

MEETING OF THE
Expert Team on WORLD DATA CENTRES
(WMO, Geneva, Switzerland, 4 May 2009)

Participants

- Ed Hare (EC, WOUDC, ET-WDC)
- Kazuto Suda (JMA, WDCGG, ET-WDC)
- Julian Wilson (JRC, WDCA, ET-WDC)
- Anatoly Tsvetkov (MGO, WRDC, ET-WDC)
- Beate Hildenbrand (DLR, WDC-RSAT, ET-WDC on behalf of Kathrin Höppner)
- Séverine Bernonville (DLR, WDC-RSAT, ET-WDC on behalf of Kathrin Höppner)
- Van Bowersox (ISWS, WDCPC, ET-WDC)
- Jörg Klausen (Empa, GAWSIS, chair ET-WDC)
- Geir Braathen (WMO, GAW)
- Oxsana Terasova (WMO, GAW, from June 2009)
- David Thomas (WMO, WIS)
- Kjetil Torseth (NILU, EMEP/EBAS, guest)

MEETING MINUTES

Meeting opened on Monday, 4 May 2009, 09:00

0. Welcome by host (Len Barrie, WMO)

Dr. Barrie welcomed the participants on behalf of WMO and highlighted that ET-WDC is a very important team within GAW. He mentioned that the development of the WMO Information System was an important strategic thrust of the Organization and welcomed the pilot project proposal out of ET-WDC as a valuable contribution to this process. He further announced a proposal by JRC and NILU to re-locate WDCA to NILU.

1. Approval of the Agenda, denoting the meeting Rapporteur

The agenda was approved and Julian Wilson denoted as meeting rapporteur.

2. News from WDCs

2.1. WDCPC

Activities of QA/SAC-Americas and WDCPC at SUNY-Albany could not be sustained; consequently NOAA ARL initiated a one-year contract with Van Bowersox to resurrect these Web sites and associated activities. This initial contract ends on 31 December 2009 and is renewable annually. Web sites for the WDCPC and QA/SAC Americas will be operational within 30 days, pending resolution of issues related to the transfer of the domain registrations to a new host (Dreamhost). QA/SAC Americas is prepared to ship test samples to laboratories that participate in the GAW Laboratory Intercomparison Study, designed to evaluate performance at precipitation chemistry laboratories. The WDCPC, in general, will post active links to networks with well-established data distribution mechanisms (e.g., NATChem, EMEP etc.) rather than duplicate these data; other data will be collected and archived. A major project is to prepare an assessment, which should be ready in approximately 1 year. A major shortcoming in preparing this assessment is the paucity of data available from Central and South America, Africa, and Oceania. Addressing this shortcoming requires capacity-building in these areas, which is a major challenge for the PC-SAG's consideration.

Discussion: Capacity-building is a big task – funding additional labs is a challenge – who is going to pay for it? The PC-SAG does not have adequate funds to support capacity-building at the present time, though several SAG members will be soliciting the interest of laboratory managers in areas where measurements are needed. Where will the WDCPC be hosted and how is it secured? At present, the site will be hosted by Dreamhost, which uses an open-

MEETING OF THE
Expert Team on WORLD DATA CENTRES
(WMO, Geneva, Switzerland, 4 May 2009)

architecture database (Drupal) at a cost ca. \$10 US per month). Data will be secured through monthly backups at NOAA, on Van's local system and at Dreamhost.

2.2. WDCA

A joint proposal by JRC and NILU foresees to move WDCA to NILU by the end of 2009. At NILU, the WDCA is not explicitly funded, however, a large fraction of the data flow is European (presently funded by EUSAAR and HTAP, data handling funded by EMEP) and the additional burden of managing other data is considered small. For regions outside HTAP/UN-ECE, NILU have limited resources for chasing non-reporting stations outside their regions of interest. JRC will continue as interested users and will finally have an opportunity to generate some products for users rather than managing data.

2.3. WRDC

WRDC started in 1964. Primary QA functions are responsibility of submitting agency (NMHI), WRDC performs additional consistency checks. After previous discussions regarding the difficulty of accessing the data, these are now available through their web site as html pages of monthly averages at all sites. WRDC holdings consist of daily data (=conventional data) as well as certain hourly and monthly data (=GAW data). WRDC maintains its JAVA interface to the database to produce graphs and tables – all data are available after registration.

Discussion: If data come from a station that is registered in GAWSIS, WRDC should discuss with data originator if data of higher resolution (in particular, hourly) are available and may be labelled GAW data. The retrieval of data as HTML pages was considered a burden on the data users; WRDC is advised to seek advice on a standard data format and make the data available in the form of data files containing adequate metadata and data **[Task 1]**.

2.4. WDCGG

WDCGG hosts data from nearly 200 stations worldwide, as well as from ships and aircraft; the latter primarily from the Pacific. WDCGG would also like to get GOSAT GHG data. Data coverage increased from 1970s to 2000-2002, however, after this, data submission shows signs of slight decline. Moreover, data submissions often lag behind by more than 2 years. WDCGG has implemented quality checks in an attempt to give user a selection of high quality data. WDCGG aspires to become a DCPC in WIS. They also need to collect data on data users, within limitations of privacy. WDCGG tries to co-ordinate with other efforts, e.g. IAGOS-ERI, with regards to GHG data.

2.5. WOUDC

Acknowledgements go to Ed Carty (programmer), Vitali Fioletev, Johannes Staehlin and others for continuous support of WOUDC activities. WOUDC is ready to add the data usage disclaimer (GAW data policy) to all files. The whole issue of citation and acknowledgement has yet to be resolved – see comments from Aerosol SAG minutes. WOUDC has documented their activities concerning the implementation of the GAW Strategic Plan. Ed highlights the importance for WOUDC of the 'data sponsorship statement' (how data are measured and collected, and by whom; how many times has the data been reprocessed; instrument histories) He raises the question if there was value in introducing tiered archives to separate long-term monitoring data sets from campaign data sets **[Task 2]**. WOUDC continues to produce weekly metadata updates for GAWSIS. He further suggests bringing together all data providers/networks to learn from one another and increase consistencies in how they all treat the data flow **[Task 3]**.

MEETING OF THE
Expert Team on WORLD DATA CENTRES
(WMO, Geneva, Switzerland, 4 May 2009)

2.6. WDC-RSAT

The MoU between WMO and DLR is due to be signed in July 2009. The WDC-RSAT strategy plan is being revised in light of comments to produce final version to WMO in the next weeks. WDC-RSAT serves as a data and service centre for UFS (Schneefernerhaus). WDC symposium (both ICSU and WMO) planned for 2010, financed by DLR **[Task 4]**.

Discussion: What satellite data sets will be used? WDC-RSAT is developing a single access point with both DLR data and data from other data providers. Initial focus is on ozone and aerosols.

2.7. GAWSIS

Since Oct 2008, a concerted effort is made to review and update GAWSIS contents. GAWSIS now accepts bibliographic references, and links back to affiliated networks (currently just those stations that are already within GAWSIS, will be expanded to entire network following negotiations with the networks concerned). Work is in process on metadata intake from NDACC and others. For some time now, GAWSIS has been using Google Analytics tools to inform about user activities.

2.8. NILU

NILU database activities build on two primary architectures, namely EBAS (relational database system, hosting primary and secondary datasets), and ESA-CDB (file collection for Envisat Cal/Val). A third system, NADIR, is used for stratospheric data, but is no longer developed. NILU is involved in a large number of projects, viz.

- GEOmon: provide a single entry point for primary and secondary data with a focus on Europe
- MACC: provide in-situ data for GMES atmospheric service
- EEA/EMEP feasibility studies: definition of responsibilities between two.

HTAP wants all relevant data put together in a harmonised and consistent fashion. The approach is to collect data from lots of networks, input to EBAS, export routines in NASA AMES, NARSTO, NETCDF, HDF. For access, a data usage protocol will need to be signed. EPA will probably pay for a development of GIOVANNI (NASA) to link to this to give access to these data. HTAP data are 2000-2006 for the moment. Heavy users could access the database directly but this is not so easy to organise.

3. ***Introduction to WIS and what we need to do to become part of it***

The WMO Information System (WIS) is presented as an evolution of the GTS, opening the GTS to all WMO programmes including GAW. At the core are a number of synchronised GISCs (central catalogues of information sources) pointing to actual data in the layer outside these, and holding NRT information for 48 hrs. Feeding in to and communicating with the GISCs are diverse national centres (NCs) and/or data collection and production centres (DCPCs). WIS follows a data-centric model rather than the communications-centric model of GTS. Once operational, WIS will end up with multiple data sources for a particular dataset, which is already the case for many data sets. This is good for redundancy/security of data suppliers but recognition for data providers/ world data centres within this system is potentially an issue.

Technically, for search ISO 23950, for metadata ISO 19139 (xml representation of ISO 19115) compliance are the keys to become part of WIS in the first instance. Where the data are available from a database, the 'service' available from the database should be described, where the files are static the metadata describe the file contents. Further integration (with GTS etc.) is possible but more involved.

MEETING OF THE
Expert Team on WORLD DATA CENTRES
(WMO, Geneva, Switzerland, 4 May 2009)

Discussion: Vocabularies: what are the standards in publications for the community? – WIS can accept many vocabularies as long as they are defined. Citation of data: ICSU are involved; is it possible to benefit from the ICSU sponsored assignment of DOI to data-sets (commercial from library world) – WDC-RSAT is the lead among the ET-WDC for this. Multiplicity of data products/versions and data lineage can be handled by the ISO 19139/19115 but there is no mechanism for communicating back up the data chain, which might be a requirement. The WIS office at WMO is ready to help with WIS implementation matters, currently until mid-2011. Presently, few WDCs have all metadata in compliant forms, what to do? First, it is recognized that it will take time for people to learn to write good metadata. For GAW, GAWSIS has already implemented WIS-compliant metadata based on metadata from the WDCs – this can be improved and serve as the ‘gateway’ for the WDCs. Security must be compliant with international law regarding privacy. WIS will allow free text searches, which is the most common search approach of users. WIS allows WDCs to participate by adding a top-layer to the data structures that is easy to update.

Data: Access, usability (Expert Team on Data Representation & Codes), archive, operational support to environmental modelling – models are getting smarter and will increasingly prefer to assimilate radiances and primary values from instruments rather than classical recognisable interpreted data fields. Thus NRT data flows will focus more on the fundamental data records (radiances etc) while the WDCs’ ‘classical’ archiving and dissemination role will be for higher level data.

4. Status and Implementation Plan for ET-WDC Pilot Project

The motivation is to improve access to the data in the WDCs and enhance discovery and usage of GAW data. In Europe, everything needs to be compliant with INSPIRE, too.

The ET-WDC pilot project proposal contains essentially three parts as follows:

Part 1: Implement and expose ISO/WIS-compliant representations of metadata for the data archived in the WDCs.

Discussion: It is recognized that a detailed discussion of the requirements of ISO19115 would be needed for a full understanding of the implications. Basically, everything one needs to make users capable of discovery and retrieval of data is a required field in the standard. As an example, GAWSIS and WOUDC together produce enough metadata to enable a compliant metadata representation. Consequently, following metadata elements are needed as a minimum (for illustration only): Title, URI, data series name, contact person for metadata, contact person for data, start and end times and dates, spatial extent of data set, frequencies of updates of both metadata and data, variable(s) in the data record(s). Data quality is defined very differently in different communities and on this, the standard is not very advanced. Data quality and data usage policy do not have defined vocabularies at the moment. However, documentation of quality-related information is possible if not in a formalized manner. Data formats are lower level info.

David Thomas reports that, as an example, ECMWF took ca. 1 day to make metadata for SIMDAT (pilot project for WMO RA VI) and implement an ISO 23950 compliant interface. This is considered a highly optimistic scenario by the group and certainly does not include the preparation and structuring of the metadata necessary in the first place.

The proposal leaves open the choice of who does it at the WDC and who does it through GAWSIS. **[Task 5]**

MEETING OF THE
Expert Team on WORLD DATA CENTRES
(WMO, Geneva, Switzerland, 4 May 2009)

Part 2: Prototype client tool to facilitate extraction of GAW data across multiple WDCs

Discussion: The idea is to facilitate access and use of GAW data. It is explicitly stated that no metadata available in the original data files will be lost. The consensus is that this part of the project is OK as long as it will be consistent with the known metadata. Opinions vary from 'overly ambitious' to 'not needed'; Ed would prefer that GAWSIS gets all its metadata up to date first. Kazuto is not happy with reformatting data files and not sure that all metadata are in the data files with the current format; he suggests changes to the wording of the proposal draft that are agreed. The decision is to ask Jörg to develop a simple demonstrator for a restricted but useful set of complementary measurements, e.g. in-situ SO₂, O₃, aerosol chem. composition, aerosol physical properties and RH. ET-WDC and the Secretariat will be given a chance to review the resulting prototype before going public.

Part 3: Toolkit providing details of satellite overpasses, trajectories etc. to help increase use of data.

Discussion: Pilot projects in WIGOS will need to deliver something to EC in July 2010. WDC-RSAT is expecting to take 2-3 years so will have to extrapolate as it won't have finished – which is, however, also true for the others. Therefore, pilot projects are advised to report on progress and problems encountered rather than providing a show and tell on finished products.

General Discussion: The proposal needs to include a statement on the costs. Based on information from the Secretariat, roughly 26,000 US\$ from WIGOS trust funds are available, to be split between the 2 CAS pilot projects, and to be used primarily for collaboration and meetings. It is agreed to ask for 13,000 CHF and include in-kind contributions from task leaders in the overall cost estimate. In terms of processing the pilot project proposal, everything is OK as far as WMO is concerned, even though it does not appear in the current list of pilot projects. It is decided to take the opportunity and present the document to of EC WG WIGOS/WIS (by way of David Thomas) who was scheduled to meet during the following days.

5. Progress on WDC-related Tasks in GAW Strategic Plan

All WDCs have responded to the call to document their progress in the web-based tool. These responses are due to be reviewed by the JSC Sub-Group on GAW Strategic Planning.

6. Controlled vocabularies for observational variables, analytical methods, sampling, etc.

There is a need for controlled vocabularies to support search and discovery. Basically, there are two options: to develop and have to maintain a vocabulary ourselves, or to use and suggest expansions of a vocabulary maintained by someone else. CF is mentioned, but it is also recognized that their efforts are modelling driven. Current vocabularies in GAWSIS are based on the data centres contents. They represent a first level of harmonization, but are clearly incomplete and in no way standardized. HTAP is mentioned as requiring CF compatibility eventually. **[Task 6]**

7. Recommendations of ET-WDC regarding its chair 2008-2010

The work of the present chair is acknowledged and a recommendation for continuation is made by the group.

8. Any other business

WORLD METEOROLOGICAL ORGANIZATION

MEETING OF THE
Expert Team on WORLD DATA CENTRES
(WMO, Geneva, Switzerland, 4 May 2009)

None.

Meeting closed on Monday, 4 May 2009, 17:59

Attachment

MEETING OF THE
Expert Team on WORLD DATA CENTRES
(WMO, Geneva, Switzerland, 4 May 2009)

LIST OF ACTION ITEMS

Task	Description	Responsible	Due Date
1	To immediately improve access to data, add a button to the WRDC web page for each data set presently displayed in HTML to allow download of these data as a file. Further, consult with expert users in the community and get advice on choice of most suitable data format for exchange of WRDC data. Report back to ET-WDC	WRDC	May 2009 June 2009
2	Please comment on this via e-mail to the group since we didn't really address this question during the meeting. Is there a consensus? What are the consequences? For example, should the campaign data not be considered 'GAW data', and hence not be visible through GAWSIS?	All	May 2009
3	Please expand on this and indicate your view on how to proceed.	WOUDC	May 2009
4	Please keep the members of ET-WDC in the loop during the development of this symposium and in particular, point out how they can contribute.	WDC-RSAT	Ongoing until 2010
5	Provide access to a user-friendly explanation of mandatory fields in ISO 19115 and possibilities to implement ISO19139, ISO23950.	David Thomas, Jörg Klausen	June 2009
6	Look at CF (http://cf-pcmdi.llnl.gov/documents/cf-standard-names/ ; and file://grib2_chem_names_CT.xls sent as e-mail attachment on 13 May 2009) and report suitability for GAW to chair ET-WDC. In particular, review vocabulary on - variable names (chemical and others) - methods of observation / analysis	All	May 2009