# IGACO-O3/UV Data Access Workshop
12 – 14 March 2007, EMPA, Duebendorf, Switzerland

## Meeting Summary

Prepared by A.Mälkki, IGACO-O3/UV coordinator, FMI
J. Kaurola, FMI
Date: 8.5.2007

## SCOPE OF THE MEETING:

- discuss different aspects of data access and availability and
- agree on practical goals and tasks for IGACO-Ozone/UV
  keywords: Integrated, Global, Access

## PARTICIPANTS

Representatives of data providers (ground-based, aircraft, satellite), database managers and maintainers, and data users.

## APPROACH FOR THE MEETING

- Plenary presentations summarise the existing situation and give background to the work from different perspectives.
- Working groups
  - Agree on which aspects of a given title need to be covered.
  - Discuss goals for each of those aspects
  - Define – as much as possible – practical tasks for implementation to reach the goals.
- Results of each working group are discussed in the plenary.

## SUMMARIES OF PRESENTATIONS, Monday 12 March 2007

- Opening – Dr. Peter Hofer, EMPA
  After opening remarks from A. Mälkki, Dr. Hofer introduced the host institution, its history and current activities to the participants.

- IGACO and GAW – Geir Braathen, WMO
  G. Braathen presented some of the current outstanding questions concerning Data exchange in atmospheric sciences. One of the most important activities at WMO is the development of the WMO Information System (WIS). The plan is that WIS will, in addition to exchange of operational meteorological data between national meteorological services, also support scientific programmes such as GAW. In the current version, metadata archives will be open to everyone (open WWW). Those directories (Global Information System Centres, GISCs) will contain links to original databases, such as the GAW World Data Centres. Some data (e.g. weather forecasts) will not be freely available, but instead require and agreement between the data provider and the user.
  The system is currently in definition and development phase, and close cooperation with key persons at WMO is important, so that aspects important to the scientific community will be well covered.
  In the following discussions it was pointed out that with limited resources, it is important that priorities are set right also in the development of WIS.

- Status of IGACO-O3/UV – Anssi Mälkki, IGACO-O3/UV Office / FMI
  A.Mälkki presented the status of IGACO-O3/UV. Main activities in 2006 have consisted of two user consultation meetings and the development of IGACO-O3/UV Implementation Plan, which will contain activities agreed for near future (2007-2010). Again, since IGACO does not have dedicated funds

available, it is important that independent institutes and especially funding agencies agree on the plans and priorities so that resources become available and the plan can be realised.

- Overview of GAW World Data Centres, including GAWSIS – Jörg Klausen, GAW QA/SAC Switzerland/EMPA
  J. Klausen's presentation covered the overall GAW data centre infrastructure and services. Within GAW, different data centres have different ancillary/metadata standards and different formats, partly due to different histories and different user communities. However, the data centre managers do discuss these aspects regularly. GAWSIS can be seen as a precursor to a future GAW data portal. The largest single group of stations reporting to GAWSIS is Ozone (including surface O3). A new feature added to the GAWSIS services is a Google Earth interface, where measurement sites can be visualised over the satellite images. There is lot of potential in GAWSIS, and for the future the challenge is to improve the information content for participating stations. Also, following the development of the WIS system is an important activity.

- Introduction to the World Data Centre for Remote Sensing of the Atmosphere (WDC-RSAT) – Frank Baier, WDC-RSAT/DLR
  F. Baier introduced the WDC-RSAT, the latest addition to the network of GAW World Data Centres. The WDC-RSAT is a World Data Centre in the ICSU (International Commission of Scientific Unions) meaning, with the difference to the GAW definition in the sense that ICSU WDC:s make data available to the world, whereas GAW data centres have by definition global coverage. The ICSU network of WDC:s consist currently roughly 50 centres. WDC-RSAT will, in addition to locally archived data, provide value adding services to the user community: data fusion, synergistic retrieval, assimilation, grid computation infrastructure (in the future), and act as a DOI publication agent. Locally WDC-RSAT archives Level 3 data from certain ESA missions, and provides an extensive link list to other remote sensing data. In the discussion that followed, important issues like version control on non-local data, coverage of alternative products (such as multiple L2 versions of same trace gas, provided by different institutions), and metadata tracking were discussed.

- Introduction of recent activities in WMO World Data Centre for Greenhouse Gases – Yukitomo Tsutsumi, WDCGG/Japan Meteorological Agency
  The most important current activities at WDCGG were presented by Y. Tsutsumi. These include Greenhouse Gas bulletins, improvements on data quality traceability by enforcing the existing guidelines, and temporally smoothing data for filling gaps of observation pauses by analytical methods. A new WDCGG Data Submission Guide is also planned, and updates on services provided by the WWW interface are also underway. A new data format has recently been adopted. This format was reviewed by GAW experts.

- The WOUDC and IGACO – Ed Hare, WOUDC/Environment Canada
  The WOUDC has been operating since 1961, with – naturally – major changes in many aspects during the years. A very important change has been method of data submission: in 1992, 70% of the data were submitted as paper copies, whereas in 2006, 98% of the data comes over ftp connection. WOUDC archives data from 470 platforms. All data, including raw (or Level 0; for some instruments) are archived, but some are behind restricted access. This also allows for re-processing and re-analysis of data if calibration and revision histories are available. An important element is also the "data passport" concept, describing (for each provider) the process from instrument raw counts to archived data. E. Hare also emphasized the importance of contact information (person) for each measurement: close collaboration with originator is key to good quality data. In the discussion that followed, it was noted that a "one-stop portal" disconnects the provider from the user, thus resulting in a possible conflict between requirements of easy access and simultaneous good quality.

- WRDC activity and updates on improved access to archived and current radiation data – Anatoly Tsvetkov, WRDC/Main Geophysical Observatory
  A. Tsvetkov summarised the history, data, and network of the World Radiation Data Centre. The network consists of several different instruments, which makes keeping datasets homogeneous extremely challenging. Sites are typically maintained by National Weather Services, and some of those do not currently report type of instrument, which reflects the complexity of maintaining a good database. This particular problem is currently being discussed with those NWSs. Near future activities at WRDC include updating visualisation software and access to data on the WRDC website.

- A user perspective – Markus Rex, Alfred Wegener Institute, Potsdam
  In his presentation, M. Rex listed – from data user point of view – key factors contributing to scientific outcome when using Ozone data. Those are: easy non-bureaucratic access, easy to find datasets, near-real-time (NRT) availability (for certain applications), and balance between global (low resolution) and high-resolution data. The success of a space mission should, in his view, be measured against scientific progress, which is currently better supported by NASA approach, funding scientific activities (including generation of Level 3 products) and implementing a completely open data policy. It was proposed that in order to reach those goals for ESA missions, maybe IGACO might contribute to the final steps. The data gateway approach (see presentation) was presented as a solution, and the Match project as a user for NRT data. High resolution data (especially high-resolution profiles) are needed to study details of ozone loss/recovery. In the discussion the implementation of a one-stop shop or data gateway was discussed. Data providers (including NASA, ESA, as well as single observatories) need to report data usage to their funding agencies, and thus a completely anonymous portal would still need protocols and reports. At the same time, referring to earlier point by E.Hare, a disconnection between the provider and the user might risk data quality. As a conclusion, stepwise implementation towards the WIS approach was seen as a feasible compromise.

- Future developments: GEO and GEOMon – Aasmund Fahre Vik, GEOMON / NILU
  A. Vik presented the GEOMON project. GEOMON can be seen as “European IGACO” or “European GEO”, as its goals bear strong similarity to those two programmes. GEOMON is a project in the EU 6$^{th}$ framework programme (FP), but continuation in the coming FP is also foreseen. GEOMON will result in a demonstration (prototype) of data management structure, collecting data from multiple sources very much in the way planned as target for IGACO. Some data types, such as in-situ surface data, will not be covered in IGACO-O3/UV. Data management will build on existing infrastructure and data flows as much as possible. This is necessary also due to the very large amount of data sources that are included in the project. In the discussion it was emphasized that due to similar goals of multiple programmes, synergy should be sought, and double work on same issues avoided. It has already been noticed that traditionally there has not been much interaction between different data groups (providers). Building a community takes thus some time. An advantage of GEOMON is that it also provides limited funding for observations, although only within Europe. This is not very common nowadays.

- A satellite viewpoint – Bojan Bojkov, CEOS and AVDC, NASA GSFC
  B. Bojkov introduced an alternative viewpoint to the discussion by presenting the needs as seen from space. The main drivers for satellite Ozone observations are monitoring the recovery of the ozone layer, (future) monitoring of tropospheric ozone, and providing O3 data for UV applications. The total ozone column (TOC) is currently well covered with satellite instrumentation, but unfortunately high-resolution profiling instruments are in danger of being discontinued. Tropospheric O3 measurement is currently not feasible from space, requiring accuracy of 1 DU from the TOC measurement (for reference: 2 % is doable at the moment, thus there is a gap of almost a factor of 10 in the error bars). For satellite in-flight calibration and validation, data sets are distributed, but access in general is not a problem. There is a more pressing need for improvement of content, quality, and continuity of ground-based data. As for the content, meta- and ancillary data provide the knowledge of quality, for quality, reliable error estimates are needed for all data, and continuity is required to cover already agreed

missions until 2020 and beyond. Examples suggestions on how to improve accuracy of Brewer data were also presented.

## SUMMARIES OF PRESENTATIONS, Tuesday 13 March 2007

- What happens to data after submission, case EUVDB – Jussi Kaurola, EUVDB, FMI
  After giving a brief overview and history of the European UV Database (EUVDB), J. Kaurola introduced the data quality flagging procedure that is in place at EUVDB. Examples of erroneous UV spectra were shown, and procedures and categories for quality flagging presented. The policy is that no corrections are made to the data at EUVDB. Instead, quality flagging information is sent to the data provider so that the provider can check and correct the data when appropriate. Flags are always available for users as well. Future planned activities at EUVDB include harmonisation with WOUDC, specifically in formats (of UV data), tools and procedures. In the discussion it was noted that a WMO data-use policy exists (available, for example, through http://www.empa.ch/gaw/gawsis/faqs.asp). As answer to a question on the flagging rules it was explained that even if the rules are pre-defined, they have originally been agreed with the users, and technically re-flagging is possibly, should the rules need to be changed.

- Significance of data protocols: Provider point of view – Esko Kyrö, FMI
  E. Kyrö gave a clear and compact presentation on the purpose and need of data protocols from a provider's point of view. He emphasized that the driver for a protocol would be an incentive for the providers to submit data to a database: "It is a policy declaration which supports timely and regular submission of data and attracts scientifically useful data from the private storages out to the public archives." Essential elements are 1) to explain the mission of the data base, 2) provide guidance for providers and users, and 3) explain access rules and rights. Examples of existing data bases and their protocols were also given. In the discussion it was also emphasized that protocols should encourage exchange of quality information (i.e., feedback from users to providers).

## GENERAL DISCUSSION

As an introduction to the general discussion, the chairman J. Staehelin presented a short summary of the aspects that had been discussed. A step-by-step implementation, with commonly agreed clearly defined goals (derived from user needs) would be the way forward. After the introduction, he asked all of the participants to specify what they would personally do (commit to) for advancing the goals of IGACO. Depending on the background of the person, answers were varied between providing error estimates on sounding data to improving cooperation between networks/databases/projects and further to improving access to existing data. In general, this kind of discussion was found useful, as it requires everyone to think about possible contributions (also in terms of resources) from his/her position.

A separate summary of the participants' commitments will be available.

## GROUP WORK

Topics for Discussion Groups:
- Data access and submission, including protocols and data rights.
- Services and tools.
- Metadata, ancillary data and quality control; Long- and short term aspects.

Tasks for the discussion groups
- Agree on which aspects of a given title need to be covered.
- Discuss goals for each of those aspects.
- Define – as much as possible – practical tasks for implementation to reach the goals.

The outcome of the group work was discussed in the plenary on Wednesday 14 March 2007 and a summary is given in a separate document.