Metadata for WIS and WIGOS: Information needed to make adequate use of observations

Jörg Klausen[(1)], Brian Howe[(2)], and the ET-WDC[(3)] and TT-WMD[(4)] Teams

[(1)] Federal Office of Meteorology and Climatology MeteoSwiss, Switzerland (joerg.klausen@meteoswiss.ch), [(2)] Environment Canada, Canada (brian.howe@ec.gc.ca), [(3)] ET-WDC (www.sites.google.com/site/wmoetwdc), [(4)] TT-WMD (http://www.wmo.int/pages/prog/www/CBS/Lists_WorkGroups/EC/ec-wg/tt-wmd)

Abstract

In the late 1990s, The World Meteorological Organization (WMO) embarked on a visionary journey intended to improve the discovery and usability of meteorological observations collected and exchanged word-wide. The two activities supporting this vision are the WMO Information System (WIS) and the WMO Integrated Observing System (WIGOS). WIGOS is a key WMO priority to underpin all WMO Programs and new initiatives such as the Global Framework for Climate Services (GFCS). Important aspects are the description of the observational capabilities of Members by way of structured metadata (WIGOS metadata), and the Observing Systems Capabilities Analysis and Requirements (OSCAR) tool in support of the Rolling Review of Requirements (RRR) process in support of rational evolution of the observing systems.

In the context of WIS, the WMO has developed the WMO Core Metadata Profile (WCMP) for the description of discovery metadata. The Commission of Atmospheric Sciences (CAS) Task Team on GAW WDCs (ET-WDC) has developed a profile of the ISO19115 metadata standard that is compliant with WCMPv1.3. It is intended to harmonize certain aspects of the documentation of observations as well as the interoperability of the WDCs. The Inter-Commission Coordination Group on WIGOS (ICG-WIGOS) Task Team on WIGOS Metadata (TT-WMD) with representation of all WMO Technical Commissions has developed the WIGOS Metadata Standard (WMDS). The current draft v0.1 is a draft semantic standard comprising of a set of metadata classes that are considered to be of critical importance for the interpretation of observations relevant to WIGOS.

The paper introduces the WMO Metadata Profile, the GAW Metadata Profile, and the WIGOS Metadata Standard, and explores how these relate to each other and to OSCAR. It touches upon a likely approach for implementing the WIGOS Metadata Standard and how the adoption by Members is foreseen.

Introduction

Since its creation in 1947, the World Meteorological Organization (WMO) has established a tradition of exchanging meteorological and related data in support of multi-hazard, multipurpose early warning systems, weather, water and climate analyses and forecasts; tsunami related information and warnings, and seismic data. In fact, the Global Telecommunications System (GTS) has been the first system ever to facilitate electronic exchange of data on a global level on a 24/7 basis, dating back to 1951 (Fenix, 2006). GTS was designed as a private, supervised, co-ordinated, point-to-point and multi-point communication network with three World Meteorological Centres (WMCs) (Melbourne, Moscow and Washington) at the core, and 15 Regional Telecommunication Hubs (RTH) to route information (cf. Figure 1). GTS has provided WMO Members with a most reliable means of sharing data for many decades. Major short-comings of the GTS include limited accessibility for providers and users outside the National Meteorological and Hydrological Services (NHMSs) and the fact that it is nearly impossible to know what information is exchanged and, hence, available through GTS.

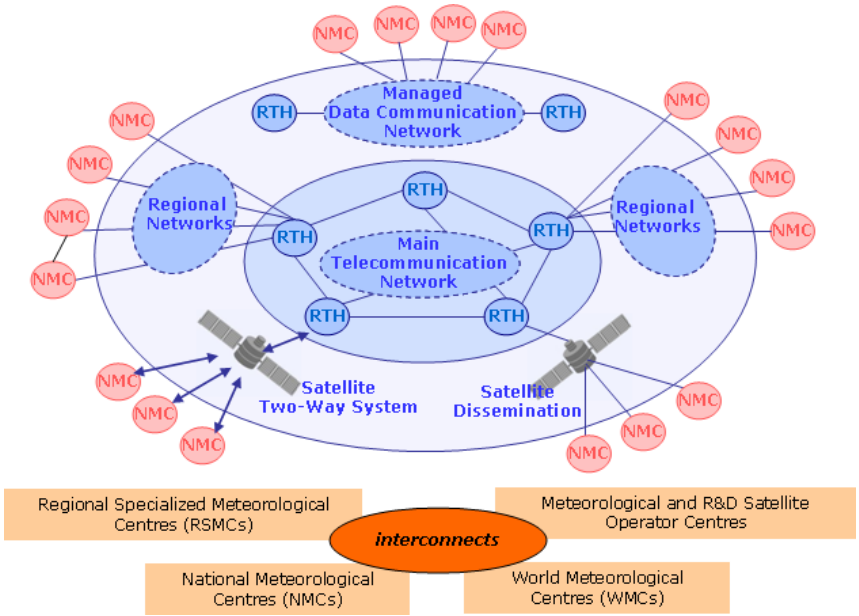

Figure 1. High-level schematic of the Global Telecommunication System (GTS) (source: http://www.wmo.int/pages/prog/www/TEM/GTS/index_en.html)

In 2003, Fourteenth WMO Congress therefore formally adopted the concepts for a WMO Information System (WIS), to be developed and implemented based on the GTS, but using industry standards for protocols, hardware and software (WMO, 2003). Important additions to the GTS include the generation and maintenance of catalogues on data available and improved access to data and information to partners outside the NMHSs, in real-time and through archives.

While WIS focuses on a mechanism for the *exchange* of data, and to facilitate search, discovery and access to data, the WMO Integrated Global Observing System (WIGOS) also governs the *generation* of (observational) data and the integration of observing systems (WMO, 2007). WIGOS is a key WMO priority to underpin all WMO Programs and new initiatives such as the Global Framework for Climate Services (GFCS). This includes all sorts of (primarily) observational data from different platforms including satellite and ground-, sea- and air-based ones. It therefore encompasses various WMO and co-sponsored observing programs, in particular World Weather Watch (WWW), Global Atmosphere Watch (GAW), Global Cryosphere Watch (GCW), Hydrology and Water Resources Programme (HWRP), Agricultural Meteorology Programme, Marine Meteorology and Oceanography Programme, Aeronautical Meteorology Programme, and others.

A common need of these programs and the scientific and operational communities supporting them is the exchange of data globally and following common approaches. As observations are increasingly automated, as modern techniques are increasingly generating larger amounts of data, and as more and more applications require robust descriptions of observations, the need for metadata also increases.

The purpose of this paper is to acquaint the reader with a basic notion of metadata in general, and the standards and profiles that have been developed or are under development by several expert and task teams operating under the WMO umbrella. It is hoped that wider awareness of the work of these teams will motivate researchers and professionals to engage in further refinement and the evolution of a metadata standard that serves data providers from different disciplines as a means of describing observational (and other) data that enables data users to make (more) adequate use of these resources.

Metadata

Metadata are data about data, and the distinction between metadata and data is blurred. Some of the confusion is due to different understanding of the term by different people (Bannerman, 2012). Moreover, metadata information is often taken for granted and considered trivial, particularly by those who are closest to the observation, namely the observer or the scientist analyzing it. For the purpose of WIGOS, metadata can be defined as “information needed to enable users to make adequate use of data – today, but also in the distant future.” In other words, metadata should enable users to judge if the data at hand are fit for purpose. This is application-dependent, and the requirements range from almost no metadata to a comprehensive description of any factor that could reasonably be assumed to have had an effect on an observation or a processed dataset in general. In particular, for climate applications, the factor ‘time’ is critical. The GCOS monitoring

principles describe some of the more pertinent measures to be taken in making climate observations. Among others, it is stated that “the details and history of local conditions, instruments, operating procedures, data processing algorithms and other factors pertinent to interpreting data (i.e., metadata) should be documented and treated with the same care as the data themselves.” (GCOS, 2003). Herein lies the difference between different flavors of metadata: they answer only some or sometimes all of the what, where, how, and even why questions related to observations or processed information and their interpretation. Thus, metadata describing collections of observations must be time-stamped, indicating the validity period of individual metadata elements. Also, as part of the life cycle of observational (and other) data, the amount of metadata needed to make adequate use of the data always increases (it never decreases), even though the amount of data may be reduced to a single aggregated value.

The WMO Core Metadata Profile

The WMO Core Metadata Profile was developed and endorsed as version 1.3 (WCMPv1.3) for the sake of facilitating discovery, access and retrieval (DAR) of data on the WIS (WMO, 2013). It comprises of a sub-set of the metadata elements specified by ISO19115 for the description of geographic information (ISO/TC 211, 2003). Figure 2 is a simplified diagram of the WCMPv1.3. The only WMO-specific aspects are amendments to code tables and new code tables. WCMPv1.3 specifies a number of constraints or conditions. In the case of observational data with a geospatial context, the specification of a geographic bounding box becomes mandatory (→EX_Extent), and in the case of global distribution of data through WIS, the specification of the data policy is mandatory (→MD_LegalConstraints). Looking at the diagram in Figure 2, one can see that the focus of this metadata model is indeed discovery (via MD_Keywords, EX_GeographicBoundingBox), access and retrieval (via gmd:fileIdentifier). Legal constraints on use must also be specified for data that are globally exchanged (via MD_LegalConstraints).

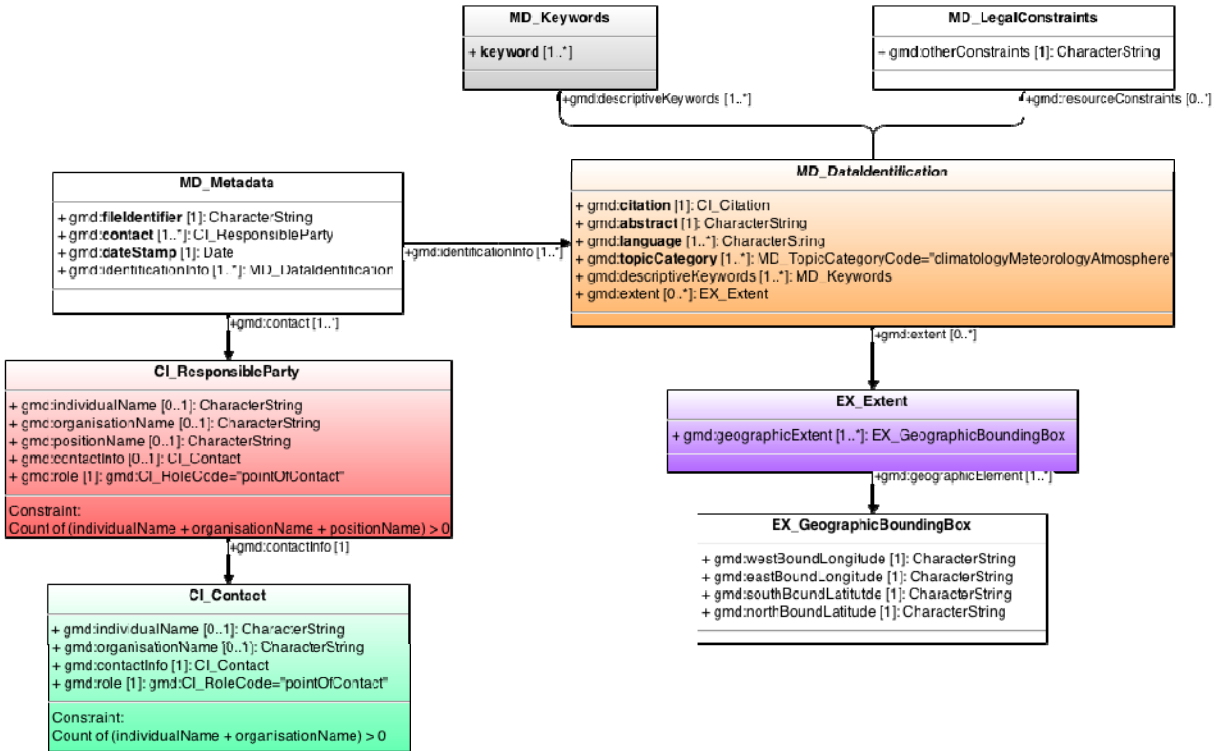

Figure 2. UML diagram of the WMO Core Metadata Profile version 1.3 (WCMPv1.3) indicating mandatory and conditional elements needed to describe geographic data for global distribution. Elements in bold define the minimum a metadata record needs to specify to be compliant.

The GAW Metadata Profile

The Global Atmosphere Watch is the WMO program addressing the chemical composition and selected physical properties of the atmosphere. Unlike the World Weather Watch, GAW is a voluntary program of WMO and still largely research-driven. In many Member countries, GAW activities are undertaken primarily by research institutions (universities, governmental organizations) outside the NMHSs. For these institutions, access to the GTS is sometimes difficult, and substantial efforts have been made in the past to develop GTS-independent data archives. Observational data of atmospheric composition and air quality are collected and archived by GAW World Data Centres (WDCs) and related data centers, including many national archives. The latter are difficult to access and certainly not interoperable. There are strong hopes that WIS and WIGOS will eventually bring about improved easier access to data and interoperability of data centers.

The Task Team on GAW WDCs (ET-WDC) has developed version 1.0 of the GAW Metadata Profile (GMPv1.0) (www.sites.google.com/site/wmoetwdc/metadata) of the ISO19115 metadata standard that is compliant with WCMPv1.3 (see Figure 3 for a UML diagram). This profile is intended to harmonize certain aspects of the documentation of observations as well as to improve the interoperability of the WDCs. Unlike WCMPv1.3, the nature of the observation can be described, and a data quality statement can be made. Ownership and data policy are very

important to some contributors of GAW data, reflecting the researchers' need to produce original scientific papers and a desire for acknowledgement of the efforts of individuals. The profile therefore includes several places where contact information can be stored for each data set, and explicitly mentions that "Data use is constrained by the data use policy of the individual measurement program."

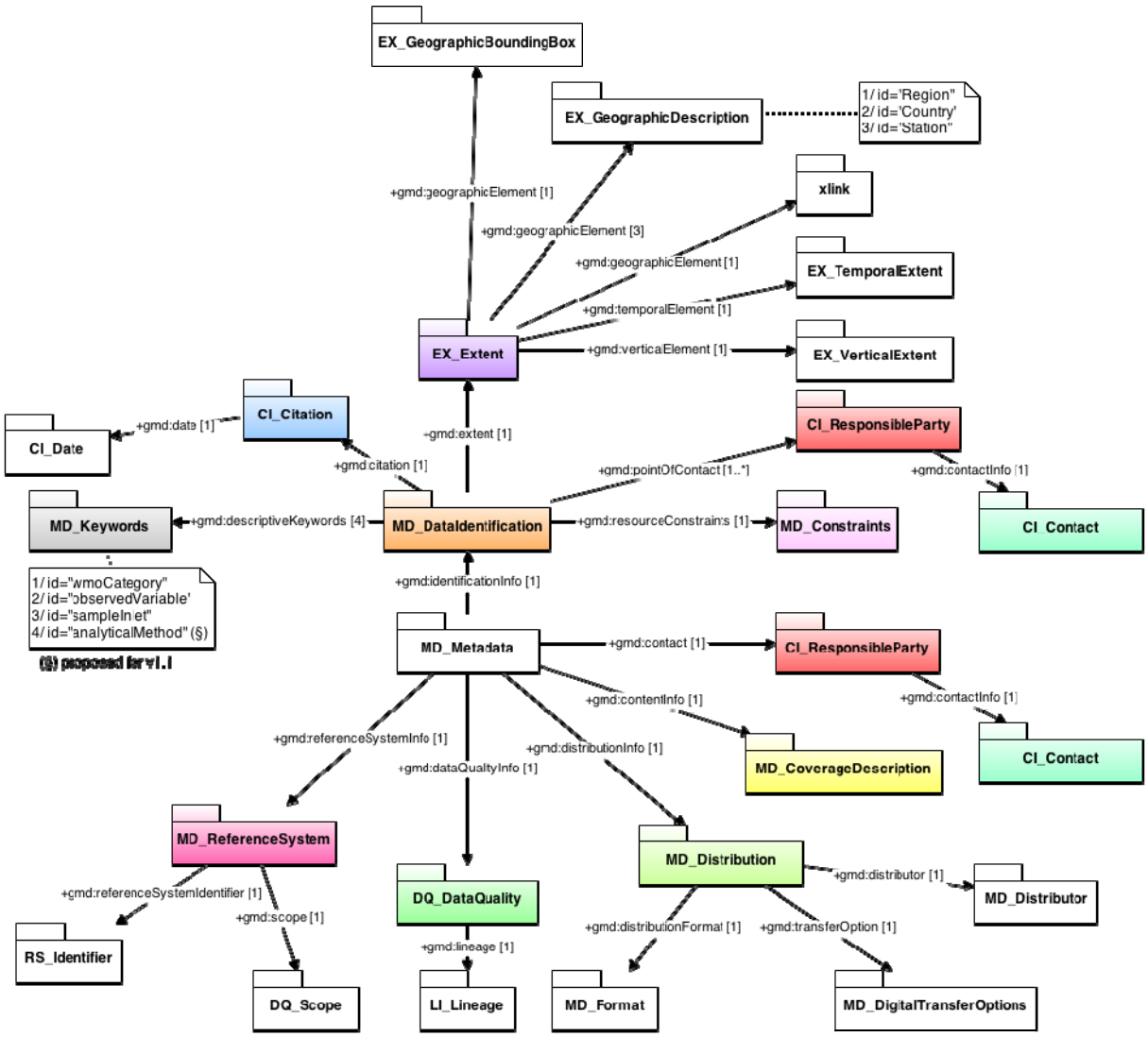

Figure 3. Simplified UML diagram of the GAW Metadata Profile v1.0 (GMPv1.0) depicting 1st through 3rd level classes. Classes defining code lists are not shown.

In version 1.0, MD_IdentificationInfo contains 3 instances of descriptiveKeywords, namely "wmoCategory", "observedVariable", and "sampleInlet". The most recent draft adds one more, namely "analyticalMethod". As demonstrated in Figure 3, a metadata record of observational data that goes beyond discovery quickly becomes fairly complex, even though the figure only lists the classes needed up to level 2 and does not mention the mandatory elements from these classes.

Implementation of the GAW Metadata Profile in GAWSIS

The GAW Station Information System (GAWSIS) is the official catalogue of stations recognized by GAW and strives to document the capabilities of global and regional monitoring networks for atmospheric composition (and some physical quantities) coordinated by WMO GAW. GAWSIS provides an interactive web interface (www.meteoswiss.ch/gawsis) supported by a database with four essential business objects: stations, measurement programs, contacts, and references. GAWSIS obtains metadata information from the GAW World Data Centers, from a number of topical or regional data archives, and through direct human interaction by registered users. Importantly, several metadata elements obtained from data centers are mapped to controlled vocabularies. This mechanism leads to significant homogenization of metadata information at the expense of precluding or delaying the import of metadata that are not found in the thesauri. The maintenance of thesauri involves a fair amount of expert judgment.

For the purpose of making metadata available through WIS, GAWSIS generates standardized metadata records in XML format, following the specifications of the GAW Metadata Profile v1.0 (GMPv1.0). At present, rules are encoded in a database table using path expressions to describe the hierarchy in the desired metadata tree and the R language (R Development Core Team, 2011) to process and generate the xml files.

The GMPv1.0 is an extension of the WCMPv1.3, but it is still a profile of the ISO19115 metadata standard. This standard is not ideally suited to describe observational data in a way to enable users to make adequate use of the data. The GMPv1.0 goes some way towards describing observational data, but is clearly insufficient. The WMO Integrated Global Observing System (WIGOS) metadata standard is being developed to alleviate these shortcomings.

The WIGOS Metadata Standard

The Inter-Commission Coordination Group on WIGOS (ICG-WIGOS) has established the Task Team on WIGOS Metadata (TT-WMD) with representation of all WMO Technical Commissions and the objective to define the WIGOS metadata standard (WDMS). The result of this effort is a draft semantic standard comprising of a set of metadata categories that are considered to be of critical importance for the interpretation of observations relevant to WIGOS. In May 2014, the WDMS version 0.1 has been submitted to the Presidents of all Technical Commission as part of the review of the WIGOS Regulatory Material (http://www.wmo.int/pages/prog/www/wigos/TT-WRM.html). The standard lists 10 categories of metadata as shown in Table 1. The elements in these categories are designated one of the obligations “mandatory”, “conditional”, or “optional”.

Conditional elements become mandatory if a certain condition is fulfilled (e.g., if station/platform is fixed, then the surface cover shall be specified.) The phases prescribe the timeline foreseen for Members to implement the standard. The phased approach takes into consideration the different traditions of Members with respect to collecting and maintaining metadata of observations.

Implementation of WIGOS Metadata Standard

Prior to implementation and use of the standard, a critical step is the translation of the semantic standard into a formal standard and associated XML schema that is machine-readable and supports automatic exchange of information. ISO19115 does not define all the entities needed to formalize certain elements in the WMDS. A valuable resource to guide further work on the WDMS is the metadata standard developed by the Global Runoff Data Centre (Dornblut, 2013). This standard focuses on hydrology observations and uses the ISO19156 standard (ISO/TC 211, 2011) for the description of the processes that were involved in generating an observation, a standard which lends itself also for transcribing WIGOS metadata describing (homogenized) datasets of observations. ISO19156 refers itself to ISO19115 and related standards for the description of datasets. The principles are depicted schematically in Figure 4. The process of formalizing the WIGOS metadata standard is expected to begin in the second half of 2014.

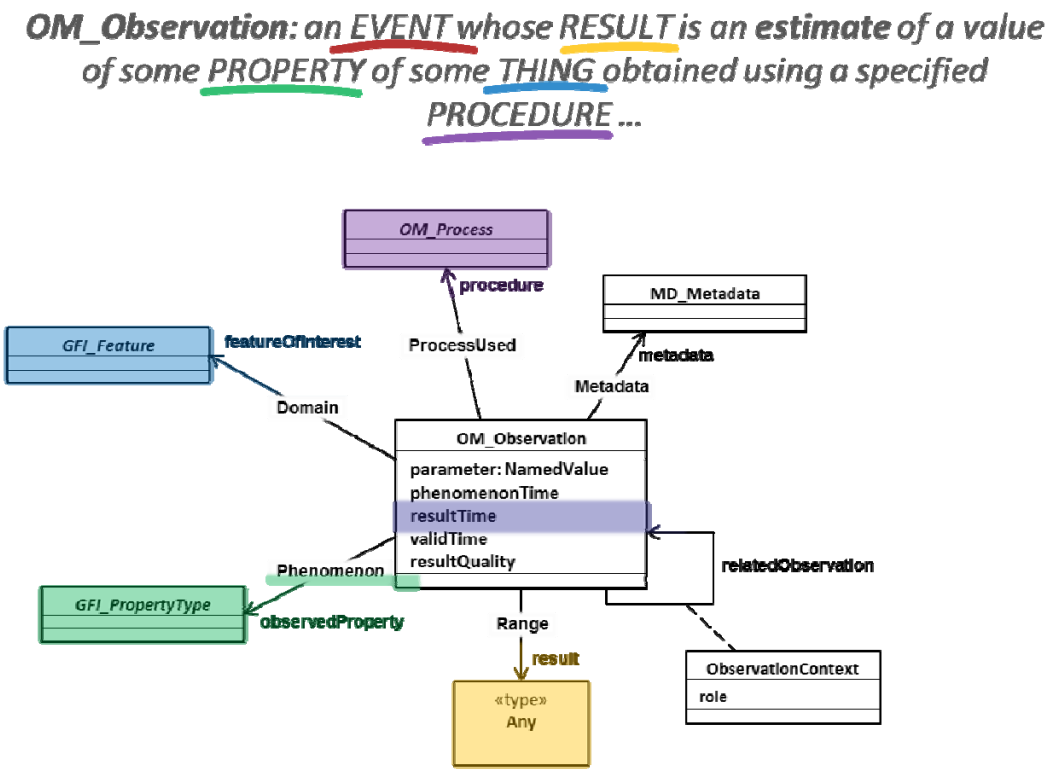

Figure 4. High-level UML diagram of the description of an observation as conceptualized in the ISO156 standard (courtesy of Aaron Braeckel, UKMO).

Adoption of WIGOS Metadata Standard

Reporting WIGOS metadata will generate substantial benefits for Members, but developing the capacity to report these metadata also requires a substantial effort on the part of (meta)data providers. To help Members comply with reporting obligations, the WIGOS Project Office will reach out to its Technical Commissions, Regional Associations, expert teams and individual Members to produce extensive guidance material, such as implementation guidelines and “best practice” guides.

Moreover, reporting obligations will be enforced in phases, in order to allow Members sufficient time to develop the capacity to comply. Balancing the effort required to generate and report individual elements, and the need to have this information to make adequate use of observations, implementation will proceed through three phases as shown in Table 1 below. Importantly, elements required by the end of Phase I are either listed as mandatory elements in WMO Pub 9 Vol A or are of critical importance for the Observing Systems Capability Analysis and Review (OSCAR) tool, and are considered of benefit for all application areas[1]. Phase II adds elements recognized to be more challenging for Members, but the knowledge of which is still of rather immediate need for the adequate use of observations, in particular for assessing quality of observations. Phase III adds the remaining metadata elements specified in this version of the standard, which are beneficial to particular application areas, but are not necessarily universally applicable and require greater effort to implement.

[1] At this stage in developing the phased adoption approach, there is a certain discrepancy between the elements identified needed for OSCAR (cf.) and the elements identified for adoption in Phase I. This mismatch will be addressed before endorsement of the WMDS. Moreover, at least one element needed for OSCAR is missing from WMDS, namely an element specifying the “distribution channel” of the data, in particular if the data are available through the GTS. Whether this information will be provided through WIS or as part of the WIGOS metadata needs to be discussed.

Table 1. List of elements specified in the WIGOS metadata standard and the phases for implementation (WMO, 2014). Elements identified to be needed for the WMO RRR process (cf.) are high-lighted in light red.

Category	Phase I	Phase II	Phase III
	2016	2017-2018	2019-2020
1. Observed Quantity	**1-01 Name of observed quantity – measurand (M)**	1-05 Representativeness of observation (O)	
	1-02 Measurement unit (M)		
	1-03 Temporal extent of observed quantity (M)		
	1-04 Spatial extent of observed quantity (M)		
	1-06 Observed medium (M)		
2. Purpose of Observation	2-01 Application area(s) (O)		
	2-02 Network affiliation (M)		
3. Data Quality		**3-01 Uncertainty of measurement (M)**	
		3-02 Procedure used to estimate uncertainty (M)	
		3-03 Quality flag (M)	
		3-04 Quality flagging system (M)	
		3-05 Traceability chain (M)	
4. Environment		*4-04 Exposure of instrument (C)*	*4-01 Surface cover (C)*
		4-05 Intervention at Station/platform (O)	*4-02 Surface Cover classification scheme (C)*
		4-06 Site information (O)	*4-03 Topography or Bathymetry (C)*
5. Data Processing and Reporting	**5-03 Reporting interval (time) (M)**	5-02 Processing/analysis centre (O)	5-01 Data processing methods and algorithms (O)
	5-04 Reporting interval (space) (C)	5-06 Level of data (O)	5-05 Software/processor and version (O)
	5-12 Reference datum (C)	**5-09 Aggregation interval (M)**	**5-07 Data format (M)**
		5-10 Meaning of time stamp (M)	**5-08 Version of data format (M)**
		5-11 Reference time (M)	5-13 Numerical resolution (O)
			5-14 Latency (of reporting) (M)
6. Sampling and Analysis	6-03 Sampling strategy (O)	**6-06 Spatial sampling resolution (M)**	6-01 Sampling procedures (O)
			6-02 Sample treatment (O)
			6-04 Sampling time period (M)
			6-05 Meaning of the time stamp (M)
			6-07 Analytical procedures (O)
7. Station/Platform	*7-01 Region of origin of data (C)*	**7-04 Station/platform type (M)**	**7-05 Station/platform model (M)**
	7-02 Territory of origin of data (C)	7-08 Data communication method (O)	
	7-03 Station/platform name (M)		
	7-06 Station/platform unique identifier (M)		
	7-07 Geospatial location (M)		
8. Method of Observation	**8-01 Source of observation (M)**	*8-12 Geospatial location (C)*	*8-06 Configuration of instrumentation (C)*
	8-02 Measurement principle (M)		*8-07 Lab calibration interval (C)*
	8-03 Observable range (M)		*8-08 Instrument lab calibration date and time (C)*
	8-04 Instrument stability (C)		*8-09 Instrument model and serial number (C)*
	8-05 Vertical distance (C)		*8-10 Instrument field maintenance (C)*
			8-11 Instrument field verification (C)
9. Ownership and Data Policy	**9-02 Data policy/use constraints (M)**	**9-01 Supervising organization (M)**	
10. Contact			**10-01 Contact (Nominated Focal Point) (M)**

WIGOS Metadata and the Rolling Review of Requirements (RRR) Process

WIGOS metadata serve various purposes. Primarily, the intention is to provide users of observational data with the information required to decide if an intended application of these data is adequate, i.e., whether the data are “fit for purpose”. This is a non-trivial decision process that may require significant expert knowledge and experience. The WMDS stipulates a relatively large number of elements, organized in categories, to support this process.

Another important intended use of WIGOS metadata is to support the assessment and rational evolution of the observing systems at the global, regional, and national scale. Requirements for observations are diverse and depend on the intended application. An analysis elaborated in the context of atmospheric composition change (Laj, et al., 2009) lists the following elements, viz.

- satisfy and verify current legislation,
- validate and help to improve our understanding of atmospheric processes,
- understand the relationship between the atmosphere, the oceans and the land masses,
- permit accurate predictions of future atmospheric states by providing inputs to forecast models,
- support answering new questions that will only emerge in future.

The WMO Rolling Review of Requirements (RRR) process uses the Observing System Capabilities Analysis and Requirements (OSCAR) tool to support an objective comparison of capabilities and requirements with the intention to develop so-called Statement-of-Guidance (SoG) documents to identify the need and scope for further development of observing systems. As shown in Figure 5, OSCAR is divided into several logical parts, namely OSCAR/Requirements, OSCAR/Capabilities (space and surface), with an OSCAR/Analysis module linking these elements. The space-related parts have already been developed and are currently operated by WMO (www.wmo.int/oscar). OSCAR/Surface and an extended OSCAR/Analysis module are under development.

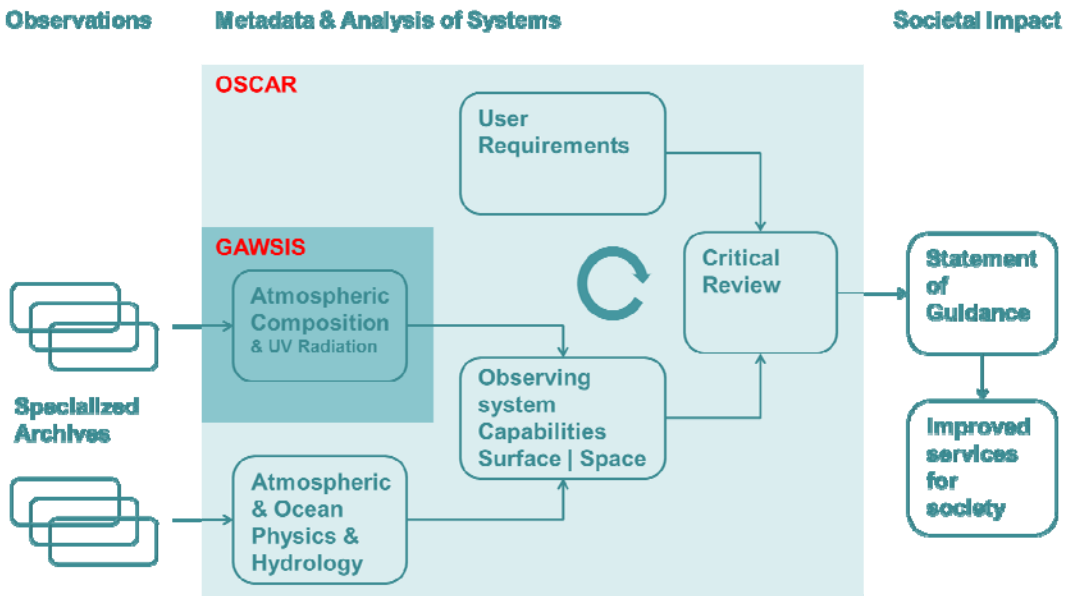

Figure 5. High-level view of the Observing System Capabilities and Requirements (OSCAR) tool.

Requirements are specified for each application area and observed property in terms of spatiotemporal coverage and resolution, as well as acceptable uncertainty. WIGOS metadata shall be used to provide the information needed to document the existing capabilities. Thus, the WDMS specifies elements that can be compared directly with corresponding elements describing requirements, or elements that can be transformed into such elements (cf. Table 1). As can be seen, the metadata requirements for RRR are quite demanding. Moreover, the intended phased adoption of the WDMS presents a challenge for the implementation of a functional OSCAR tool.

Conclusions and Outlook

The WMO Core Metadata Profile v1.3 (WCMPv1.3) as a profile of the ISO19115 metadata standard for geographical information provides a mechanism to facilitate search and discovery of observational data in WIS. WCMPv1.3 does not provide any information on the data quality or the processing chain involved in generating observational data.

Within GAW, the ISO19115 standard was explored in that direction. The GAW Metadata Profile v1.0 (GMPv1.0) extends the capabilities of WCMPv1.3 by including more keywords to describe the sampling and analysis used. It also specifies additional elements to describe the origin of the observations and the temporal extent of a data set to support search and discovery. Further, it allows the specification of several contacts for a data set to acknowledge the intellectual property rights involved. Data quality is still described in a more than rudimentary manner, pending the availability of more information from the data originators and/or data archives. Data use policy is referenced but not explicitly stated.

The WIGOS Metadata Standard is currently under review as version 0.1 (WMDSv0.1) and is a semantic standard. While WIS metadata are there to facilitate discovery, access and retrieval of data, the purpose of WIGOS metadata is to facilitate adequate use of observational data. To this end, the WMDS stipulates several dozen elements to describe the nature, origin and processing chain of observational data. The formalization of WMDSv0.1 is a challenging task still to be tackled. The WMDS will be adopted by Members in three phase between 2016 and 2018. This phased implementation presents a challenge for the development and operation of the WIGOS Observing Systems Analysis and Requirements (OSCAR) tool in support of the Rolling Review of Requirements (RRR) process. Once implemented and adopted, the WMDS is expected to be a powerful extension of the WCMP with direct benefits for all Members and all WMO Application Areas with a need for observational data that are documented sufficiently well to assess their fitness for purpose.

References

Bannerman, B. (2012). *Stations Metadata and WMO Core Profile, A Way Forward (discussion paper).* Retrieved May 20, 2014, from http://www.wmo.int/pages/prog/wcp/wcdmp/documents/etcdms-metadata-discussionpaper.pdf

Dornblut, I. (2013). *Hydrologic Information - Metadata. Semantic structure for the description of hydrologic data (GRDC Hydrologic Metadata).* Koblenz, Germany: Federal Institute of Hydrology (BfG).

Fenix, J. L. (2006, June 8). *The National Weather Service Gateway: A history in communications technology evolution.* Retrieved May 20, 2014, from NOAA History: http://www.history.noaa.gov/stories_tales/gateway.html

GCOS. (2003). *GCOS Monitoring Principles*. Retrieved May 20, 2014, from http://www.wmo.int/pages/prog/gcos/index.php?name=ClimateMonitoringPrinciples

ISO/TC 211. (2003). *ISO 19115:2003, Geographic information – Metadata.* Oslo, Norway: International Standards Organization.

ISO/TC 211. (2011). *ISO19156:2011, Geographic Information - Observations and Measurements.* Geneva, Switzerland: International Standards Organization.

R Development Core Team. (2011). A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

WMO. (2003). *Fourteenth World Meteorological Congress: Abridged Final Report with Resolutions (WMO-No. 960).* Geneva, Switzerland: World Meteorological Organization.

WMO. (2007). *Fifteenth World Meteorological Congress: Abridged Final Report with Resolutions (WMO-No. 1026).* Geneva, Switzerland: World Meteorological Organization.

WMO. (2013). *Manual on the WMO Information System: Annex VII to the WMO Technical Regulations.* Geneva, Switzerland: World Meteorological Organization.

WMO. (2013, January 15). *WMO Core Metadata Profile version 1.3: Specification. Part 1 – Conformance Requirements. C.1.3-Part 1 to the Manual on the WMO Information System (WMO-No. 1060).* Geneva, Switzerland: World Meteorological Organization.

WMO. (2014, May 19). *Appendix 2.3 (The WIGOS Metadata Standard) to the Manual on WIGOS, draft version 2014-05-19.* Retrieved May 26, 2014, from Task Team on WIGOS Regulatory Material (TT-WRM) of ICG-WIGOS (Updated 20 May 2014): http://www.wmo.int/pages/prog/www/wigos/TT-WRM.html