ANNOTATING CLIMATE DATA WITH COMMENTARY: THE CHARME PROJECT

Debbie Clifford, Jon Blower, Raquel Alegre, Rhona Phipps
Department of Meteorology
University of Reading

Victoria Bennett, Philip Kershaw
Centre for Environmental Data Archival
Science and Technology Facilities Council

ABSTRACT
The CHARMe project enables the annotation of climate data with key pieces of supporting information that we term commentary. Commentary reflects the experience that has built up in the user community, and can help new or less-expert users (such as consultants, SMEs, experts in other fields) to understand and interpret complex data. In the context of global climate services, the CHARMe system will record, retain and disseminate this commentary on climate datasets, and provide a means for feeding back this experience to the data providers. Based on novel linked data techniques and standards, the project has developed a core system, data model and suite of open-source tools to enable this information to be shared, discovered and exploited by the community.

Index Terms—Linked data, climate services, data integrity, data sharing, Big Data

1. INTRODUCTION
Users of climate data and services are highly diverse, ranging from research scientists (for example, searching for signals of long-term climate change) through government policymakers (for example, setting caps on carbon dioxide emissions) to operational decision-makers (for example, planning construction of flood defences). To be able to quickly determine what information is needed would be invaluable for climate services. Ideally these users would have access to a range of additional information - that we term “commentary” - to judge whether a particular dataset is fit for their purpose. Measurements from space are an important component of these climate services, and it is recognized that there is a need for both the satellite data and its metadata to be curated and shared in a systematic manner, including user feedback [1, 2]. The capture, discovery and preservation of diverse and disparate commentary metadata is a Big Data problem, and part of the data lifecycle that has not been significantly addressed previously.

∗On behalf of the CHARMe consortium. CHARMe has been funded by the European Union’s Seventh Framework Programme for research, technological development and demonstration under grant agreement No. 312541.
and more informal material such as websites, blog entries and
ad-hoc comments. It complements existing metadata (such
as the spatio-temporal coverage and resolution and the data
format) that is known by the originator and is already pro-
vided through many data infrastructures. A taxonomy of such
metadata is provided by [3]; what we term “commentary” is
analogous to “character” in this taxonomy. Commentary in-
formation is useful for several reasons, for instance:

- It helps new users to select between apparently similar
datasets to choose the best dataset for their purpose, in
a similar manner to the use of reviews on a shopping or
travel website.
- It increases the probability that vital results and lessons
concerning the strengths and weaknesses of datasets are
retained by the community, avoiding reinvention.
- It provides another view of data quality (in the sense of
“fitness for purpose”).
- It increases the traceability of conclusions in the liter-
ature back to their source data and increases the repro-
ducibility of results (e.g. the draft 3rd US National Cli-
mate Assessment [4], refers to the importance of the
“line of sight between conclusions and data”).
- It provides a new route to data discovery, particularly
where users record information about how datasets re-
late to each other.
- It provides valuable feedback to data providers, as it
helps them to improve their data and report back to their
own funding agencies.

Although many types and sources of commentary meta-
data currently exist, there has been no mechanism to provide
unambiguous links back to the source data, and make this in-
formation discoverable alongside it. A flexible, extendable
system to provide this functionality is the key innovation of
the CHARMe project. Further discussion of commentary and
potential CHARMe users can be found in [5].

3. THE CORE CHARME SYSTEM

The core CHARMe system consists of a specialised data
store commentary metadata (the CHARMe “node”) and a
data model that describes the key concepts, structure and
vocabulary of commentary metadata.

A central challenge to CHARMe is the variety and com-
plexity of climate data, which makes it impossible to repre-
sent every possible use case in one model. The approach fol-
lowed here has been to develop a data model which is flex-
ible enough to support a broad scope, and can be supported
through specialisations to meet the needs of individual use
cases. The model is based on W3C’s Open Annotation stan-
dards [6], and a number of data formats for exchanging in-
formation in this data model are also defined. Items of com-
mentary are modelled as annotations, which simply attach new in-
formation (the piece of commentary, or “body”) to an existing
resource (the “target”), such as a climate dataset. In this way,
anything that has a unique identifier (for example, a Digital
Object Identifier (DOI) or persistent URL) can be annotated
with commentary.

The CHARMe node is a server for hosting this comment-
ary information, consisting of a triplestore that is accessed
via Web Service APIs (OpenSearch, REST, SPARQL) to-
gether with a user interface for user management and mod-
eration of submitted annotations. The node is hosted by the
Centre for Environmental Data Archival in Harwell, UK. The
tools described in the following section are examples of client
programs, hosted elsewhere, which use the APIs to add and
retrieve commentary information from this central repository.

4. CHARME TOOLS

CHARMe has developed a suite of tools and applications that
demonstrate different ways in which commentary metadata
can be used, including a “significant events” viewer (which
matches timeseries of climate data with events in time that
might have affected the data), a plugin for data providers,
and the CHARMe Maps tool, which examines fine-grained
commentary and supports data and metadata intercompari-
son. The plugin and Maps tools will be described further in
this section, while the significant events viewer, as applied to
climate reanalyses, is the subject of a separate paper in this
issue.

The CHARMe plugin is a Javascript component that is de-
dsigned to be integrated into existing data-provider websites,
providing an interface for viewing and entering commentary
metadata. The results of a user’s search are augmented with a
“C” icon, which is coloured when commentary informa-
tion has already been recorded for that search result. Figure
2 shows a screenshot of the plugin being tested at ECMWF’s
data archive. Within the project, the plugin is also being tested
deployed at KNMI (the European Climate Assessment and
Dataset archive), DWD and CEDA. In this way, we are allow-
ing users to discover commentary via the websites that they
are already using to access climate data. The plugin has a
faceted search interface to search for existing annotations and
functionality for adding new annotations.

The CHARMe Maps tool is an experimental interactive
map interface for browsing datasets, and creating commen-
tary information attached to subsets of datasets. For exam-
ple, a user might want to highlight an interesting feature in
a satellite image, such as a dust storm or volcanic ash cloud,
or flag up a potential problem with a processing algorithm or
sensor, which may affect all data in a certain geographic re-

Fig. 2. Screenshot of the CHARMe plugin being tested at the ECMWF archive. The “C” icon is coloured-in to indicate that a dataset has already been annotated with commentary.

Fig. 3. Screenshot of the CHARMe Maps tool, showing a user browsing several datasets at once via an interactive map, with intercomparison of their associated commentary information (in this case comprising publications and technical reports) in the right-hand panel.
tists working on projects within ESA's Climate Change Initiative, which is producing long-term, high-quality climate data records. CHARMe Maps includes functionality for data intercomparison: users can load several datasets in parallel and visualize the available commentary annotations at the same time. A screenshot of the tool is shown in figure 3.

The CHARMe Maps tool is being developed as a proof-of-concept for fine-grained annotations and the ability of the data model to support geographical information, and will not be fully operational at the end of the project. However, since the early design stages of the tool, different international science and user groups have showed an interest in testing the tool for future integration in their work, including ESAs Climate Change Initiative (CCI) Sea Surface Temperature group at University of Reading (UK), the CCI Clouds group at DWD (Germany) and the US National Climate Predictions and Projections Platform, formed by scientists from NOAA, NASA and JPL (USA).

5. HOW DOES CHARME HELP IN A “BIG DATA” FUTURE?

The climate science community has to deal with many issues relating to Big Data, including volume (e.g. the state of the art climate model output database is of petabyte scale), velocity (e.g. 8TB/day from ESA’s Sentinel series of climate monitoring platforms), variety (e.g. satellite, in situ and model output) and veracity (i.e. data quality). This project is making particular contributions to the understanding of Big Data variety and veracity, by linking disparate information and enabling users to make judgments about the applicability of datasets to different problems.

CHARMe is harnessing the power of the Semantic Web and Linked Data, which enables us to publish commentary metadata widely in a way that can be interpreted both by humans and by automated software. The project is not attempting to alter the entire approach, formats and standards used by the climate and EO community, but rather to engender a different way of working so that the vital commentary metadata can be understood from a common perspective, allowing users (from whatever origin) to be able choose data appropriate to their needs. Although CHARMe has a particular focus on products derived from Earth observation, the open-source technologies developed in the project could readily be applied to other fields. All CHARMe software will be open-source, released under a liberal licence, permitting future projects to re-use the source code as they wish.

The CHARMe system provides a means of making climate data comprehensible to new communities, as well as serving the existing user community with tools for the inter-comparison of metadata records and best practice for their generation and preservation. In future, climate data users should start to expect a CHARMe-button at their data provider, giving access to the diverse commentary relevant to their chosen dataset. Behind this small button, that at first may not look like much, is the start of a new functionality serving the rapidly-growing area of climate services.

6. REFERENCES


