

2nd GRUAN Implementation-Coordination Meeting (ICM-2)

Payerne, Switzerland
2-4 March 2010

Item 8.2

Report from the GRUAN Data Management Coordination Meeting, 28-29 Sept 2009 – NOAA/NCDC, Asheville, North Carolina USA

(Submitted by GRUAN Lead Centre)

Summary and Purpose of Document

This document outlines possibilities for GRUAN data dissemination and data management practices. It gives an overview on the Sept 2009 meeting between representatives from the GRUAN Lead Centre, the NOAA National Climatic Data Centre (NCDC), and the US Department of Energy's Atmospheric Radiation Measurement (ARM) program Climate Research Facility (ACRF) to discuss the distributed process that will be employed for GRUAN data management. Specifically:

- Data archiving capabilities at NCDC and ARM
 - Capabilities of the Meta-database (located at the Lead Centre)
 - Development of a data disseminating portal that links the Meta-database and the GRUAN data archive with data users
 - The role of NCDC and ARM in radiosonde QA/QC
 - Organizing data flows from the GRUAN site through a QA/QC institution to the GRUAN data archive
-

Final Version [15 January 2010]

GCOS Reference Upper Air Network (GRUAN) Data Management Coordination Meeting Report 28-29 September 2009 – NOAA/NCDC, Asheville, North Carolina USA

Background: At the first GRUAN Implementation and Coordination meeting (Oklahoma City, March 2009), it was decided that representatives from the GRUAN Lead Center, NCDC, and the DOE ARM Climate Facility (ACRF) Program should meet in the fall of 2009 to discuss the distributed process that will be employed for GRUAN data management. The general areas of discussion would include the following topics:

1. Data archiving capabilities at NCDC and ARM
2. Capabilities of the Meta-database (located at the lead center)
3. Development of a data dissemination portal that links the Meta-database and the GRUAN data archive with the data users
4. The role of NCDC and ARM in radiosonde QA/QC
5. Organizing data flows from the GRUAN sites, through a QA/QC institution to the GRUAN data archive.

NOTE: The meeting presentations are all posted at the following FTP site at:
<ftp://dossier.ogp.noaa.gov/GCOS/GRUAN-DM-Meeting/>

Monday, 28-September

Session 1. Lindenberg Data Management Plans and Expectations, Presentations by Holger Volmer and Michael Sommer

The session started with a brief review of the Altmann Laboratory history, operational practices, capabilities, and other program involvements. The review also covered the GRUAN focus on reference observations for research purposes, the mandate to provide long-term high quality climate records, initially of the priority 1 variables (P, T, H – UT/LS), the need to constrain and calibrate data from more spatially-comprehensive global observing systems (including satellites and current radiosonde networks), and the requirement to fully characterize the properties of the atmospheric column.

The challenges of assuring data continuity in the face of inevitable instrument changes and factors such as instrument performance changes with altitude were discussed. It was emphasized that GRUAN must quantify measurement uncertainty and distinguish between systematic and random errors.

Particular emphasis was given to the characteristics of a measurement that uses an instrument to compare the measured quantity against a recognized standard. A measurement must be traceable to that standard, but a reference measurement also specifies the quality of the comparison against the recognized standard, i.e., the measurement uncertainty. It also gives the best estimate for the quantity to be measured, not just the instrument error, but also the total measurement system uncertainty. The point was emphasized that in order to provide a reference standard, GRUAN observations must include the measurement uncertainty.

The GRUAN approach to these requirements was summed up as follows: **Analyze** all sources of uncertainties; **Synthesize** one uncertainty estimate for each data point; and **Verify** this estimate in comparisons. Of course, there is the need to distinguish between systematic and random errors, correct for systematic errors, quantify remaining random error, and estimate total measurement uncertainty as well as possible. The sources of uncertainty in taking measurements were reviewed in detail, as well as methods of testing these sources through laboratory work, ground checks, in situ comparisons, etc. Data archiving needs to capture all these uncertainties and to capture metadata needed to keep track of data changes and be able to communicate those changes to users.

The session concluded with a discussion of data management plans from the Lindenberg perspective. Points covered included the need for long-term stability and reference quality measurements and the requirements this places on GRUAN data processing practices. The characteristics of the required engineering and raw data as well as the needed metadata were discussed in detail. Plans for dissemination of data and the monitoring of network performance were also covered.

The discussion also touched upon the likely need for the metadata to be stored (and controlled) at Lindenberg, the establishment of an eventual single GRUAN portal, and the need to develop a “branding label” for GRUAN data that allows users to understand the value and contents that have gone into it – e.g., uncertainty estimates. The point was also made that it will be necessary to register users so that “recall” notices can be sent when errors in the GRUAN data are discovered. The Lindenberg staff also indicated that they would also be satisfied with the idea of NCDC providing simple (e.g., FTP) access to the data initially, and the NCDC staff agreed that this was probably quite realistic given their other data management commitments.

Session 2. NCDC Data Management Plans and Expectations for GRUAN, Presentations by Russ Vose, Matt Menne, and Imke Durre

NCDC representatives noted their commitment to the WMO policy on the exchange of meteorological and related data and products, and expressed their desire to assist GRUAN in whatever way they realistically can with data handling and dissemination. They did, however, state that their contributions might be constrained by mission

and budget realities. As the world's largest climate data archive (land, oceans, satellites, etc.) is maintained by NCDC, providing access to data in as-received state should present no problems, but any more sophisticated role may be problematic.

NCDC would be particularly interested in the radiosonde data as NCDC already maintains and handles most quality control issues for the Integrated Global Radiosonde Archive (IGRA) that consists of radiosonde and pilot balloon observations at over 1500 globally distributed stations. IGRA access and update procedures were outlined in view of potential GRUAN support activities.

Issues that will have to be defined include identifying the GRUAN users and determining what they will want from the archive. That will define the user interface. It is probable that GRUAN will not be catering to simple user requirements, but will need some decision-making tools that allow users to pull data (data deli approach). The possibility of starting with the commonality of measurements/instruments at the current GRUAN sites and then adding in their unique measurements was raised.

The role of GRUAN in the WMO organizational and data management structures following the revised 2008 WMO strategic plan was reviewed. It was noted that GRUAN has been endorsed by the AOPC and that GRUAN has been accepted as a pilot project under the WMO Integrated Global Observing System (WIGOS) framework. This should lead GRUAN to develop operational procedures in consultation with appropriate CBS and CIMO expert teams, GSICS and other relevant partners. As part of the vision for the Global Observing System in 2025, it is possible that endorsements from the Commission on Basic Systems may lead to some limited funding as well as support expertise.

Session 3. DOE Data Management Plans and Expectations for GRUAN,
Presentations by Raymond McCord, Randy Peppler, Mike Ritsche, Ken Kehoe, and Doug Sisterson

The role of sonde mentors to provide highly specialized performance monitoring in the DOE system was reviewed. The mentors are responsible for data from beginning to end. Specific areas include establishing and monitoring procedures such as prelaunch and launch routines, performing daily data quality checks, ensuring calibration of all critical components, and the collection of metadata and historical documentation. The discussion covered aspects of quality control and procedures for flagging questionable data. The biggest contributions have been in long-term trend and problem analysis, continuous improvement of procedures, and quality of metadata.

The value of ground checks was covered including the question of whether they may occasionally degrade the quality of soundings. An example of problems that may be introduced was the testing of the quality of the desiccant rather than the sonde itself when doing ground humidity checks. Such an event would necessitate adding a

correction into the error bars. The point was also made that a basic underlying issue potentially affecting sounding quality is the amount of human interaction involved in radiosonde operation.

ACRF Data Review and Documentation procedures were reviewed, and it was pointed out that strong points that underlie the ACRF quality assurance process are the end-to-end nature of the system, the availability of data and related quality information to anyone, and the ability to ingest data from other programs.

The structure and operations of the ARM data archive were also reviewed. These include the differences between the continuous data and special data characteristics, and the archive design around a “cluster” of systems with each focused one “role” and mostly operating in a parallel manner (extensible) with the database serving as the central connection. The ARM data include raw data files that are available upon request but not accessible from the user interface, are accompanied by minimal documentation, and are provided in a wide variety of formats, many of which are binary. As opposed to the raw data, processed data files are accessible from user interfaces and come in common formats including NetCDF and HDF. In addition to the data, value added products (VAPs) such as advanced algorithms, multiple data inputs, and inputs from long-time periods are available. ARM produces some VAPs to improve the quality of existing measurements. In addition, when more than one measurement is available, ARM also produces “best estimate” VAPs.

The Session concluded with a discussion of challenges that will be faced in operating the GRUAN archive. These include questions such as the process to use in combining data of different dimensionality, the collection and review of metadata, how to “brand” GRUAN data so that users understand the homogeneity, traceability, and connection to references/standards. Other issues include identifying and understanding the user community, deciding how to “roll up” metadata, i.e. how to QA the information as measured values are “rolled up”, and the establishment of measures of performance.

Tuesday, 29-September

Session 4. Use of the Global Observing Systems Information Center (GOSIC) as a tool with respect to GRUAN data access, Presentation by Christina Leif

The history of GOSIC was briefly reviewed. Established in 1997, GOSIC is a common point of access for climate data from the Atmosphere (GCOS), Land (GTOS), and Oceans (GOOS). The GCOS Joint Data and Information Management Panel (JDIMP) established GOSIC in 1997. Development work was completed at the University of Delaware and GOSIC was transitioned to operational status at NCDC in 2007. Links and information are kept up to date and are quality controlled with input from the secretariats and the World Data Centers.

The current breakdown of domain types using GOSIC is 60% academia, 20% government (includes foreign), and 20% military. The GOSIC cooperates with NASA's Global Change Master Directory (GCMD) to provide metadata for the global observing systems datasets via a portal on the GCMD. In addition to documentation of datasets and program elements, GOSIC can provide value added products, data matrices, data flow diagrams, and several search mechanisms. The GOSIC currently supports regional observing systems (develop web pages, data flow diagrams, etc., and links) and has links to GRUAN information, but could also do a central GRUAN web page at a higher level if that proves desirable. See GOSIC page at <http://gosis.org>.

Session 4. Brief discussion on issues related to the launch of the cryogenic frost point hygrometers (CFH) at the ACRF SGP site for GRUAN

The transition of RS-80 to InterMet sonde to interface the CFH has no impact on CFH or ground receiver. Only a simple receiver (not a DigiCora) is needed. Also, the CFH is not a fully operational instrument, but in transition from being a scientific instrument to becoming an operational instrument, ACRF can do the ingest and simple level quality checks and then pass the data on to the Lead Center for the detailed analyses. There is a need for an independent surface reference for the CFH, but nothing has yet been determined or evaluated. It might be possible to use SURTHREF for the CFH as well as the RS-92, but probably not. Another possibility would be the 100% RH method at Lindenberg, or to do what we do for RS-92s. The question was raised about flying ozone sondes with the CFH. They would be optional and convenient to launch with CFH, but are not necessary at this time. The focus should be on the CFH.

Points of Agreement:

1. ACRF will go ahead with its own plans for RS-92 baseline check improvements (SURTHREF). The Lead Center likes the idea of a 100% RH check (they have a simple one-point method). Combined with SURTHREF, this gives a two point check. Mike Ritsche will do what's best for ACRF sites and not an issue for GRUAN.
2. ACRF will assume CFH "ownership" (once per month flight with our RS-92) up to the point: QA/QC that can be automated (not visual inspections or VAPS) will be done by ACRF. Then, through the Archive, the data will be sent to the GRUAN Lead Center for further visual inspection and error bars on the data.
3. Ozone sonde remains an option that can be flown with CFH and RS-92 (CFH with InterMet sonde interface can interface with ozone sondes as well). GRUAN would like to see this done, but it not a requirement. ACRF will follow its new instrument/data stream recommendation process of vetting through the Working Groups (with regard to monthly ozone sondes).

4. Safety issues: ACRF's participation with CFH and ozone sondes requires FAA approval for the larger balloon and CFH and RS-92 sonde launches. (Parachute needed?)
5. Howard Diamond will provide all GRUAN incremental funding for materials needed for CFH, ozone sondes, and any other GRUAN requirements for U.S. GRUAN stations, as budgets allow. Initially in 2010, this will only be able to be done at the DOE ACRF site at Southern Great Plains (SGP) in Oklahoma. The GRUAN operations at the SGP site will serve as a prototype for establishing operational procedures and protocols for other U.S. GRUAN stations, and possibly as a pathfinder activity for other national GRUAN stations to benefit from.
6. Key configuration files for RS-92 DigiCoras need to be determined. (There are settings to input surface temp and RH or use sonde values.)
7. Mike Ritche (ACRF CFH Mentor) is single point of contact.
8. ARM should not take ownership of the data stream at this time. The ARM archive facility at Oak Ridge (e.g., Raymond McCord) will manage through delivery but not into value added activities/ i.e., ARM is first level.
9. QA/QC – ARM does not have other instrumentation to use for validation.
10. GRUAN has no ground check for CFH at this time
11. Issues remain related to what it really means to be reference site and what ground check procedures will be used, e.g., 100% RH check as opposed to ambient RH check. ARM sites operate differently in some respects but this should not be an issue with what is being done for CFH. Dual sonde launch information must be included in the archive, and when a CFH flown is with RS92, it must show that data come from both.

Wrap-up Session. Suggestions and discussions for path forward.

- Lindenberg will work with NCDC (Russ Vose) on GRUAN data flow with respect to NCDC data dissemination (possibly via IGRA).
- RS-92 data from all GRUAN sites will be sent to Lindenberg first, put into the ACRF NetCDF format, and then sent to ACRF DMF for ACRF standard processing. (Proposed -- details need to be resolved)
- ACRF will contact Lindenberg when a data problem is observed, but the communications will have to be determined (DQPR, DQRs, etc.).
- While ACRF will pass the processed RS-92 data on to Lindenberg, Lindenberg would add the error bars and synthesize the data with other GRUAN related data (e.g., Raman Lidar); then send it to NCDC for official posting as a GRUAN dataset with access via GOSIC.
- Michael Sommer (Lead Center) and Christina Lief (NCDC) to collaborate with respect to utilizing GOSIC for GRUAN data access.

- Establish a GRUAN RS-92 Data Registry for GRUAN sites using ACRF (e.g., NetCDF template) – perhaps GOSIC could assist in that effort (registry forms).
- Jeff Arnfield (NCDC) to work with Lindenberg on metadata definitions and standards for GRUAN. Once established, GOSIC to work with GCMD on loading the information into the database.
- GOSIC to create a GRUAN Portal on NASA Global Change Master Directory (GCMD) – Action to Christina Lief.
- ACRF to work with Lindenberg on providing NetCDF file templates for sondes, and eventually all other GRUAN datasets.
- Lindenberg to develop Data Flow Diagram that includes the above assignments. (suggest using diagram from Michael Sommer's presentation)
- Acknowledge joint partnerships of DOE, NOAA, and DWD.
- Develop a GRUAN DM Plan that documents this meeting and our path forward in time for the GRUAN ICM-2 Meeting in March 2010. Could be a report of this meeting in Asheville with components documented here by February 2010. Howard Diamond will take responsibility for first draft, and NCDC, ACRF, NCAR (Steve Williams), and Lead Center can iterate on that draft. The Final version of the report will be submitted to the sponsors for concurrence in lieu of developing a formal MOU.

Appendix I - Participants

DOE/ACRF

Randy Peppler

Ken Kehoe

Doug Sisterson

Raymond McCord

Jimmy Voyles

Mike Ritsche

Lindenberg Observatory

Michael Sommer

Holger Vömel

UCAR

Steve Williams

NCDC and NOAA

Russ Vose

Imke Durre

Matt Menne

Neal Lott

Tom Peterson

Stuart Hinson

Dave Easterling

Christina Lief

Jeff Arnfield

Bill Murray

Howard Diamond

APPENDIX II – MEETING AGENDA

FINAL DRAFT AGENDA (22-September-2009)

**GCOS Reference Upper Air Network (GRUAN) Data Management Coordination
Meeting 28-29 September 2009 – NCDC, Asheville, North Carolina
151 Patton Avenue, Room 400 (Federal Building) – PLEASE NOTE NEW ROOM**

Purpose:

At the first GRUAN Implementation and Coordination meeting that took place in Oklahoma City, OK in March 2009, there was a brief side meeting between the folks from the GRUAN Lead Center in Lindenberg, Germany and their colleagues from NCDC and the DOE ARM Climate Facility (ACRF) Program. The Lindenberg group believed that it would be a good idea to come over to NCDC in the Fall 2009 timeframe and meet with their GRUAN data management colleagues in order to ensure that GRUAN employs a consistent data management approach to what will be a distributed process for GRUAN data management.

The general areas of discussion would include the following topics:

1. Data archiving capabilities at NCDC and ARM
2. Capabilities of the Meta-database (probably located at the lead center)
3. Development of a data dissemination portal that links the Meta-database and the GRUAN data archive with the data users
4. The role of NCDC and ARM in radiosonde QA/QC
5. Organizing data flows from the GRUAN sites, through a QA/QC institution, To the GRUAN data archive.

Monday, 28-September

Moderators: Howard Diamond and Russ Vose

Rapporteur: Bill Murray

8:00 am – Visitors met in the lobby of NCDC

8:30 am – Introductions

9:00 am – Presentations from Lindenberg on their Data Management Plans and Expectations for GRUAN

- The Observatory at Lindenberg – Holger Vömel
- Background and Basics about the GRUAN - Holger Vömel
- GRUAN Data Management Plans from the Lindenberg Perspective – Michael Sommer

10:30 am – Break

10:45 am – Presentations from NCDC on their Data Management Plans and Expectations for GRUAN

- Overview of the NCDC Archive – Russ Vose

- Relationship to WMO CBS – Matt Menne
- Options to assist GRUAN – Russ Vose/Imke Durre

Noon – Lunch

1:30 pm – Presentations from DOE on their Data Management Plans and Expectations for GRUAN

- Overview of the ACRF Archive – Raymond McCord
- Overview of ACRF Data Quality – How ARM sondes are processed – Randy Pepler/Mike Ritsche/Ken Kehoe
- Options to assist GRUAN – Doug Sisterson

3:00 pm – Break

3:15 pm – Discussions on developing a consolidated data management approach to GRUAN and initial work on a plan for that from an overarching and more general perspective.

5:00 pm – Adjourn

Tuesday, 29-September

Moderators: Howard Diamond and Russ Vose

Rapporteur: Bill Murray

8:30 am – Recap of Day 1 Discussions

9:00 am – Use of the Global Observing Systems Information Center (GOSIC) as a tool with respect to GRUAN data access – Christina Lief [see GOSIC page at <http://gosic.org>]

9:45 am - Break

10:00 am – Discussion - Developing a way forward for a consolidated data management plan for GRUAN – Detailed technical details.

11:30 am – Brief discussion on issues related to the launch of the cryogenic frost point hygrometers at the ACRF SGP site for GRUAN – Lindenberg Staff, Mike Ritsche, and Doug Sisterson

Noon – Follow-up actions required

12:30 pm – Lunch and Adjournment of Plenary Session

1:30 pm – Individual discussions amongst experts as required

3:00 pm – Optional Tour of NCDC [Sign-Up Sheet will be available]