Guide to the WMO Information System Volume II (DRAFT) - WMO Information System 2.0

World Meteorological Organization

Date: 2024-07-22

Version: 2024.dev

Document location: https://community.wmo.int/wis2-guide

Standing Committee on Information Management and Technology (SC-IMT)[1]

Commission for Observation, Infrastructure and Information Systems (INFCOM)[2]

Copyright © 2024 World Meteorological Organization (WMO)

Table of Contents

Introduction . . . 4
Purpose . . . 4
PART I . . . 5
1.1 Introduction to WIS2 . . . 5
1.1.1 Leveraging open standards . . . 5
1.1.2 Simpler data exchange . . . 5
1.1.3 Cloud-ready solutions . . . 6
1.1.4 Why are datasets so important? . . . 6
1.2 Data consumer . . . 8
1.2.1 How to search the Global Discovery Catalogue to find datasets . . . 8
1.2.2 How to subscribe to notifications about the availability of new data . . . 8
1.2.3 How to use a notification message to decide whether to download data . . . 9
1.2.4 How to download data . . . 10
1.2.5 How to use data . . . 11
1.2.6 Further reading for data consumers . . . 11
1.3 Data publisher . . . 12
1.3.1 How to get started . . . 12
1.3.2 How to provide discovery metadata to WIS2 . . . 12
1.3.3 How to provide data to WIS2 . . . 13
1.3.4 Further reading for data publishers . . . 19
PART II . . . 22
2.1 WIS2 architecture . . . 22
2.2 Roles in WIS2 . . . 22
2.2.1 Data publisher . . . 22
2.2.2 Global coordinator . . . 23
2.2.3 Global Service operator . . . 23
2.2.4 Data consumer . . . 23
2.3 Specifications of WIS2 . . . 23
2.4 Components of WIS2 . . . 24
2.4.1 WIS2 Node . . . 24
2.4.2 Global Broker . . . 24
2.4.3 Global Cache . . . 25
2.4.4 Global Discovery Catalogue . . . 25
2.4.5 Global Monitor . . . 25
2.5 Protocols configuration . . . 26
2.5.1 Publish-subscribe protocol (MQTT) . . . 26
2.5.2 Download protocol (HTTP) . . . 26
2.6 Implementation and operation of a WIS2 Node . . . 26

2.6.1 Practices and procedures . . . 26
2.6.2 Performance management . . . 30
2.6.3 WIS2 Node reference implementation: WIS2 in a box . . . 30
2.7 Implementation and operation of a Global Service . . . 31
2.7.1 Procedure for registration of a new Global Service . . . 31
2.7.2 Performance management and monitoring of a Global Service . . . 32
2.7.3 Global Broker . . . 34
2.7.4 Global Cache . . . 35
2.7.5 Global Discovery Catalogue . . . 38
2.7.6 Global Monitor . . . 40
2.8 Operations . . . 40
2.8.1 Interoperability with external systems . . . 41
PART III . . . 49
3.1 Information management . . . 49
3.1.1 Introduction . . . 49
3.1.2 Principles of information management . . . 50
3.1.3 The information management life cycle . . . 52
3.1.4 Other considerations . . . 57
PART IV . . . 58
4.1 Security . . . 58
PART V . . . 59
5.1 Competencies . . . 59

Introduction

Purpose

In conjunction with the *Manual on the WMO Information System* (WMO-No. 1060), Volume II – WMO Information System 2.0 (*Manual on WIS*, Volume II), the present Guide to the WMO Information System, Volume II – WMO Information System 2.0 *(Guide to WIS*, Volume II) is designed to ensure adequate uniformity and standardization in the data, information and communication practices, procedures and specifications employed by Members of the World Meteorological Organization (WMO) in the operation of the WMO Information System WIS 2.0 as it supports the mission of the Organization. The Manual on WIS, Volume II contains standard and recommended practices, procedures and specifications. The Guide to WIS contains additional information concerning practices, procedures and specifications that Members are invited to follow or implement in establishing and conducting their arrangements in compliance with the WMO technical regulations and in developing meteorological and hydrological services.

[1] https://community.wmo.int/governance/commission-membership/commission-observation-infrastructures-and-information-systems-infcom/commission-infrastructure-officers/infcom-management-group/standing-committee-information-management-and-technology-sc-imt

[2] https://community.wmo.int/governance/commission-membership/infcom

PART I

1.1 Introduction to WIS2

Since the Global Telecommunication System (GTS) entered operational life in 1971, it has been a reliable real-time exchange mechanism of essential data for WMO Members.

In 2007, the WMO Information System (WIS) entered operations to complement the GTS, providing a searchable catalogue and a Global Cache to enable additional discovery, access and retrieval. The success of WIS was limited as the system only partially met the requirement of providing simple access to WMO data. Today's technology developed for the Internet of Things (IoT) opens the possibility of creating a WIS2 that is able to stand to its expectations of delivering an increasing number and volume of real-time data to WMO centres in a reliable and cost -effective way.

WIS2 has been designed to meet the shortfalls of the current WIS and GTS, support Resolution 1 (Cg-Ext(2021)) – WMO Unified Policy for the International Exchange of Earth System Data (*World Meteorological Congress: Abridged Final Report of the Extraordinary Session* (WMO-No. 1281)), support the Global Basic Observing Network (GBON) and meet the demand for high data volume, variety, velocity and veracity.

WIS2 technical framework is based around three foundational pillars: leveraging open standards, simpler data exchange and cloud-ready solutions.

1.1.1 Leveraging open standards

WIS2 leverages open standards to take advantage of the ecosystem of technologies available on the market and avoid building bespoke solutions that can force National Meteorological and Hydrological Services (NMHS) to procure costly systems and equipment. In today's standards development ecosystem, standards bodies work closely together to minimize overlap and build on one another's areas of expertise. For example, the World Wide Web Consortium provides the framework of web standards, which the Open Geospatial Consortium and other standards bodies leverage. WIS2 leverages open standards with industry adoption and wider, stable and robust implementations, thus extending the reach of WMO data sharing and lowering the barrier to access by Members.

1.1.2 Simpler data exchange

WIS2 prioritizes public telecommunication networks, unlike private networks for GTS links. As a result, using the Internet will enable the best choice for a local connection, using commonly available and well-understood technology.

WIS2 aims to improve the discovery, access and utilization of weather, climate and water data by adopting web technologies proven to provide a truly collaborative platform for a more participatory approach. Data exchange using the Web also facilitates easy access mechanisms. Browsers and search engines allow web users to discover data without specialized software. The Web also enables additional data access platforms, such as desktop Geographical Information Systems (GIS), mobile applications and forecaster workstations. The Web provides access control and security mechanisms that can be utilized to freely share core data as defined by Resolution 1

(Cg-Ext(2021)) and protect the data with more restrictive licensing constraints. Web technologies also allow for authentication and authorization for the provider to retain control of who can access published resources and to request users to accept a license specifying the terms and conditions for using the data as a condition for providing access to them.

WIS2 uses a "publish-subscribe" pattern where users subscribe to a topic to receive new data in real-time. The mechanism is like WhatsApp and other messaging applications. It is a reliable and straightforward way to allow the user to choose her data of interest and to receive them reliably.

1.1.3 Cloud-ready solutions

The cloud provides reliable platforms for data sharing and processing. It reduces the need for expensive local IT infrastructure, which constitutes a barrier to developing effective and reliable data processing workflows for some WMO Members. WIS2 encourages WMO centres to adopt cloud technologies where appropriate to meet their users' needs. While WMO technical regulations will not mandate cloud services, WIS2 will promote a gradual adoption of cloud technologies that provide the most effective solution.

The cloud-based infrastructure allows easy portability of technical solutions, ensuring that a system implemented by a specific country can be packaged and deployed easily in other countries with similar needs. In addition, using cloud technologies allows WIS2 to deploy infrastructure and systems efficiently with minimum effort for the NMHSs by shipping ready-made services and implementing consistent data processing and exchange techniques.

It should be clear that hosting data and services on the cloud does not affect data ownership. Even in a cloud environment, organizations retain ownership of their data, software, configuration and change management as if they were hosting their infrastructure. As a result, data authority and provenance stay with the organization, and the cloud is simply a technical means to publish the data.

1.1.4 Why are datasets so important?

WMO enables the international exchange of observations and model data for all Earth-system disciplines.

Resolution 1 (Cg-Ext(2021)) describes the Earth system data that are necessary for efforts to monitor, understand and predict the weather and climate – including the hydrological cycle, the atmospheric environment and space weather.

WIS is the mechanism by which this Earth system data is exchanged.

A common practice when working with data is to group them into "datasets". All the data in a dataset share some common characteristics. The Data Catalog Vocabulary (DCAT) defines a dataset as a "collection of data, published or curated by a single agent, and available for access or download in one or more representations" [1].

Why is this important? The "single agent" (such as a single organization) responsible for managing the collection ensures consistency among the data. For example, in a dataset:

- All the data should be of the same type (for example, observations from weather stations).

- All the data should have the same license and/or usage conditions.
- All the data should be subject to the same quality management regime - which may mean that all the data is collected or created using the same processes.
- All the data should be encoded in the same way (such as, using the same data formats and vocabularies).
- All the data should be accessible using the same protocols - ideally from a single location.

This consistency means that one can predict what data is in a dataset, at least as far as the common characteristics, making it easier to write applications to process the data.

A dataset might be published as an immutable resource (such as, data collected from a research programme), or it might be routinely updated (for example, every minute as new observations are collected from weather stations).

A dataset may be represented as a single, structured file or object (for example, a CSV file where each row represents a data record) or as thousands of consistent files (for example, output from a reanalysis model encoded as many thousands as possible of General Regularly-distributed Information in Binary form (GRIB) files). Determining the best way to represent a dataset is beyond the scope of this Guide – there are many factors to consider. The key point here is that we consider the dataset to be a single, identifiable resource irrespective of how it is represented.

Because we group data into a single, conceptual resource (that is, the dataset) we can: * Give this resource an identifier and use this identifier to unambiguously refer to collections of data; * Make statements about the dataset (that is, metadata) and infer that these statements apply to the entire collection.

The dataset concept is central to WIS: * We publish discovery metadata about datasets, as specified in the *Manual on WIS*, Volume II – Appendix F: WMO Core Metadata Profile; * We can search for datasets that contain relevant data using the Global Discovery Catalogue (see 2.4.4 Global Discovery Catalogue); * We can subscribe to notifications about updates about a dataset via a Global Broker (see 2.4.2 Global Broker); * We can access the data that comprises a dataset from a single location using a well -described mechanism.

It is up to the data publisher to decide how their data is grouped into datasets – effectively, to decide what datasets they publish to WIS. That said, we recommend that, subject to the consistency rules above, data publishers should organize their data into as few datasets as possible.

For a data publisher, this means fewer discover metadata records to maintain. For a data consumer this means fewer topics to subscribe to and fewer places to access the data.

There are some things that are fixed requirements for datasets:

1. All data in the dataset must be accessible from a single location.
2. All data in the dataset must be subject to the same license or usage conditions.

Here are some examples of datasets:

- The most recent 5-days of synoptic observations for an entire country or territory [2].

- Longterm record of observed water quality for a managed set of hydrological stations.
- Output from the most recent 24-hours of operational numerical weather prediction model runs.
- Output from 6-months of experimental model runs. It is important to note that output from the operational and experimental should not be merged into the same dataset because they use different algorithms - it is very useful to be able to distinguish the provenance (or lineage) of data.
- A multipetabyte global reanalysis spanning 1950 to present day.

In summary, datasets are important because they are how data is managed in WIS.

1.2 Data consumer

As a data consumer wanting to use data published via WIS2 you should read the guidance presented here. In addition, a list of references to informative material in this Guide and elsewhere is provided at the end of this section.

1.2.1 How to search the Global Discovery Catalogue to find datasets

The first step to using data published via WIS2 is to determine which dataset or datasets contains the data that is needed. To do this, a data consumer may browse discovery metadata provided by the Global Discovery Catalogue. Discovery metadata follows a standard scheme (see *Manual on WIS*, Volume II – Appendix F: WMO Core Metadata Profile). A data consumer may discover a dataset using keywords, geographic area of interest, temporal information, or free text. Matching search results from the Global Discovery Catalogue provide high-level information (title, description, keywords, spatiotemporal extents, data policy, licensing, contact information), from which a data consumer can assess and evaluate their interest in accessing/downloading data associated with the dataset record.

A key component of dataset records in the Global Discovery Catalogue is that of "actionable" links. A dataset record provides one to many links that clearly identify the nature and purpose of the link (informational, direct download, API, subscription) so that the data consumer can interact with the data accordingly. For example, a dataset record may include a link to subscribe to notifications (see 1.2.2 How to subscribe to notifications about the availability of new data) about the data, or an API, or an offline archive retrieval service.

The Global Discovery Catalogue is accessible via an API and provides a low barrier mechanism (see [_2_2_4_global_discovery_catalogue]). Internet search engines are able to index the discovery metadata in the Global Discovery Catalogue, thereby providing data consumers with an alternative means to search for WIS2 data.

1.2.2 How to subscribe to notifications about the availability of new data

WIS2 provides notifications about updates to datasets; for example, when a new observation record from an automatic weather station is added to a dataset of surface observations. Notifications are published on message brokers. Where data consumers need to use data rapidly once it has been published (such as input to a weather prediction model), they should subscribe to one or more Global Broker to get notifications messages using Message Queuing Telemetry Transport (MQTT)

protocol[3].

In WIS2, notifications are republished by Global Brokers to ensure resilient distribution. Consequently, there will be multiple places where one can subscribe. Data consumers requiring real-time notifications must subscribe to Global Brokers. A data consumer should subscribe to more than one Global Broker, thereby ensuring that notifications continue to be received if a Global Broker instance fails.

A dataset in WIS2 is associated with a unique *topic*. Notifications about updates to a dataset are published to the associated topic. Topics are organized according to a standard scheme (see the *Manual on WIS*, Volume II - Appendix D: WIS2 Topic Hierarchy).

A data consumer can find the appropriate topic to subscribe to either by searching the Global Discovery Catalogue, using an Internet search engine[4], or by browsing the topic hierarchy on a Message Broker.

WIS2 uses Global Caches to distribute core data, as defined in the Unified Data Policy (Resolution 1 (Cg-Ext (2021))). Each Global Cache republishes core data on its own highly available data server and publishes a new notification message advertising the availability of that data from the Global Cache location.

Notifications from WIS2 Nodes and Global Caches are published on different topics: The root topic used by WIS2 Nodes is `origin`, while the root topic used by Global Caches is `cache`. Other than the root, the topic hierarchy is identical. For example, for synoptic weather observations published by Environment Canada:

- Environment and Climate Change Canada, Meteorological Service of Canada's WIS2 Node publishes to: `origin/a/wis2/ca-eccc-msc/data/core/weather/surface-based-observations/synop`
- Global Caches publish to: `cache/a/wis2/ca-eccc-msc/data/core/weather/surface-based-observations/synop`

As per clause 3.2.13 of the *Manual on WIS*, Volume II, data consumers should access core data from the Global Caches. Consequently, they need to subscribe to the `cache` topic hierarchy to receive the notifications from Global Caches, each of which provides a link (that is, URL) to download from the respective Global Cache's data server.

1.2.3 How to use a notification message to decide whether to download data

On receipt of a notification message, a data consumer needs to decide whether to download the newly available data. The content of the notification message provides the information needed to make this decision. For details of the specification, see the *Manual on WIS*, Volume II - Appendix E: WIS2 Notification Message.

In many cases, data consumers will use a software application to determine whether or not to download the data. This section provides insight about what happens.

When subscribing to multiple Global Brokers the data consumer will receive multiple copies of a notification message. Each notification message has a unique identifier, defined using the `id` property. Duplicate messages should be discarded.

The core data will be available from both a WIS2 Node and Global Caches, each of which publishes a different notification message advertising an alternative location from where the data can be downloaded. Because these are different messages, they will have different identifiers. However, each of these messages refers to the same data object, which is uniquely identified in the notification message using the data_id property. Notification messages from different sources can easily be compared to determine if they refer to the same data. By subscribing to the cache root topic, data consumers will only receive notifications about data available from the Global Caches. The origin root topic should be used when subscribing to notifications about recommended data. Data consumers should not subscribe to the origin root topic for notifications about core data because notification messages provided on these topics will refer to data published directly on the WIS2 Nodes (referred to as, the "origin").

Data consumers need to consider their strategy for managing these duplicate messages. From a data perspective, it does not matter which Global Cache instance is used – they will all provide an identical copy of the data object published by the originating WIS2 Node. The simplest strategy is to accept the first notification message and download it from the Global Cache instance that the message refers to by using a URL for the data object at that Global Cache instance. Alternatively, a data consumer may have a preferred Global Cache instance, for example, that is located in their region. Whichever Global Cache instance is chosen, data consumers will need to implement logic to discard duplicate notification messages based on `id` and duplicate data objects based on `data_id`.

A notification message also provides a small amount of metadata about the data object it references such as location and time. Data consumers can use this metadata to decide if the data object referenced in the message should be downloaded. This is known as client-side filtering.

The notification message should also include the metadata identifier for the dataset to which the data object belongs. A data consumer can use the metadata identifier to search the Global Discovery Catalogue and discover more about the data - in particular, whether there are any conditions on the use of this data.

1.2.4 How to download data

Links to where data can be accessed are made available through dataset discovery metadata (via the Global Discovery Catalogue) and/or data notification messages (via Global Brokers). Links can be used to directly download the data (according to the network protocol and content description provided in the link) using a mechanism appropriate to the workflow of the data consumer. This could include web and/or desktop applications, custom tooling, or other approaches.

A discovery metadata record or notification message may provide more than one download link. The preferred link will be identified as "canonical" (link relation: "rel": "canonical" [5]).

Where data is provided through an interactive web service, a canonical link that provides a URL where one can directly download a data object may be complemented with an additional link providing the URL for the root of the web service from where one can interact with or query the entire dataset.

If a download link implements access control (for example, the data consumer needs to take some additional action(s) to download the data object), the download link will contain a security object that provides the pertinent information (such as the access control mechanism used and

where/how a data consumer would need to register to request access).

1.2.5 How to use data

Data is shared on WIS2 in accordance with the Unified Data Policy (Resolution 1 (Cg-Ext (2021))). This data policy describes two categories of data: core and recommended.

- Core data is considered essential for the provision of services for the protection of life and property and the well-being of all nations. Core data is provided on a free and unrestricted basis, without charge and with no conditions on use.
- Recommended data is exchanged on WIS2 in support of Earth system monitoring and prediction efforts. Recommended data *may* be provided with conditions on use and/or subject to a license.

Furthermore, the Unified Data Policy (Resolution 1 (Cg-Ext (2021))) encourages attribution of the source of the data in all cases. In this way, credit is given to those who have expended effort and resources in collecting, curating, generating, or processing the data. Attribution provides visibility of who is using data which, for many organizations, provides necessary evidence to justify continued provision of and updates to the data.

Details of the applicable WMO data policy and any rights or licenses associated with data are provided in the discovery metadata that accompanies the data. Discovery metadata records are available from the Global Discovery Catalogue.

The *Manual on WIS*, Volume II – Appendix F: WMO Core Metadata Profile, section 1.18 Properties / WMO data policy provides details on how data policy, rights and/or licenses are described in the discovery metadata.

When using data from WIS2, data consumers:

- Shall respect the conditions of use applicable to the data as expressed in the WMO data policy, rights statements, or licenses.
- Should attribute the source of the data.

1.2.6 Further reading for data consumers

As a data publisher planning to operate a WIS2 Node, as a minimum you should read the following sections:

- 1.1 Introduction to WIS2
- 2.1 WIS2 architecture
- 2.2 Roles in WIS2
- 2.4 Components of WIS2

The following specifications in the *Manual on WIS*, Volume II are useful for further reading:

- Appendix D: WIS2 Topic Hierarchy
- Appendix E: WIS2 Notification Message

- Appendix F: WMO Core Metadata Profile

1.3 Data publisher

As a data publisher with authoritative Earth system data that you want to share with the WMO community you should read the guidance presented here. In addition, a list of references to informative material in this Guide and elsewhere is provided at the end of this section.

1.3.1 How to get started

The first thing you need to do is consider your data, how it can be conceptually grouped into one or more datasets (see [_1_1_4_why_are_datasets_so_important?]), and whether it is core or recommended data, as per the Unified Data Policy (Resolution 1 (Cg-Ext(2021))) .

Next, you need to consider where it is published. If your data relates to your country or territory, you need to publish it through a National Centre (NC). If your data relates to a region, programme, or other specialized function within WMO, you need to publish it through a Data Collection or Production Centre (DCPC). The functional requirements for NC and DCPC are described in the *Manual on WIS*, Volume II - Part III Functions of WIS.

All NCs and DCPCs are affiliated with a Global Information System Centre (GISC) that has a responsibility to help establish efficient and effective data sharing on the WIS. Your GISC will be able to help you in getting your data onto WIS2.

You may be able to identify an existing NC or DCPC that can publish your data. Alternatively, you may need to establish a new NC or DCPC. The main difference is that an NC is designated by a Member, whereas a DCPC is designated by a WMO or related international programme and/or a regional association.

Both NC and DCPC require the operation of a WIS2 Node (see [_2_4_2_wis2_node]). The procedure for registering a new WIS2 Node is provided in 2.6.1.1 Registration and decommissioning of a WIS2 Node.

Once you have determined the scope of your datasets, the data policy that applies, and have a WIS2 Node ready for data publication, you are ready to progress to the next step: providing discovery metadata.

1.3.2 How to provide discovery metadata to WIS2

Discovery metadata is the mechanism by which you tell potential consumers about your data, how it may be accessed, and any conditions you may place on the use of the data.

Each dataset you want to publish must have an associated discovery metadata record. This record is encoded as GeoJSON (RFC 7946[6]) must conform to the specification given in the *Manual on WIS*, Volume II - Appendix F: WMO Core Metadata Profile.

Copies of all discovery metadata records from WIS2 are held at the Global Discovery Catalogues, where data consumers can search and browse to find data that is of interest to them.

Depending on local arrangements, your GISC may be able to help you transfer your discovery metadata record(s) to the Global Discovery Catalogues. If this is not the case, you will need to publish the discovery metadata record(s) yourself[7] using one of two ways:

- The simplest method is to encode the discovery metadata record as a file and publish it to an HTTP server where it can be accessed with a URL.
- Alternatively, you may operate a local metadata catalogue through which discovery metadata records can be shared using an API (for example, Open Geospatial Consortium (OGC) API – Records[8]). Each discovery metadata record can be accessed with a unique URL via the API (for instance, an item that is part of the discovery metadata catalogue).

In both cases, a notification message needs to be published on a Message Broker that tells WIS2 there is a new discovery metadata to upload and that it is accessed at the specified URL[9]. The notification messages shall conform to the specification given in the *Manual on WIS*, Volume II - Appendix E: WIS2 Notification Message. Furthermore, the notification message must be published on a topic that conforms to the specification given in *Manual on WIS*, Volume II - Appendix D: WIS2 Topic Hierarchy. For example, metadata published by Deutscher Wetterdienst would use the following topic: `origin/a/wis2/de-dwd/metadata/core`.

These discovery metadata records are then propagated through the Global Service components into to the Global Discovery Catalogue where data consumers can search and browse for datasets of interest.

Upon receipt of a new discovery metadata record, a Global Discovery Catalogue (see 2.4.4 Global Discovery Catalogue) will validate, assess, ingest, and publish the record. Validation ensures that your discovery metadata record complies with the specification. The assessment examines your discovery metadata record against good practice. The Global Discovery Catalogue will notify you if your discovery metadata record fails validation and provide recommendations for improvements for you to consider.

Discovery metadata must be published in the Global Discovery Catalogues before you begin publishing data.

1.3.3 How to provide data to WIS2

WIS2 is based on the web architecture[10]. As such it is *resource oriented*. Datasets are resources; the "granules" of data grouped in a dataset are resources; the discovery metadata records that describe datasets are resources. In web architecture, every resource has a unique identifier (such as a URI[11]), and the unique identifier can be used to resolve the resource identified and interact with it (for example, to download a representation of the resource over an open standard protocol such as HTTP).

Simply, you provide data (and metadata) to WIS2 by assigning it a unique identifier, in this case a URL[12], and make it available via a data server - most typically a web server using the HTTP protocol[13]. It is up to the data server to decide what to provide when resolving the identifier: the URL of a data granule may resolve as a representation encoded in a given data format, whereas the URL of a dataset may resolve as a description of the dataset (that is, metadata) that includes links to access the data from which the dataset is comprised - either individual files (that is, the data granules) or an interactive API that enables a user to request just the parts of the dataset they need

by specifying query parameters.

The following sections cover specific considerations relating to publishing data to WIS2.

1.3.3.1 Data formats and encodings

Whether providing data as files or through interactive APIs you need to decide which encodings (in other words, *data formats*) to use. WMO technical regulations may require that data be encoded in specific formats. For example, synoptic observations must be encoded in Binary universal form for the representation of meteorological data (BUFR). The *Manual on Codes* (WMO-No. 306) provides details of data formats formally approved for use in WMO. However, technical regulations do not cover all data sharing requirements. In such cases, you should select data formats that are open, non-proprietary, widely adopted and understood in their target user community. In this context, “open” means that anyone can use the format without needing a license to do so – either to encode data in that format or write software that understands the format.

1.3.3.2 Providing data as files

The simplest way to publish data through WIS2 is to persist your data as files and publish those files on a web server. All these files need to be organized somehow – perhaps in a flat structure or grouped into collections that resemble folders or directory structures.

To make your data usable, your users need to be able to find the specific file (or files) they need.

Naming conventions for files and/or directories are useful – but only when the scheme is understood. If users do not understand the naming convention, it will be a barrier to widespread reuse; many users will simply treat the filename as an opaque string. Where communities commonly use file-naming conventions (such as names with embedded metadata), these should only be used when adequate documentation is provided to users.

WIS2 does not require the use of specific naming conventions.

Another mechanism to consider is complementing the collections (such as, directories or folders in which files are grouped) with information that describes their content. Then users,both humans and software agents, can browse the structure and find what they need. Examples of this approach include:

- Web Accessible Folders (WAF) and "README" files: a web-based folder structure listing the data object files by name, where each folder contains a formatted "README" file describing the folder contents.
- SpatioTemporal Asset Catalog (STAC)[14]: a community standard based on GeoJSON to describe geospatial data files that can be easily indexed, browsed and accessed. Free and open source tools present STAC records (one for each data object file) through a web-based, browseable user interface.

When publishing collections of data it is tempting to package content into zip or submission information package (SIP)[15] resources - perhaps even packaging the entire collection, complete with folders, into a single resource. Similarly, WMO formats such as GRIB and BUFR allow multiple data objects (such as, fields or observations) to be packed into a single file. Only having to download a single resource is convenient for many users, but the downside is that the user must download

the entire resource and then unpack/decompress it. The convenience of downloading fewer resources needs to be balanced against the cost of forcing users to download data they may not need. Whatever your choice, you should be guided by common practice in your domain - for example, only zip, SIP, or pack if your users expect it.

1.3.3.3 Providing interactive access to data with APIs

Interactive data access aims to support efficient data workflows by enabling client applications to request only the data that they need. The advantage of interactive data access is that it provides more flexibility. Data publishers can offer an API structured around how users want to work with the data rather than forcing them to work with the structure that is convenient for you as a data publisher.

But it is more complex to implement. You need a server running software that can:

1. Interpret a user's request;
2. Extract the data from wherever it is stored;
3. Package that data up and send it back to the user.

Importantly, when considering the use of interactive APIs to serve your data you need to plan for costs: every request to an interactive API requires computational resources to process.

Based on the experience of data publishers who have been using web APIs to serve their communities, this Guide makes the following recommendations about interactive APIs:

- First, interactive APIs should be self-describing. A data consumer should not need to know, a priori, how to make requests from an API. They should be able to discover this information from the API endpoint itself – even if this is just a link to a documentation page they need to read.
- Second, APIs should comply with OpenAPI[16] version 3 or later. OpenAPI provides a standardized mechanism to describe the API. Tooling (free, commercial, etc.) is widely available that can read this metadata and automatically generate client applications to query the API.
- Third, the OGC has developed a suite of APIs[17] (called "OGC APIs") that are designed specifically to provide APIs for geospatial data workflows (discovery, visualization, access, processing/exploitation) – all of which build on OpenAPI. Among these, OGC API – Environmental Data Retrieval (EDR)[18], OGC API – Features[19], and OGC API - Coverages[20] are considered particularly useful. Because these are open standards, there is an ever-growing suite of software implementations (both free and proprietary) that support them. We recommend that data publishers assess these open-standard API specifications to determine their suitability for publishing their datasets using APIs.

Finally, you should consider versioning your API to avoid breaking changes when adding new features. A common approach is adding a *version number* prefix into the API path; for example, `/v1/service/{rest-of-path}` or `/service/v1/{rest-of-path}`.

More guidance on the use of interactive APIs in WIS2 is anticipated in future versions of this Guide.

1.3.3.4 Providing data in (near) real-time

WIS2 is designed to support the data sharing needs of all WMO disciplines and domains. Among these, the World Weather Watch [21] drives specific needs for the rapid exchange of data to support weather forecasting.

To enable real-time data sharing[22], WIS2 uses notification messages to advertise the availability of a new resource, either data or discovery metadata, and how to access that resource. Notification messages are published to a queue on a Message Broker in your WIS2 Node[23] using the MQTT protocol and immediately delivered to everyone subscribing to that queue. A queue is associated with a specific *topic*, such as a dataset.

For example, when a new temperature profile from a radio sonde deployment is added to a dataset of upper-air data measurements, a notification message will be published that includes the URL used to access the new temperature profile data. Everyone subscribing to notification messages about the upper-air measurement dataset would receive the notification message, identify the URL and download the new temperature profile data.

Optionally, data may be embedded in a notification message using a `content` object *in addition* to publishing via the data server. Inline data must be encoded as `UTF-8`, `Base64`, or `gzip`, and must not exceed 4096 bytes in length once encoded.

Notification messages are encoded as GeoJSON (RFC 7946) and must conform to the *Manual on WIS*, Volume II, Appendix E: WIS2 Notification Message.

The URL used in the notification message should refer only to the newly added data object (for example, the new temperature profile) rather than the entire dataset. However, the WIS2 Notification Message specification allows for multiple URLs to be provided. If you are providing your data through an interactive API, you might provide a "canonical" link (designated with link relation: `"rel": "canonical"`[24]), and an additional link providing the URL for the root of the web service from where one can interact with or query the entire dataset.

You should include the dataset identifier in the notification message (`metadata_id` property). This allows data consumers receiving the notification to cross reference with information provided in the discovery metadata for the dataset, such as the conditions of use specified in the data policy, rights, or license.

Furthermore, if you have implemented controlled access to your data (such as, the use of an API key), you should include a security object in the download link that provides the pertinent information (for example, the access control mechanism used, and where or how a data consumer would need to register to request access).

To ensure that data consumers can easily find the topics they want to subscribe to, data publishers must publish to an authorized topic, as specified in the *Manual on WIS*, Volume II, Appendix D: WIS2 Topic Hierarchy.

If your data seems to relate to more than one topic, select the most appropriate one. The topic hierarchy is not a knowledge organization system - it is only used to ensure the uniqueness of topics for publishing notification messages. Discovery metadata is used to describe a dataset and its relevance to additional disciplines; each dataset is mapped to one, and only one, topic.

If the WIS2 Topic Hierarchy does not include a topic appropriate for your data, you should publish on an *experimental* topic. This allows for data exchange to be established while the formalities are considered[25]. Experimental topics are provided for each Earth-system discipline at level eight in the topic hierarchy (for example, `origin/a/wis2/{centre-id}/data/{earth-system-discipline}/experimental/`). Data publishers can extend the experimental branch with subtopics as they deem appropriate. Experimental topics are subject to change and will be removed once they are no longer needed. For more information, see *Manual on WIS*, Volume II, Appendix D: WIS2 Topic Hierarchy, section 1.2 Publishing guidelines.

Whatever topic is used, the discovery metadata provided to the Global Discovery Catalogue must include subscription links using that topic[26]. The Global Broker will only republish notification messages on topics specified in the discovery metadata records.

1.3.3.5 Considerations when providing core data in WIS2

Core data, as specified in the Unified Data Policy (Resolution 1 (Cg-Ext 2021)) is considered essential for the provision of services for the protection of life and property and for the well-being of all nations. Core data is provided on a free and unrestricted basis, without charge and with no conditions on use.

WIS2 ensures highly available, rapid access to *most* core data via a collection of Global Caches (see 2.4.3 Global Cache). Global Caches subscribe to notification messages about the availability of new core data published at WIS2 Nodes, download a copy of that data and republish it on a high-performance data server and then discard it after the retention period expires (normally after 24-hours[27]). Global Caches do not provide sophisticated APIs. They publish notification messages advertising the availability of data on their caches and allow users to download data via HTTPS using the URL in the notification message.

The URL included in a notification message that is used to access core data from a WIS2 Node, or the "canonical" URL if multiple URLs are provided, must:

1. Refer to an individual data object; and
2. Be directly resolvable, such that the data object can be downloaded simply by resolving the given URL without further action.

A Global Cache will download and cache the data object accessed via this URL.

The Global Caches are designed to support Members efficiently share real-time and near real-time data; they take on the task of making sure that core data is available to all on a free and unrestricted basis, as required by the WMO Unified Data Policy (Resolution 1 (Cg-Ext(2021))).

Unfortunately, Global Caches cannot republish *all* core data: there is a limit to how much data they can afford to serve. Currently, a Global Cache is expected to cache about 100 GB of core data each day.

If frequent updates to your dataset are very large (for example, weather prediction models or remote sensing observations) you will need to share the burden of distributing your data with the Global Cache operators. You should work with your GISC to determine the highest priority elements of your core datasets that will be republished by the Global Caches.

For core data that is not to be cached, you must set the `cache` property in the notification message to `false`[28].

You must ensure that core data that is not cached is publicly accessible from your WIS2 Node, that is, with no access control mechanisms in place.

A Global Cache operator may choose to disregard your cache preference - for example, if they feel that the content you are providing is large enough to impede the provision of caching services for other Members[29]. In such cases, the Global Cache operator will log this behaviour. In collaboration with the Global Cache operators, your GISC will work with you to resolve concerns.

Finally, please note that Global Caches are under no obligation to cache data published on *experimental* topics. For such data, the `cache` property should be set to `false`.

1.3.3.6 Implementing access control

Recommended data, as defined in the WMO Unified Data Policy (Resolution 1 (Cg-Ext(2021)), is exchanged on WIS2 in support of Earth system monitoring and prediction efforts and *may* be provided with conditions on use. This means that you may control access to recommended data.

Access control should use only the "security schemes" for authentication and authorization specified in OpenAPI[30].

Where access control is implemented, you should include a `security` object in download links provided in discovery metadata and notification messages that provide the user with pertinent information about the access control mechanism used and where/how they might register to request access.

Recommended data is never cached by the Global Caches.

Use of core data must always be free and unrestricted. However, you may need to leverage existing systems with built-in access control when implementing the download service for your WIS2 Node.

Example 1: API key. Your data server requires a valid API key to be included in download requests. The URLs used in notification messages should include a valid API key.[31][32]

Example 2: Presigned URLs. Your data server uses a cloud-based object store that requires credentials to be provided when downloading data. The URLs used in notification messages should be *presigned* with the data publisher's credentials and valid for the cache retention period (for example, 24-hours).[33]

In both cases, the URL provided in a notification message can be directly resolved without a user, or Global Cache, needing to take additional action such as providing credentials or authenticating.

Finally, note that if you are only publishing core data, you may be able to entirely rely on the Global Caches to distribute your data. In such cases, your WIS2 Node may use Internet Protocol (IP)filtering to allow access only from Global Services. For more details, see section 2.6 Implementation and operation of a WIS2 Node.

1.3.3.7 Providing access to data archives

There is no requirement for a WIS2 Node to publish notification messages about newly available data, however, the mechanism is available if needed (for instance, for real-time data exchange). Data archives published through WIS2 do not need to provide notification messages for data unless the user community has expressed a need to be rapidly notified about changes (for example, the addition of new records into a climate observation archive).

However, notification messages must still be used to share discovery metadata with WIS2. Given that the provision of metadata and subsequent updates is likely to be infrequent, it may be sufficient to "handcraft" a notification message and publish it locally on an MQTT broker[34] or with help from a GISC. See above for more details on publishing discovery metadata to WIS2.

Note that some data archives are categorized as core data; for example, Essential Climate Variables. Core data may be distributed via the Global Caches. However, given that these provide only short-term hosting of data (for instance, 24-hours), Global Caches are not an appropriate mechanism to provide access to archives of core data. The archive must be accessed directly via the WIS2 Node.

1.3.4 Further reading for data publishers

As a data publisher planning to operate a WIS2 Node, as a minimum you should read the following sections:

- 1.1 Introduction to WIS2
- 2.1 WIS2 architecture
- 2.2 Roles in WIS2
- 2.4 Components of WIS2
- 2.6 Implementation and operation of a WIS2 Node

The following sections are useful for further reading:

- 3.1 Information management
- 4.1 Security
- 5.1 Competencies

Note that sections *4.1. Security* and *5.1. Competencies* reference content originally published for WIS1. These remain largely applicable and will be updated in subsequent releases of this Guide.

If you are publishing aviation weather data via WIS2 for onward transmission through the International Civil Aviation Organization (ICAO) System Wide Information Management (SWIM), you should also read: * [_2_8_8_1_publishing_aviation_weather_data_through_wis2_into_icao_swim].

Finally, you should also review the specifications in the *Manual on WIS*, Volume II:

- Appendix D: WIS2 Topic Hierarchy
- Appendix E: WIS2 Notification Message

- Appendix F: WMO Core Metadata Profile 2

[1] Data Catalog Vocabulary (DCAT) - Version 2, W3C Recommendation 04 February 2020 https://www.w3.org/TR/vocab-dcat-2/#Class:Dataset

[2] Why 5-days in this example? Because the system used to publish the data in this example only retains data for 5-days

[3] Subscribing to notifications about newly available data means that you don't need to continually to poll the data server to check for updates.

[4] Internet search engines allow data consumers to discover WIS2 datasets by indexing the content in the Global Discovery Catalogues.

[5] IANA Link Relations https://www.iana.org/assignments/link-relations/link-relations.xhtml

[6] RFC 7946 - The GeoJSON Format: https://datatracker.ietf.org/doc/html/rfc7946

[7] In the future, WIS2 may provide metadata publication services (such as, through a WIS2 metadata management portal) to assist with this task. However, such a service is not currently available.

[8] OGC API - Records - Part 1: Core https://docs.ogc.org/DRAFTS/20-004.html

[9] Both data and metadata publication use the same notification message mechanism to advertise the availability of a new resource.

[10] Architecture of the World Wide Web https://www.w3.org/TR/webarch/

[11] RFC 3986 - Uniform Resource Identifier (URI) - Generic Syntax: https://datatracker.ietf.org/doc/html/rfc3986

[12] The term "Uniform Resource Locator" (URL) refers to the subset of URIs that, in addition to identifying a resource, provide a means of locating the resource by describing its primary access mechanism (such as its network "location"). RFC 3986

[13] WIS2 strongly prefers secure versions of protocols (such as HTTPS) wherein the communication protocol is encrypted using Transport Layer Security (TLS)

[14] Spatio Temporal Asset Catalogue (STAC) https://stacspec.org/en

[15] See https://www.iasa-web.org/tc04/submission-information-package-sip or end of https://www.eumetsat.int/formats

[16] OpenAPI Specification https://spec.openapis.org/oas/v3.1.0

[17] Open Geospatial Consortium OGC API https://ogcapi.ogc.org/

[18] OGC API - Environmental Data Retrieval (EDR) https://ogcapi.ogc.org/edr

[19] OGC API - Features https://ogcapi.ogc.org/features

[20] OGC API - Coverages https://ogcapi.ogc.org/coverages

[21] WMO World Weather Watch https://wmo.int/world-weather-watch

[22] In the context of WIS2, real-time implies anything from a few seconds to a few minutes - not the milliseconds required by some applications.

[23] WIS2 ensures rapid global distribution of notification messages using a network of Global Brokers which subscribe to message brokers of WIS2 Nodes and republish notification messages (see [_2_4_2_Global_Broker]).

[24] IANA Link Relations https://www.iana.org/assignments/link-relations/link-relations.xhtml

[25] The "experimental" topic is necessary for the WIS2 pre-operational phase and future pre-operational data exchange in test mode.

[26] The Global Discovery Catalogue will reject discovery metadata records containing links to topics outside the official topic-hierarchy.

[27] A Global Cache provides short-term hosting of data. Consequently, it is not an appropriate mechanism to provide access to archives of core data, such as Essential Climate Variables. Providers of such archive data must be prepared to serve such data directly from their WIS2 Node.

[28] Default value for the `cache` property is `true`; omission of the property will result in the data object being cached.

[29] Excessive data volume is not the only reason they may refuse to cache content. Other reasons include: too many small files, unreliable download from a WIS2 Node, etc.

[30] OpenAPI Security Scheme Object: https://spec.openapis.org/oas/v3.1.0#security-scheme-object

[31] A specific API key should be used for data publication via WIS2 so that usage can be tracked.

[32] Given that users are encouraged to download Core data from the Global Cache, there will likely be only a few accesses using the WIS2 account's API key. If the usage quota for the WIS2 account is exceeded (for instance, further data access is blocked) then this should encourage users to download via the Global Cache as mandated in the *Manual on WIS*, Volume II.

[33] Working with presigned URLs on Amazon S3 https://docs.aws.amazon.com/AmazonS3/latest/userguide/using-presigned-url.html

[34] MQTT broker managed services are available online, often with a free (no cost) starter plan sufficient for infrequent publications of notifications about metadata. These provide a viable alternative to implementing an MQTT broker instance

yourself.

PART II

2.1 WIS2 architecture

WIS2 is a federated system of systems based on web-architecture and open standards, comprising of many WIS2 Nodes for publishing data and Global Services that enable fault tolerant, highly available, low latency data distribution.

National Centres (NC), Data Collection or Production Centres (DCPC), and Global Information System Centres (GISC) are all types of WIS centre.

NCs and DCPCs operate WIS2 Nodes.

GISCs coordinate the operation of WIS within their Area of Responsibility (AoR) and ensure the smooth operation of the WIS2 system.

A WIS centre may also operate one or more Global Services.

WIS centres shall comply with the technical regulations defined in the *Manual on WIS*, Volume II.

2.2 Roles in WIS2

When describing the functions of WIS2 there are four roles to consider:

1. Data publisher
2. Global coordinator
3. Global Service operator
4. Data consumer

These roles are outlined below.

2.2.1 Data publisher

- This role is fulfilled by NC and DCPC.
- Data publishers operate a WIS2 Node to share their data within the WIS2 ecosystem.
- Data publishers manage, curate, and provide access to one or more datasets.
- For each dataset, a data publisher provides:
    1. "Discovery metadata" to describe the dataset, provide details on how it can be accessed, and under what conditions.
    2. An API or web-service to access (or interact with) the dataset.
    3. Notification messages advertising the availability of new data and metadata.

2.2.2 Global coordinator

- This role is exclusive to GISCs.
- All GISCs supporting WMO Members in their AoR fulfill their data sharing obligations via WIS2.

2.2.3 Global Service operator

- To ensure highly available global data exchange, a WIS centre may operate one or more Global Services:
  1. Global Discovery Catalogue: enables users to search all datasets provided by data publishers and discover where and how to interact with those datasets (for example, subscribe to updates, access/download/visualize data, or access more detailed information about the dataset).
  2. Global Broker: provides highly available messaging services where users may subscribe to notifications about all datasets provided by data publishers.
  3. Global Cache: provides highly available download service for cached copies of core data downloaded from data publishers' web-services.
  4. Global Monitor: gathers and displays system performance, data availability, and other metrics from all WIS2 Nodes and Global Services.

2.2.4 Data consumer

- This role represents anyone wanting to find, access, and use data from WIS2 – examples include (but are not limited to): NMHS, Government agencies, research institutions, private sector organizations, and so on.
- Searches or browses the Global Discovery Catalogue (or other search engine) to discover the dataset(s) that meet their needs (namely, "datasets of interest").
- Subscribes via the Global Broker to receive notification messages about the availability of data or metadata associated with datasets of interest.
- Determines whether the data or metadata referenced in notification messages is required.
- Downloads data from Global Cache or WIS2 Node.

2.3 Specifications of WIS2

Leveraging existing open standards, WIS2 defines the following specifications in support of publication, subscription, notification and discovery:

Table 1. WIS2 Specifications

Specification	Granularity	Primary WIS2 Component(s)
WMO Core Metadata Profile 2 (WCMP2)	Datasets	Global Discovery Catalogue (GDC)
WIS2 Topic Hierarchy (WTH)	Dataset granules	Global Broker, WIS2 Nodes

Specification	Granularity	Primary WIS2 Component(s)
WIS2 Notification Message	Dataset metadata, dataset granules	Global Broker, WIS2 Nodes

Please refer to the *Manual on WIS*, Volume II for details.

2.4 Components of WIS2

2.4.1 WIS2 Node

- WIS2 Nodes are central to WIS2. These are operated by National Centres (NC) and Data Collection or Production Centres (DCPC) to publish their core and recommended data.
- WIS2 adopts web technologies and open standards enabling WIS2 Nodes to be implemented using freely-available software components and common industry practices.
- WIS2 Nodes publish data as files of a web server or using an interactive web service.
- WIS2 Nodes describe the data they publish using discovery metadata. See the *Manual on WIS*, Volume II, Appendix F: WMO Core Metadata Profile.
- WIS2 Nodes generate notification messages advertising the availability of new data. These notification messages are published to a Message Broker. The WIS2 Topic Hierarchy is used to ensure that all WIS2 Nodes publish to consistent topics. The information in the notification message tells the data consumer where to download data from. Notification messages are also used to advertise the availability of discovery metadata. See the *Manual on WIS*, Volume II - Appendix D: WIS2 Topic Hierarchy and Appendix E: WIS2 Notification Message.
- WIS2 Nodes may implement controlled access for the data they publish. Global Services will operate with fixed IP addresses, enabling WIS2 Nodes to easily distinguish their requests.

2.4.2 Global Broker

- WIS2 incorporates several Global Brokers, ensuring highly resilient distribution of notification messages across the globe.
- A Global Broker subscribes to the Message Broker operated by each WIS2 Node and republishes notification messages.
- A Global Broker subscribes to notifications from other Global Brokers to ensure it receives a copy of all notification messages.
- A Global Broker republishes notification messages from every WIS2 Node and Global Service.
- A Global Broker operates a highly available, high-performance Message Broker.
- A Global Broker uses the WIS2 Topic Hierarchy enabling a data consumer to easily find topics relevant to their needs.
- Data consumers should subscribe to notifications from a Global Broker not directly to the Message Brokers operated by WIS2 Nodes.

2.4.3 Global Cache

- WIS2 incorporates several Global Caches, ensuring highly resilient distribution of data across the globe.
- A Global Cache provides a highly available data server from which a data consumer can download core data, as specified in Resolution 1 (Cg-Ext(2021)).
- A Global Cache subscribes to notification messages via a Global Broker.
- On receipt of a notification message, the Global Cache downloads from the WIS2 Node a copy of data referenced in the notification message, makes this copy available on its data server, and publishes a new notification message advertising the availability of this data at the Global Cache.
- A Global Cache will subscribe to notification messages from other Global Caches enabling it to download and republish data it has not acquired directly from WIS2 Nodes. This ensures that each Global Cache holds data from every WIS2 Node.
- A Global Cache shall retain a copy of core data for a duration compatible with the real-time or near real-time schedule of the data and not less than 24-hours.
- A Global Cache will delete data from the cache once the retention period has expired.
- Data consumers should download data from a Global Cache when available.

2.4.4 Global Discovery Catalogue

- WIS2 includes several Global Discovery Catalogues.
- A Global Discovery Catalogue enables a data consumer to search and browse descriptions of data published by each WIS2 Node. The data description (that is, discovery metadata) provides sufficient information to determine the usefulness of data and how one may access it.
- A Global Discovery Catalogue subscribes to notification messages via a Global Broker about the availability of new (or updated) discovery metadata. It downloads a copy of the discovery metadata and updates the catalogue.
- A Global Discovery Catalogue will amend discovery metadata records to add details of where one can subscribe to updates about the dataset at a Global Broker.
- A Global Discovery Catalogue makes its content available for indexing by search engines.

2.4.5 Global Monitor

- WIS2 includes a Global Monitor service.
- The Global Monitor collects metrics from WIS2 components.
- The Global Monitor provides a dashboard that supports the operational management of the WIS2 system.
- The Global Monitor tracks:
  1. What data is published by WIS2 Nodes.
  2. Whether data can be effectively accessed by data consumers.
  3. The performance of components in the WIS2 system.

2.5 Protocols configuration

2.5.1 Publish-subscribe protocol (MQTT)

- The MQTT protocol[1] is to be used for all WIS2 publish-subscribe workflows (publication and subscription).
- MQTT v3.1.1 and v5.0 are the chosen protocols for the WIS2 Notification Messages publication and subscription.
  - To connect to Global Brokers, MQTT v5.0 is preferred as it provides additional features such as the ability to use shared subscriptions.
- The following parameters are to be used for all MQTT client/server connectivity and subscription:
  - Message retention: false
  - Quality of Service (QoS) of 1
  - A maximum of 2000 messages to be held in a queue per client
- In order to permit authentication and authorization for users, WIS2 Node, Global Cache, Global Discovery Catalogue and Global Brokers shall use a user and password based mechanism.
- To improve the overall level of security of WIS2, the secure version of the MQTT protocol is preferred. If used, the certificate must be valid.
- The standard Transmission Control Protocol (TCP) ports to be used are 8883 for Secure MQTT (MQTTS) and 443 for Secure Web Socket (WSS).

2.5.2 Download protocol (HTTP)

- The HTTP protocol (RFC 7231[2]) is to be used for all WIS2 download workflows.
- To improve the overall level of security of WIS2, the secure version of the HTTP protocol is preferred. If used, the certificate must be valid.
- The standard TCP port to be used is 443 for Secure HTTP (HTTPS).

2.6 Implementation and operation of a WIS2 Node

2.6.1 Practices and procedures

2.6.1.1 Registration and decommissioning of a WIS2 Node

Registration and decommissioning of WIS2 Nodes must be approved by the Permanent Representative (PR) with WMO for the country or territory in which the WIS centre resides. The WIS National Focal Point can register a WIS2 Node on behalf of the PR for an official NC or DCPC listed in the *Manual on WIS*, Volume I. Where the WIS2 Node is part of a DCPC, the sponsoring WMO Programme or Regional Association shall be consulted.

A WIS2 Node can be registered to exchange data concerning a WMO project or campaign, for a limited time. The National Focal Point can register such a project-related WIS2 Node in

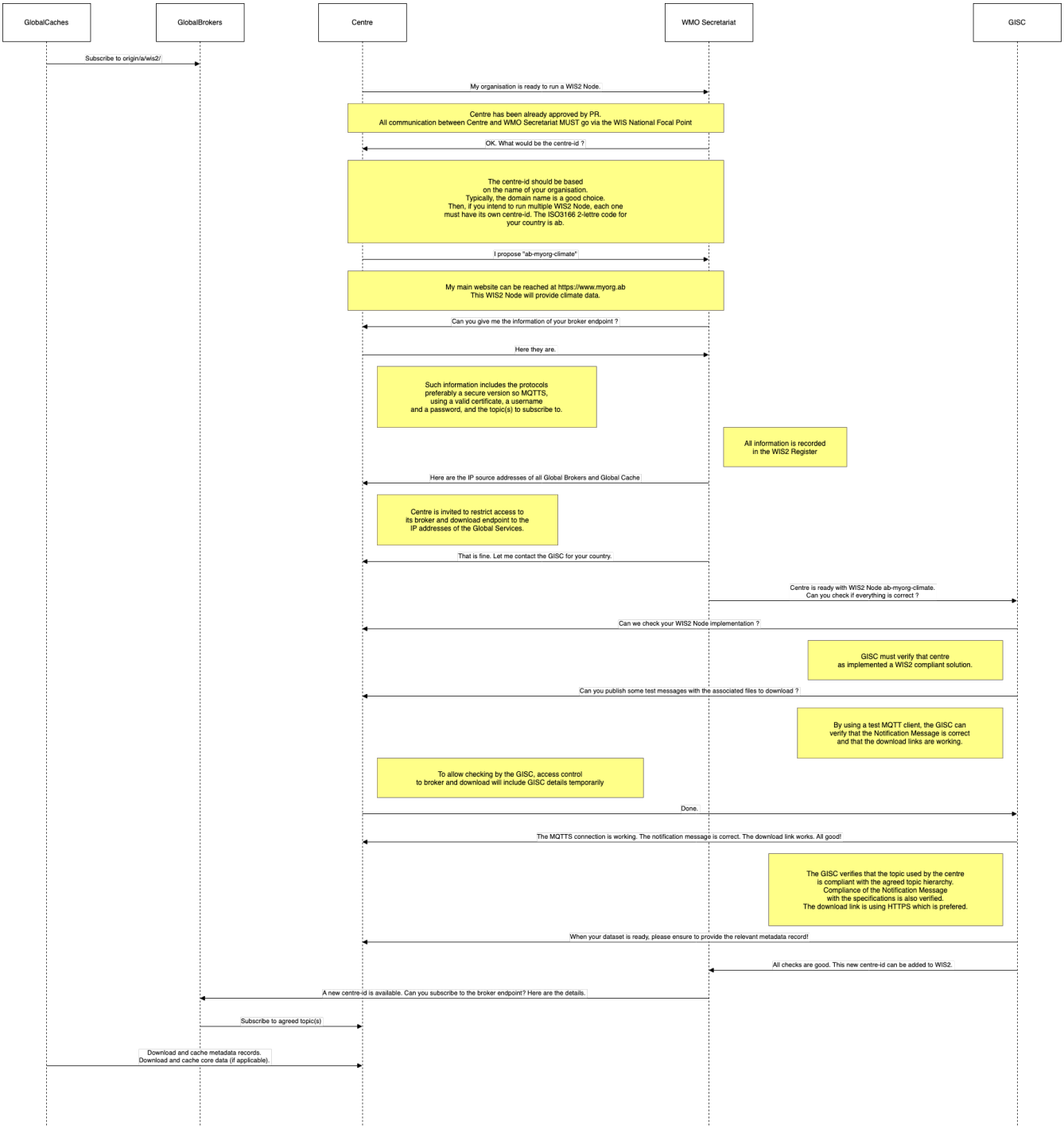
coordination with the Secretariat.

WMO Secretariat will operate a WIS2 register as an authoritative list of WIS2 Nodes and Global Services.

The registration of a WIS2 Node involves the following steps:

- Request hosting a WIS2 Node: A request for hosting a WIS2 Node shall be put forward by the WIS National Focal Point (NFP) of the country of the WIS2 Node host centre, or, in the case of international organizations, by either the Permanent Representative (PR) of the country or territory where the WIS2 Node host centre is located or the president of the relevant organization in case of WMO partner or programme designated as DCPC.
- Assign a centre-id: The centre identifier (centre-id) is an acronym as proposed by the Member and endorsed by the WMO Secretariat. It is a single identifier comprised of a top -level domain (TLD) and centre -name, and represents the data publisher, distributor or issuing centre of a given dataset or data product/granule (see the Manual on WIS, Volume II – Appendix D: WIS2 Topic Hierarchy). See below for guidance on assigning a centre identifier (2.6.1.2 Guidance on assigning a centre identifier for a WIS2 Node).
- Complete the WIS2 register: The WIS NFP shall complete the WIS2 register operated by the WMO Secretariat.
- Provide Global Service details: The WMO Secretariat provides connection details for the Global Services (such as, IP addresses) so that the WIS2 Node can be configured to provide the access.
- WIS2 Node assessment: The principal GISC verifies that the WIS2 Node is compliant with WIS2 requirements. The assessment includes:
  - Verification of the compliance of the topics used by the centre with the WIS2 Topic Hierarchy (WTH) specification.
  - Verification of compliance of notification messages with the WIS2 Notification Message (WNM) specification.
  - Verification that the data server is correctly configured and properly functioning.
  - Verification that the Message Broker is correctly configured and properly functioning.
- Add new centre to WIS2: Upon completion of this verification, and confirmation that it satisfies all conditions for operating a WIS2 Node, GISC notifies WMO Secretariat and confirms that this WIS2 Node can be added to WIS2.
- Communicate details to the Global Services: WMO Secretariat provides the WIS2 Node details to the Global Brokers to subscribe to the WIS2 Node.

A diagram of the process of registering a WIS2 Node is presented below.



Once a WIS2 Node has been registered and connected with Global Services, it can proceed to register the datasets it will publish via WIS2. To register a dataset, the authorized WIS2 Node publishes discovery metadata about the new dataset. Validation of the discovery metadata is completed by the Global Discovery Catalogues and Global Brokers automatically subscribes to the topics provided in the discovery metadata record. For more information, see 1.3.2 How to provide discovery metadata to WIS2.

Once the dataset has successfully been registered, the WIS2 Node can proceed to exchange data - see [_1_3_3_how_to_provide_data_in_wis2].

When decommissioning a WIS2 Node operators must ensure that obligations relating to data sharing within WIS continue to be met after the WIS2 Node is decommissioned, for example, by migrating these data sharing obligations to another WIS2 Node. In the case of a DCPC, this may mean the responsibilities are transferred to another Member.

2.6.1.2 Guidance on assigning a centre identifier for a WIS2 Node

The centre identifier (`centre-id`) is used in WIS2 to uniquely identify a participating WIS2 Node. The `centre-id` must conform to the specification given in the *Manual on WIS*, Volume II - Appendix D: WIS2 Topic Hierarchy, section 7.1.6 Centre identification.

The `centre-id` comprises two dash-separated tokens.

Token 1 is a *Top Level Domain* (TLD) defined by the Internet Assigned Numbers Authority (IANA)[3].

This is usually a simple choice for a Member. However, overseas territories require some thought. The recommended approach depends on the governance of the overseas territory. Take some French examples. Réunion is a French Department – it is considered part of France, it uses the Euro. Here, we would use the “fr” TLD. New Caledonia is a French overseas territory with top-level-domain of “nc”. It has separate, devolved governance. The recommendation is to use “nc”. All that said, it is a national decision which TLD to use.

Token 2 is a descriptive name for the centre and this may contain dashes (but not other special characters).

The descriptive name should be something recognizable - not only by our community, but by other users too. Basing things on the web domain name is likely to ensure that centre identifiers remain unique within a particular country/territory. A UK example this time: the UK’s National Meteorological Service is the Met Office (http://www.metoffice.gov.uk), so “metoffice” is better than “ukmo”[4]. Using the 4-letter GTS centre identifiers (CCCC) is not recommended because people unfamiliar with the GTS do not understand them.

The centre identifier specification says that larger organizations operating multiple centres may wish to register separate centre-ids for each centre. This is good practice. Keeping with the UK example, Met Office operates a National Meteorological Centre (NMC), 9 DCPCs (for example, a Volcanic Ash Advisory Centre) and a WIS2 Global Service, so it is important to split them out. For example:

- `uk-metoffice-nmc`
- `uk-metoffice-vaac`
- `uk-metoffice-global-cache`

Using a system name in the centre-id is not a good idea because these may change over time. Functional designations are longterm durable. Appending `-test` may be used to designate test WIS2 Nodes.

2.6.1.3 Authentication, authorization, and access control for a WIS2 Node

When configuring your WIS2 Node you need to consider how it will be accessed by Global Services and data consumers.

Global Brokers must authenticate when they connect to the MQTT Message Broker in your WIS2 Node. Username and password credentials are used[5]. When registering your WIS2 Node with the WMO Secretariat, you will need to provide these credentials. The WMO Secretariat will share these credentials with the Global Service operators and store them in the WIS register. You should not

consider these credentials as confidential or secret.

Given that Global Brokers will republish notification messages provided by your WIS2 Node, you may decide to restrict access to the MQTT Message Broker. Global Brokers operate using a fixed IP address which allows you to permit them access using IP filtering[6]. You must ensure that your MQTT Message Broker is accessible for more than one Global Broker to provide resilient transmission of notification messages to WIS2.

If your WIS2 Node is only publishing core data[7], you may also restrict access to your data server - instead, relying on the Global Caches to distribute your data. Similarly, Global Caches also operate on fixed IP addresses allowing connections from them to be easily identified. Again, you must ensure that access is given to more than one Global Broker to ensure resilience.

During registration, the WMO Secretariat will provide host names and IP addresses of the Global Services to enable configuration of access control.

Access controls may be implemented for recommended data. You should use only the "security schemes" for authentication and authorization specified in OpenAPI[8].

2.6.2 Performance management

2.6.2.1 Service levels and performance indicators

A WIS2 Node must be able to:

- Publish datasets and compliant metadata and discovery metadata
  - Publish metadata to the Global Data Catalogue
  - Publish core data to the Global Cache
  - Publish data for consumer access
  - Publish data embedded in a message (such as, Common Alerting Protocol (CAP) warnings)
  - Receive metadata publication errors from the Global Data Catalogue
  - Provide metadata with topics to Global Brokers

2.6.2.2 System performance metrics

If contacted by the Global Monitor via GISC for a performance issue, the WIS2 Node should provide metrics to the GISC and Global Monitor when service is restored to indicate the resolution of the issue.

2.6.3 WIS2 Node reference implementation: WIS2 in a box

To provide a WIS2 Node, Members may use whichever software components they consider most appropriate to comply with WIS2 technical regulations.

To assist Members participating in WIS2, a free and open-source reference implementation is available for use. WIS2 in a box (wis2box) implements the requirements of a WIS2 Node in as well as additional enhancements. The wis2box builds on mature and robust free and open-source

software components that are widely adopted for operational use.

The wis2box provides functionality required for both data publisher and data consumer roles. It provides the following technical functions:

- Configuration, generation and publication of data (real-time or archive) and metadata to WIS2, compliant to WIS2 Node requirements
- MQTT Message Broker and notification message publication (subscribe)
- HTTP object storage and raw data access (download)
- Station metadata curation/editing tools (user interface)
- Discovery metadata curation/editing tools (user interface)
- Data entry tools (user interface)
- OGC API server, providing dynamic APIs for discovery, access, visualization and processing functionality (APIs)
- Extensible data "pipelines", allowing for transformation, processing and publishing of additional data types
- Provision of system performance and data availability metrics
- Access control for recommended data publication, as required
- Subscription to notifications and and download of WIS data from Global Services
- Modular design, allowing for extending to meet additional requirements or integration with existing data management systems

Project documentation can be found at https://docs.wis2box.wis.wmo.int.

The wis2box is managed as a free and open source project. Source code, issue tracking and discussions are hosted openly on GitHub: https://docs.wis2box.wis.wmo.int.

2.7 Implementation and operation of a Global Service

2.7.1 Procedure for registration of a new Global Service

Successful operations of WIS will depend on having a set of Global Services running well-managed IT environments with a very high level of reliability so that all WIS Users and WIS2 Nodes will be able to access and provide the data they need for their duties.

Depending on the nature of the Global Service, the following is the minimum capability of Global Service operation, so that collectively, the level of service is 100% (or very close):

- Three (3) Global Brokers: Each Global Broker is connected to at least two (2) other Global Brokers
- Three (3) Global Caches: Each Global Cache is connected to at least two (2) Global Brokers and should be able to download the data from all WIS2 Nodes providing core data
- Two (2) Global Discovery Catalogues: Each Global Discovery Catalogue is connected to at least one (1) Global Broker

- Two (2) Global Monitors: Each Global Monitor should scrape the metrics from all other Global Services

In addition to the above, WIS architecture can accommodate adding (or removing) Global Services. Candidate WIS centres should inform their WIS National Focal Point and contact the WMO Secretariat to discuss their offer to provide a Global Service.

Running a Global Service is a significant commitment for a WIS centre. To maintain a very high level of service of WIS, each Global Service will have a key role to play.

On receipt of an offer from a Member to operate a Global Service, the WMO Secretariat will suggest which Global Service the Member may provide to improve WIS2. This suggestion will be based on the current situation of WIS2 (such as, the number of existing Global Brokers, whether an additional Global Cache is needed, and so forth).

The *Manual on WIS*, Volume II, this Guide and other material available will help WIS centres in deciding the best way forward.

When decided, the WIS National Focal Point will inform the WMO Secretariat of its preference. Depending on the type of Global Service, the WMO Secretariat will provide a checklist to the WIS centre so that the future Global Service can be included in WIS operations.

A WIS centre must commit to running the Global Service for a minimum of four years.

The WMO Secretariat and other Global Services will make the required changes to include the new Global Service in WIS operations.

2.7.2 Performance management and monitoring of a Global Service

2.7.2.1 Monitoring and metrics for WIS2 operations

The availability of data and performance of system components within WIS2 are actively monitored by GISCs and the Global Monitor service to ensure proactive response to incidents and effective capacity planning for future operations.

WIS2 requires that metrics are provided using OpenMetrics – the de-facto standard [9] for transmitting cloud-native metrics at scale. Widely adopted, many commercial and open-source software components already come preconfigured to provide performance metrics using the OpenMetrics standard. Tools such as Prometheus and Grafana provide aggregation and visualization of metrics provided in this form, making it simple to generate performance insights. The OpenMetrics standard can be found at openmetrics.io [10]].

The WIS2 Global Services, namely the Global Broker, Global Cache, and Global Discovery Catalogue expose monitoring metrics on their respective service to the Global Monitor.

There is no requirement on WIS2 Nodes to provide monitoring metrics. However their WIS2 interfaces may be queried remotely by Global Services, which in turn can provide metrics on the availability of WIS2 Nodes.

Metrics for the WIS2 monitoring should follow the naming convention:

```
wmo_<program>_<name>
```

Where program is the name of the responsible WMO Programme and name is the name of the metric. Examples for WIS2 metrics can look like:

```
wmo_wis2_gc_downloaded_total
```

```
wmo_wis2_gb_messages_invalid_total
```

The full set of the WIS2 monitoring metrics is given in WMO: WIS2 Metric Hierarchy [11]

2.7.2.2 Service levels, performance indicators, and fair-usage policies

- Each WIS centre operating a WIS2 Node will be responsible for achieving the highest possible level of service based on their resources and capabilities.
- All Global Services, in particular Global Brokers and Global Caches, are collectively responsible for making the WIS a reliable and efficient means to exchange data required for the operations of all WIS centres. The agreed architecture provides a redundant solution where the failure of one component will not impact the overall level of service of WIS.
- Each Global Service should aim at achieving at least 99.5% availability of the service they propose. This is not a contractual target. It should be considered by the entity providing the Global Service as a guideline when designing and operating the Global Service.
- A Global Broker:
  - Should support a minimum of 200 WIS2 Nodes or Global Services
  - Should support a minimum of 1000 subscribers.
  - Should support processing of a minimum of 10000 messages per second
- A Global Cache:
  - Should support a minimum of 100 GB of data in the cache
  - Should support a minimum of 1000 simultaneous downloads
  - Could limit the number of simultaneous connections from a user (known by its originating source IP) to 5
  - Could limit the bandwidth usage of the service to 1 Gb/s
- A Global Monitor:
  - Should support a minimum of 50 metrics providers
  - Should support 200 simultaneous access to the dashboard
  - Could limit the bandwidth usage of the service to 100 Mb/s
- A Global Discovery Catalogue:
  - Should support a minimum of 20000 metadata records

- Should support a minimum of 50 requests per second to the API endpoint

2.7.2.3 Metrics for Global Services

In the following sections and for each Global Service, a set of metrics is defined. Each Global Service will provide those metrics. They will then be ingested by the Global Monitor.

2.7.3 Global Broker

2.7.3.1 Technical considerations

- As detailed above, there will be at least three instances of Global Broker to ensure highly available, low latency global provision of messages within WIS.
- A Global Broker instance subscribes to messages from WIS2 Nodes and other Global Services. The Global Broker should aim at subscribing to all WIS centres. If this is not possible, for whatever reason, the Global Broker should inform the WMO Secretariat so that the situation is documented.
- Every WIS2 Node or Global Service must have subscriptions from at least two Global Brokers.
- For full global coverage, a Global Broker instance will subscribe to messages from at least two other Global Brokers.
- When subscribing to messages from WIS2 Nodes and other Global Services, a Global Broker must authenticate using the valid credentials managed by the WIS centre and available at WMO Secretariat.
- A Global Broker is built around two software components:
  - An off the shelf broker implementing both MQTT 3.1.1 and MQTT 5.0 in a highly-available setup, typically in a cluster mode. Tools such as EMQX, HiveMQ, VerneMQ, RabbitMQ (in its latest versions) are compliant with these requirements. It must be noted that the open source version of Mosquitto cannot be clustered and therefore should not be used as part of a Global Broker.
  - Additional features including anti-loop detection, notification message format compliance, validation of the published topic, and provision of metrics are required.
- When receiving a message from a WIS centre or Global Service broker, The metric `wmo_wis2_gb_messages_received_total` will be increased by 1.
- A Global Broker will check if the topic on which the message is received is valid (in particular, a discovery metadata record must exist with a corresponding topic so that data can be made available using this topic). If the topic is invalid, the Global Broker will discard non-compliant messages and will raise an alert. The metric `wmo_wis2_gb_messages_no_metadata_total` will be increased by 1. Global Broker should not request Global Discovery Catalogue for each notification message but should keep a cache of all valid topics for every `centre-id`.
- During the pre-operational phase (2024), Global Broker will not discard the message but will send a message on the `monitor` topic hierarchy to inform the originating centre and its GISC.
- A Global Broker will validate notification messages against the standard format (see *Manual on WIS*, Volume II – Appendix E: WIS2 Notification Message), discarding non-compliant messages and raising an alert. The metric `wmo_wis2_gb_messages_invalid_total` will be increased by 1.

- A Global Broker instance will republish a message only once. Using the message id as defined in the WIS2 Notification Message, the Global Broker will record the id of messages already published and will discard subsequent identical (with the same message id) messages. This is the anti-loop feature of the Global Broker.
- When publishing a message to the local broker, the metric `wmo_wis2_gb_messages_published_total` will be increased by 1.
- All above defined metrics will be made available on HTTPS endpoints that the Global Monitor will ingest from regularly.
- As a convention Global Broker centre-id will be `tld-{centre-name}-global-broker`.
- A Global Broker should operate with a fixed IP address so that WIS2 Nodes can permit access to download resources based on IP address filtering. A Global Broker should also operate with a public resolvable DNS name pointing to that IP address. The WMO Secretariat must be informed of the IP address and/or hostname, and any subsequent changes.

2.7.4 Global Cache

In WIS2 Global Caches provide access to WMO core data for data consumers. This allows for data providers to restrict access to their systems to Global Services and it reduces the need for them to provide high bandwidth and low latency access to their data. Global Caches work transparent for end users in that they resend notification messages from data providers which are updated to point to the Global Cache data store for data, they copied from the original source. Additionally, Global Caches also resend notification messages from data providers for core data, that is not stored on the Global Cache, for instance if the originator indicates that a certain dataset should not be cached in the notification message. In the latter case, the notification messages that a Global Cache resends are unchanged and point to the original source. Data consumers should subscribe to the notification messages from Global Caches instead of the notification messages from the data providers for WMO core data. When data consumers receive a notification message they should follow the URLs from that message which either point to a Global Cache holding a copy of the data, or - in case of uncached content - point to the original source.

2.7.4.1 Technical considerations

- A Global Cache is built around three software components:
  - A highly available data server allowing data consumers to download cache resources with high bandwidth and low latency.
  - A Message Broker implementing both MQTTv3.1.1 and MQTTv5 for publishing notification messages about resources that are available from the Global Cache
  - A cache management implementing the features needed to connect with the WIS ecosystem, receive data from WIS2 Nodes and other Global Caches, store the data to the data server and manage the content of the cache (such as, expiration of data, deduplication, and so forth).
- The Global Cache will aim at containing copies of real-time and near real-time data designated as "core" within the Unified Data Policy (Resolution 1 (Cg-Ext(2021))).
- A Global Cache instance will host data objects copied from NC/DCPCs.
- A Global Cache instance will publish notification messages advertising availability of the data

objects it holds. The notification messages will follow the standard structure (see *Manual on WIS*, Volume II -Appendix E: WIS2 Notification Message).

- A Global Cache instance will use the standard topic structure in their local message brokers (see *Manual on WIS*, Volume II -Appendix D: WIS2 Topic Hierarchy).
- A Global Cache instance will publish on topic `cache/a/wis2/⋯`.
- There will be multiple Global Cache instances to ensure highly available, low latency global provision of real-time and near real-time "core" data within WIS2.
- There will be multiple Global Cache instances that may attempt to download cacheable data objects from all originating centres with "cacheable" content. A Global Cache instance will also download data objects from other Global Cache instances. This ensures the instance has full global coverage, mitigating where direct download from an originating centre is not possible.
- A Global Cache instance will operate independently of other Global Cache instances. Each Global Cache instance will hold a full copy of the cache – albeit that there may be small differences between Global Cache instances as "data availability" notification messages propagate through WIS to each Global Cache in turn. There is no formal ‘synchronization’ between Global Cache instances.
- A Global Cache will temporarily cache all resources published on the `metadata` topic. A Global Discovery Catalogue will subscribe to notifications about the publication of new or updated metadata, download the metadata record from the Global Cache and insert it into the catalogue. A Global Discovery Catalogue will also publish a metadata record archive each day containing the complete content of the catalogue and advertise its availability with a notification message. This resource will also be cached by a Global Cache.
- A Global Cache is designed to support real-time distribution of content. Data consumers access data objects from a Global Cache instance by resolving the URL in a "data availability" notification message and downloading the file to which the URL points. Apart from the URL it is transparent to the data consumers from which Global Cache they download the data. There is no need to download the same data object from multiple Global Caches. The data id contained within the notification messages is used by data consumers and Global Services to detect such duplicates.
- There is no requirement for a Global Cache to provide a "browseable" interface to the files in its repository allowing data consumers to discover what content is available. However, a Global Cache may choose to provide such a capability (for example, implemented as a WAF) along with adequate documentation for data consumers to understand how the capability works.
- The default behaviour for a Global Cache is to cache all data published under the `origin/a/wis2/data/+/core` topic. A data publisher may indicate that data should not be cached by adding the `"cache": false` assertion in the WIS2 Notification Message.
- A Global Cache may decide not to cache data. For example, if the data is considered too large, or a WIS2 Node publishes an excessive number of small files. Where a Global Cache decides not to cache data it should behave as though the `cache` property is set to false and send a message on the `monitor` topic hierarchy to inform the originating centre and its GISC. The Global Cache operator should work with the originating WIS2 Node and their GISC to remedy the issue.
- If core data is not cached on a Global Cache (that is, if the data is flagged as `"cache": false` or if the Global Cache decides not to cache this data), the Global Cache shall nevertheless republish

the WIS2 Notification Message to the `cache/a/wis2/⋯` topic. In this case the message id will be changed and the rest of the message will not be modified.

- A Global Cache should operate with a fixed IP address so that WIS2 Nodes can permit access to download resources based on IP address filtering. A Global Cache should also operate with a public resolvable DNS name pointing to that IP address. The WMO Secretariat must be informed of the IP address and/or hostname, and any subsequent changes.
- A Global Cache should validate the integrity of the resources it caches and only accept data that matches the integrity value from the WIS2 Notification Message. If the WIS2 Notification Message does not contain an integrity value, a Global Cache should accept the data as valid. In this case a Global Cache may add an integrity value to the message it republishes.
- As a convention Global Cache centre-id will be `tld-{centre-name}-global-cache`.

2.7.4.2 Practices and procedures

- A Global Cache shall subscribe to the topics `origin/a/wis2/#`, `cache/a/wis2/#`.
- A Global Cache shall ignore all messages received on the topics `origin/a/wis2/+/data/recommended/#` and `cache/a/wis2/+/data/recommended/#` [12]
- A Global Cache shall retain the data and metadata they receive for a minimum period of 24 hours. Requirements relating to varying retention times for different types of data may be added later.
- For messages received on topic `origin/a/+/data/core/#` or `cache/a/+/data/core/#`, a Global Cache shall:
  - If the message contains the property `"properties.cache": false`
    - Republish the message at topic `cache/a/wis2/⋯` matching `+/a/wis2/⋯` where the original message has been received after having updated the id of the message.
  - Else
    - Maintain a list of data_ids already downloaded.
    - Verify if the message points to new or updated data by comparing the pubtime value of the notification message with the list of data_ids.
    - If the message is new or updated
      - Download only new or updated data from the `href` or extract the data from the message content.
      - If the message contains an integrity value for the data, verify the integrity of the data.
      - If data is downloaded successfully, move the data to the HTTP endpoint of the Global Cache.
      - Wait until the data becomes available at the endpoint.
      - Modify the message identifier and the canonical link's `href` of the received message. Leave all other fields untouched.
      - Republish the modified message to topic `cache/a/wis2/⋯` matching the `+/a/wis2/⋯` where the original message has been received.

        - The metric `wmo_wis2_gc_downloaded_total` will be increased by 1. The metric `wmo_wis2_gc_dataserver_last_download_timestamp_seconds` will be updated with the timestamp (in seconds) of the last successful download from the WIS2 Node or Global Cache.
    - Else
        - Drop the messages for data already present on the Global Cache.
- If the Global Cache is not able to download the data the metric `wmo_wis2_gc_downloaded_error_total` will be increased by 1.
- A Global Cache shall provide the metric defined in this Guide at an HTTP endpoint
- A Global Cache should make sure that data is downloaded in parallel and downloads are not blocking each other
- The metric `wmo_wis2_gc_dataserver_status_flag` will reflect the status of the connection to the download endpoint of the Centre. Its' value will be 1 when the endpoint is up and 0 otherwise.

2.7.5 Global Discovery Catalogue

2.7.5.1 Technical considerations

- The Global Discovery Catalogue provides data consumers with a mechanism to discover and search for datasets of interest, as well as how to interact with and find out more information about those datasets.
- The Global Discovery Catalogue implements the OGC API – Records – Part 1: Core standard[13], adhering to the following conformance classes and their dependencies:
    - Searchable Catalog (Deployment)
    - Searchable Catalog - Sorting (Deployment)
    - Searchable Catalog - Filtering (Deployment)
    - JSON (Building Block)
    - HTML (Building Block)
- The Global Discovery Catalogue will make discovery metadata available via the collection identifier of `wis2-discovery-metadata`.
- The Global Discovery Catalogue advertises the availability of datasets and how to access them or subscribe to updates.
- The Global Discovery Catalogue does not advertise or list the availability of individual data objects that comprise a dataset (that is, data files).
- A single Global Discovery Catalogue instance is sufficient for WIS2.
- Multiple Global Discovery Catalogue instances may be deployed for resilience.
- Global Discovery Catalogue instances operate independently of each other; each Global Discovery Catalogue instance will hold all discovery metadata records. Global Discovery Catalogues do not need to synchronize between themselves.
- A Global Discovery Catalogue is populated with discovery metadata records from a Global Cache instance, receiving messages about the availability of discovery metadata records via a

Global Broker.

- A Global Discovery Catalogue should connect and subscribe to more than one Global Broker instance to ensure that no messages are lost in the event of a Global Broker failure. A Global Discovery Catalogue instance will discard duplicate messages as needed.
- A Global Discovery Catalogue will validate that a discovery metadata record identifier's `centre-id` token (see *Manual on WIS*, Volume II -Appendix F: WMO Core Metadata Profile) matches against the `centre-id` level of the topic from which it was published (see *Manual on WIS*, Volume II -Appendix D: WIS2 Topic Hierarchy), to ensure that discovery metadata is published by the authoritative organization.
- A Global Discovery Catalogue will validate discovery metadata records against the WMO Core Metadata Profile (WCMP2). Valid WCMP2 records will be ingested into the catalogue. Invalid or malformed records will be discarded and reported to the Global Monitor against the centre identifier associated with the discovery metadata record.
- A Global Discovery Catalogue will only update discovery metadata records to replace links for dataset subscription and notification (origin) with their equivalent links for subscription at Global Broker instances (cache).
- A Global Discovery Catalogue will periodically assess discovery metadata provided by NCs and DCPCs against a set of key performance indicators (KPIs) in support of continuous improvement. Suggestions for improvement will be reported to the Global Monitor against the centre identifier associated with the discovery metadata record.
- A Global Discovery Catalogue will remove discovery metadata that is marked for deletion as specified in the data notification message.
- A Global Discovery Catalogue should apply faceting capability as specified in the cataloguing considerations of the WCMP2 specification, as defined in OGC API - Records.
- A Global Discovery Catalogue will provide human-readable web pages with embedded markup using the schema.org vocabulary, thereby enabling search engines to crawl and index the content of the Global Discovery Catalogue. Consequently, data consumers should also be able to discover WIS content via third party search engines.
- A Global Discovery Catalogue will generate and store a zip file of all WCMP2 records once a day, that will be made be accessible via HTTP.
- A Global Discovery Catalogue will publish a WIS2 Notification Message of its zip file of all WCMP2 records on its centre-id's metadata topic (for example, `origin/a/wis2/centre-id/metadata`, where `centre-id` is the centre identifier of the Global Discovery Catalogue).
- A Global Discovery Catalogue may initialize itself (cold start) from a zip file of all WCMP2 records published.
- As a convention Global Discovery Catalogue centre-id will be `tld-{centre-name}-global-discovery-catalogue`.

2.7.5.2 Global Discovery Catalogue reference implementation: wis2-gdc

To provide a Global Discovery Catalogue, Members may use whichever software components they consider most appropriate to comply with WIS2 technical regulations.

To assist Members' participation in WIS2, a free and open-source Global Discovery Catalogue

reference implementation is made available for download and use. The wis2-gdc builds on mature and robust free and open-source software components that are widely adopted for operational use.

The wis2-gdc provides the functionality required Global Discovery Catalogue, providing the following technical functions:

- Discovery metadata subscription and publication from the Global Broker
- Discovery metadata download the Global Cache
- Discovery metadata validation, ingest and publication
- WCMP2 compliance
- Quality assessment (KPIs)
- OGC API - Records - Part 1: Core compliance
- Metrics reporting
- Implementation of metrics

The wis2-gdc is managed as a free and open source project. Source code, issue tracking and discussions are hosted in the open on GitHub.[14]

2.7.6 Global Monitor

2.7.6.1 Technical considerations

- WIS standardizes how system performance and data availability metrics are published from WIS2 Nodes and Global Services.
- For each type of Global Service, a set of standard metrics has been defined. Global Services will implement those metrics and provide an endpoint for those metrics to be scraped by the Global Monitor
- The Global Monitor will collect metrics as defined in the OpenMetrics standard.
- The Global Monitor will monitor the 'health' (namely, performance) of components at NC/DCPC as well as Global Service instances.
- The Global Monitor will provide a web-based ‘dashboard’ that displays the WIS2 system performance and data availability.
- As a convention Global Monitor centre-id will be `tld-{centre-name}-global-monitor`.

```
The main task of the Global Monitor is to regularly query the provided metrics from
the relevant WIS2 entities, aggregate and process the data and then provide the
results to the end user in a suitable presentation.
```

2.8 Operations

2.8.1 Interoperability with external systems

The WIS2 principles enable lowering the barrier to weather/climate/water data for WMO Members. Lowering the barrier is driven by international standards for data discovery, access, and visualization. In addition to Member benefits, a by-product of utilizing standards is being able to provide the same data and access mechanisms to external systems at no extra cost of implementation.

WIS2 standards are based on industry standards (OGC, W3C, IETF) and allow for broad interoperability. This means that non-traditional users can also use data from WIS2 data in the same manner, without the requirement for specialized software, tools, or applications.

2.8.1.1 Publishing meteorological data through WIS2 into ICAO SWIM

Meteorological data is an essential input for public weather services and aviation services alike. WIS2 provides the mechanism for data exchange in WMO, while SWIM is the ICAO initiative to harmonize the provision of aeronautical, meteorological and flight information to support air traffic management (ATM).

Both WIS2 and SWIM support similar outcomes relating to data exchange. However, there are differences in both approach and implementation.

Specifications for WIS2 are defined in the *Manual on WIS*, Volume II, and further elaborated in this Guide. Specifications for SWIM will be defined in the Procedures for Air Navigation Services –Information Management (PANS-IM) (ICAO Doc. 10199)[15].

During the WIS2 transition phase (2025-2033), meteorological data published via WIS2 will automatically be published to the GTS via the WIS2-to-GTS gateways.

WIS2	**SWIM**
Earth-system scope: weather, climate, hydrology, atmospheric composition, cryosphere, ocean and space weather data	ATM scope: aeronautical, meteorological and flight information
Data centric - a consumer discovers data and then determines the services through which that data may be accessed	Service centric - a consumer discovers a service (or service provider) and determines what resources (that is, information) is available therein
Technical protocols: MQTT, HTTP	Technical protocols: AMQP[16]

An organization (for example, the National Meteorological Service) that is responsible for providing meteorological data to WIS2 may be designated by the ICAO Contracting State as a responsible entity to provide aeronautical meteorological information into SWIM. Where requirements dictate, the organization may provide regional capability on behalf of a group of countries or territories.

This section of the Guide outlines how such an organization may efficiently fulfil the requirements in providing required data/information to the two systems. It proposes an interoperability approach between WIS2 and SWIM where meteorological data published via WIS2 can be automatically propagated to SWIM.

This Guide covers only how data from WIS2 can be published into SWIM. Consumption of information from SWIM services is not in scope.

This Guide also does not cover implementation details of the SWIM service - including, but not limited to:

- Mechanisms used by SWIM to discover service providers and services.
- Specification of the SWIM data message.
- AMQP message broker configuration.
- Operation, logging and monitoring.
- Cybersecurity considerations for the provision of SWIM services.

This Guide will be updated as more information is made available from ICAO and/or recommended practices are updated.

Finally, it should be noted that the provision of aeronautical meteorological information and its exchange via the ICAO Aeronautical Fixed Service (AFS) is also out of scope as they are solely defined under the ICAO regulatory framework.

2.8.1.1.1 WIS2 to SWIM gateway

The WIS2 to SWIM interoperability approach employs a gateway component (as per the figure below):

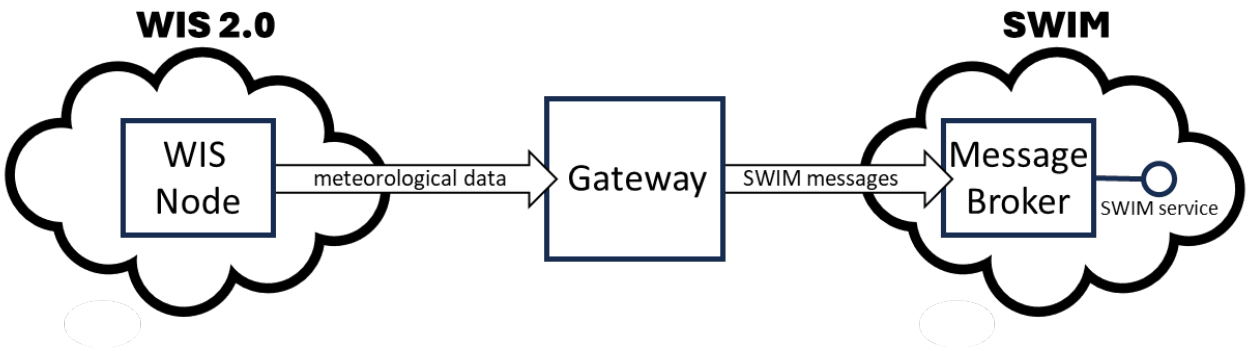


Figure 1. Schematic of an interoperability approach

The gateway can operate as an "adapter" between WIS2 and SWIM by pulling the requisite meteorological data from WIS2 and re-publishing it to SWIM.

2.8.1.1.2 Data types and format

Specifications for aeronautical meteorological information are provided in ICAO Annex 3 and other relevant guidance materials. The ICAO Meteorological Information Exchange Model (IWXXM) format (FM 205)[17] is to be used for encoding aeronautical meteorological information in SWIM.

2.8.1.1.3 Publishing meteorological data via WIS2

For meteorological data to be published from WIS2 to SWIM, the organization responsible for this provision will need to operate a WIS2 Node and comply with the pertinent technical regulations as specified in the *Manual on WIS*, Volume II. Onward distribution of the data by the Message Broker

over SWIM can be handled by the respective Information Service Provider in accordance with ICAO Standards and Recommended Practices (SARPs).

The responsible organization should consider whether this data should be published via an existing WIS2 Node, or whether a separate WIS2 Node should be established. For example, the data may be provided by a separate operational unit, or there may be a requirement to easily distinguish between data for SWIM and any other meteorological data.

Where a new WIS2 Node is needed, the responsible organization must establish a new WIS2 Node and register it with the WMO Secretariat. For more information, see 2.6 Implementation and operation of a WIS2 Node.

Datasets are a central concept in WIS2. Where meteorological data is published via WIS2, it will be packaged into “datasets”. The data should be grouped at the country/territory level; (for instance, datasets should be published for a given country/territory), one for each datatype (for example, aerodrome observation, aerodrome forecast and quantitative volcanic ash concentration information).

For the purposes of publishing through WIS2, datasets containing aeronautical meteorological information should be considered as "recommended data", as described in Resolution 1 (Cg-Ext(2021)). The recommended data category of the policy is intended to cover data that should be exchanged by Members to support Earth system monitoring and prediction efforts.

Recommended data:

- May be subject to conditions on use and reuse.
- May have access controls[18][19] applied at the WIS2 Node.
- Are not cached within WIS2 by the Global Caches[20].

Resolution 1 (Cg-Ext(2021)) requires transparency on the conditions of use for recommended data. Conditions regarding the use of aeronautical meteorological information are specified in ICAO Annex 3 and, optionally, by the ICAO Contracting State. Such conditions of use should be explicitly stated in the discovery metadata for each dataset as described below.

- The attribute `wmo:dataPolicy` should be set to `recommended`.
- Information about conditions of use should be specified using a `rights` property (see example below) and/or a `link` object with a relation `license`.
- Information about access control should be specified using a `security` object in the `link` object describing the data access details.

Example expression of conditions relating to the use of aeronautical meteorological information:

```
"properties": {
  ...
  "rights": "This information is freely disseminated for the purposes of safety of
international air navigation. ICAO Annex 3"
  ...
}
```

For more information on the WMO Core Metadata Profile version 2, see the *Manual on WIS*, Volume II, Appendix F.

On receipt of new data, the WIS2 Node will:

1. Publish the data as a resource via a Web server (or Web service).
2. Publish a WIS2 Notification Message to a local message broker that advertises the availability of the data resource.

Note that, in contrast to the GTS, WIS2 publishes data resources individually, each with an associated notification message. WIS2 does not group data resources into bulletins.

The data resource is identified using a URL. The notification message refers to the data resource using this URL[21].

For more details on the WIS2 Notification Message, see the *Manual on WIS*, Volume II, Appendix E: WIS2 Notification Message.

The notification message must be published to the proper topic on the Message Broker. WIS2 defines a standard topic hierarchy to ensure that data is published consistently by all WIS2 Nodes. Notification messages for aviation data should be published on a specific topic allowing a data consumer, such as the gateway, to subscribe only to aviation-specific notifications. See the example below:

Example topic used to publish notifications about Quantitative Volcanic Ash Concentration Information

```
origin/a/wis2/{centre-id}/data/recommended/weather/aviation/qvaci
```

For more details of the WIS Topic Hierarchy, see the *Manual on WIS*, Volume II, Appendix D: WIS2 Topic Hierarchy.

WIS Global Brokers subscribe to the local message brokers of WIS2 Nodes and republish notification messages for global distribution.

As a minimum, the WIS2 Node should retain the aviation data for a duration that meets the needs of the gateway. A retention period of at least 24 hours is recommended.

2.8.1.1.4 Gateway implementation

The potential interactions between the gateway component, WIS2 and SWIM are illustrated in the figure below[22]

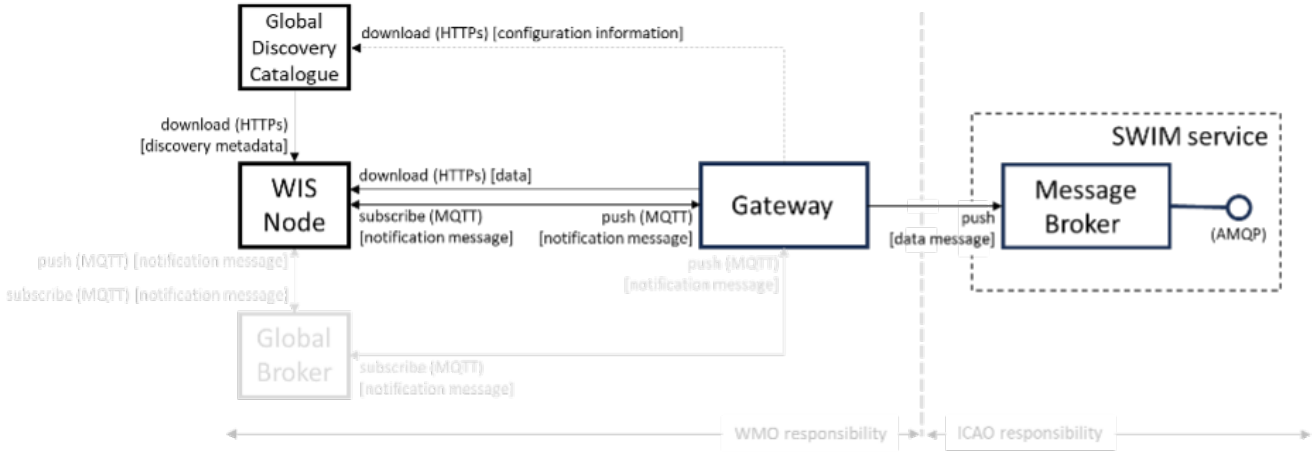


Figure 2. Interactions between the gateway and components of WIS2 and SWIM

Configuration

Dataset discovery metadata will provide useful information that can be used to configure the gateway, for example, the topic(s) to subscribe to plus various other information that may be needed for the SWIM service.

Discovery metadata can be downloaded from the Global Discovery Catalogue.

Functions

The gateway component implements the following functions:

- Subscribe to the pertinent topic(s) for notifications about new aeronautical meteorological information[23].
- On receipt of notification messages about newly available data:
  - Parse the notification message, discarding duplicate messages already processed previously;
  - Download the data resource from the WIS2 Node[24] using the URL in the message - the resource should be in IWXXM format;
  - Create a new "data message" as per the SWIM specifications, including the unique identifier extracted from the data resource[25], and embedding the aviation weather data resource within the data message;
  - Publish the data message to the appropriate topic on the SWIM Message Broker component of the SWIM service.

The choice of protocol for publishing to the SWIM Message Broker should be based on bilateral agreement between operators of the gateway and SWIM service.

The gateway should implement logging and error handling as necessary to enable reliable operations. WIS2 uses the OpenMetrics standard[26] for publishing metrics and other operating information. Use of OpenMetrics by the gateway would enable monitoring and performance reporting to be easily integrated into the WIS2 system.

Operation

The gateway may be operated at national or regional level depending on the organizational

governance in place.

2.8.1.1.5 SWIM service

The SWIM aviation weather information service may comprise of a Message Broker component which implements the AMQP 1.0 messaging standard[27].

The Message Broker publishes the data messages provided by the gateway.

The Message Broker must ensure that data messages are provided only by authorized sources such as a gateway and should validate incoming messages as aeronautical meteorological information.

2.8.1.2 The Ocean Data and Information System

The Ocean Data and Information System (ODIS) is a federation of independent data systems coordinated by the International Oceanographic Data and Information Exchange (IODE) of IOC-UNESCO. This federation includes continental-scale data systems as well as those of small organizations. ODIS partners use web architectural approaches to share metadata describing their holdings, services, and other capacities. In brief, IODE publishes guidelines on how to share metadata as linked open data, serialized in JSON-LD using schema.org[28] semantics. ODIS nodes use these guidelines to publish their metadata catalogues on the Web. This allows all systems with web connectivity to harvest and merge these catalogues, creating a global map of the ocean data. IODE harvests all metadata shared by ODIS partners, combines it as a knowledge graph, and processes this to export derivative products (for example, diagnostic reports and cloud-optimized data products). The Ocean InfoHub (OIH) system is IODE's reference implementation of a discovery system leveraging ODIS. The ODIS architecture and tools are free and open-source software (FOSS), with regular releases published for the community.

To reach beyond the oceans domain, ODIS works with other data systems and federations, such as WIS2, to define sustainable data and metadata exchanges and - where needed - translators or converters. The resources needed to convert between such systems are developed in the open and in close collaboration with staff from those systems. These exchanges include extract transform load (ETL) functions, to ensure that the bilateral exchange is mutually beneficial.

2.8.1.2.1 Cross system interoperability

Given the strong support for standards and interoperability by both WIS2 and ODIS, data and metadata exchange is realized using web architectural principles and approaches. The ability to discover ODIS data on WIS2 (as well as the inverse) is a goal in extending the reach of both systems and data beyond their primary communities of interest.

The WIS2 Global Discovery Catalogue will provide discovery metadata records using the OGC API - Records standard. This will include schema.org and JSON-LD annotations on WCMP2 discovery metadata in the GDC, to enable cross-pollination and federation.

ODIS dataset records will be made available using the WCMP2 standard and provided as objects available via HTTP for ingest, validation and publication to the GDC as a federated catalogue. ODIS data will be published as recommended data as per the Unified Data Policy (Resolution 1 (Cg-Ext(2021))).

[WIS2 and ODIS metadata and catalogue interoperability] | *images/wis2-odis-metadata-discovery-interop.png*

Figure 3. WIS2 and ODIS metadata and catalogue interoperability

As a result, federated discovery will be realized between both systems, allowing for use and reuse of data in an authoritative manner, closest to the source of the data.

[1] MQTT Specifications: https://mqtt.org/mqtt-specification/

[2] TFC 7231 - Hypertext Transfer Protocol (HTTP/1.1): https://datatracker.ietf.org/doc/html/rfc7231

[3] IANA Top Level Domains https://data.iana.org/TLD

[4] The ".gov" part of the domain name is superfluous for the purposes of WIS2. There is nothing preventing its use, but it does not add any value.

[5] The default connection credentials for a WIS2 Node Message Broker are username `everyone` and password `everyone`. WIS2 Node operators should choose credentials that meet their local policies (for example, password complexity).

[6] In WIS2 the IP addresses are used to determine the origin of connections and therefore confer trust to remote systems. It is well documented that IP addresses can be hi-jacked and that there are alternative, more sophisticated, mechanisms available for reliably determining the origin of connections requests, such as Public Key Infrastructure (PKI). However, the complexities of such implementation would introduce a barrier to Member's participation in WIS2. IP addresses are considered to provide an adequate level of trust for the purposes of WIS2: distributing publicly accessible data and messages.

[7] In some cases, WIS2 Nodes will need to serve core data directly (see 1.3.3.5 Considerations when providing core data in WIS2). In these situations, the WIS2 Node data server must remain publicly accessible.

[8] OpenAPI Security Scheme Object: https://spec.openapis.org/oas/v3.1.0#security-scheme-object

[9] OpenMetrics is proposed as a draft standard within IETF.

[10] cncf-openmetrics[https://openmetrics.io

[11] https://github.com/wmo-im/wis2-metric-hierarchy

[12] It is also technically possible to filter recommended data by using a wildcard subscription such as `origin/a/wis2/+/data/core/#`. However, avoiding wildcard subscription is generally considered good practice as it limits the burden of the broker operated by Global Brokers.

[13] OGC-API Records - Part 1 https://docs.ogc.org/DRAFTS/20-004.html

[14] https://github.com/wmo-im/wis2-gdc

[15] The PANS-IM is expected to available on ICAO NET by July 2024 and become applicable in November 2024. Information provided in herein is based on best understanding of draft proposals from ICAO.

[16] AMQP 1.0 is one of the protocols proposed in the draft PANS-IM

[17] IWXXM (FM205) is defined in the *Manual on Codes* (WMO-No. 306), Volume I.3 – International Codes

[18] WIS2 follows the recommendations from OpenAPI regarding choice of security schemes for authenticated access - a choice of HTTP authentication, API keys, OAuth2 or OpenID Connect Discovery. For more information see OpenAPI Security Scheme Object: https://spec.openapis.org/oas/v3.1.0#security-scheme-object

[19] WIS2 does not provide any guidance on use of Public Key Infrastructure (PKI).

[20] Global Caches enable highly available, low-latency distribution of core data. Given that core data is provided on a free and unrestricted basis, Global Caches do not implement any data access control.

[21] Where the data resource does not exceed 4 Kb, it may additionally be embedded in the notification message.

[22] Note that the figure simplifies the transmission of discovery metadata from WIS2 Node to the Global Discovery Catalogue. The WIS2 Node publishes notification messages advertising the availability of new discovery metadata resource at a given URL. These messages are republished by the Global Broker. The Global Discovery Catalogue subscribes to a Global Broker and downloads the discovery metadata from the WIS2 Node using the URL supplied in the message.

[23] WIS2 recommends that one subscribes to notifications from a Global Broker. However, where both gateway and WIS2 Node are operated by the same organization, it may be advantageous to subscribe directly to the local message broker of WIS2 Node, for example, to reduce latency.

[24] The WIS2 Node may control access to data - the gateway will need to be implemented accordingly.

[25] In case a unique identifier is required for proper passing of an aviation weather message to the gateway, one can use the GTS abbreviated heading (TTAAii CCCC YYGGgg) in the COLLECT envelop (available in IWXXM messages having a corresponding TAC message), or content in attribute `gml:identifier` (available in newer IWXXM messages like WAFS SIGWX Forecast and QVACI), for such purpose. There is currently no agreed definition for unique identifier of IWXXM METAR and TAF reports of individual aerodrome.

[26] OpenMetrics: https://openmetrics.io

[27] AMQP 1.0: https://www.amqp.org/resources/specifications
[28] https://schema.org

PART III

3.1 Information management

3.1.1 Introduction

3.1.1.1 Background

The efficient and effective provision of services relying on meteorological, climatological, hydrological and oceanographic information depends on a reliable information infrastructure. This infrastructure should be guided by community best practices and standards, including recommendations and requirements on sourcing, securing, managing, archiving, exchanging and providing easy access to information. This part of the Guide aims to provide high-level guidance on these activities, which can be grouped under the term “information management”. This is done by identifying and describing the fundamental principles of good information management and by highlighting the different stages of the information management life cycle.

Note: The term "information" is used in a general sense and includes data and products.

3.1.1.2 Scope

High-level guidance on information management practices that apply in the context of information related to the Earth system is provided in this part of the Guide. Detailed technical information, such as the specification of data formats or quality control and assurance methods, is provided in other parts of the Guide and in other WMO publications. These are referenced where applicable.

The principles of information management are described below. Five focus areas are described in 3.1.3 The information management life cycle. These are:

1. Planning, information creation and acquisition. The creation of information using internal and external data sources and the acquisition of information from various sources.
2. Representation and metadata. The use of standards to represent metadata, data and information; This is of primary importance to enable the interoperability and longterm usability of the information.
3. Publication and exchange of information. The creation and publication of discovery metadata in a standardized format enabling users to discover, access and retrieve the information.
4. Usage and communication. The publication of guidance material on the use of published information, including on the limitations and suitability of the information and any licensing terms.
5. Storage, archival and disposal. The policies and procedures for business continuity and disaster recovery, as well as retention and disposal.

3.1.1.3 Intended audience

This guidance is primarily aimed at personnel within WMO centres, who are responsible for planning and undertaking the creation or acquisition, stewardship, exchange and provision of

information related to the Earth system.

Specifically, the guidance has five main target audiences across the information life cycle:

1. Information producers or creators (those who produce or acquire the information) - they need to ensure the scientific quality of the underpinning information;
2. Information managers (those who manage information);
3. Information providers or publishers (those who publish the information) - they are responsible for the provision of the information, and for ensuring that appropriate access is enabled, licensing agreements are in place, and so forth;
4. Service providers (those who disseminate the information) - they are responsible for ensuring information availability and maintaining capability for easy and secure access to the information;
5. Information consumers (those who utilize the information) - they need to understand the restrictions, rights, responsibilities and limitations associated with the information together with the suitability for the intended usage or purpose.

3.1.2 Principles of information management

The effective management of information is essential for WMO centres to deliver operational services and information that is authoritative, seamless, secure and timely. The principles below underpin and provide a framework for information management across the full information life cycle. These principles are independent of the information type and are largely independent of technology; they are therefore expected to remain stable over time life cycle.

3.1.2.1 Principle 1: Information is a valued asset

- An information asset is information that has value. This value may be related to the cost of generating and collecting the information, it may be associated with the immediate use of the information, or it may be associated with the longer -term preservation and subsequent reuse of the information.
- This value should be recognizable and quantifiable and the asset should have an identifiable life cycle. Risks associated with, and to, an information asset should also be identified. As such, information management must be considered an integral part of a WMO centre's responsibilities and needs to be adequately resourced over the full life cycle of the information.

3.1.2.2 Principle 2: Information must be managed

- An information asset must be managed throughout its life cycle, from creation to use to eventual disposal, in a way that makes it valuable, maximizes its benefits and reflects its value in time and its different uses.
- Information managers must consider the entire information life cycle, from identifying needs and business cases to creation, quality assurance, maintenance, reuse, archival and disposal. Careful consideration must be given to disposal, ensuring that information is destroyed only when it has ceased to be useful for all categories of users.
- Professionally qualified and adequately skilled staff with clear roles and responsibilities should

apply a sound custodianship framework concerning security, confidentiality and other statutory requirements of different types of information.

3.1.2.3 Principle 3: Information must be fit for purpose

- Information should be developed and managed in accordance with its function and use for internal and external users.
- WMO centres should regularly assess information to ensure that it is fit for its purpose and that the related processes, procedures and documentation are adequate.
- Processes should be consistent with the general provisions and principles of quality management as described in the *WMO Technical Regulations* (WMO-No. 49).

3.1.2.4 Principle 4: Information must be standardized and interoperable

- Information must be stored and exchanged in standardized formats to ensure wide usability in the short and longterm. It is essential for longterm archiving that information be stored in a form that can be understood and used after several decades.
- Standardization is essential for structured information such as dataset definitions and metadata to support interoperability.
- Interoperability is essential for users to be able to utilize information through different systems and software. Open standards help ensure interoperability with their openness and wide adoption across various communities.
- Which standards to use depends on the user community and organizational policies. Interoperability requirements should be considered when selecting the standard for internal use and broader dissemination.
- The use of closed and proprietary standards is strongly discouraged.

3.1.2.5 Principle 5: Information must be well documented

- WMO centres should comprehensively document information processes, policies, and procedures to facilitate broad and longterm use.
- WMO centres should keep documentation up to date to ensure full traceability of processes along the information life cycle, particularly for its creation.
- Previous versions of the documentation should be retained, versioned, archived and made readily available for future use. In addition, versions should be assigned a unique and persistent identifier for future unambiguous identification.

3.1.2.6 Principle 6: Information must be discoverable, accessible and retrievable

- Information should be easy to find through the web, and for this purpose, the publisher should share discovery metadata with a catalogue service. The catalogue service should include a web application programming interface (API) to be used by other applications in order to offer user-tailored search portals.
- For information to be easily retrievable once discovered, it should be accessible using standard data exchange protocols.

3.1.2.7 Principle 7: Information should be reusable

- In order to maximize the economic benefits of an information asset it should be made as widely available and as accessible as possible.
- Resolution 1 (Cg-Ext(2021)) encourages the reuse of data and information through the open and unrestricted exchange of core WMO data. WMO encourages the free and unrestricted exchange of information in all circumstances.
- The publisher should provide an explicit and well-defined licence for each information item or dataset as part of the associated metadata.
- The Findable, Accessible, Interoperable and Reusable (FAIR) data principles promote open data with the ultimate goal of optimizing the reuse of data. These principles should be followed where possible.

Note: Information on the FAIR data principles can be found at: FAIR Principles - GO FAIR[1]

3.1.2.8 Principle 8: Information management is subject to accountability and governance.

- Information management processes must be governed as the information moves through its life cycle. All information must have a designated owner, steward, curator and custodian. These roles may be invested in the same person but should be clearly defined at the time of creation. A WMO centre with responsibility of managing information must ensure:
  - The implementation of general information management practices, procedures and protocols, including well-defined roles, responsibilities and restrictions on managing the information;
  - The definition and enforcement of appropriate retention policy, taking into account stakeholder needs and variations in value over the information life cycle;
  - The establishment of licensing and the definition and enforcement of any access restrictions.
  - The designated owner should have budget and decision-making authority with respect to preservation and data usage, including the authority to pass ownership to another entity.

3.1.3 The information management life cycle

3.1.3.1 Overview

All information should be subject to a well -defined and documented life cycle. The governance of this process is often referred to as the information management life cycle; it helps organizations manage information from planning, creation and acquisition through usage and exchange to archival and disposal.

The following sections describe two overarching themes, governance and documentation, which apply to all stages of the information life cycle; these sections provide high -level guidance and are split into five aspects:

- Planning, creation and acquisition;
- Representation and metadata;

- Publication and exchange;
- Usage and communication;
- Storage, archival and disposal.

Governance covers the rules that apply to managing information in a secure and transparent manner; documentation covers the act of recording the reasons for, and details of, all operations in the information management process.

3.1.3.2 Overarching requirements

3.1.3.2.1 Governance

- Information management governance defines a set of organizational procedures, policies and processes for the management of information. This includes defining accountabilities and compliance mechanisms.
- Effective governance helps ensure that all aspects of the information management process are conducted in a rigorous, standardized and transparent manner and that the information is secure, accessible and usable.
- WMO centres should establish a board or leadership group to develop and regularly review such a governance structure and ensure compliance with its requirements.

3.1.3.2.2 Documentation

- Documentation describing the who, what, why, when, where and how with respect to the various actions that are undertaken in the management of information is required to ensure the traceability and integrity of the information and to ensure operations can continue if key staff leave.
- This documentation is required for all aspects of the information life cycle and should be clear, well -communicated, regularly updated and easy to find. Guidance relating to the documentation should be provided to new staff taking on responsibilities for information management and be a key component of training.

3.1.3.3 Aspects of the information management life cycle

3.1.3.3.1 Planning, information creation and acquisition

Before the creation or acquisition of new information a business case and an information management plan should be developed, covering both the input information sources and any derived information. The plans should include:

- Why the information is required;
- How it will be collected or created;
- How it will be stored;
- Whether it will be exchanged with other users and under what policy;
- Where it should be submitted for longterm archival;
- Key roles and responsibilities associated with the management of the information.

For externally sourced data the plans should include where the information has come from and what the licensing terms are.

Once information has been acquired it should be checked to ensure that the contents and format are as expected. This may be done using a compliance checker or validation service. Once these checks have been performed the information content should also undergo quality control checks using well documented procedures to identify any issues. A record of the checks should be kept and any issues detected should be documented and sent back to the originators. It is also important to subscribe to updates from originators so any issues identified externally can be taken into account.

Information created rather than acquired should undergo the same processes as the acquired information. Information created should undergo quality control and the resulting files should be checked against the specified format requirements. The results of the processes and checks should be documented.

To ensure traceability and reproducibility the information and documents at this, and subsequent stages, should be version controlled and clearly labelled with version information. Similarly, software, or computer code, used to generate or process information should be version controlled with the version information recorded in the documentation and metadata. Where possible, software should be maintained within a code repository.

3.1.3.3.2 Representation and metadata

The formats used to store and exchange information should be standardized to ensure its usability in both the short and the longterm. It is essential that the information be accessible many years after archival if required. To ensure this usability, the format and version of the information should be recorded in the information metadata record and included within the information itself where the format allows.

Information exchanged on the WIS and between WMO centres is standardized through the use of the formats specified in the *Manual on Codes* (WMO-No. 306), Volume I.2 and the *Manual on WIS*, Volume II. These include the GRIB and BUFR formats for numerical weather prediction products and observational data and the WMO Core Metadata Profile for discovery, access and retrieval metadata. The format for the exchange of station and instrumental metadata, WIGOS Metadata Data Representation, is defined in the *Manual on Codes* (WMO-No. 306), Volume I.3.

These formats have been developed within the WMO community to enable the efficient exchange of information between WMO centres and to enable the information to be interoperable between centres and systems. The formats, including detailed technical information, have also been published openly through the WMO manuals, permitting other communities to use the formats and the information and promoting the reuse of the information.

The WMO formats specified in the manuals are subject to strong governance processes, and changes to the formats can be traced through the versions of the manuals. The code tables and controlled vocabularies are also maintained in a code repository. To enable future reuse, the technical information, including detailed format specifications, should be archived alongside information for future access. This includes any controlled vocabulary, such as BUFR tables or WIGOS metadata code lists, associated with the format.

3.1.3.3.3 Publication and exchange of information

To maximize the benefits and return on investment in the acquisition and generation of information there needs to be a clear method as to how the information will be published, exchanged and accessed by users.

Information is published on WIS through the creation of discovery metadata records. These records are publicly searchable and retrievable via WMO cataloguing services, providing access to the records via the web and via a web application programming interface (API). The metadata records should include information on how to access the described datasets and services (see *Manual on WIS*, Volume II – Appendix F: WMO Core Metadata Profile) and how to subscribe to receive updates and new data.

Technical regulations are provided in the *Manual on WIS*, Volume II. Before exchange and publication, the metadata should be assessed using the WMO Core Metadata Profile KPIs to ensure usable and high -quality metadata in addition to metadata that conforms to the technical standard.

The web standards and protocols used should be adequately documented to enable users to find and retrieve the information. This should be possible both manually and automatically via machine-to-machine interfaces and should be standardized between centres.

Updates to the information exchanged on WIS, including the publication of new information or the cessation of previously exchanged information, is published in the WMO Operational Newsletter.

Note: The newsletter is available from: https://community.wmo.int/news/operational-newsletter

3.1.3.3.4 Usage and communication

For information to have value, it must inform users, aid knowledge discovery and have an impact through informed decision -making. Ensuring that the user can make effective use of the information is an important step in the information management life cycle. This is accomplished in two ways:

1. By providing suitable information within the discovery metadata, enabling users to discover and access the information - including licensing information - and to assess whether it meets their requirements;
2. By providing user guides and documentation on the suitability of the information for different uses, including any technical caveats or restrictions on the use of the information.

For common types of information, the guides may be generic or link to standard documentation. Information on the observations available from the WIGOS is provided within the *Manual on the WMO Integrated Global Observing System* (WMO-No. 1160) and the *Guide to the WMO Integrated Global Observing System* (WMO-No. 1165). This includes information on the expected uses and quality of the data. Similarly, information on the data and products available through the WMO Integrated Processing and Prediction System is provided in the *Manual on the WMO Integrated Processing and Prediction System* (formerly Manual on the Global Data Processing and Forecasting System) (WMO-No. 485).

For non-standard and specialist products targeted user guides may be more appropriate. These should be accessible and retrievable via a link within the discovery metadata and should include a

plain text summary for the non-technical user. Any user guide should be in addition to the technical documentation described under 3.1.3.3.1 Planning, information creation and acquisition.

Updates and the availability of new information should be announced and published via the WMO Operational Newsletter (see 3.1.3.3.3 Publication and exchange of information). Other communication methods may also be used but these should not be in place of the operational newsletter. It is also recommended that users be allowed to subscribe to the newsletter to receive updates directly.

The discovery metadata should include a valid point of contact, enabling users to provide feedback and ask questions about the information provided.

3.1.3.3.5 Storage, archival and disposal

The type of storage used should be appropriate to the type of information stored. Core information exchanged operationally should be stored and made available via high-availability and low latency media and services. For some operation critical information, such as hazard warnings, there is a requirement for the end-to-end global distribution of the information to be completed in two minutes. For other operational data there is a requirement for the global exchange to be completed in 15 minutes.

The storage requirements for non-operational services and information may be different but the guidance provided in this section applies equally. Further information on the performance requirements is provided within the WIS2 technical specifications listed in the *Manual on WIS*, Volume II.

Backup policies and data recovery plans should be documented as part of the information management plan. They should be implemented either before or when the information is created or acquired and should include both the information and the associated metadata. The backup and recovery process should be routinely tested.

Business rules governing access to, and modification of the information should be clearly documented in the information management plan. These must include the clear specification of the roles and responsibilities of those managing the information. Information on who can authorize the archival and disposal of the information and the processes for doing so should be included. The roles associated with an information resource are standardized as part of the WMO Core Metadata Profile.

The archival and longterm preservation of an information resource should be identified and included in the information management plan. This may be at a national data centre and/or a WMO centre. WMO centres are recommended for globally exchanged core data and include those centres contributing to the Global Atmosphere Watch, the Global Climate Observing System and Marine Climate Data System (see *Manual on Marine Meteorological Services* (WMO-No. 558), Volume II, as well as the WMO World Data Centres and those defined in the *Manual on WIS*, Volume II and those defined in the *Manual on the WMO Integrated Processing and Prediction System* (formerly *Manual on the Global Data Processing and Forecasting System*) (WMO-No. 485).

Earth system information, especially observational data, is often irreplaceable. Other information, while technically replaceable, is often costly to produce and therefore not easily replaceable. This includes outputs from numerical models and simulations. Before an information resource is

marked for disposal careful consideration must be given to whether longterm archival or disposal is more appropriate. This consideration must follow a clearly defined process documented in the information management plan.

When an information resource is marked for disposal the reasons for disposal, including the outcome of the consultation with stakeholders and users, must clearly be documented. The disposal must be authorized by the identified owner and custodian of the information. Information relating to the disposal must be included in the metadata associated with the information resource. The metadata must be retained for future reference.

3.1.4 Other considerations

3.1.4.1 Technology and technology migration

Information managers must be aware of the need to ensure that the technologies, hardware and software used do not become obsolete and must be aware of emerging data issues. This topic is discussed further in the *WMO Guidelines on Emerging Data Issues* (WMO-No. 1239).

3.1.4.2 Information security

Further information on information security and best practices can be found in the *Guide to Information Technology Security* (WMO-No. 1115).

[1] https://go-fair.org

PART IV

4.1 Security

For this initial version of the Guide to WIS, Volume II, existing guidance on information technology security (also known as cybersecurity) remains largely applicable. Please refer to:

- *Guide to Information Technology Security* (WMO-No. 1115)
- *Guide to the WMO Information System* (WMO-No. 1061), Vol I, Appendix E - Annex To Paragraph 7.8 (ICT Service Incident Management), and Appendix F - WIS IT Security Incident Response Process

PART V

5.1 Competencies

For this initial version of the *Guide to WIS*, Volume II, existing guidance on competencies remains largely applicable. Please refer to *Guide to the WMO Information System* (WMO-No. 1061), Volume I, Appendix A - WIS training and learning guide.