-RC3>, 2021

This manuscript describes the schema and strategy for compiling the Active Faults of Eurasia Database (AFEAD). It also provides a link to the database itself. The database can be accessed freely through a web mapper interface and downloaded as images (jpg) with topographic background or as vectors (kmz or shapefile) of predefined tiles. The shapefile of the entire collection of faults and an Excel file with the list of references are also available for download through the ResearchGate link.

I commend the authors for the great effort in putting together such an extensive compilation of faults. I am also aware that there is a need for earth scientists to get hold of this type of data through a single access point. Nonetheless, I'm afraid that at the moment, this collection of data suffers from a few weaknesses. In brief, they are: 1) most scientific content is outdated; 2) the database design and organization of the data is technically poor. I elaborate on these aspects in the following.

Scientific weaknesses

The data collection is based on bibliographical investigations, but most of the bibliographic references are quite outdated. Out of the 657 references (in the Excel file), only 13 are post-2010. Of these 13, three are classified as unpublished information. Of all 657, 55 are classified as unpublished information, most of which are as old as 1996. How reliable can be a piece of information supplied to the authors 25 years ago and never published since then?

In the last decade, several active fault databases have been published containing updated information. Below I list some of them (not necessarily exhaustively) that have significant geographical overlap with AFEAD and contain more up-to-date data than AFEAD.

- Europe (Atanackov et al., 2021; Caputo & Pavlides, 2013; DISS Working Group, 2018; European Geological Data Infrastructure, 2021; Ganas, 2021; Jomard et al., 2017; Vanneste et al., 2013)
- Middle East (Danciu et al., 2018)
- Central Asia (Mohadjer et al., 2016)
- Georgia (Onur et al., 2019, 2020)
- Japan (National Institute of Advanced Industrial Science and Technology, 2012)

- Africa (Williams et al., 2021)
- World (Christophersen et al., 2015; Styron & Pagani, 2020)

Apart from those compilations released in the last year, most of these have been around for quite a long time now. In addition to this lack of data, the relationship between the fault representation in AFEAD and the fault representation in the source dataset is not clear. This is of particular concern for the blind faults since only criteria associated with the topographic signature are recalled. On the one hand, not considering the latest fault compilations prevents AFEAD from listing the newly recognized active faults. On the other hand, it also prevents AFEAD from eliminating those faults that were once considered active but are currently considered not active based on new evidence. Unfortunately, the CONF parameter does not consider the recency of the information.

The compilation of the fault parameters also remains rather obscure in several aspects. For example, of the 47,363 faults, 22,270 (47%) have no parameter assigned (field "Parm" is NULL). Of the 25,093 faults with the field "Parm" not NULL, only 6,849 reports a "Rate=" value; how was then the Rate (rank) parameter assigned to the remaining faults?

Technical weaknesses

The AFEAD is distributed as a single shapefile. Technically speaking, it is not even a database apart from the implicit relation between geographic features and their attributes. No relational table is provided between AFEAD and any of its linked information. In other words, it should be classified as a geographical flat-file, not a proper database.

The fields in the shapefile attribute table are very poorly organized.

First of all, none of the fields can be identified as a primary key. The lack of a primary key prevents the user from uniquely identifying any records and establishing their possible relations with external information. Also, the user cannot make an explicit reference to an individual AFEAD record when using it, including this review.

Both the "Auth" and "Parm" fields contain long text strings that, in the next update, could become even longer and easily exceed the limitations imposed by the shapefile format.

Notice that the maximum number of characters in a text field of a shapefile is 254, see Attribute limitations in ESRI documentation at: <https://desktop.arcgis.com/en/arcmap/latest/manage-data/shapefiles/geoprocessing-considerations-for-shapefile-output.htm#GUID-A10ADA3B-0988-4AB1-9EBA-AD704F77B4A2>

or

<https://support.esri.com/en/technical-article/000012081>

These two fields are also very difficult to explore, especially the Parm field that contains very heterogeneous parameters. This poor organization makes it hard for the user to use the database. For example, selecting the faults that have a certain "depth" information would require a very complex query, which would discourage the non-experts in SQL and expose the users to uncertain results. Also, the Parm field takes up more bytes than needed by repeating within the field the word to identify the parameter type, such as "Sense=" or "Rate=" or "Depth=", occasionally also including the reference to the parameter itself.

The use of the "+" (plus) sign in the "Side" field is unnecessary because all the non-null values are a plus. It could also be troublesome because the plus sign can be automatically converted when importing the data in other systems (try saving the attribute table into the Microsoft Excel format, for example).

Other issues (listed by line "L" number)

L1: The name of the database does not reflect its abbreviation "AFEAD" should be "Active Faults of Eurasia Database," not "Database of the Active Faults of Eurasia." Please make a choice and stick to it.

L14: In the file provided, the sources are 657, not 612. The difference is 55, which

corresponds to the number of unpublished work. Rephrase to make this clear for the readers.

L25: Unclear reference to "Geologische Rundschau, 1955"; see also L327.

L72-74: This statement is unclear, or it is at least quite questionable. Linear landforms created by nontectonic processes are not rare, and several earthquakes have reactivated faults with very complex patterns. Also, cases of tectonic inversion are known. Maybe the authors can expand this paragraph to make it clearer and more documented for what they want to say.

L91-92: Is the fold axis represented? Otherwise, which element of the structure is represented? And how can the user be aware of that?

Table 1: Is the strike-slip with unknown sense contemplated?

L166: Unclear to whom "our team" is referring.

Recommendations

The following technical fixes are necessary to make AFEAD suitable for using it in a proper DBMS.

- Establish a primary key that uniquely identifies each record (fault) of the shapefile.
- Separate the "Parm" attributes into different columns, paying attention to storing single numerical values in individual columns.
- Establish a primary key for the table of bibliographic references.
- Create a relational table (many to many) that connects the fault table primary keys with the bibliographic reference table primary keys.
- Once the relational table is created, the column "Auth" can be deleted from the shapefile.
- Remove all "+" "-" "=" and similar signs/symbols from all columns. Use the "+" or "-" sign only with numerical values.

The European plate boundary along the Mid-Atlantic Ridge should be completed to make AFEAD adhere to its name (it could be disappointing for the AFEAD user to find data in the African plate and not the complete European plate).

More explanations are needed to make the user understand the source of information used to assign the Rate ranks.

A justification is needed for not considering all the recent fault data compilations published in the last decade. The authors should also discuss the implications due to the lack of updated information and warn the users about the limitations in using AFEAD instead of more up-to-date regional/local data.

References

- Atanackov, J., Jamšek Rupnik, P., Ježek, J., Celarc, B., Novak, M., Milanič, B., et al. (2021). Database of Active Faults in Slovenia: Compiling a New Active Fault Database at the Junction Between the Alps, the Dinarides and the Pannonian Basin Tectonic Domains. *Frontiers in Earth Science*, 9, 604388. <https://doi.org/10.3389/feart.2021.604388>
- Caputo, R., & Pavlides, S. (2013). Greek Database of Seismogenic Sources (GreDaSS): A compilation of potential seismogenic sources ($M_w > 5.5$) in the Aegean Region [Text/html,application/vnd.google-earth.kml+xml,image/jpg]. University of Ferrara, Italy. <https://doi.org/10.15160/UNIFE/GREDASS/0200>
- Christophersen, A., Litchfield, N., Berryman, K., Thomas, R., Basili, R., Wallace, L., et al. (2015). Development of the Global Earthquake Model's neotectonic fault database. *Natural Hazards*, 79(1), 111–135. <https://doi.org/10.1007/s11069-015-1831-6>
- Danciu, L., Akyuz, K., Demircioglu, M., Gulen, L., Zare, M., Basili, R., et al. (2018). The 2014 Earthquake Model of the Middle East: seismogenic sources. *Bulletin of Earthquake Engineering*, 16(8), 3465–3496. <https://doi.org/10.1007/s10518-017-0096-8>
- DISS Working Group. (2018, April 26). Database of Individual Seismogenic Sources

(DISS), version 3.2.1. Istituto Nazionale di Geofisica e Vulcanologia (INGV), DOI: 10.6092/INGV.IT-DISS3.2.1. Retrieved July 14, 2020, from <http://diss.rm.ingv.it/diss/> European Geological Data Infrastructure. (2021). HIKE European Fault Database. Retrieved from <https://geoera.eu/projects/hike10/faultdatabase/>

Ganas, A. (2021). NOAFAULTS KMZ layer Version 3.0.1 (2021 update) (Version V3.0.1) [Data set]. Zenodo. <https://doi.org/10.5281/ZENODO.4897894>

Jomard, H., Cushing, E. M., Palumbo, L., Baize, S., David, C., & Chartier, T. (2017). Transposing an active fault database into a seismic hazard fault model for nuclear facilities – Part 1: Building a database of potentially active faults (BDFa) for metropolitan France. *Natural Hazards and Earth System Sciences*, 17(9), 1573–1584. <https://doi.org/10.5194/nhess-17-1573-2017>

Mohadjer, S., Ehlers, T. A., Bendick, R., Stübner, K., & Strube, T. (2016). A Quaternary fault database for central Asia. *Natural Hazards and Earth System Sciences*, 16(2), 529–542. <https://doi.org/10.5194/nhess-16-529-2016>

National Institute of Advanced Industrial Science and Technology. (2012). Active Fault Database of Japan, February 28, 2012 version. Research Information Database DB095, National Institute of Advanced Industrial Science and Technology. Retrieved from https://gbank.gsj.jp/activefault/index_e_gmap.html

Onur, T., Gok, R., Godoladze, T., Gunia, I., Boichenko, G., Buzaladze, A., et al. (2019). Probabilistic Seismic Hazard Assessment for Georgia (No. LLNL-TR--771451, 1511856) (p. LLNL-TR--771451, 1511856). <https://doi.org/10.2172/1511856>

Onur, T., Gok, R., Godoladze, T., Gunia, I., Boichenko, G., Buzaladze, A., et al. (2020). Probabilistic Seismic Hazard Assessment Using Legacy Data in Georgia. *Seismological Research Letters*, 91(3), 1500–1517. <https://doi.org/10.1785/0220190331>

Styron, R., & Pagani, M. (2020). The GEM Global Active Faults Database. *Earthquake Spectra*, 36(1_suppl), 160–180. <https://doi.org/10.1177/8755293020944182>

Vanneste, K., Camelbeeck, T., & Verbeeck, K. (2013). A Model of Composite Seismic Sources for the Lower Rhine Graben, Northwest Europe. *Bulletin of the Seismological Society of America*, 103(2A), 984–1007. <https://doi.org/10.1785/0120120037>

Williams, J. N., Wedmore, L. N. J., Scholz, C. A., Kolawole, F., Wright, L. J. M., Shillington, D., et al. (2021). The Malawi Active Fault Database: an onshore-offshore database for regional assessment of seismic hazard and tectonic evolution (preprint). *Geophysics*. <https://doi.org/10.1002/essoar.10507158.1>