(N)RT data services from GAW

Jörg Klausen (chair ET-WDC)

Øystein Godøy, Vincent-Henri Peuch, Martin Schultz, Ellsworth Judd Welton (Members ET-WDC)

Markus Fiebig, David Gay, Jörg Klausen, Tom Kralidis, Julian Meyer-Arnek, Kjetil Tørseth, Anatoly Tsvetkov, Mikio Ueno (ex-officio Members ET-WDC)

Geir Braathen (WMO Secretariat)

Revision history

2018-04-09	v0.0	initial draft by Jörg Klausen
2018-04-17	v0.1	cross-references to Figure 1 corrected.
2018-04-26	v0.2	comments by Martin Schultz and Øystein Godøy considered
2018-06-25	v0.2.1	comments by Tom Kralidis
2018-06-27	v0.3	changes after ET-WDC telco from 2018-06-26, considering feedback by Judd Welton and David Gay
2018-06-28	v0.3.1	comments by Tom Kralidis
2018-06-28	v0.3.2	comments by Martin Schultz
2018-06-28	v0.3.3	comments by Markus Fiebig
2018-06-28	v0.3.4	comments by Kjetil Torseth
2018-06-28	v0.3.5	comments by Vincent-Henri Peuch merged in (separate branch from v0.3)
2018-06-29	v0.4	consolidation by Jörg Klausen, additions to table on behalf of Mikio Ueno, considering further feedback by Judd Welton, Geir Braathen
2018-09-26	v1.0	version number set to 1.0 after approval by SAG APP

Terminology

- **Real-time (RT) data service**: delivery of data to an end user within three (3) hours of observation
- **Near-real-time (NRT) data service**: delivery within seventy-two (72) hours of observation. The acronym "(N)RT" is used whenever the requirements are either for RT or NRT, or somewhere in between
- **Operational data service**: end-to-end service that is provisioned and monitored on a 24/7 basis with sufficient redundancy to achieve at least 95% availability

Motivation and Objectives

At the origin of this document is a funding opportunity brought by ECMWF in the context of the the Copernicus Atmospheric Monitoring Service (CAMS) to support enhancement of RT data delivery of GAW stations outside of Europe, in addition of efforts by ECMWF to promote RT data delivery from

mostly European programs/networks/research infrastructures such as EMEP, ACTRIS, ICOS and NDACC. The specific use of ECMWF of these data is assimilation in forecast models for atmospheric composition and verification/validation[1]. While there is thus a specific need to answer this request, this document nevertheless takes a more holistic perspective in reflection of the WMO GAW Implementation Plan [1].

In response to this initiative by ECMWF, the President of CAS (Øystein Hov) and the chair of the WMO OPAG EPAC SSC (Greg Carmichael) requested ET-WDC to advise on how to promote and pursue the following:

1. (N)RT data delivery of quality observations from the global GAW network to ECMWF/CAMS [and other entities that have similar interests and capabilities (in particular represented in the SAG-Applications]
2. (N)RT data delivery mechanisms in existence at the WDCs for RG, PM, GHGs and possibly deposition observations
3. further development of the value chain of (N)RT data from the GAW network, with an emphasis on improving weak links in the value chain.

The GAW Implementation Plan defines "science for services" (N)RT as a strategic priority of GAW [1]. To meet this objective, data need to be made available to the user in known quality and with short lead times. As a vision for the future, operational data services (see definition above) are established for atmospheric chemical composition data in the same way as they exist for weather data. Aspects to consider include

- desired data quality,
- data production, processing, and delivery
- data policy,
- data dissemination (discovery, access, visualization)
- archiving of (N)RT data,
- documentation of processes, workflows and data,
- the role of existing GAW Central Facilities, and
- adequate recognition and/or compensation of stations and processing centers engaged in providing or handling (N)RT data.

The primary perspective of this document is the one of the WMO GAW Programme. However, it is acknowledged with appreciation that a number of independent, well established atmospheric composition monitoring programs/networks exist that are not governed by WMO GAW but that contribute substantially to the objectives of GAW. WMO GAW has no mandate to specify the data management of these contributors beyond an invitation to enable interoperability and possibly to adopt the approaches described here.

This paper briefly recalls guidance from the GAW Implementation Plan and the GAW data policy in place. It suggests a number of GAW principles for (N)RT data and briefly addresses data policy and licenses. It uses a diagram (Figure 2) to identify important steps in the end-to-end process from observation to

[1] Verification = automated validation. For time-critical assimilation and verification, RT data are needed, while NRT data are also useful for delayed mode assimilation and validation.

service delivery, as a blueprint for existing and future GAW (N)RT data services, and for comparison with current implementations at NILU (WDCA, WDCRG), Environment Canada (WOUDC), the North American Deposition Program (NADP), and the general provision of weather data by operational weather services, in an attempt to analyze the potential role of WDCs and other data centers. It formulates terms of reference of any center interested in operating (N)RT data services for GAW. Finally, conclusions and recommendations to the attention of the GAW and CAS governing bodies are made.

Background

The GAW Implementation Plan [4, also Appendix] defines the ambitions of GAW with respect to (N)RT [data] delivery services, while maintaining the emphasis on data of known quality as a foundation of GAW:

A-O-8. Support the development of increased capacity to deliver near-real-time (N)RT data and improve their accuracy through establishing standards and best practices, sharing experiences, and training.

A-DM-6. Promote delivery of those variables pertinent to air quality and forecasting in (N)RT, using WMO GTS/WIS as it evolves into an open, decentralized and node-oriented structure. Continue to seize opportunities to expand the (N)RT delivery services for GAW variables.

In support of these ambitions, the SAG on Applications (SAG-Apps) will focus on applications that use (N)RT data delivery on scales larger than urban, including the development of boundary conditions for local modelling and improvement of models and development of services related to dust, volcanic ash and biomass burning plumes, and health applications. In addition, in collaboration with WIS/WIGOS which are ensuring dissemination of observational data, the SAG-Apps will review data systems that ensure possibility of service delivery.

The Terms of Reference of the Expert Team on World Data Centres (ET-WDC) [4, also see Appendix] support this strategic orientation of the GAW Programme in a general way:

ET/WDC-1. Take responsibility for metadata and data management issues in support of the scientific and operational objectives of GAW. [Note: This includes the question of QA/QC of data]

ET/WDC-2. Work with the SAGs, WMO expert teams and partners to establish harmonized data management guidelines, including standardized data formats to allow for adequate (seamless) interoperability.

ET/WDC-3. Guide and support the further development of GAWSIS as the central catalogue of observing facilities and observations supporting GAW, linking the WDCs and Contributing Data Centres.

ET/WDC-4. Keep abreast of and recommend best use of changing technologies [and evolving standards] affecting information management within GAW.

GAW Principles for (N)RT Data

1. GAW RT data refers to data with a delay between observation and delivery of data to the intended user of no more than three (3) hours. GAW NRT data refers to data with a delay between observation and delivery of data to the intended user of no more than seventy-two (72) hours.
2. Instruments used to produce GAW (N)RT data shall be operated according to the principles of the GAW program. In particular, a traceability chain shall be maintained and documented that permits the estimation of measurement uncertainty according to the GUM [2].
3. GAW (N)RT data are not fully quality controlled, but are subjected to automated quality control that is capable of identifying invalid data and suspect data with high confidence while not rejecting an unreasonable proportion of valid data.
4. All automated procedures applied to raw data are fully traceable, publicly documented and implemented as open-source algorithms using open-source implementations as far as possible.
5. (N)RT data are accompanied with sufficient metadata to support the intended use. For internationally exchanged observations, WIGOS-compliant metadata are collected and stored in OSCAR/Surface (or GAWSIS). GAW WDCs handling (N)RT data may impose additional metadata requirements.
6. GAW (N)RT data are generated, processed and distributed by WMO Members in a manner similar to weather data, i.e., generation and processing by the Member and making use of the established distribution mechanisms of the GTS/WIS. Ideally, processes are implemented as "operational data services."
7. By default, GAW (N)RT data are made available to specific (named) users for a specific purpose only. However, data originators are encouraged to any of the Creative Commons licenses [3]. Alternatively, but not recommended, they may require compliance with the GAW Data Policy. **[NB: This principle needs more discussion. There are issues with data use, (long-term) archiving, and re-distribution that need to be agreed.]**

End-to-end Approach to GAW (N)RT data services

For operational data generation, processing and distribution of weather data in the WMO context, the GTS/WIS infrastructure is used. As shown in Figure 1, this is a distributed infrastructure, where Members take responsibility for the generation and processing of observational data and use National Centers (NC) to feed data to a GISC. This infrastructure is already in use for some chemical composition data such as aerosol, total ozone, ozone soundings, and others. IAGOS uses the AMDAR system to push chemical composition data in RT to the GTS/WIS. In future, the NCs shall be used also for surface observations made at GAW stations whenever possible. , as already demonstrated for aerosol and reactive gas observations.

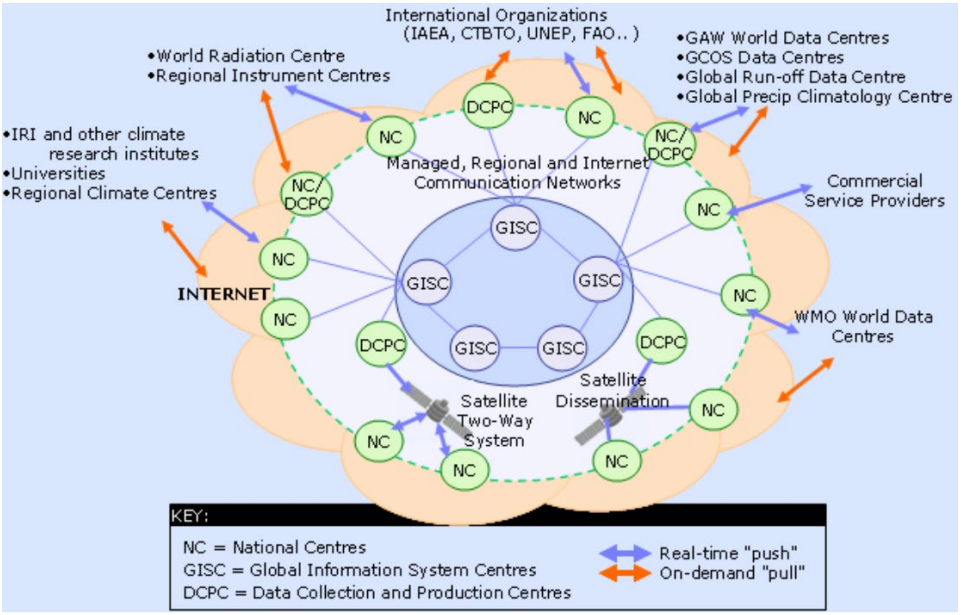
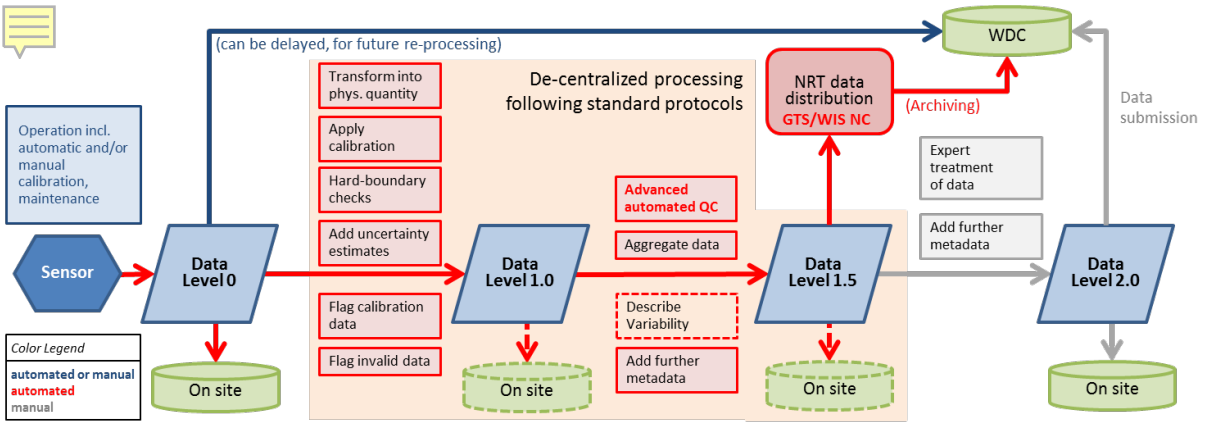


Figure 1. Diagram showing WIS core components and Information Exchange (source: https://www.wmo.int/pages/prog/www/WIS/centres_en.html)

To achieve consistency in the data products generated, it is critical to apply uniform processing of the data and adequate sufficient metadata to document the observations before they are distributed through the GTS/WIS. All instrumental observations undergo a work flow that starts with collection of data from an instrument on site and ends with the delivery of data to a user as shown in Figure 2. An important feature of this figure is the modular, service-oriented architecture that only requires interfaces between the various process steps to be defined but does not anticipate or prescribe any specific implementation. In this way, existing process steps ("modules") can be agreed and further improved within a specific community of practice (such as the WMO GAW Programme) without impacting on the preceding or following steps in the end-to-end chain. Of course, particular implementations may choose to combine several of these modules in a combined work-flow.



Scope	Raw instrument data Auxiliary control data	All Observations (Uncertainties) Data quality flags	Valid observations (Uncertainties) Data quality flags	Valid observations (Uncertainties) Data quality flags
Format IN:	Native (depending on instrument or DAQ)	Unified, self-descriptive	Unified, self-descriptive	Unified, self-descriptive WDC-specific (?!)
OUT:	Unified, self-descriptive	Unified, self-descriptive	BUFR/CREX	
Resolution	Native	Native	Hourly (?)	Hourly
Metadata	Native as provided from instrument	Station, instrument, process description or reference	distribute WIS metadata	WIGOS metadata

2018-06-27 / joerg.klausen@meteoswiss.ch

Figure 2. Generic data generation, processing and distribution chain for GAW (N)RT data services.

Shown in the blue parallelograms in Figure 2 are the intermediate endpoints of the work-flow, expressed as data of a certain quality level. Shown in the red boxes are individual steps ("modules") involved in moving from one data level to the next. A few examples exist where this (or something very similar) has been implemented as an end-to-end solution. One such example is the GAW aerosol programme, whose RT data service has global coverage and is supported by the European research infrastructure ACTRIS[2] and the U.S. NOAA. It has been online since 2011. Here, RT data production is centralized at the data centre, and an operational data subscription service to ECMWF has been established in 2012. The centralized architecture takes advantage of the instrument competence present at the topical WDC, resulting in observations with various physical principles being handled by the service with consistent high quality. It also helps RPOs with taking their observations RT, which otherwise wouldn't be possible for capacity reasons at the RPOs. This is a particular focus of GAW, which is including non-NMHSs. Another example is ICOS, where an entire chain is now operational. Documentation on the ACTRIS work-flow can be found at https://sites.google.com/view/nrt-docs/home. ICOS can be seen as an example for an end-to-end workflow, but the ICOS approach may not be directly transferable for two main reasons: (1) ICOS only supports the use of a particular brand of instruments, and (2) the data collection and distribution happens centralized outside the GTS/WIS infrastructure. Other examples are AERONET and MPLNET, where similar work-flows apply. All of these implementations work well, but lack open interfaces between the various processing steps, which makes it harder to de-centralize the work-flow to reduce the risks of single points of failure. IAGOS, an aircraft program, uses the GTS/WIS to disseminate observations of individual aircraft in RT. For GAW in situ observations, no such de-centralized end-to-end

[2] ACTRIS homepage: https://www.actris.eu; ACTRIS Data Centre: http://actris.nilu.no/

work-flows presently exist on the global scale. The break-down of the work-flow into a sequence of steps/modules offers a number of advantages, viz.

- Any step/module can easily be exchanged for another (mutually agreed) process, provided the interfaces are respected.
- Different instruments generating data in specific formats can more easily be integrated.
- The documentation of each step is manageable and critical steps can easily be referenced in the metadata.
- The workflow or any part of it can be implemented on a single machine, on several machines (potentially distributed across service providers), or even in the cloud.
- Expertise of different communities such as instrument experts, data analysts can be included for specific sub-processes.
- Observation-specific approaches, e.g., calibration, automated QC procedures can be separated.
- Different approaches, in particular for “Advanced automated QC” can be developed and tested before integration in the end-to-end solution (this could allow involving other academic communities, e.g. machine learning, that can bring substantial advances on this aspect).
- Existing proven end-to-end workflow implementations can be integrated seamlessly.

A few challenges exist, namely

- Parts of this work-flow that already exist somewhere may need to be adapted for subsequent integration.

The specification of the interfaces, data formats between all the (atomic) steps shown in Figure 2 would create an excessive work-load without substantial benefits, provided they haven’t been implemented yet. It is therefore suggested to only specify a common data model for the final data product delivered by the workflow. ET-WDC should agree on this data model. Once agreed, converters for the level 1.5 data and a specific BUFR template can be implemented.

GAW Data Policy vs Creative Commons licenses

WMO data policy is governed by several resolutions, in particular Resolutions 40, 25 and 60 [ref]. These policies divide data in three categories, namely WMO Essential, WMO Additional, and WMO Other. GAW data qualify as “WMO Additional”, so certain restrictions on their use may be applied. As regards the dissemination and use of GAW data, GAW WDCs are bound to the following GAW Data Policy, which was formulated with a perception that data are submitted by the providers after full (and often manual) quality control:

"For scientific purposes, access to these data is unlimited and provided without charge. By their use you accept that an offer of co-authorship will be made through personal contact with the data providers or owners whenever substantial use is made of their data. In all cases, an acknowledgement must be made to the data providers or owners and the data centre when these data are used within a publication. [4]"

The GAW SAG Aerosol has adopted a separate data policy for aerosol RT data, which is clearly distinguished from the GAW data policy applicable for fully quality assured data (NB: This policy uses “NRT” in the sense of RT):

"Data delivery on a near-real-time (NRT) schedule is a WMO-GAW pilot project. Near-real-time aerosol data are raw, automatically processed data that have not passed through any review by aerosol scientists. They are intended for applications where gaps and glitches in the data are expected and accepted, for example, assimilation by air quality forecast models. Near-real-time data should not be used in lieu of final, 'clean', regularly reported data, and the data providers expressly deny permission for users to publish near-real-time data. Access to WDCA NRT data requires a registration and the user's consent not to distribute or publish data without agreement of [the] data provider. In the pilot project phase, requests for registration will be evaluated by the GAW Scientific Advisory Group for aerosol. Contact between data users and providers is strongly encouraged. Data providers are very interested in learning about how the data are being used."

Data centers of contributing networks have their own unique data policies that are typically similar to the GAW Data Policy.

In addition to the WMO and GAW data policies, and to the specific data policies applied by some contributing programs, the so-called Creative Commons licenses [3] are receiving increasing attention. Several Creative Commons licenses exist. The CC BY license is one of the least restrictive licenses offered by Creative Commons. It only requires proper attribution of the originator of the data (BY), but allows re-distribution, re-processing and even commercial use of the data. More restrictive variants of CC licenses prohibit commercial use (NC) or distribution of the data itself or any material derived from the data (ND). For (N)RT data services from GAW, the CC-BY or the CC-BY-NC licenses are recommended to encourage wide use of GAW data. The CC-BY-NC-ND license is recommended as the default license to restrict use of (N)RT data to specific (named) users for a specific purpose.

The current GAW Data Policy also encourages unlimited distribution of the data and requires proper attribution of the originator (BY), it also prohibits commercial use (NC), but in addition requires an offer of co-authorship in publications whenever *substantial use* is made of the data. In the context of (N)RT data services, where data from many individual data providers are typically used, the current GAW data policy is in general inadequate or simply impractical. One of the problems with this policy is that there is no unambiguous definition what *substantial use* means.

While GAW should promote an open data policy in general, there may be use cases where providers wish to restrict the distribution of GAW (N)RT data to specific (named) users for a specific purpose to limit unintended use and to protect the intellectual property rights of the data originator to the maximum degree possible, while still allowing useful applications. In such cases, providers should choose the CC-BY-NC-ND license model. In the long-term, more and more services are expected that will need (N)RT observations and more flexible licenses are probably required. For example, both the US EPA, and the EEA operate under an open license with minimal restrictions. The CC license model offers a range of possibilities that support the dissemination of (N)RT data for use in operational services.

Terms of Reference for (N)RT data service providers from GAW

The exact system architecture for (N)RT data services from GAW has not been decided . The requirement to treat chemical data in a similar way as weather data strongly argues for generation and processing of observations locally by Members and international dissemination by way of the GTS/WIS infrastructure whenever possible. This implies the following Terms of Reference:

1. Operate processing tools (including automated QA/QC) to process Level 1.0 data to Level 1.5 data with particular emphasis on achieving data quality that meets the "GAW Principles for (N)RT Data". Processing tools can be implemented differently, but must adhere to the methodologies endorsed by the Scientific Advisory Groups of GAW and approved by WMO. Publish, as part of the service, documentation on the tools / procedures used.
2. Clearly specify the data policy/license regulating use of the data. The CC-BY 4.0 license is recommended to encourage the broadest reach with the least restrictions. The CC-BY-NC-ND is recommended where data providers wish to maintain tighter control of the (N)RT data use.
3. Disseminate Level 1.5 data in (N)RT, using GTS/WIS. Additional distribution channels are encouraged.

There may be data (service) providers who are unable to get sufficient support from their NC for distribution of their (N)RT data through GTS/WIS. In these cases, facilities operated by some other Member can be requested to handle the RT distribution to the GTS/WIS and optionally, the processing of the data prior to distribution. There are also regional arrangements, e.g. the European Research Infrastructures ICOS and ACTRIS, or other, well established programs/networks such as EMEP, NDACC, MPLNET, NADP or EANET (list not exhaustive), where distribution, and maybe also processing, of observations are handled centrally. In these cases, those central facilities should either apply to become a DCPC, or establish operational links to an existing DCPC that can serve as a (N)RT data service provider under the ToRs above.

Role of WDCs and other data centres in (N)RT data management

Due to their observation specific competence, the WDCs will have a key role in applying data QC and curation. The quality of automatically processed (N)RT data is inherently different from manually treated observational data. Recognizing the superior knowledge of the measurement PI with regard to site and instrument specifics, GAW has always insisted on the responsibility of the data originator for "signing off" on the quality of data submitted to the GAW WDCs, and the expectation remains that data submitted to them are of the highest quality possible. Similar requirements have been put in place by data centers of GAW contributing networks. As a consequence, the emphasis and primary role and responsibility of GAW WDCs as outlined in the Terms of Reference of WDCs [4, listed in the Appendix], is to archive and disseminate fully quality controlled data for certain groups of observed quantities. To this end, several of the WDCs have established tools to perform quality checks of the data they receive. WDCs will usually inform data originators if their data fail any such tests. For (N)RT processes, the automated tools applied to the data before they are transmitted to the GTS/WIS will block further transmission of the data, or at least flag such data that do not pass these standard quality tests.

Historically, (N)RT data handling was not part of the responsibilities of WDCs. Today, this has changed, and the current GAW Implementation Plan urges WDCs to also consider the increasing needs for (N)RT

data services. Indeed, as DCPCs in WIS, the various GAW WDCs can play an important role for observations of quantities in their remit, and they should attend to a number of important tasks in (N)RT data management:

1. Act as (N)RT data service providers as described above if a Member's NC is unable to support the needs of (N)RT data providers.
2. Monitor the GTS/WIS, harvest and archive (N)RT data for re-distribution in accordance with the data policy/license attached.
3. Operate web-based tools to display recent observations in support of data providers and users in accordance with the data policy/license attached.
4. In addition, and not restricted specifically to (N)RT data service requirements, archive level 0 data for any future re-processing of data.

Contributing networks usually maintain their own data management infrastructure. As stated upfront in this document, WMO GAW has no mandate to specify data management practices for these programs/networks, but they are encouraged to consider providing similar services.

Table 1 summarizes the current capabilities.

Host Institution (WDC hosted)	**Capability**	**Data Submission**	**Data Delivery**	**References**	**Users**
EC (WOUDC)	supports (N)RT data submissions on Ozone and UV. level 0 data are received as is and made available to contributors on the WOUDC WAF to contributors only (password protected)	ftp	http/https	https://guide.woudc.org/en/#231-upload-guidelines	WOUDC contributors
JMA (WDCGG)	receives NRT data as event data	email, ftp	http		
NILU (WDCA, WDCRG)	Supports (N)RT data submissions of a number	ftp	ftp	https://sites.google.com/view/nrt-docs/home	ECMWF, scientific community, policy

	of aerosol and reactive gases data with global coverage. Operational data subscription service available.				frameworks (CLRTAP, others)ECMWF, ACTRIS, GAW aerosol stations
MGO (WRDC)	No (N)RT capability	n/a	n/a	n/a	n/a
DLR (WDC-RSAT)	?	?	?	?	?

Conclusions and Recommendations

This document was reviewed and endorsed by its co-authors, including ET-WDC regular and ex-officio members. It will be discussed by the SAG Apps during their July 2018 Geneva meeting. The conclusions and recommendations below are directed to the GAW SSC OPAG EPAC as well as to the CAS MG for consideration.

1. ET-WDC endorses the "GAW Principles for (N)RT Data" as strategic guidance for future development of (N)RT data services from GAW.
2. ET-WDC recommends that chemical data are treated using similar infrastructure as weather data and endorses use of the WMO GTS/WIS infrastructure as a mechanism for dissemination of GAW (N)RT data.
3. ET-WDC endorses a modular (service-oriented architecture) approach as depicted in Figure 2 for interoperable implementation(s) of end-to-end GAW (N)RT data delivery services.
4. ET-WDC endorses the terms of reference for (N)RT data service providers of GAW. Of the existing WDCs, WDCA, WDCRG, X Y Z have ambitions to take responsibility to become such a (N)RT data service provider for GAW.
5. ET-WDC has identified the need to adopt one or several suitable (N)RT data policies/licenses as the current GAW data policy is inadequate for (N)RT data services. In light of the general movement towards open data, CC-BY is the recommended license, but other variants of CC, in particular CC-BY-NC-ND, shall be accepted if data providers wish to limit unintended use of (N)RT data.

References

[1] WMO (2017), WMO Global Atmosphere Watch (GAW) Implementation Plan: 2016-2023, GAW Report No. 228

[2] ISO/BIPM (2008), Evaluation of measurement data – Guide to the expression of uncertainty in measurement, https://www.bipm.org/utils/common/documents/jcgm/JCGM_100_2008_E.pdf

[3] https://creativecommons.org/licenses/by-nc-nd/4.0/

[4] WMO (2007), WMO Global Atmosphere Watch (GAW) Strategic Plan: 2008 – 2015 , GAW Report No. 172; WMO, http://www.wmo.int/pages/prog/arep/gaw/world_data_ctres.html, accessed: 4 April 2018

Glossary

BUFR	Binary Universal Form for the Representation of meteorological data) is a flexible binary format mainly used to encode in situ and satellite observations. It is a table-driven code form that is used by the GTS/WIS. Tables are operationally maintained by WMO. Templates exist for various types of observations (e.g., https://www.wmo.int/pages/prog/www/WMOCodes/WMO306_vI2/TemplateExamples.html), and new templates can be created.
CREX	See BUFR
DAR	(Data) Discovery, Access and Retrieval.
GISC	Global Information Service Center. See WIS
GTS	Global Telecommunication System. The GTS is an operational distributed infrastructure used globally by NMHSs to distribute data in RT. Best known for weather data, the GTS also handles other data, including chemical composition.
WIS	WMO Information System. The WIS evolves from the GTS (which is also continuously evolving), adding in particular DAR capability by way of catalogues maintained at the GISCs. For more information on WIS, cf. https://www.wmo.int/pages/prog/www/WIS/index_en.html.
WDC	World Data Center. WMO GAW has a number of WDCs that have assumed responsibility for specific groups of atmospheric composition variables
GAWSIS and OSCAR/Surface	Official repositories for metadata on (non-satellite) observing facilities and observations made world-wide.
WIGOS	https://www.wmo.int/pages/prog/www/wigos/index_en.html
WMDS	WIGOS Metadata standard. Expressed as in XML based on a GML (Geography Markup Language, an OGC standard) schema, defines the metadata exchange format for observations.
NMHS	National Meteorological and Hydrological Service. In most countries, the NMHS represents a country/territory (the 'WMO Member') in WMO. NMHSs also operate the GTS/WIS for WMO, sometimes in consortia.

Appendix

(excerpts from [4])

5.1.9 Expert Team on World Data Centres (ET-WDC)

Specific activities:

ET/WDC-1. Take responsibility for metadata and data management issues in support of the scientific and operational objectives of GAW.

ET/WDC-2. Work with the SAGs, WMO expert teams and partners to establish harmonized data management guidelines, including standardised data formats to allow for adequate (seamless) interoperability.

ET/WDC-3. Guide and support the further development of GAWSIS as the central catalogue of observing facilities and observations supporting GAW, linking the WDCs and Contributing Data Centres.

ET/WDC-4. Keep abreast of and recommend best use of changing technologies affecting information management within GAW.

5.2.4 World Data Centres (WDCs) and Contributing Data Centres (CDCs)

Specific activities:

WDC-1. Provide adequate archiving facilities for observational data for which GAW has global coordination responsibilities.

WDC-2. Check submitted data for necessary format elements and the availability of comprehensive metadata and reject the submission of data that do not meet these formal criteria.

WDC-3. Perform plausibility and consistency checks on submitted data, flag data problems, and provide feedback to the data providers, when necessary.

WDC-4. Continually improve the ease of access to data of known quality by evolving WDC operations in line with the development of WIGOS and with particular attention to the increasing needs for NRT data services.

WDC-5. Contribute to the agreement of standards for interoperability of data archives through the Expert Team on GAW World Data Centres (ET-GAW WDCs). This also includes support for the establishment of harmonized guidelines and data formats for the submission and dissemination of atmospheric composition data, metadata and products.

WDC-6. Support and participate in the establishment of a distributed data management system involving all WDCs, the archives of contributing networks, and GAWSIS as the central metadata repository for discovery and access purposes.