WMO/IOC/UNEP/ICSU
GLOBAL CLIMATE OBSERVING
SYSTEM (GCOS)

Doc. 3.5
(15.II.2013)

5th GRUAN Implementation-Coordination Meeting (ICM-5)
De Bilt, Netherlands
25 February – 1 March 2013

Session 3

GRUAN Metadata Task Team Terms of Reference (DRAFT v4)

(Submitted by Peter Thorne)

Summary and Purpose of Document

This document constitutes the basis for the metadata task team. The annex provides a detailed discussion of what is meant by the term 'metadata' in a GRUAN context and should be considered (and amended) by members of the task team on a regular basis.

GRUAN metadata task team terms of reference

7th February 2013

DRAFT

This document constitutes the basis for the metadata task team. The annex provides a detailed discussion of what is meant by the term 'metadata' in a GRUAN context and should be considered (and amended) by members of the task team on a regular basis.

Task team roles and responsibilities

Under the auspices of the Working Group on GRUAN (WG-GRUAN), and in liaison with the Lead Centre and remaining task teams the metadata task team has been established to:

- Identify the information that is needed to allow the majority of GRUAN data users to use GRUAN data products in appropriate contexts and in a defensible way.
- Identify potential sources of metadata for specific measurement systems and for GRUAN sites as a whole.
- Create the GRUAN Core Metadata Standard, and a mechanism for maintaining that standard, that allows the essential information to be exchanged unambiguously.
    - Develop and implement protocols to ensure that metadata collection and archiving is done consistently (e.g. including consistent naming of metadata fields) across all measurement programmes within GRUAN. This will likely require liaison across a number of task teams.
    - Implement within the GRUAN Core Metadata Standard a method for providing users with any necessary additional indication of the quality of the data in addition to the measurement uncertainties provided with every GRUAN datum.
    - Regularly investigate new metadata sources, information management technologies and information sharing capabilities.
    - Ensure a realizable and resilient metadata retention and storage architecture that enables long-term association of GRUAN data with the relevant metadata.
    - Ensure that the GRUAN Core Metadata Standard is consistent with and inter-operable with the congruent WIGOS metadata standards which are under development.
- Liaise with measurement system task teams to ensure that the necessary metadata are captured as part of the data collection process.
- Review Technical Documents prior to publication to ensure that relevant consideration has been given to aspects pertaining to metadata in the measurement specification.
- Undertake the preparation of a Technical Document that describes the instrument-independent metadata that are required to be archived by sites. This should include photographic, geodetic and any other information deemed necessary and specify the frequency with which such data must be archived. These metadata need not be entirely machine-readable but must be readily available to help interpret the data streams from a given site.
- Regularly review metadata content for confirmation and accuracy to the extent that task team resources permit.
- Regularly review the use of metadata by the GRUAN user community and ensure that all metadata required are being provided in an appropriate and consistent way to fully meet the needs of the various user communities.

- Liaise with the task team of site representatives, Lead Centre and WG-GRUAN to ensure that metadata requirements are acheivable, affordable and practical for the field-based nature of the GRUAN network. In particular:
    - Work to ensure that metadata requirements as stipulated are realizable within the GRUAN network given site resources before their roll-out.
    - Provide help and support to individual sites on an on-request basis from any of the WG-GRUAN, Lead Centre, site representative or task team of site representatives to address perceived metadata collection and retention issues arising.

Membership, reporting and governance

- The team is led by two co-chairs at least one of whom is a member of the WG-GRUAN.
- The team shall include at least one representative from each task team with a specific measurement system focus.
- The team shall include at least one representative from the WIGOS metadata task team to be nominated by the Commission for Basic Systems.
- The task team co-chairs shall report on a six-monthly basis in February and August to the WG-GRUAN a brief written progress report.
- At least one task team co-chair, and potentially one additional task team member, shall attend as deemed appropriate and affordable the annual implementation and coordination meetings to report in person on progress.
- The task team is expected to respond to all reasonable formal requests for advice from the WG-GRUAN, Lead Centre, other task teams or sites made on an ad-hoc basis in a timely manner.
- Reports to the WG-GRUAN, or any kind of recommendation or publication issued by the task team, need to be approved by at least the co-chairs and all core members OR their alternates.
- The task team is expected as its first deliverable to create a work plan consistent with the expectations laid out in the terms of reference and GRUAN Implementation Plan that is realisable given available resources and prioritizes activities accordingly. This timeline need not be bound by the 2017 time horizon of the Implementation Plan.

Operation

- Means of communication are primarily electronic (email, blog, wiki or other as deemed necessary by task team members) and telephone-conference. The telephone-conferences will be organized by the co-chairs at least twice a year. To undertake in-person meetings, the task team should take advantage of other meetings where a sufficient number of members are in attendance.
- Task team chairs will seek funding for dedicated meetings if deemed appropriate.
- The task team will exist until such time as its duties are deemed to have been completed by the WG-GRUAN.
- Task team terms of reference and membership will be revised at the very latest two years from the date of this version, or earlier if requested by either party, by the task team members in consultation with WG-GRUAN.

Annex

Background

Collection of metadata within GRUAN measurement programmes, and provision of those metadata to data users, is the foundation stone for the successful operation of GRUAN. Metadata are essential for reprocessing historical raw measurements to create updated GRUAN products. Such reprocessing is usually required after some change in instrumentation, improved understanding of the observations, or change in some other aspect of a measurement system. In all cases, comprehensive metadata must be available to tie the measurements, perhaps made using several different instruments operating over the lifetime of a site via a comparable traceability chain back to the same recognized standard. It is particularly important that metadata identify any sources of uncertainty that could not be quantified when instrument systems were changed. Metadata collection and archiving is therefore essential to change management within GRUAN.

It must be possible to completely recreate the entire measurement process from the metadata available. Since it is not always known in advance which metadata are likely to be required for reprocessing at a later date, collection of metadata should err on the side of collecting and collating as much metadata as possible even if no immediate use for those metadata can be envisaged.

Care must be taken to guard against collection and retention of incomplete or incorrect metadata. Incomplete, outdated, or inaccurate metadata can be as detrimental, indeed in some cases worse, than no metadata at all.

In particular a distinction must be made between more ‘static’ metadata that may pertain to a given site or measurement type for an extended period of time and ‘transient’ metadata which are unique to a given individual measurement or restricted set of measurements. Different protocols will be required for these two classes. A further distinction is between measurement stream metadata and more generic site metadata which may include such aspects as the geographic configuration of the instrument suite at the site, its locale and changes within a reasonable proximity through time.

The availability of sufficiently detailed metadata is also a vital component in evaluating measurement uncertainty budgets as detailed in Immler et al., 2010[1]. The more detailed the metadata the 'deeper' the measurement uncertainty can be diagnosed. The approach that should be followed is that a reference standard, application of an operating procedure, use of a data processing algorithm, and all evaluated uncertainty sources must be available through the metadata tagged to that measurement.

Complete metadata should include a full account of the operation of the site from as early as possible but at the very least from the date of its certification as a GRUAN site to the present. Measurements from the period prior to site certification may be able to be processed as GRUAN data if sufficient metadata can be associated with them and submitted to the data processing centre.

---

[1] Immler, F. J., J. Dykema, et al. (2010). "Reference Quality Upper-Air Measurements: guidance for developing GRUAN data products." Atmospheric Measurement Techniques **3**(5): 1217-1231.

Management and maintenance of metadata requires the investment of resources. Present day technology for database warehousing of digitized metadata has the added benefit that metadata can be accessed, linked to measurements, and easily transferred. Metadata needs to have the same level of commitment as observed data. Given the relatively small network size and the attention to detail paid by sites to their measurements, these metadata goals should be achievable in a way that is currently impractical for much of the remainder of the Global Climate Observing System (GCOS). Given GRUAN's intended role as a reference standard this is entirely appropriate. That said, metadata requirements should still be stipulated and collected in as cost-effective a manner as possible to avoid unacceptably burdening the GRUAN sites. Furthermore, the benefits must be clearly stipulated.

Finally, GRUAN metadata should include all information related to acknowledgements and/or co-authorship on publications making use of the data. Certain sites will require specific acknowledgements as detailed in the GRUAN data policy guidance[2].

Metadata sources

Potential sources of metadata might include, depending upon the nature of the data themselves:

- Instrument logbooks, including all maintenance actions on instruments.
- Manufacturer's specifications of instrument characteristics.
- Data provided in calibration and other certificates.
- Key steps in instrument calibration procedures.
- Uncertainties assigned to reference data taken from handbooks.
- Key steps in the measurement process.
- Data processing algorithms, including computer source code.
- GRUAN measurement system change evaluation reports.
- GRUAN measurement system change impact reports.
- Laboratory studies of measurement systems.
- Previous measurement data.
- Experience with or general knowledge of the behaviour and properties of relevant materials and instruments.
- Ancillary observations that provide an environmental context for primary data e.g. a pointer to cloud cover observations when surface radiation is the primary data.
- Outcomes from instrument intercomparisons.
- Instrument operators. Metadata should include codes (not names to protect the privacy of operators) to denote where different operators have been responsible for measurements.
- Geo-tagged and time stamped digital images of the instrumentation used, the measurement environment, and the measurement site. Pictures may capture information not initially considered to be relevant but later found to be useful in assessing causes of changes.
- Feedback files of observation minus background from NWP centres.

[2]http://www.wmo.int/pages/prog/gcos/documents/GRUAN_Data_Policy_(Feb2009).pdf

- Metadata should not preclude information derived from historical documents such as observing practices manuals, site inspection reports, government policies, resource and funding programs, even local newspapers

Not all of the above classes will apply to all measurement streams within GRUAN and some may be specific to a given site e.g. only some sites may have access to certain laboratory facilities. The task team will work with the task team of site representatives, WG-GRUAN, Lead Centre and other task teams to define that subset of metadata classes appropriate for GRUAN purposes on a data stream by data stream basis. They will develop and / or review metadata collection tools to ensure that all relevant metadata can be obtained and retained in a cost-effective and realizable manner.

Key attributes of metadata

- Discoverable: Every effort should be made to facilitate the discovery of metadata. The data files providing the GRUAN data products must include a tree of pointers to relevant metadata such that all metadata associated with the data product are fully discoverable. This will likely be most easily achieved if centralized storage of GRUAN metadata is undertaken.
- Breakpoint identification: Metadata need to define where changes to any aspect of a measurement system require potential reprocessing of the raw data. For example at a point where a new batch of radiosondes was first deployed at a site.
- Tight binding: Metadata data need to be tightly bound to the raw data they describe such that access to the metadata occurs seamlessly when the original data are accessed and analysed.
- Complete: For metadata associated with individual measurements, in addition to the measurement uncertainty associated with that datum, metadata such as the exact date and time associated with the datum, as well as the exact altitude, latitude and longitude must be directly available or easily derivable from other metadata. The provision of such metadata recognises the fact that e.g. balloon-borne instruments drift in latitude and longitude during a flight.