



Characterisation of metadata to enable high quality climate applications and services

Deliverable 800.3

Report on first stakeholder workshop

Held on 14 March 2013, Reading, UK

Partners providing input: Met Office, University of Reading, Infoterra, DWD, KNMI, SIH

version number	date	comments
0.1	29 April 2013	First draft by Met Office
0.2	31 May 2013	Includes input by WP800 partners
1.0	15 July 2013	First version



CHARMe is funded by the EC under its FP7 Research Programme

Contents

1	Introduction and purpose of the workshop	3
2	Structure of the workshop	3
3	Workshop outputs	4
	3.1 Current view of metadata	4
	3.2 Judging a dataset's fitness-for-purpose (BOG1)	5
	3.3 Using Commentary metadata (BOG2)	10
4.	Conclusions	13
	Annex 1: Workshop programme	15
	Annex 2: Workshop attendance list	16
	Annex 3: CHARMe scope and rationale	17

Report on CHARMe first stakeholder workshop

14 March 2013, Reading, UK

1 Introduction and purpose of the workshop

CHARMe is an EC FP7 research project whose aim is to develop and implement an online tool for characterising metadata to enable high quality climate applications and services. It is a short, focused project running for two years which ends in December 2014. At the core of the project are four Work Packages (WPs) with the technical task of developing the CHARMe tool, while other WPs in the project will support this core through engaging with users and stakeholders at all levels.

The success of CHARMe thus depends equally upon developing the system and successful engagement with the wider climate research community to understand how their data needs can be better served, and then to demonstrate the value of CHARMe. To address the former, a stakeholder group was identified at the start of the project. A user workshop was organised to achieve the following:

1. Inform a key stakeholder group about the work and aims of CHARMe
2. Discuss the stakeholders' requirements from data sets and data application
3. Discuss, as far as possible, how the putative CHARMe system might fulfil those requirements

The workshop was held on 14 March 2013 at the Holiday Inn Hotel in Reading, UK. This report is a brief description of the workshop and its outputs, especially those which will act as input to the user requirements document (D300.1). This report is available on the deliverables page of the CHARMe project website¹.

2 Structure of the workshop

Management of the workshop was by the University of Reading (UREAD), the lead partner for CHARMe, who constructed the programme and stakeholder invitation list in consultation with other project partners. The workshop was divided in to two parts, a morning and an afternoon session, with both split in to three sections, as: Introduction; Breakout; and Feedback sessions. The workshop programme is given in **Error! Reference source not found.** The Introduction and Feedback sessions were in plenary and were chaired by the CHARMe Project Manager. The breakout groups were organised as three

¹ <http://www.charme.org.uk>

parallel groups of about ten experts each, comprising consortium members and stakeholders. They were chaired by consortium partners.

From a group of over twenty five stakeholders, identified for their interest in data development and use, ten attended the meeting as ‘users’ to provide an independent external perspective for the discussions aimed at developing the CHARMe system. The full list of stakeholders will be further expanded in the project for further engagement to develop the CHARMe system. The stakeholders who attended the meeting are listed in **Error! Reference source not found.**

Prior to the arrival of the stakeholders at the workshop a ninety minute session was held with consortium members. This pre-meeting was to agree the detail of the scope and rationale of the project, to discuss what the user requirements might look like, and to agree how the breakout groups could be best employed to inform the user requirements for the project.

3 Workshop outputs

3.1 Current view of metadata

WP300 is directed towards gathering user experiences in using climate datasets for informing the system development team. The WP leader, John Blower (UREAD) gave a presentation on existing commentary metadata and described the key attributes of useful metadata and also what potentially could comprise a successful enhanced metadata system. Examples cited included the Tim Berners-Lee five star scheme (**Table 1**) and research work published by Parsons and Fox² and Bryan Lawrence³. In summary it was noted that accessibility and openness are essential for the success of an enhanced metadata system.

Table 1: The Tim Berners-Lee five star deployment scheme for open data

- ★ Available on the web (whatever format) *but with an open licence, to be Open Data*
- ★★ Available as machine-readable structured data (e.g. Excel instead of image scan of a table)
- ★★★ as (2) plus non-proprietary format (e.g. CSV instead of Excel)
- ★★★★ All the above plus, use open standards from W3C (RDF and SPARQL) to identify things, so that people can point to your data
- ★★★★★ All the above, plus: link your data to other people’s data to provide context

The NOAA Maturity Index was also mentioned as useful metadata descriptor, while the ESGF was noted as having metadata links on the data serving page which were thought to be useful. From the wide array of existing metadata types currently in use consideration would need to be given as to how and which

² Parsons, Mark A., P. Fox 2011, Is data publication the right metaphor? *Data Science Journal*.

³ Lawrence, Bryan, et al 2009, doi:10.1098/rsta.2008.0237

ones a CHARMe system could interact with. A fundamental question which arose was whether the CHARMe information should be embedded in the data, or should it be separate. It was agreed that CHARMe should always be descriptive and not used as a 'quality stamp' for data.

The *Faceted search* as used on many web sites to filter the data was shown as a tool for enhancing the utility of data, and features of the Google Finance data tool were also demonstrated. A time series of a climate datasets with significant events (e.g. new satellite observations, volcanic eruption) flagged on the plot was shown as a good example of how CHARMe can add value. The importance of knowing external experiences, for example on how to make datasets 'discoverable' was discussed (CMSAF, CCI, EURO4M, QA4EO, EUMETSAT and GCOS are projects which already are 'discoverable'). ISO standards and related projects were also briefly reviewed, including: GeoViQua, MOLES, Metafor and ESA LTDP. Open annotation was proposed as a way of linking documents and other things to datasets.

A summary document describing some of the pertinent features of metadata and its uses was circulated to delegates as pre-meeting reading. It is reproduced in **Error! Reference source not found..**

3.2 Judging a dataset's fitness-for-purpose (BOG1)

The stakeholders then joined the workshop and a summary description of the project's aims and plan was given for them. This was followed by the brief below to structure the discussion in the breakout groups:

1. What climate data do you use?
2. What is your relationship with data?
 - are you a data user, provider or service?
3. What sort of climate data do you use?
 - model or observation climate data?
 - processed or raw data?
4. What supporting information do you need?
5. Judging fitness for purpose
 - is it easy to find information about the usefulness of the data?
 - is that information complete/appropriate to your needs?
6. Where/how should CHARMe information appear

The attendees then broke in to three pre-arranged parallel break out groups of one hour duration to discuss and answer the above questions. These groups were labelled as 'blue', 'orange' and 'green'. For ease of understanding the results from all three breakout groups are presented here in an integrated format.

SUMMARY RESULTS FROM BREAKOUT GROUPS

What climate data do you use?

- *Climate data, IT data management, Data Provider and service provider, in situ, stations from all over Europe, climate indices, ground based and airborne - KNMI*

- *Climate data, Observations (all!), model data – Met Office*
- *CM SAF – long term record, polar orbiters, research, met services, commercial users (solar energy) – DWD*
- *User of a hierarchy of data, low level L1,L2, ancillary cal-val – NCEO*
- *Satellite observations - ESA*
- *EO sat images, from ESA - Magellium*
- *UK observations, in situ*

What is your relationship with data?

- *Producing forecast data - ECMWF*
- *Climate monitoring - EUMETSAT*
- *Make climate maps - KNMI*
- *CMSAF – operational centre – use operational data – DWD*
Climate modellers use it for different purposes: model initialisation, attribution studies, boundary conditions, process studies, model validation – Met Office
Producer/supplier - ESA

What sort of climate data do you use?

- *NOAA, NASA data for in situ data for validation*
Discover and interpret (tests on incoming data)
Ancillary data, reanalyses, sat data from ECMWF and USA sources
- *Data from ESA, EUMETSAT, NASA - NCEO*
Climate quality data – higher level.

What supporting information do you need?

- *Ancillary data used in the algorithm to process satellite data*
- *Land , Sea, Cloud, Ice masks.*
- *Uncertainty characterisation – accuracy, stability, precision – is this described per datum, spatially or temporally – it needs to be described in a clear, open and accessible way.*
- *Multi-media learning materials, not only docs or pictures*
- *Compliance flags; cost / availability / policy; popularity; amount of supporting information e.g. no. of citations*
- *Scope: target with id?, version, modified by... creator?, time, geography, sensor, phenomenon*

Judging fitness for purpose

- *Are climate trends seen in the data real or false? Is the time series long enough for climate studies? Trends are what are important therefore: Events are critical, but difficult to bring together*
- *Homogeneity – gaps, changes in algorithms, change in platform*

- *Uncertainty characterisation – accuracy, stability, precision – is this described per datum, spatially or temporally – it needs to be described in a clear, open and accessible way.*
- *Ancillary data used in the algorithm to process satellite data*
- *What about the fitness of supporting information like land masks?*
- *It can be difficult to get calibration of raw satellite data – esp. older sensors, changes in calibration*
- *L1 calibration drift, L2 retrieval method – difficult to communicate to user, L3 what sampling, averaging*
- *Very hard to make a judgement without expert knowledge Are the right people making the assertions?*
- *Info about jumps and changes in the data – why are jumps there?*
- *Published in peer reviewed papers*
- *CMIP archives – papers helpful if data is old enough for papers to have been written*
- *Information that is provided on a website but users can't find it*
- *ATB, user manuals*
- *Context is important, direct links needed to see contextual information*
- *Credibility of asserted precision – find independent info through calval activities*
- *Log of changes, evolve then pass back to the data provider*
- *StackOverflow – vote up comments (endorsements) – you need to have a certain amount of 'credit' on the system, it **is** moderated.*
- *Info about SW tools (processing, exploring, etc) related to datasets: pros/cons, links to where to download or user guides...*
- *[Tool to find] info about provenance tree, calibration, validation, etc*
- *Contact info to trace user comments on data*
- *Data comparison information*
- *Issue on allowing user comments: this risks the image of the provider. Users should have different permissions depending on them being newbies, experienced users, providers, people able to qualify a dataset as ISO (or else) compliant, etc.*
- *Discussion on should then CHARMe allow user comments? Publications more trustable (“important”) than user comments.*
- *Categorize level of data or group data so there's a quick view which can be taken in by user rather than an overwhelming list, then expand as needed (use expanding drop-down boxes)*
- *Bar chart on citations in categories? This would require keyword harvest, thus asking users who put in metadata to populate some pre-defined bins*
- *Maps with location of commented observation stations, so users' feedback can be assessed for distribution*
- *User feedback – when would you harvest the feedback... at point of download? One could get a measure of data use, but the usefulness would have to wait a long time and may never materialise. So data providers may have to take the initiative to populate the metadata. Search results could be presented in order of investment in feedback... ? Could we offer prizes for entering metadata?*
- *Data provider can make a choice to order by a metadata criteria e.g. no. of citations. On CHARMe's own portal, can allow the user to do that, or comments, or active thread, ordering/archiving comments.*

- *Data provider should be allowed to flag things of note.*
- *TripAdvisor as another analogous system; the hotel owner is recognised and able to respond to comments.*
- *The CHARMe system would also be useful for gathering evidence of “impact”, if it’s tracking data usage. A user could press a button to discover how a dataset has been used, and by who.*
- *Which is the best set for my needs? There is an order in which users select – geo-temporal?, features and coverage are of fundamental consideration. Start with CHARMe parameter search - OR - start with geo-temporal search?*
- *Need to flag some issues upfront before proceed with refining search*
- *Augment search terms with CHARMe ones e.g. data complete or not, coverage not complete*
- *Timeliness of updates – climate data long time span*

Where/how should CHARMe information appear?

- *Data provider websites (archives, sources) e.g. ESGF. Should CHARMe information appear similar to the ‘Model Metadata’ link on ESGF?*
- *Online databases of sensors, missions etc.: CEOS handbook, ESA, Metafor? There are international initiatives such as GEO and CEOS – would the link appear there?*
- *Service providers: CEMS*
- *Specialised tools : DMITRI (sensor comparison, overlaps/overpasses)*
- *Dedicated CHARMe: search engine website.*
- *What would I have to host as a data provider to make such a link appear?*
- *The CHARMe link should appear consistently in these different places, i.e. there should be some degree of branding so the user can quickly learn what the button means.*
- *Should the CHARMe system expose other non-C metadata? What overhead would this put on archive holders?*
- *Can imagine accessing the CHARMe information through a wiki page (user guide?).*
- *Will CHARMe have its own access portal, somewhere where all the CHARMe records can be accessed?*
- *Need to understand what is hosted by CHARMe and what is hosted elsewhere (deployment diagram).*
- *Would like to have the option of sharing the CHARMe information with a colleague -> this could be particularly useful for on-the-fly comparisons of data. Bryan Lawrence suggested that for the Metafor pop-up in ESGF, the link required for this is available, it is just not presented to the user.*

Other points

- *Problem: No maintenance budget: what if issues from users come up? What if new ideas/suggestions arise after the end of CHARMe?*
- *Often users do not know where to go to find the data in the first place – can CHARMe help here? Possibly not in this first phase, but maybe CHARMe will be a way of helping users find data across archives in the future.*

-
- *From the point of view of a data provider, CHARMe could allow you to ‘crowdsource’ the verification of your data. How would you want to capture/view/filter this information as a data provider? User scenario required*
 - *What if a user wants to select data based on a feature within the data, e.g. whether a MODIS scene is cloud-free. Could the CHARMe system provide a tool to process the data to discover this? Probably not, and definitely not in scope at this stage, but perhaps the output from such a tool could be captured and used in the repository.*
 - *It would be useful to track usage of datasets not just by application, but by location. E.g. is MODIS LAI used to study parts of the world more than others? It would be useful to apply this in reverse too, e.g. I am interested in the Tibetan Plateau, which datasets are used most often to study this region?*
 - *External events: these could be quite different for global datasets and in situ data. For the latter, a change in the local measurement environment (e.g. a new building, a tree chopped down, change in land cover) will be important.*
 - *Any data entry portal should allow users to input just a bit of info at a time and the option of returning to it later (i.e. no requirement for completeness in one go).*
 - *Could comments be ‘voted up’ rather than given a star rating? E.g. ‘I found this comment useful/unhelpful’*
 - *Data providers are likely to be the earliest users of the CHARMe system.*
 - *The system will potentially be holding lots of information – will need to think about how to filter this for the user rather than showing them everything all at once.*
 - *A data provider may want to host their own ‘in-house’ version of the CHARMe system to harvest feedback on its archive but keep this feedback private.*
 - *Ideas for visualization of the metadata – bar graph showing how many comments? Requires a thematic classification of comments, and probably classifications of comments in time and space. Could this be done at data-entry stage with a drop-down menu?*
 - *Can journal articles be harvested automatically? Are data providers expected to populate the repository or will it all come from users?*
 - *Where data providers are already obliged to report data use to other authorities, e.g. Eumetsat, this should be injected straight into CHARMe.*
 - *The project should provide a standardized look-and-feel to the CHARMe button for a data archiver website, perhaps with a few options on size etc. Also a choice of what happens when the button is pressed – pop-out, redirection to another page, embedded text on current page, etc.*
 - *Option of tailoring this look-and-feel for specialized end users?*
 - *For looking at the CHARMe metadata held for in situ sites, it would be good to have a map interface that shows the location of measurement stations and highlights the ones that have user comments. Then it would be good to discover other datasets that depend on that station (e.g. reanalysis) or have used that station for validation.*
 - *It would also be good to show the geographic distribution of CHARMe users/contributors.*
 - *Highlighting of ‘most active discussion’ threads?*
 - *We need to be aware of language barriers – will comments only be allowed in English?*
 - *The comment portal should be linked to the discover and download interface, to encourage users to submit something while they’re thinking about the data.*

- *Forms already exist in archives asking for users to give feedback on data and they are little used.*
- *Could the system provide metrics (e.g. summaries of existing entries, most comments per user etc) that entice people to leave comments?*
- *Could discovered results in an archive be ranked in terms of the amount of data that exists in the CHARMe system?*
- *Privacy issues.*

3.3 Using commentary metadata (BOG2)

For the second session of breakout groups the same membership was retained, and a new set of questions posed. These questions concentrated on metadata and its properties, feasibility and implementation, and were as follows:

1. What properties of a data set are inherently external to that set?
2. What corrections, comments or additions would be useful content for metadata?
3. What hard to find information would be useful content for metadata?
4. What aspects of well handled or best practice metadata should be considered for inclusion?
5. What is out of scope for inclusion in CHARMe?
6. If you are a data user, how would you like to discover data?
7. What apps would be helpful in your work?

An example of the draft output from brainstorming in the 'blue' group in the second breakout session is shown in Figure 1.

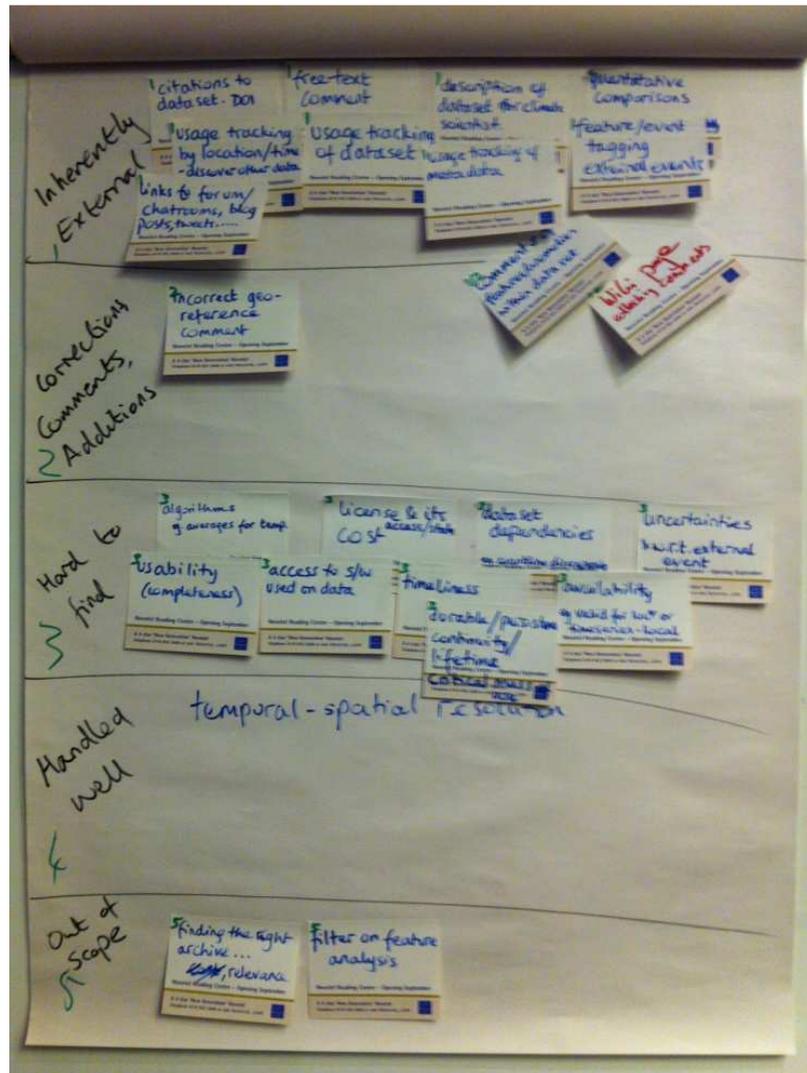


Figure 1. Example of the draft output from brainstorming in the 'blue' group with information about commentary metadata which they considered useful.

What properties of a data set are inherently external to that set?

- Citations to dataset (DOI),
- Free text comment,
- Description of dataset for climate scientist,
- Usage tracking by location, time → discover other data,
- Links to forum, chatroom, blog, post, tweet...,
- Usage tracking of dataset,
- Usage tracking of metadata,
- Quantitative comparisons,
- Feature/event tagging external events

What corrections, comments or additions would be useful content for metadata?

-
- comments of features/anomalies within dataset, wiki page collection comments (which also cross into previous category)
 - Incorrect geo-reference comment
 - Changes in instrumentation,
 - Time series of ECVs corrections

What hard to find information would be useful content for metadata?

- Algorithms e.g. average calculation for temperature
- Usability (completeness)
- Access to s/w used on data
- Licence, its access, status & cost
- Timeliness
- Dataset dependencies e.g. coastline discrepancies
- Uncertainties especially with reference to external events
- Availability e.g. valid for location or time series
- Durable/persistent/continuous/lifetime, critical mass of usage
- Provenance – traceability
- Information about versioning
- Mapping between datasets & journals
- Journal/papers/tech notes don't report problems
- Means to feed back that an issue has been fixed
- Attribution & contact point for further info.
- Changes: calibration, external (climate) events
- Validity time of info provided

What aspects of well handled or best practice metadata should be considered for inclusion?

- Temporal-spatial resolution
- Journal, papers, tech notes
- Assessment and validation reports should be in the metadata.

What is out of scope for inclusion in CHARMe?

- Finding the right archive, relevance, filter on feature analysis
- Versioning of dataset,
- Given parameter search across data providers,
- Static instrumentation specifications

If you are a data user how would you like to discover data?

Dataset intercomparison vs Numerical Intercomparison

Dataset discovery/selection intercomparison	Numerical Intercomparison
Citations	Cal/val results
Compliance	Errors/uncertainties
Spatial/Temp coverage	Traceability (measurement)
Cost	Repeatability
Availability	Coincidence of criteria
Policy	Ground measurements to compare with EO data
Popularity	
Algorithm inheritance	
Amount of info	
Resolution	

What apps would be helpful in your work?

- Feedback by e-mail
- Subscription systems to enable notifications.
- Applying tags to threads of discussion, get daily digest of updates
- Exec quality checks => automatically dump into CHARMe system
- automated dumping of data
- mapping between CHARMe and existing repositories – rest interface fronting existing RDBMS
- UI for ordering data, link within search
- Central repository UI is still useful
- Happy to have search plug at data provider but yields results from all data providers

Other points

- What is most important and what is easy to implement
- Provenance has come up many times
- Events – scope of events – need to way to filter over them
- Exportability of information e.g. Google Reader is closing down
- Text mining – need to look at what is possible
- Portal for CHARMe system – makes it a GEOSS system in its own right – better to work with other project e.g. GMES-CLIP
- OpenAIRE – as way to fit journals to datasets
- Anonymous feedback should not be allowed – it's an attack vector
- Need to consider the automated capture of metadata

4 Conclusions

It was agreed that the meeting achieved the aims set out in advance. The discussions with stakeholders provided valuable insight on their requirements, which will feed in to the user specification of the CHARMe system. The delegate discussion in the final plenary is portrayed in the photo shown in **Error!**
Reference source not found..

Some of the follow on points and open questions arising from the workshop include:

1. Next stakeholder meeting should be held in Europe, but not in the UK, as the first meeting was hosted there.
2. Need to engage with stakeholders more widely across Europe. (Note that actions to address this point are included as part of the Deliverable 800.1: Communication and outreach strategy.)
3. What will happen after the project ends? A successful system should become self-sustaining, especially if it is adopted across the climate research community as an open source tool.



Figure 2. Photo of delegates in the final plenary.

Annex 1: Workshop programme



CHARMe user workshop, 14th March 2013, Reading, UK

Agenda

09:30	CHARMe project team meeting
10.30	Stakeholder arrival and coffee
11:00	Introduction to CHARMe project and Commentary metadata
11.30	First breakout session: Judging a dataset's fitness-for-purpose
12.30	Feed back to group
1:00	Lunch
2:00	Introduction to CHARMe system and applications
2.30	Second breakout session: Using Commentary metadata
3.30	Feed back to group and wash up
4.30	Close

Annex 2: Workshop attendance list

Stakeholders are marked in italic text, while CHARMe project researchers are in plain text.

Raquel Alegre (UREAD)
Helen Beadman (UKEOF)
Jon Blower (UREAD)
Ruth Boumphrey (UKSA)
Sarah Callaghan (STFC)
Hannah Clarke (Vega)
Debbie Clifford (UREAD)
Francis Colledge (UK Met Office)
Dimitra Founda (NOA)
Petra Fuchs (DWD)
Phil Harwood (CGI)
Lewis Keen (Magellium)
Phil Kershaw (STFC)
Tom Lankester (Infoterra)
Bryan Lawrence (STFC)
Jane Lewis (UREAD)
Kevin Marsh (STFC)
Jan Willem Noteboom (KNMI)
Alan O'Neill (UREAD)
Maria Pahoula (SIH)
Rhona Phipps (UREAD)
Baudouin Raoult (ECMWF)
Roger Saunders (UK Met Office)
Nathalie Selbach (DWD)
Garin Smith (Magellium)
Ian Sykes (Magellium)
Paul van der Linden (UK Met Office)
Spiros Venturas (STFC)
Ge Verver (KNMI)
Joanne Walker (ESA)
Emma Woolliams (NPL)

Annex 3: CHARMe scope and rationale

This document is intended to provide the background information required to set the scope for the CHARMe project. CHARMe will build a system that highlights to users the existence of pieces of information that inform their judgement of whether a dataset is fit for their purpose. Often, the information that users need to judge whether a dataset is fit for their purpose is not discoverable in the same search that discovers the dataset itself. The information may come from different sources and may change independently of the dataset (e.g. journal articles, reports of validation campaigns). This set of information (what we term ‘Commentary’ metadata) has not been significantly addressed by previous efforts to standardise metadata capture and discovery.

Scope of the CHARMe system

The CHARMe system is intended to allow users to discover and enter Commentary metadata. The system does not guarantee access to that piece of information (e.g. a published paper, the contents of a data policy), but it should show a user that the information exists, and how to obtain it. It also does not offer an assessment of the fitness-for-purpose of a dataset for a particular application (a ‘quality mark’ etc). The exception to this is found in the applications in WP7: here, the example applications that we develop will show how these functions can be built on top of the CHARMe system.

Three types of requirement are recognised within the CHARMe project:

1. Requirements on metadata for a user to judge a dataset’s fitness for their purpose
2. Requirements on the system that holds, links to, serves that metadata information
3. Requirements for an application that uses the metadata in some way, e.g. to help the user make a value judgement on a dataset

Types of metadata

The general term “metadata” is very broad and encompasses many different types of information. We follow the terminology of Lawrence et al (2009⁴) and classify metadata into the following categories: Archive, Browse, Commentary (Character in the original paper), Discovery and Extra.

- **Archive** metadata is required for almost all usage of the data. It encompasses description of the spatial and temporal referencing, the units of measure and the physical/biological/chemical phenomena being measured. The Climate and Forecast conventions and HDF-GEOMS are well-established examples.
- **Browse** metadata is used to distinguish between datasets in the context of browsing through a catalogue of data. It encompasses information about the platform or sensor, measurement mode and the general geographic area to which the dataset pertains. For climate data, Browse metadata is addressed systematically in the Metafor and ESA

⁴ B.N Lawrence, R Lowry, P Miller, H Snaith and A Woolf. Information in environmental data grids, Phil. Trans. R. Soc. A 13 March 2009 vol. 367 no. 1890 1003-1014

Heterogeneous Mission Accessibility projects.

- **Character/Commentary** metadata, to be defined below
- **Discovery** metadata consist of high-level descriptions that a user might employ in a web search and encompass parts of other metadata types. The Directory Interchange Format and ISO19115(-2) are well-established models for Discovery metadata and are applied in many projects including the GMES core services (e.g. MyOcean).
- **Extra** metadata are specific to a particular location, campaign or discipline and may consist of arbitrary notations or documents, often in ad-hoc formats.

At proposal stage, we defined the scope of Commentary metadata to include:

1. Post-fact annotations, e.g. citations, ad-hoc comments and notes;
2. Results of assessments, e.g. validation campaigns, intercomparisons with models or other observations, reanalysis, quantitative error assessments;
3. Provenance, e.g. dependencies on other datasets, processing algorithms and chain, data source;
4. Properties of data distribution, e.g. data policy and licensing, timeliness (is the data delivered in real time?), reliability;
5. External events that may affect the data, e.g. volcanic eruptions, El-Nino index, satellite or instrument failure, operational changes to the orbit calculations.

An alternative list of information required to judge 'fitness-for-purpose' was drawn up in the CEMS project data integrity study:

Data quality – based on the results of routine Quality Control activities;

Data accuracy – based on more complex data evaluation and comparison exercises;

Data availability – considering the length, coverage and operational scenario of each mission and the performance of data production within the ground segment;

Data accessibility – considering data policy issues and the practical implications of accessing data;

Data provenance – considering the origin, evolution and status of the algorithms and processes applied to the data;

Quantitative error/uncertainty – derived through error modelling and ideally provided on a pixel by pixel basis, as an integral part of the data set

A further perspective on the usability of climate data is provided by NOAA's maturity index, first proposed for satellite Climate Data Records in 2006 by Bates and Barkstrom⁵.

⁵ Bates, J. J., Barkstrom, B. R., 2006: A maturity model for satellite-derived climate data records, manuscript P2.11, proceedings of the 14th Conference on Satellite Meteorology and Oceanography, 86th AMS Annual Meeting (Atlanta, US) 28-30 January 2006.

Recommendations and feedback from international working groups and space agencies were included in a version-controlled Maturity Matrix V4 by NOAA's Climate Data Records Program, and this version was published by Bates and Privette in EOS Transactions AGU in the Fall of 2012 (DOI: 10.1029/2012EO440006). The maturity index addresses the following:

1. Reference – Where can products easily be found?
2. Provenance – What original observations were used in the product?
3. Context – What methods were used to create the product?
4. Fixity – Digital signature is needed to prove authenticity of product
5. Access rights – Are the data publicly available?

The maturity index is being applied to NOAA CDRs, Eumetsat CDRs and ESA CCI.

Are all of these categories listed above C-metadata? What is a useful definition of C-metadata?

The following concepts may be useful:

1. Fragility – can the metadata change without implying an essential change to the data itself?
2. Timing – can this piece of metadata be determined when a dataset is 'first published' or is it only through later assessments?
3. Externality – can this piece of metadata be defined by someone other than the original data producer? Can it be held separately from the data?

Annotation on the metadata itself (any category) would also be C-metadata, e.g. corrections to the geographic scope as recorded in the A-metadata.

What is in scope for CHARMe?

Many things users are interested in do not come under C-metadata. E.g.

Spatial and temporal coverage – this is D or A

Provenance – B?

Data policy – D?

Certain non-C types of metadata can be chosen to be within scope for CHARMe where:

- The metadata is clearly required by users

and

- The metadata is currently not easily accessible to a user when they search for a dataset.