

Sharing Knowledge about climate data using Open Annotation: the CHARMe project

Raquel Alegre (University of Reading, UK), on behalf of the CHARMe Consortium.

CHARMe (<http://www.charme.org.uk>) is a European collaborative project that is coordinated by the University of Reading and has nine partners from both industry and public institutions from different European countries. It has a duration of two years and commenced in January 2013.

Participant's background and point of view

Climate data are derived from a variety of sources, and have a wide range of users and applications. Ideally, users require easy access to detailed and comprehensive supporting information about the datasets. This supporting information – the metadata – is as vital for the exploitation of that data as the data themselves. **The core concept of the CHARMe project is to ensure that climate data is usable to a wide interdisciplinary user community, by providing users with access to the information they require to judge whether a climate dataset is fit for their purpose.** This information will take many forms, such as publications in scientific journals, technical reports, user feedback and other related information –we term this “commentary metadata”. Commentary metadata comes from different sources: it is primarily provided by users and complements existent metadata generated by climate data providers. However, these metadata are often very disperse and difficult to find. The CHARMe project will provide this missing linkage.

The project applies the principles of Linked Data and adopts the Open Annotation standard to record and publish commentary information. Open Annotation maps directly to CHARMe requirements: a target could be a climate dataset and the annotation body some user comment about that dataset, or a reference to a publication, or a question from the user community (see Figure 1. Example of usage of the Open Annotation model in the CHARMe project. The CHARMe project benefits from the Open Annotation model to record linkage between climate datasets and commentary metadata. However, annotation targets won't always be climate datasets; annotations can also refer to satellites that recorded the imagery, scientific programs, algorithms used to produce the data, or other annotations, as long as they can be represented by an URI.). This way, the CHARMe project will allow users to benefit from the existing expertise accumulated within the climate data community.

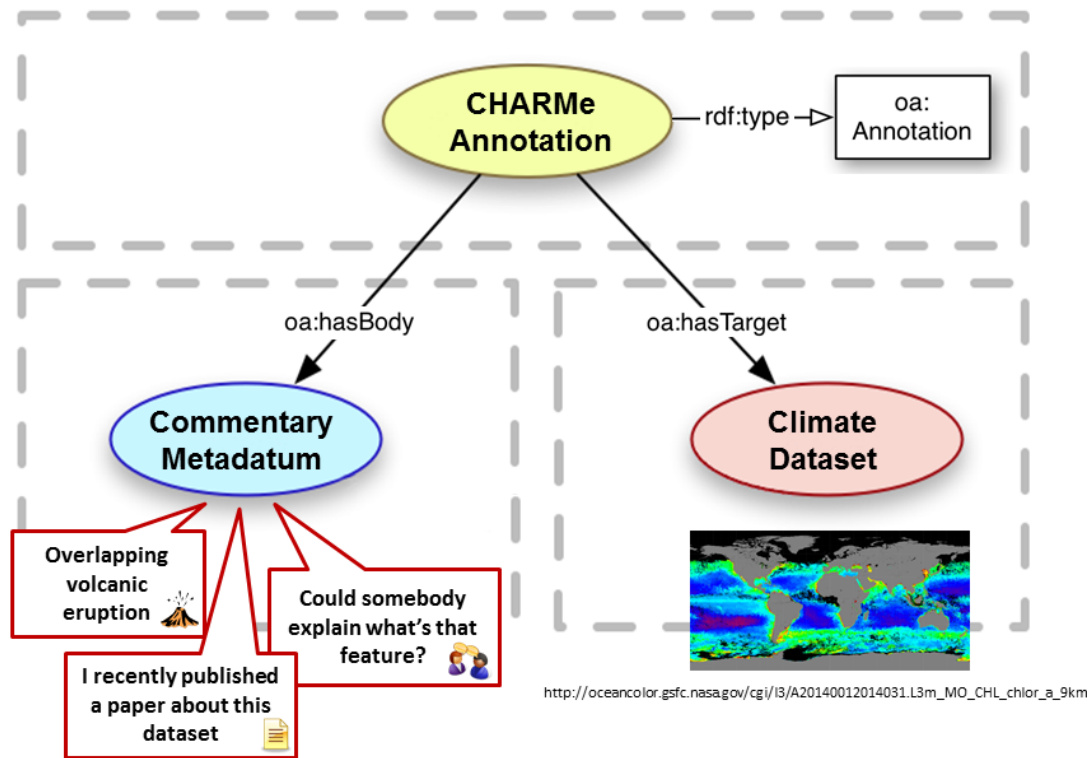


Figure 1. Example of usage of the Open Annotation model in the CHARMe project. The CHARMe project benefits from the Open Annotation model to record linkage between climate datasets and commentary metadata. However, annotation targets won't always be climate datasets; annotations can also refer to satellites that recorded the imagery, scientific programs, algorithms used to produce the data, or other annotations, as long as they can be represented by an URI.

We are following an iterative development methodology and we therefore need a flexible data model that encompasses the general notion of Commentary Metadata, permitting further sophistication to be added gradually as our developments and experiments progress. We can also benefit greatly from other mechanisms supported by the Open Annotation model, like recording an annotation motivation.

Examples and suggestions

As part of CHARMe's work, we will develop and deliver a suite of tools to enable Commentary Metadata to be associated with climate related datasets. We will combine interactive web-based visualizations with commentary information, enabling the user to see data in its full context. This will allow users to **search and insert** annotations about climate data, **compare** annotations of different datasets to aid their judgement on fitness for purpose, or develop their own customized tools on top of the CHARMe system appropriate to their specific field of study.

The main use cases satisfied by the advanced tools that we are currently developing within CHARMe are:

- **Visualization of climate data user's commentary related to a subset of a dataset.** For many applications it is very useful to be able to provide commentary about specific geographic regions, time windows or even specific pixels within a dataset. For example: recording instances of poor data quality, inaccurate pixel classification or effects of cirrus cloud and sun-glint, etc. This tool is aimed mainly at scientific users and analysts who wish to understand climate datasets in detail, from new users who want to find out more about a feature found in a dataset to experts who want to record, disseminate and evaluate the results of their research work, especially when related to a specific area of the world.
- **Visual intercomparison of data and metadata, allowing synchronized interaction with several comparable climate datasets.** This will assist users in judging which dataset is best fitted for their purpose as well as provide users with a better understanding of a dataset by measuring observational difference between datasets.
- **Plotting time series data alongside "significant events" that may have affected the quality of the data.** These include natural phenomena (e.g. hurricanes or volcanic eruptions) and features of the observing system (instrument failures, changes to processing algorithms) that might affect the quality of the data. This will help the user access uncertainties in climate products to determine whether the climate signals represented by the product are real or artefacts.

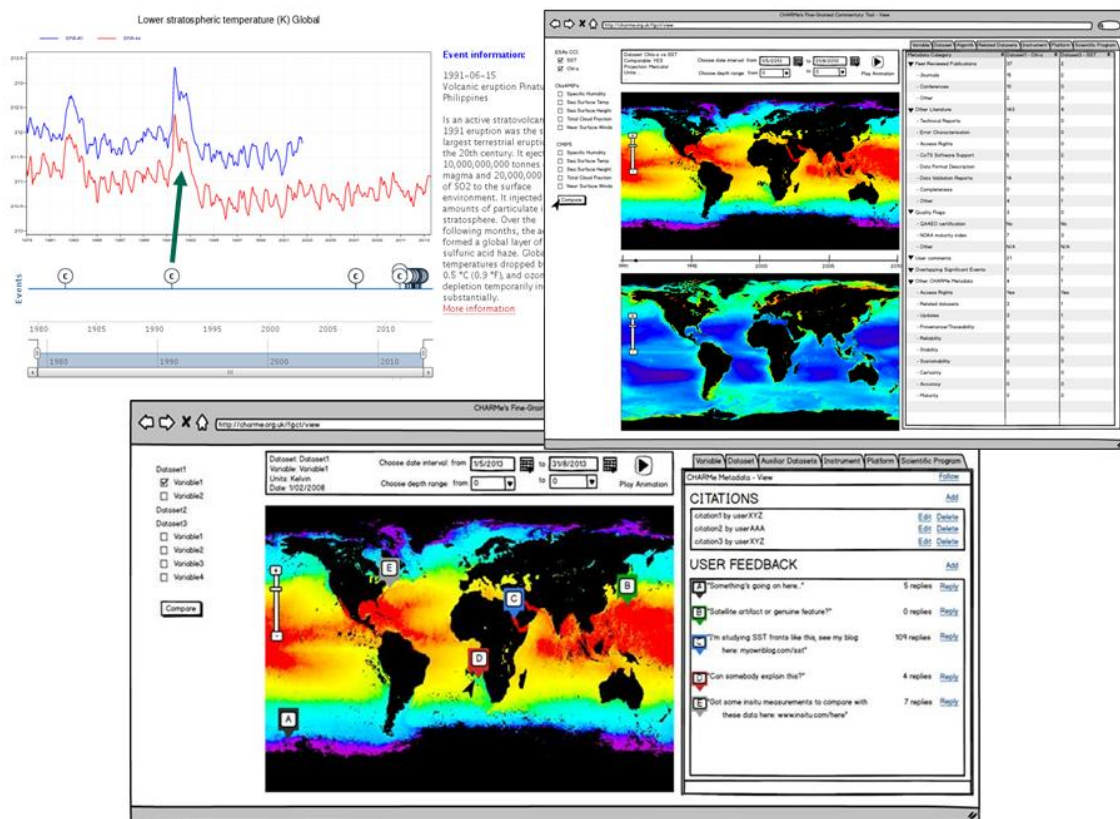


Figure 2. CHARMe data visualization tools. CHARMe will not only provide linkage between datasets and metadata, but also a system on which to build advanced applications fit for purpose. These are a few mock ups showing what we plan to develop within the CHARMe project to demonstrate it.

To satisfy the user requirements derived from these examples, we need to apply the Open Annotation data model to cope with annotation targets that could be subsets of datasets. For that, we consider necessary that Open Annotation is able to handle the following constraints or a combination of several of them:

- **Geographic regions** from individual coordinates within a dataset to transects, bounding boxes and polygons as well as depth or height.
- **Time-related targets** such as time instants or intervals;

To achieve this, we suggest expanding the actual data model for the Open Annotation “selectors” to cope, not only with regions inside images, but provide it with a way of representing time and geographic references, as shown in the following diagram:

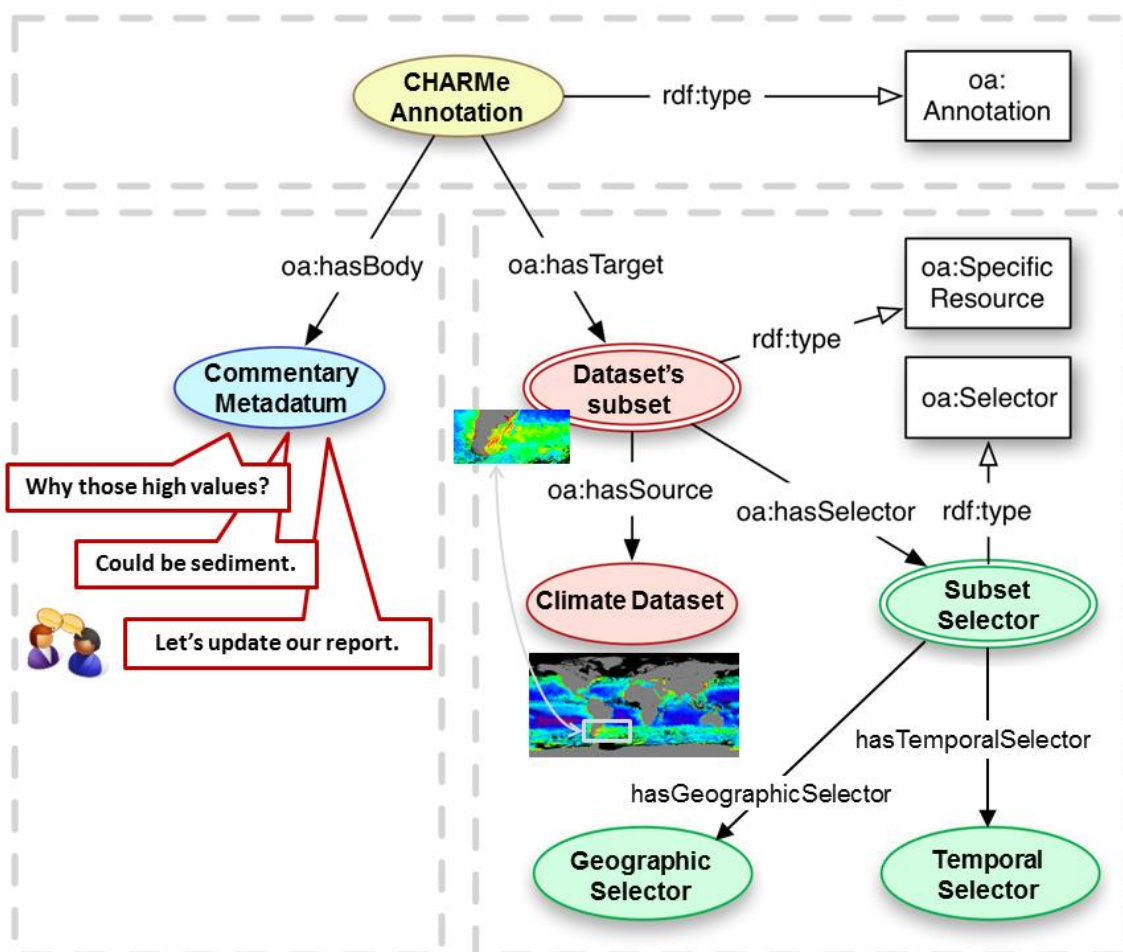


Figure 3. Simplified view of an extension for Open Annotation. This diagram presents CHARMe’s suggestion for an Open Annotation model that copes with annotation of subsets of gridded data. The target of the annotation would be a temporal or spatial subset from a climate dataset, and the selector will indicate the specific temporal and geographic boundaries of the subset.

Conclusions and statement of interest

The CHARMe project has been running for more than a year now. During this time, we have held a user workshop to complete the gathering of user requirements, designed CHARMe's data model using an Open Annotation approach, and working in the development of a plug-in tool that will be available for climate data providers to allow user consultation and insertion of annotations about the data. We are currently starting the development of several prototypal interactive web-based tools that show how the CHARMe information model can be used to tackle real scientific problems, and we think our efforts during the project can also be reused and extrapolated to other fields beyond climate data.

There are still a few technical challenges that we need to address as part of our work. At this stage of the project, **we would value greatly the opportunity that this workshop affords to present our plans and interact with the wider Open Annotation community, raising new questions and extending its data model to cope with a wider range of user scenarios. We consider attending the W3C workshop on Open Annotation would benefit significantly CHARMe's development, and we would like to explore the opportunity to contribute to the next steps in the evolution of the Open Annotation standards.**

CHARMe consortium: University of Reading, UK's National Centre for Earth Observation, European Center for Medium-Range Weather Forecasts (ECMWF), Deutscher Wetterdienst (DWD), Terraspatium, UK's Science and Technology Facilities Council (STFC), Astrium, UK's Met Office, CGI, Royal Netherlands Meteorological Institute (KNMI), Terra Spatium.