

Interactive comment on “DamaGIS: a multisource geodatabase for collection of flood-related damage data” by Clotilde Saint-Martin et al.

Anonymous Referee #4

Received and published: 29 March 2018

The manuscript by Clotilde Saint-Martin et al. presents a new database that attempts to fill the gap of organized and collected data on flooding damages. In addition, the authors propose a method based on the functionality of elements to assess the severity of identified damages. Although the idea of this database is not new, according to the review done by the authors, it is apparently needed at a country level, and indeed might be relevant at any other country, if inexistent, as it compiles information that might be very important for future risk mitigation and preparedness. However, the database itself, as well as the manuscript, present several points that could be further considered. Regarding the database: First of all, it is not clear what the authors mean by “conceptual data model”. The way it is presented, it seems that the different tables or shapes are actually linked, through spatial joins but, as the reader advances in the

C1

explanation, it turns apparent that they are not. It is not clear why the authors have chose a polygon feature class for the EVENT shapefield, as they do not explain the criteria to define its size or extent. I wonder why the authors have not included any sort of additional information relative to the magnitude of the event, at least an indicator (e.g. flooding levels, precipitation, ...), I believe that if the final reason of this database is to support modelling, that sort of information associated with the event table, would help to better understand the actual relation between damage, and events. I have checked the geodatabase but I was not able to find any DETAIL field, do the authors actually mean field, or shapefield, and if any, why is not included in the sample geodatabase? Even if I find this work of interest, and the idea of the database very valuable, I would appreciate if the authors further explain how they plan to maintain it over time? Who is going to take care of uploading the data and doing the search of information after this manuscript is published? Indeed, it is not clear at all how the geodatabase can be updated. Is that possible through the zenodo open access site? who is going to run the quality control of the database if opened to all public? The authors should give details about this in order to actually explain the relevance of the database other than the 23 events already collected. In relation to the source of data. I find it very challenging as nowadays the number of information within social platforms and media is overwhelming, so how do the authors suggest to set a limit? I mean, when it is enough to avoid overrepresentation of certain events just because there was a lot of people present or observing. On the other hand, how do the authors run the quality control of the data from social/media sources? The authors suggest that the implementation of a quality index would improve the database. Why this was not already included? Why this index should only apply to assess the geographic accuracy? I understand that the data on damage is even more important to asses. From what I understood from the manuscript, the event with larger and better information was one that was complemented with information from fieldwork. . .this is from my point of view a bad signal as it indicates that at the end, the database should be implemented or filled with fieldwork? Regarding the manuscript: In general terms, the manuscript is well written, though it

C2

could be improved to make its reading easier. Regarding the structure of the paper, I understand very well the organization of the sections, however, the nature and some more details of the database should be presented earlier in the text, otherwise many questions arise during the reading. I would recommend the authors to explain what they mean by assessment of the damage already in the abstract. From the abstract, it is not clear to whom this database is directed, research community? I would expect this to have a broader audience; stakeholders from different sectors. It is important as well, to explain what the authors mean by local scale (already in the abstract). The last sentences of the abstract are repetitive with the first introductory sentence without adding extra information. In general terms, the authors spend a lot of time talking about the database without giving details of how it works, how it will work and what do they mean by assessment. In the introduction, lines 20 to 21 page 1; is the situation related to flood damage going to get worse only by the effect of climate change? Nothing to add related to occupation? I would recommend the authors to introduce a bit more details about the structure of the database and how they will assess the severity already also in the introduction, last paragraph. Line 20 page 3; "DamaGIS provides an increased database for..." could the authors explain this? Increased relative to what? Figure 1 shows a graph with the % distribution for 8 classes, but the legend includes only 7. Could you fix this or explain it? Line 5 page 5; I cannot actually understand how the authors expect to counterweigh the overrepresentation of large or more impacted cities, first, if more impacted, it is normal to be more represented, then if more people have seen the flood, you will have more sources of information, but not necessarily ensures that this will be proportionally distributed. Lines 1,2 page 6; what do the authors mean by conceptual data model? I would expect that then they explain better how these are actually connected, from what I understand the only connection is the fact that they have a common field. How do you determine the region field within the event shapefile? Line 7 page 7; before presenting the fields within the DAMAGE shape I would recommend to state that the estimation of the severity field will be explained later. Line 11 page 7; DATE: date of the occurrence of the damage; delete entry. TYPE: type of

C3

damaged element (please add here the element to connect with what is explained later and to inform that here you will include the element and not the damage itself Line 27 page 7; does your database have a common field with the Vict-In database? Lines 2 to 3 page 9; please rewrite; suggestion: the severity of a damage is classified or rated following the method presented in the next section. Section 3 is a bit confusing, you spend too much time introducing again the database without explain so the reader gets a bit confused. I would recommend to move the paragraph at line 25 to line 17 after the point behind damaged element. There is a general need to explain what you have decided and how you make it rather than talking about others works, this can be done and discussed later, once the reader understand your choice. At the conclusions, the authors state the accuracy of their method, but there is not a section or explanation in the text about how this is done. How the media data are contracted with reliable data. They also state the simplicity of the database, but they never explained how this is going to be maintained. Through volunteers?

Interactive comment on Earth Syst. Sci. Data Discuss., <https://doi.org/10.5194/essd-2018-28>, 2018.

C4