Geophysical Research Abstracts
Vol. 16, EGU2014-14574-1, 2014
EGU General Assembly 2014

CHARMe Commentary metadata for Climate Science: collecting, linking and sharing user feedback on climate datasets

Jon Blower (2), Bryan Lawrence (1,2,4), Philip Kershaw (1,3,4), Maurizio Nagni (1,3,4)

(2) National Centre for Atmospheric Science, (1) University of Reading, (4) STFC: Centre for Environmental Data Archival, (3) STFC: Rutherford Appleton Laboratory

The research process can be thought of as an iterative activity, initiated based on prior domain knowledge, as well on a number of external inputs, and producing a range of outputs including datasets, studies and peer reviewed publications. These outputs may describe the problem under study, the methodology used, the results obtained, etc. In any new publication, the author may cite or comment other papers or datasets in order to support their research hypothesis. However, as their work progresses, the researcher may draw from many other *latent channels* of information. These could include for example, a private conversation following a lecture or during a social dinner; an opinion expressed concerning some significant event such as an earthquake or for example a satellite failure. In addition, other sources of information of grey literature are important public such as informal papers such as the arxiv deposit, reports and studies.

The climate science community is not an exception to this pattern; the CHARMe project, funded under the European FP7 framework, is developing an online system for collecting and sharing user feedback on climate datasets. This is to help users judge how suitable such climate data are for an intended application. The user feedback could be comments about assessments, citations, or provenance of the dataset, or other information such as descriptions of uncertainty or data quality. We define this as a distinct category of metadata called *Commentary* or C-metadata**.** We link C-metadata with target climate datasets using a Linked Data approach via the *Open Annotation* data model. In the context of Linked Data, C-metadata plays the role of a resource which, depending on its nature, may be accessed as simple text or as more structured content. The project is implementing a range of software tools to create, search or visualize C-metadata including a JavaScript plugin enabling this functionality to be integrated in situ with data provider portals.

Since commentary metadata may originate from a range of sources, moderation of this information will become a crucial issue. If the project is successful, expert human moderation (analogous to peer-review) will become impracticable as annotation numbers increase, and some combination of algorithmic and crowd-sourced evaluation of commentary metadata will be necessary. To that end, future work will need to extend work under development to enable access control and checking of inputs, to deal with scale.