

## ***Interactive comment on “Data rescue of daily climate station-based observations across Europe” by Joan Ramon Coll et al.***

### **Anonymous Referee #1**

Received and published: 27 March 2019

The paper describes the digitisation and examination of daily climate data (610K observations of maximum and minimum temperature, rainfall, sunshine duration, snow depth) that have been digitised in the framework of a European project and which will be included in the ECA&D data set. The paper accompanies a publication of the data on PANGEA. This falls within the scope of the journal. The paper is short (well, it is not a huge data set), focused on the digitising process, and generally well written. However, I have a few comments that I would like to authors to address.

General: The English is sometimes a bit awkward - please have your manuscript checked by a native speaker.

General: I found the project in I-DARE (<https://www.idare-portal.org/>), but unfortunately not in the State of Data Rescue Assessment  
C1

(<https://rmets.onlinelibrary.wiley.com/doi/10.1002/gdj3.56>)

Data on PANGEA: I did not find a legend for the "Reference" column. What does so33 mean?

Abstract: Please mention somewhere the time period covered.

Abstract: Can you say something on metadata in the abstract? The systematic collection with the form given in the paper is worth mentioning.

Introduction: Perhaps the paper could cite Thorne et al. (BAMS, 2017) as a vision where all global terrestrial data set eventually could end up.

L. 151: No pressure measurements were digitised, although they appear on the form. It is not clear whether this has already been done (it is somehow implied, but not explicit). Pressure is an important variable. Even though it is not stated in the manuscript as a focus variable, it would still have been good to collect it along with the other data.

The abstract mentions "preliminary quality control". I find this very useful. However, in my view this is not sufficient for calculating extreme indices. The QC is focused only on digitising errors, not on many other possible error sources. Also, no break-point detection is applied. However, extreme indices may be sensitive to outliers as well as to biases. The authors are clearly aware of that and mention it on lines 301 and following, but then they still go on. I know that this is only for illustration of what the data could deliver, what they add to the already existing material. But perhaps this should be phrased even more carefully.

Table 1: It would be good to have more information such as number of stations, provenance (repository/archive), perhaps number of pages/images.

Interactive comment on Earth Syst. Sci. Data Discuss., <https://doi.org/10.5194/essd-2019-6>, 2019.