Taylor & Francis
Taylor & Francis Group

Multi-scale hydrological system-of-systems realized through WHOS: the brokering framework

Enrico Boldrini, Stefano Nativi, Silvano Pecora, Igor Chernov & Paolo Mazzetti

Published online: 21 Jul 2022.

Submit your article to this journal ⌇

Article views: 1764

View related articles ⌇

View Crossmark data ⌇

Citing articles: 4 View citing articles ⌇

Multi-scale hydrological system-of-systems realized through WHOS: the brokering framework

Enrico Boldrini ®[a], Stefano Nativi ®[a], Silvano Pecora[b,c], Igor Chernov[b] and Paolo Mazzetti ®[a]

[a]Institute of Atmospheric Pollution Research (CNR-IIA), National Research Council of Italy, Sesto Fiorentino (FI), Italy; [b]World Meteorological Organization (WMO), Geneva, Switzerland; [c]Ministero della Transizione Ecologica (MITE), Roma, Italy

ABSTRACT

Global Change challenges are now systematically recognized and tackled in a growingly coordinated manner by intergovernmental organizations such as the United Nations. Heterogeneous observing networks provide the founded data sources to assess the Earth environmental status and take sound decisions to achieve a sustainable development. WMO Hydrological Observing System (WHOS) allows to discover and access historical and near real time hydrological observations. WHOS represents the hydrological contribution to the wider WIGOS-WIS system of WMO. It is a digital ecosystems framework contributed by a set of data providers and technical support centers. In this framework, three regional pilots were successfully completed. The WHOS architecture applies the services brokering style, implemented through the Discovery and Access Broker technology. A brokering approach makes a global system of systems possible and sustainable, where the different enterprise systems are enabled to interoperate, despite they implement heterogeneous communication interfaces and data models. In this manuscript, the WHOS brokering solution is detailed by recurring to the definition of a set of transversal viewpoints to describe the important aspects of the complex ecosystem –namely: enterprise, information, computational, engineering, and technological views. Finally, the three regional pilot ecosystems are described as successful cases of WHOS implementation.

1. Introduction

1.1. System-of-systems and SDG agenda implementation

Environmental changes and their resulting impacts on our Society are key challenges for humankind. Collectively known as Global Change, these challenges are even more important today than in 2003, when governments and international organizations committed to a vision of a future wherein decisions and actions for the benefit of humankind had to be informed by coordinated, comprehensive, and sustained Earth observations (GEO 2016). Consistent with that, on 25 September 2015, the United Nations General Assembly formally adopted the universal, integrated, and transformative 2030 Agenda for Sustainable Development (United Nations 2015), along with a set of 17 Sustainable Development Goals (SDG) and 169 associated targets (United Nations 2016b).

The SDG initiative requires the development of a knowledge platform (United Nations 2016a) to support the implementation of the introduced goals and their related targets. Enabled by advanced

technological frameworks (harnessing the digital transition of society), a knowledge platform is required to facilitate multi-stakeholder collaboration and partnerships (Nativi, Mazzetti, and Craglia 2021) through the sharing of data, analytical models/tools, information, best practices, and policy advice (Mazzetti et al. 2022; Guo et al. 2020; Nativi et al. 2020; Domenico et al. 2006). These platforms (also identified as geospatial digital ecosystems (Nativi and Mazzetti 2021)) are supposed to be utilized by the United Nations Member States, the civil society, the private sector, the scientific community, and the other stakeholders (United Nations 2016a).

1.2. The brokering pattern

1.2.1. The key role of architectural styles

In information engineering, an architectural style is defined as 'a coordinated set of architectural constraints that restricts the roles/features of architectural elements and the allowed relationships among those elements within any architecture that conforms to that style' (Fielding 2000). An architectural style can be designed based on other styles by adopting multiple architectural constraints to reach a compromise on the resulting characteristics. The importance of the architectural style concept lies in the fact that many characteristics of information systems (i.e. efficiency, scalability, evolvability, etc.) do not depend on the specific technological architecture adopted but rather on the architectural style. Therefore, the concept of architectural style introduces a useful level of abstraction above the concept of system architecture (Nativi, Mazzetti, and Craglia 2021). Defining an architectural style as an invariant assures that the system evolution will respect those characteristics related to that specific style. However, the system will keep a great degree of freedom, with possible changes in the system architecture to respond to new requirements and changes, as far as the resulting architecture conforms to the architectural style (Nativi and Craglia 2021a). Therefore, for making a Global Change system of systems that can evolve along with its stakeholders' requirements, the choice of an invariant and flexible architectural style is a good compromise between an invariant system architecture (greatly limiting evolvability), and an unrestricted architecture (which is not able to avoid disruptive changes in fundamental characteristics). The brokering style is a good example of invariant and flexible architectural style.

1.2.2. The brokering style

In early 2000, the technological scenario of geospatial system-of-systems was characterized by the attempt to port the principles of enterprise Service-Oriented Architecture (SOA) to the Web. Later, for the emergence of data variety challenges, the geospatial system of systems moved to a Layered-Client-Server (LCS) style (see Figure 1). This architectural style adds proxy and gateway components to the well-known Client-Server style (Fielding 2000). By adding 'proxy and gateway components', which are specifically dedicated to implement mediation and harmonization tasks, we realize a Brokered style. Finally, the appliance of caching solutions accomplishes the Brokering style – as depicted in Figure 1. In keeping with the Brokering style, a geospatial System-of-Systems can introduce new intermediate components to provide added-value services, including semantic components for enhancement discovery and use services.

The explicit choice of a brokered architecture style vs. a basic Client-Server style assures a greater scalability and evolvability due to a better separation of concerns dedicating specific resources to (the complex and evolvable) interoperability tasks. It is worth noting that the brokered (layered) architecture style constrains service clients to connect servers through brokers (i.e. brokering components), largely decoupling their dependency and, hence, facilitating the system-of-systems evolution by including new architectural styles. For example, recently, new challenges on data volume and velocity (i.e. Big Data and analytical services) introduced the Mobile Code constraint – i.e. the basic architectural style that requires to move the analytical code where Big Data is, and not vice-versa, as traditionally done with the Spatial Data Infrastructures (SDIs) paradigm.
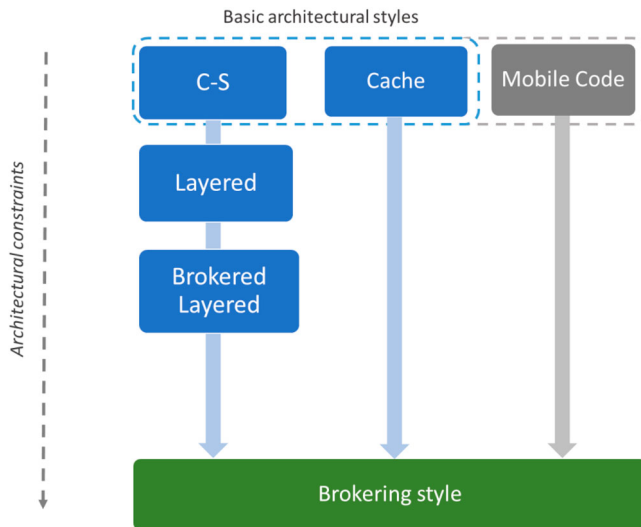
Figure 1. Architectural constraints of the Brokering Style, applied by geospatial System-of-Systems.

Indeed, brokered layered architectures are more viable because of increased flexibility and evolvability. On the other hand, these architectural constraints commonly require a more complex governance as a result of the introduction a new layer and the related components and services. This challenge is becoming increasingly topical with the emergence of the paradigms of Digital Transformation (see, for example, the Web-as-a-Platform and Digital Twins patterns (Nativi and Craglia 2021b)), and the Mobile Code architectural constraint (Nativi and Craglia 2021a).

2. The WHOS case

The goal of observations of the hydrological cycle is to collect reliable data for use in water resources planning and decision-making, including for managing flood and drought conditions, integration into hydrological and climate applications and services, and for research […] Hydrological datasets have intrinsic value and are worth the huge human and financial commitment required to collect them over long periods of time. (WMO 2022)

The concept of developing a WMO Hydrological Observing System (WHOS) was first proposed (by the WMO Commission for Hydrology) in 2013. In 2015, the World Meteorological Congress urged the promotion of WHOS among National Hydrological Services (NHSs) and the hydrological community. The Congress advocated for a full implementation of WHOS.

The present paper describes the original contributions of the authors to WHOS focusing on the design and technical implementation of its brokering component – the WHOS brokering framework. At the same time, useful background on WHOS is provided to frame the contributions in the general picture.

2.1. A contribution to WIGOS-WIS

The WMO Hydrological Observing System (WHOS[1]) is the hydrological component of WIGOS[2] (the WMO Integrated Global Observing System, a top priority of the new overarching framework for all WMO observing systems). In addition, WMO is a demonstration project for the WMO Information System 2.0 (WIS2.0[3]). Consequently, WHOS aims to illustrate, evolve, validate and/ or refine the concepts, solutions, and implementation approach of WIS 2.0, especially in hydrology. In keeping with that, WHOS is implemented by applying the WIS2.0 principles. Comprehensively,

the following WIS 2.0 (present and next future) principles have been considered in the WHOS project – especially, in the design of the WHOS brokering framework:

Mandatory principles

Principle 1: to adopt Web technologies and leverages industry best practices and open standards;

Principle 2: to utilize Uniform Resource Locators (URL) to identify resources;

Principle 3: to prioritize use of public telecommunications networks (i.e. Internet) when publishing digital resources;

Principle 4: to require provision of Web service(s) to access or interact with digital resources (e.g. data, information, products) published using WIS;

Principle 5: to encourage National Centers (NCs) and Data Collection or Production Centers (DCPCs) to provide 'data reduction' services via WIS that process 'big data' to create results or products that are small enough to be conveniently downloaded and used by those with minimal technical infrastructure;

Optional principles (mandatory in the next future)

Principle 6: to add open standard messaging protocols that use the publish-subscribe message pattern to the list of data exchange mechanisms approved for use within WIS and GTS;

Principle 7: to require all services that provide real-time distribution of messages to cache/store the messages for a minimum of 24-h and allow users to request cached messages for download;

Principle 8: to adopt direct data exchange between provider and consumer;

Principle 9: to phase out the use of routing tables and bulletin headers

Principle 10: to provide a Catalogue containing metadata that describes both data and the service(s) provided to access that data;

Principle 11: to encourage data providers to publish metadata describing their data and Web services in a way that can be indexed by commercial search engines.

2.2. The WHOS implementation strategy

The implementation of WHOS is being carried out in two phases (WMO 2022):

- Phase I: providing a map interface with links to those NHSs that make their real-time and/or historical hydrological data available online.
- Phase II: providing a fully WIS-compliant services-oriented framework linking hydrologic data providers and users through a hydrologic information system-of-systems enabling data registration, discovery, and access.

In keeping with that, the strategy to implement WHOS has applied a bottom-up approach (from regional systems to a global one). Indeed, an important WHOS objective is to address multi-scale hydrological data sharing. Because of this aim, the implementation plan has consisted of identifying potential regional systems which could have gained benefits from WHOS, being characterized by a shared governance to ease the implementation process. Presently, three WHOS regional prototypes have been successfully implemented, namely in the La Plata Basin, the Arctic Region, and the Dominican Republic. While further regional systems are planned to be added in the future, this strategy is now also being assisted by a top-down approach: the WHOS Operational Plan (2024-2029) is currently being developed building on the achievements, the experience, and the lessons learned from the regional prototypes.

The Operational Plan will cover multiple objectives, also thanks to the work of dedicated Support Centers, including:

- Assuring a continued operativity of WHOS infrastructure components (mainly WHOS broker) leveraging a cloud-based deployment
- Assuring advancements to the WHOS infrastructure components as required (e.g. addition of new accessor and profiler components to the WHOS broker to support new data publication systems and user tools)
- Providing support and training to organizations willing to join WHOS with their systems (both data providers and data users)

3. The brokering architecture of WHOS

WHOS applied a brokering approach. To describe the WHOS architecture, in the following sections a viewpoints framework formalism is used as a guideline: the *RM-ODP*.[4] This is a standard specification (i.e. ITU-T Rec. X.901-X.904 and ISO/IEC 10746) for describing distributed software systems, by defining transversal viewpoints: different views are used to represent the whole system from the perspective of a set of different concerns.

The *RM-ODP* is a well-used formalism to define architectures, which is compliant with IEEE 1471, and it is freely available (ISO/IEC 1998). In the rest of the manuscript, we will use the standard Unified Modeling Language (UML) for representing the architecture views graphically – see UML4ODP (ITU-T 2014). For the readers' convenience, a legend, comprising the most common UML symbols used in the view diagrams, is represented in Figure 2.

Five viewpoints will be defined to describe the WHOS brokering framework, namely:

- **Enterprise view:** purpose, scope and policies governing the activities of the WHOS broker, including requirements and typical processes;
- **Information view:** the types of information handled by the WHOS broker system and its processing;
- **Computational view:** the WHOS broker components and their interfaces;
- **Engineering view:** the infrastructure required to provide the WHOS broker functionalities;
- **Technology view:** technologies used in WHOS and by WHOS broker.

Each viewpoint addresses a set of different concerns that commonly interest diverse system contributors, like system administrators, software developers, production engineers, and interoperability experts.
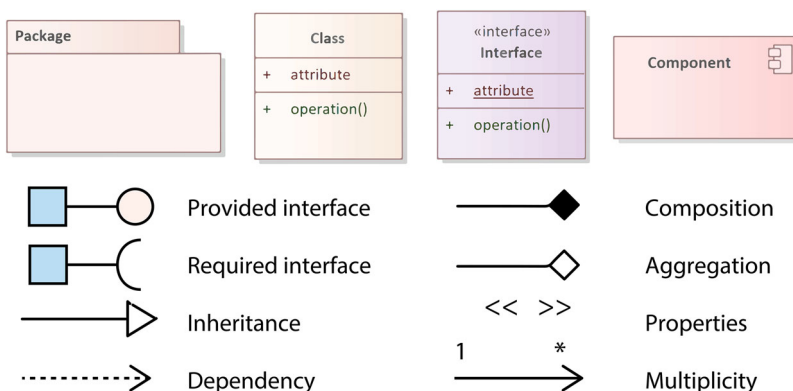

Figure 2. Common symbols used in the UML diagrams.

3.1. Enterprise view

The main WHOS stakeholders and/or administrators and their respective objectives/roles are specified as follows:

(a) **World Meteorological Organization (WMO)** – leads, guides and coordinates the WHOS development and implementation as the hydrological component of the WMO Information System (WIS);
(b) **Support Centers** – manage, maintain, and deploy the WHOS technology;
(c) **Data Providers** – publish hydrological data using web services and make them available through WHOS;

(d) **Data Users** – discover and access data available through WHOS using available tools and applications;
(e) **Tools Developers** – develop and/or maintain tools, applications and modeling systems used by data users;
(f) **WHOS Broker community** – evaluates and advances the WHOS brokering framework technology.

3.1.1. WMO's role in WHOS implementation

The WMO Constituent Bodies lead and guide the WHOS development and implementation as a component of the WIGOS and WIS. The WMO Secretariat supports the work of the WMO Constituent Bodies and facilitates the WHOS development and implementation.

3.1.2. Support Centers

The main role of the Support Centers is to manage, maintain, and deploy WHOS infrastructure technologies (WHOS broker being the core one) as well as to support WHOS implementation in countries and regions. Currently, there are two operational Support Centers:

- **WHOS broker Support Center**, based in the Institute of Atmospheric Pollution Research of the National Research Council of Italy (CNR-IIA). CNR-IIA has envisioned the brokering approach to realize a distributed system of systems and is the main designer and maintainer of the DAB software framework[5] powering the WHOS broker. The Center's main tasks are:
  (a) To host and optimize the WHOS Broker infrastructure;
  (b) To develop and implement the WHOS Broker components;
  (c) To support the implementation of the WHOS Hydrological Ontology;
  (d) To provide support and training for the WHOS Broker operation, also using instruments such as the WHOS distance learning courses and webinars.[6]
- **La Plata River Basin Support Center**, based in the Brazil National Institute of Meteorology (INMET).[7] INMET is also a Global Information System Center (GISC) for WIS. The Center's main tasks are:
  (a) To host and maintain the WHOS Broker infrastructure as well as the WHOS web portal for the local basin. In order to operate and maintain WHOS at the regional level, the Support Center has received face-to-face training;
  (b) To provide support and assistance regarding WHOS implementation to the participating countries.

Within each Support Center, the following roles are assigned regarding the WHOS implementation:

- **Focal point**, consisting in the primary contact point of the hosting organization regarding governance and administrative matters.
- **Technical expert(s)**, deemed responsible for the WHOS Broker administration and advancements.

3.1.3. Data providers

Data providers are institutions – e.g. National Meteorological and Hydrological Services: (NMHSs), hydropower companies – that publish their data through web services and make them available via WHOS. The publication system of each data provider should offer discovery and access functionalities by means of communication protocols, which are described by international standards (e.g. ISO, OGC, W3C), or by documented/open community standards.

Within each data provider institution, the following roles are assigned regarding the WHOS implementation:

- **Focal point**, the primary contact point of the institution responsible for data sharing via WHOS. In addition, in his/her institution, the focal point is often called to coordinate the activities regarding the WHOS implementation.
- **Technical experts**, the institution's staff responsible for technical-related issues (e.g. web-service availability; data catalogue development; data, metadata, and semantic services)
- **Infrastructure administrator(s)** are the expert(s) responsible for managing the data publishing infrastructure (servers, DBs, network)

In some institutions, multiple roles can be assigned to the same person.

3.1.4. Data users

Data users are public and private entities, as well as individuals discovering and accessing data that are available through WHOS by using published tools, applications, and modeling systems. For hydrological data, the requirements of data users include:

- to produce hydrological forecasts and early warnings;
- to continuously monitor the status of specific areas;
- to develop and innovate methods and models for hydrological analysis and forecasting;
- to produce short-term simulations and longer-term climate change scenarios for decision support;
- to assess global environmental issues and carry out environmental monitoring, including in the framework of international treaties and agreements.

Generally, WHOS users need near real-time discovery and access of hydrological data at different levels on a local, national, regional, and global scale. WHOS must support the tools that are commonly used by the WHOS community. According to the *Terms of Use* of WHOS:

- The user accepts all risks which may occur by using the data available on WHOS and accepts to not use the data for commercial purposes without the prior consent of the original data provider, noting that specific conditions of uses and licenses might apply.
- The user may not copy content available on WHOS to create a database in electronic or any other format, or publicly use and distribute it to third persons without prior consent of the original data provider.
- The user will attribute the source of the data for scientific publications and for operational products and services.

3.1.5. Tool developers

Various tools and applications are designed and implemented by different communities to discover, access, and perform user-specific functionalities (e.g. filtering, downloading, analyzing, modeling, alerting) by utilizing data accessible through WHOS – as depicted in Figure 3.

Traditionally, user applications and tools can be either web- or desktop-based. However, also due to cloud storage and increased availability, they are increasingly shifting towards a web environment. Tools and applications can be either proprietary or open-source. Proprietary tools and applications include those that require licensing fees and those that can be used for free. Open-source tools and applications come with a free license and grant users the rights to modify and redistribute the tool, as shown in Figure 4.

Tools can be distinguished in terms of their functionalities. For example, web portals can be used for data consultation; interactive maps and plots can be used for data visualization; GIS applications can be used for data visualization and analysis; statistical libraries can be used for data processing; and custom model software can be used for data processing and modeling. All tools and applications are developed and implemented by different communities. To guide the development of different tools and applications, various standards and best practices have been developed and published by the standardization bodies at different levels (international or community). These standards are used by different developers to design their tools and applications. Globally, it leads to the availability of numerous tools and applications, each of them implementing one or more standards. It is important to note that implementations for the same specific standard may differ.

To summarize, Tools Developers are institutions and individuals developing and/or maintaining applications, tools, and/or modeling systems that are used by Data Users to discover, access, and use the data which are available through WHOS.

3.1.6. WHOS broker community

The WHOS broker is the component allowing a harmonized discovery and access to specific hydrology data that is published online through distributed and heterogeneous services, which are managed by the different WHOS data provider (see Figure 5). As such, the WHOS broker adopts different web technologies, such as different messaging protocols, data-interchange formats, backend and front-end languages, as needed to implement diverse web service interfaces, also according to the community's
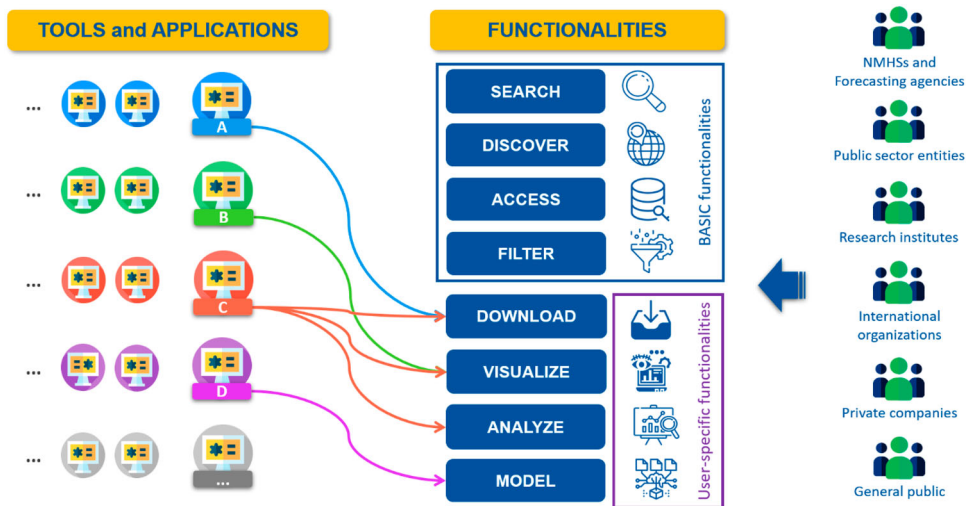

Figure 3. Different tools and applications associated with user-specific functionalities are made available to different user types. They all leverage data searched and accessed from WHOS.
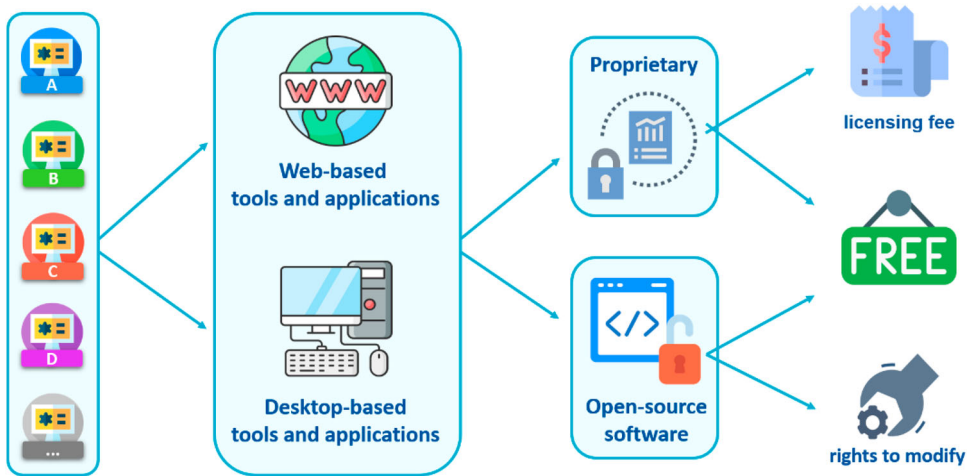
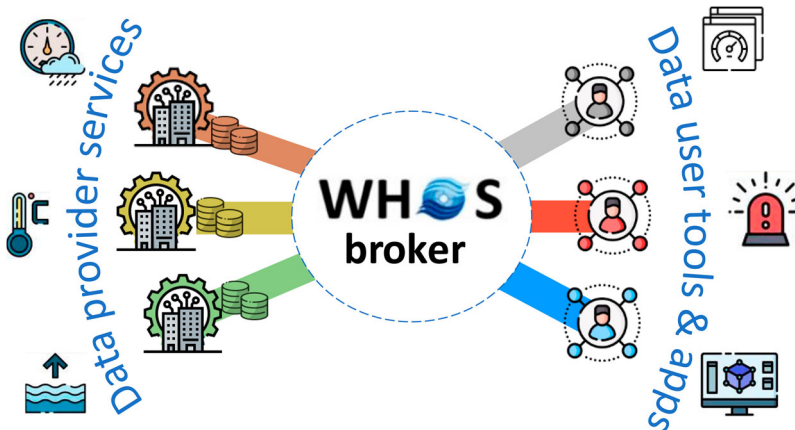**Figure 4.** Types of tools and applications.

Figure 5. The WHOS broker system connects the systems made available by its data providers and user tools.

best practices (for example, Linked Data principles and FAIR principles), international directives (for example, the European directive INSPIRE), open standards (ISO, OGC, and W3C), and in accordance also with **WIS 2.0 Principles 1, 2, 3 and 4** (see Section 2.1).

WHOS developers are responsible for implementing the requirements specified and indicated by WMO, by adding new software components (modules) to the WHOS broker - or advancing the existing ones. **WHOS administrators** are instead responsible for operationally managing a deployment of the WHOS broker, for instance, by installing and configuring it on a specific cloud environment. The WHOS broker should operatively implement the requirements coming from WMO, including brokering new data/semantics providers, publishing new service interfaces, and extending the metadata model to add new queryables/filters. For the WHOS broker, the main functional and system requirements implemented by the WHOS broker are:

- **Harmonized dataset discovery**: to enable queries for datasets against a heterogeneous set of data providers. Each query is characterized by an extensible set of user constraints, including keywords, spatial temporal extents, data provider.

- **Support publish/subscribe message pattern**: to inform interested users of the availability of updated resources of their interest, in accordance with WIS 2.0 Principles 6 and 9 (see Section 2.1).
- **Dataset discovery protocols (data sources)**: heterogeneous data publication systems are made available by data providers, and additional ones will join WHOS in the course of time. The WHOS broker is required to implement support for each communication protocol and enable mapping from the data provider data models towards a harmonized internal one.
- **Dataset discovery protocols (user tools)**: heterogeneous data tools are used by WHOS users and additional ones will join WHOS in the course of time. The WHOS broker is required to implement support for each communication protocol by publishing the required discovery service interfaces. Discovery service interfaces can be used to realize catalogs in accordance with WIS 2.0 Principle 10 or indexes for commercial search engines in accordance with WIS 2.0 Principle 11 (see Section 2.1).
- **Interoperability APIs**: Application Programming Interfaces for tool developers (web portal developers) enhance the connection to the WHOS broker functionalities in a specific language (e.g. JavaScript). Additional APIs should be supported by WHOS in the course of time.
- **Ranking metrics**: to have results ordered by importance (the definition of importance is based as a customizable formula, which is dependent on a query matching score and a quality of results score).
- **Paging**: to browse big result sets page by page.
- **Views**: a view is a virtual instance of WHOS having all the service interfaces and functionalities available but working only on a specific subset of all the available resources: the ones that are of interest for a particular community or objective (for example, the WHOS-Plata and WHOS-Arctic views containing the acquisition respectively on the Plata River basin and the Arctic Region). A view is defined with an identifier, a description and a query identifying the subset of interest.
- **Semantically enhanced discovery**: obtain enhanced query results by expanding the user's query terms by consulting ontologies available, as semantics service (like the WMO Hydrological Ontology web service). Such ontologies define hydrological concepts and the relations between them, including equivalent relations supporting multilingual applications.
- **Ontology providers protocols**: support the communication protocol needed to access the semantics capabilities of a given semantics service (for example, the WMO Hydrology Ontology is based on a SPARQL endpoint).
- **Filters/Facets discovery**: faceted search consists in presenting the actual values documenting a specific metadata element in a set of resources, whose objective is to have the user to select one of the values to act as a result-set filter.
- **Harmonized dataset access**: it enables seamlessly access of heterogeneous data sources to download data in common standard formats and having it transformed by means of simple transformations.
- **Implement specific data transformations**: to support a specific simple transformation (e.g. data format conversion, CRS reprojection, interpolation, subsetting). Implementation of subsetting and downsampling functionalities is in accordance with **WIS 2.0 Principle 5** (see Section 2.1).
- Other significant non-functional requirements considered by the WHOS broker implementation are:
- **High reliability**: since the broker is a central component of the WHOS system of systems, it should be operationally maintained in a suitable environment to assure high reliability and fault tolerance (e.g. to single node failures).
- **High performance**: best-optimized data should be dispatched to users (as near real-time data is usually requested), also in accordance with **WIS 2.0 Principle 7 and 8** (see Section 2.1).
- **Scalability**: to assure the correct functioning for an increasing number of data providers, users, requests.

- **Security**: to enable different levels of data sharing based on user credentials and data provider policies.
- **Accuracy**: it should not introduce loss of data/metadata quality (e.g. in metadata mappings and data transformation).
- **Flexibility**: to be capable of supporting existing and emerging standard (e.g. new data publication system or application types)
- **Modularity**: to easily support extensions through additional modules (opposite to a *monolithic* architecture)
- **Sustainability:** example-given, the technology should be based on a community-maintained open-source code, in order to accept future contributions from the WHOS community and to avoid lock-ins towards a specific vendor.

3.1.7. Primary WHOS enterprise processes

3.1.7.1. Brokering a new data source. When a data provider decides to share data through WHOS, the process requires five main steps, as shown in Figure 6.

Step 1: The data provider focal point requests to participate in WHOS by sending a request (e.g. by email) to the WMO, providing a general description of the motivation and the contact information of the technical experts who are going to be responsible, in the institution, for data publication.

Step 2: WMO contacts the technical experts of the data provider requesting the technical details necessary to connect the provider's data publication services to WHOS.

Step 3: If the data provider's service type is already supported by the WHOS broker, WMO configures a new data source within WHOS Broker. Otherwise, WMO must first request one of the Support Centers to implement a new WHOS Broker component (i.e. a new *accessor*).

Step 4: Troubleshooting and feedbacks take place between the Support Center and the Data Provider technical experts, until the component is fully operational.

Step 5: A final implementation feedback report is provided to the Data Provider Focal Point.

In Step 2, if the hydrological data are not yet published online through web services, the Support Center can help to identify potential standard technologies and solutions. On the other hand,
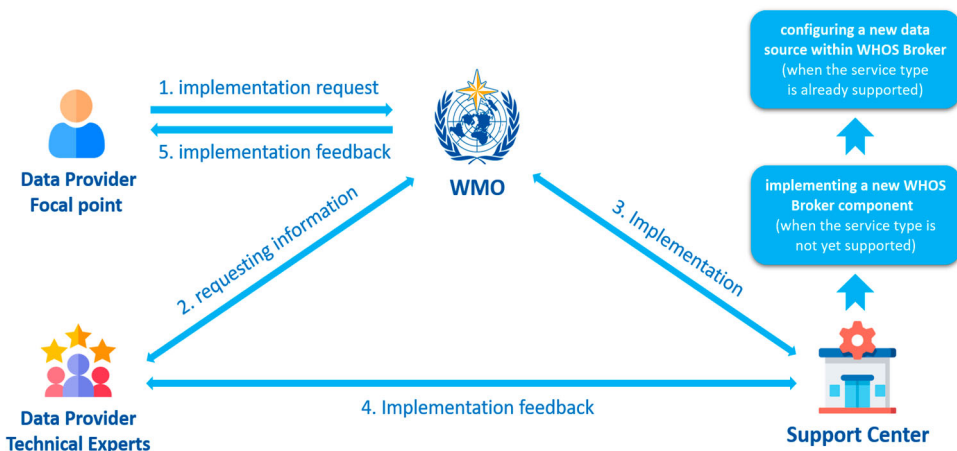

Figure 6. Brokering a new data source diagram.

when hydrological data is already published online through web services, the provider's technical experts should produce information on:

(1) the web service(s) endpoint (URL) to discover and access the published data.
(2) the credentials (e.g. specific IP or login details), in the case of restricted access.
(3) the documentation on the web services (if available), and the description of the metadata and data models applied by the provider.

Based on this information, the relevant Support Center will carry out service protocol(s) interoperability – i.e. interface(s) and metadata/data models interoperability.

The WHOS interoperability test will run checks on the following:

(1) Service(s) connectivity: the data publication system is reachable online.
(2) Data discoverability: the data can be discovered by means of Internet requests.
(3) Data Accessibility: the data can be accessed by means of Internet requests.

\If needed, one or more cycles of tuning & testing can take place as shown in Figure 7. Once the data provider passes the interoperability test, the published data becomes part of WHOS.

If any improvement is required, implementation feedback is provided back to the data provider. Improvements are often required at the level of metadata model: during the process of metadata harmonization, the WHOS experts work closely with the data provider to enrich metadata where necessary. Improvements may also be required at the encoding and web service levels.

When data successfully becomes part of WHOS, it is readily available through one or more WHOS *views*. A *view* represents a subset of the entire WHOS data content. For example, a *view* can be defined by specifying clauses on a given area, institution, temporal extent, and/or observed variable. *Views* are defined by a specific set of users (community of practice), as shown in Figure 8. The data provider can decide, along with WMO, which *views* will include its data.
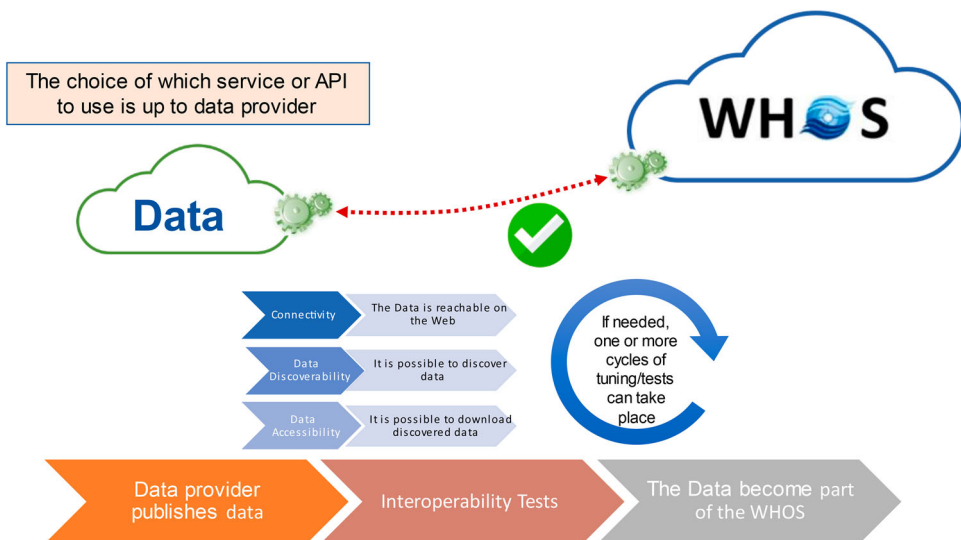

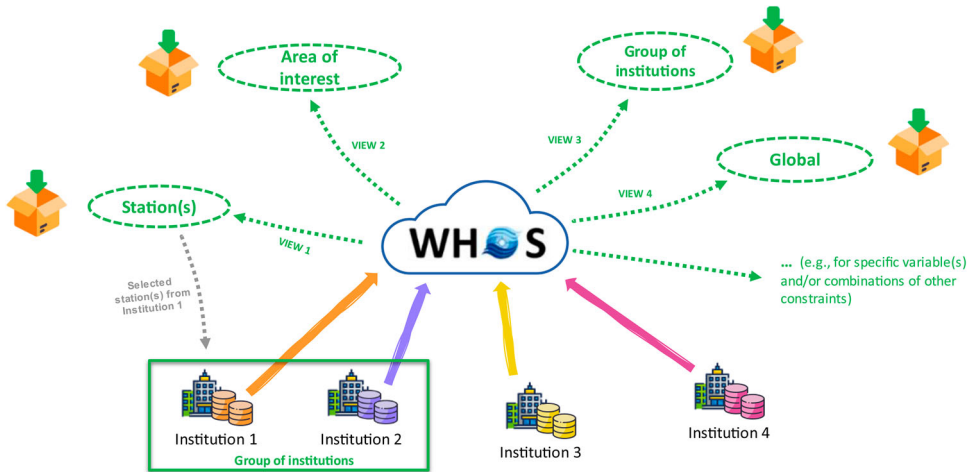**Figure 7.** Implementation cycles to broker a new data source.

Figure 8. WHOS broker views functionality provides logical subsets of the entire WHOS data offering tailored for specific user communities.

3.1.7.2. Brokering a new user application/tool. This section describes the process of connecting a new application or tool that is not yet supported by WHOS, as depicted in Figure 9:

Step 1: The application/tool developer (or the application/tool user) requests to participate in WHOS by sending a request (e.g. by email) to WMO and providing the application/tool documentation (including, where available, details on the web service interface required to support the application/tool). If possible, the request should also include a link for downloading the tool for its testing.

Step 2: If the application/tool is already supported by the WHOS broker, the relevant Support Center configures the publication of the required service interface within the WHOS Broker. Otherwise, the Support Center must first implement a new WHOS Broker component (i.e. a new *profiler*)

Step 3: Implementation testing and feedback take place between the Support Center and the application/tool developer, until the component is fully operational.
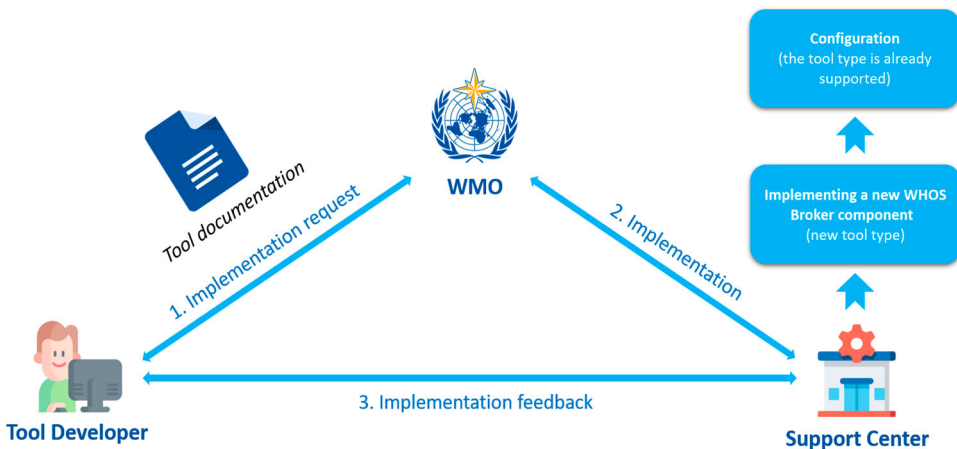

Figure 9. Brokering a new tool diagram.

3.2. Information view

This view deals with the information that is managed by the WHOS broker including its processing. The information model of the WHOS broker is quite complex, being composed of many interconnected **packages**, which contain, in turn, related sets of information objects – see Figure 10. Each object (as the occurrence of a class) is characterized by a set of properties, which the WHOS broker handles to implement its requirements. Referring to Figure 10, the main packages are:

- **Core**: holding information about the main objects handled by the brokering system, including the classes: Resource, **Resources Collection**, **Source**, **Relation**
- **Query & View**: required to realize discovery and view functionalities, including classes such as: **User**, **Query (including Constraint)**, **Request**, **View**
- **Result Set**: holding information about results of a discovery process, including the classes: **Result Set**, **Count Set, Element Value Frequency**
- **Metadata**: holding information to describe resources, including the classes: **Metadata Element, Core element, Augmented element, Extended element, Original metadata, Identifier**
- **Semantics**: semantics-related classes, such as: **Ontology**
- **Service**: holding information to describe different geo information services, including the classes: **Service**, **Access Service**, **Discovery Service**, **Processing Service**
- **O&M**: to model concepts from Observation and Measurement model, such as: **FOI**, **Observation**, **Sensor**
- **Dataset**: to handle information needed to realize the access (download) functionality, such as: **Dataset**, **Encoding**, **Thumbnail**, **Variable**
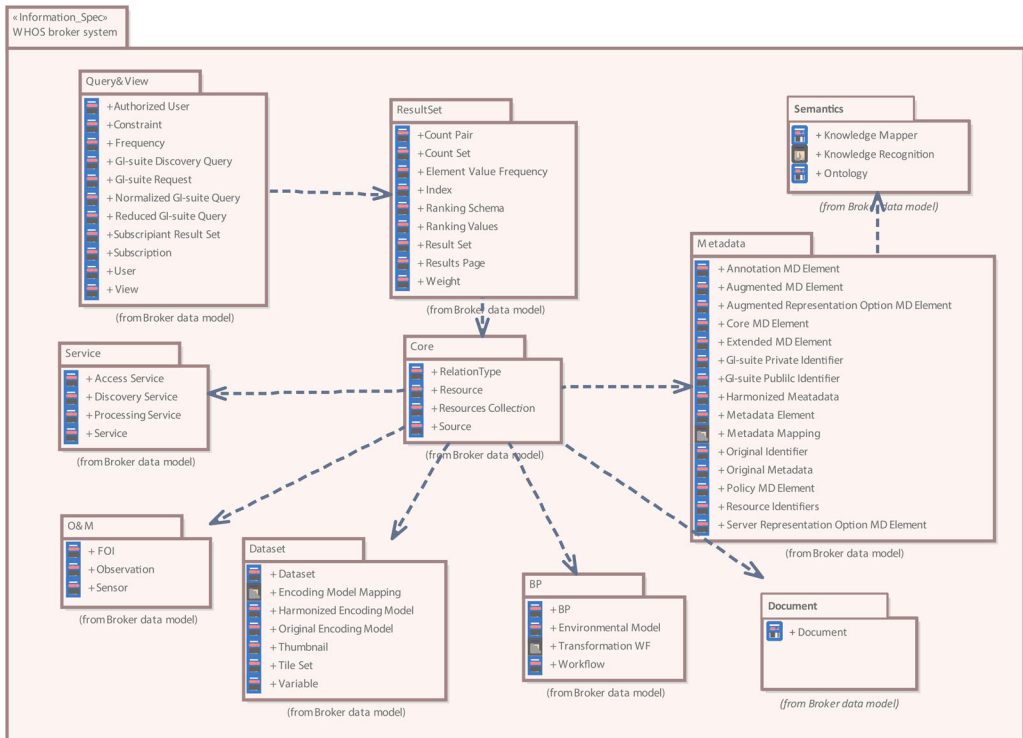

Figure 10. Information schema diagram depicting WHOS broker interconnected packages.

- **BP**: to handle information for the execution of business processes, such as: **BP**, **Workflow**, **Environmental Model**
- **Document**: to handle descriptive resources

3.2.1. Data and metadata modeling

The two most important information packages (handled by the WHOS broker) are the **metadata** and the **data** packages. They deal with the information describing hydrological and meteorological data being shared by the WHOS data providers and accessed and processed by the WHOS users. Generally, data is either acquired by hydrological and meteorological sensors (e.g. gauges, pluviometers, remote sensing instruments) or is produced by forecasting and simulation models. User tools and applications can further elaborate and process it to gain insights or produce forecasts and simulations. Indeed, it is possible to distinguish:

- **Data**: the actual measurements (i.e. the ordered entries 'timestamp, observed quantity' representing a time series).
- **Metadata**: information about the observed data that is produced by the data provider. Metadata describes data by a set of elements that can be further differentiated depending on their function:
  - **discovery** (e.g. observed parameter, spatial–temporal extent, originator),
  - **evaluation** (e.g. spatial–temporal resolution, quality, license information)
  - **use** (e.g. data encoding, units of measurements, online distribution information).

This information is being autonomously structured by each data provider, using metadata and data elements defined by different abstract models, which are afterwards encoded by using different schemas. Then, this information is published by the data providers through their autonomous data publication systems. Data user applications and tools can access and further use (hydrological and meteorological) data only if it is structured according to the data model and encoding expected by them. Therefore, in general, the data published by a specific data provider cannot operate with all the tools in use by the WHOS users. Standardization efforts from international standardization bodies (notably, ISO and OGC) have significantly helped to reduce this heterogeneity of information models and schemas. However, it seems unlikely that a single model and/or schema will prevail and be used by all the actors in a short time.

To fill this gap, the WHOS broker processes the information coming from the different data providers in two steps, as illustrated in Figure 11:

(1) the heterogeneous **original information**, which comes from the diverse data provider, is first mapped to the correspondent **harmonized information**, which is compliant with the internal WHOS broker information model;
(2) then, the **harmonized information** is mapped to the **target information** model, which is required by the specific user tools and applications.

3.2.2. Metadata mapping and augmentation

As depicted in Figure 12, the original metadata (published by a data provider) consists of multiple elements that are described by the original metadata information model. The original metadata is mapped by the WHOS broker to the harmonized metadata model, which is composed by a set of core elements (taken from ISO 19115 metadata model) and a group of extended elements (taken from other standard models, such as the WIGOS metadata model) whenever the core elements are not sufficient. The mapping procedure is lossless, as the WHOS broker information model
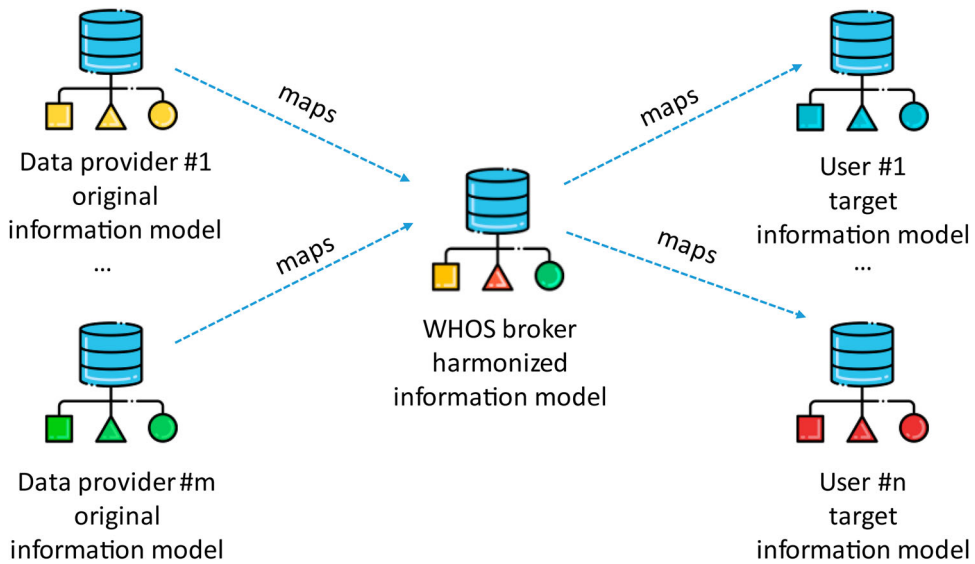
Figure 11. Information models and mappings.

was designed to be extensive enough to accommodate the most common elements shared by hydrological and meteorological data providers, being based on more than 400 metadata elements from ISO-19115 and can be as well extended through the extended schema mechanism. Finally, the harmonized metadata elements can be mapped to the target metadata ones, which can be further used by the end-users in their applications. Once again, each element from the harmonized metadata is mapped to the correspondent element in the target metadata schema.

As an example, a data provider might publish its original metadata according to the CUAHSI WaterML 1.0 schema, where the metadata element 'variable name', from the 'variable info' section
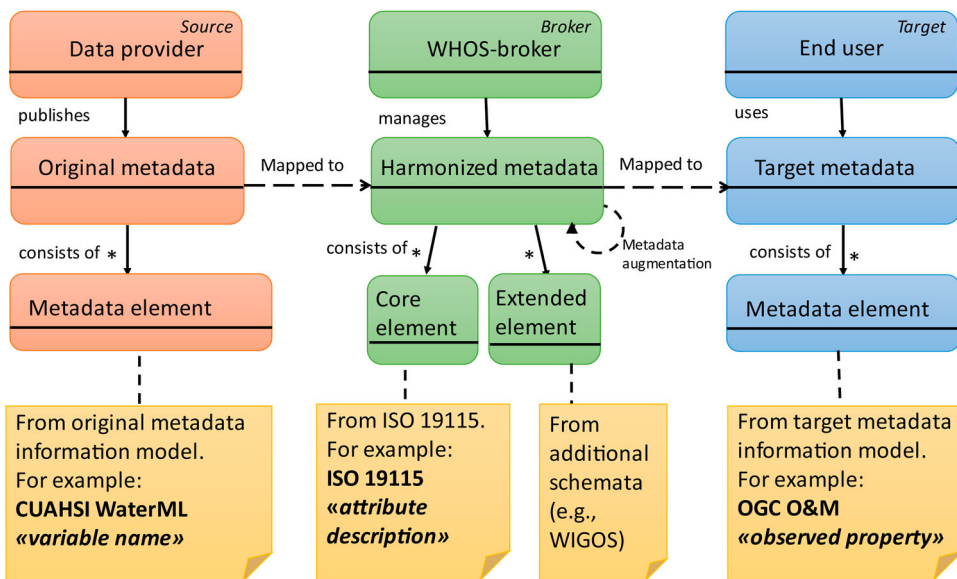

Figure 12. Metadata processing model.

is used to describe the measured parameter (e.g. discharge, precipitation and so on). During the mapping process the 'variable name' element is mapped to ISO 19115 'attribute description' metadata element, from the 'content information' section of ISO 19115. Finally, the 'attribute description' element is mapped to the 'observed property' metadata element from the 'observation member' section of O&M target schema.

Referring to Figure 12, the harmonized metadata class is characterized by a self-relationship, due to the existence of the metadata augmentation process. This practice starts from a harmonized metadata instance and aims at modifying its metadata elements by adding new ones and/or improving the existing elements. WHOS broker can be configured to perform one or more metadata augmentation processes on the metadata published by the data providers. By instance, the following are a common couple of possible augmentation processes:

- **Keyword augmenter**: if no keywords are present in the keyword section, this augmenter can automatically extract (e.g. by ignoring stop words) relevant keywords from other elements (e.g. title, abstract, responsible organization, attribute description, …). For example, in case the title is 'Precipitation at Golden Gate bridge' the keyword augmenter could extract 'precipitation' and 'Golden Gate bridge' as additional keyword elements.
- **Linked data augmenter**: this augmenter checks a specified set of metadata elements (for example responsible organization, attribute description, attribute units) to check if the free text value corresponds to a term from a specified controlled vocabulary. If the response is affirmative, then the correspondent concept URI is added as an additional metadata element. This augmenter is useful for end user applications based on linked data. For example, in case the controlled vocabulary to be checked is the 'WMO Codes Registry' and the 'attribute description' element has value 'River discharge', then there is an exact match with the title used by the concept in the 'WMO Codes Registry' identified by the URI http://codes.wmo.int/wmdr/ObservedVariableTerrestrial/171 and as such, an additional metadata element 'attribute description concept URI' can be added with the specified URI as its value unambiguously describing the attribute as a known concept from a structured domain ontology.

3.2.3. Data mapping and processing

WHOS data mapping process is approached in a way like the metadata mapping process previously described – see Figure 13. In general, each data provider encodes and publishes original data according to different data models. The WHOS broker converts the original data encoding to a
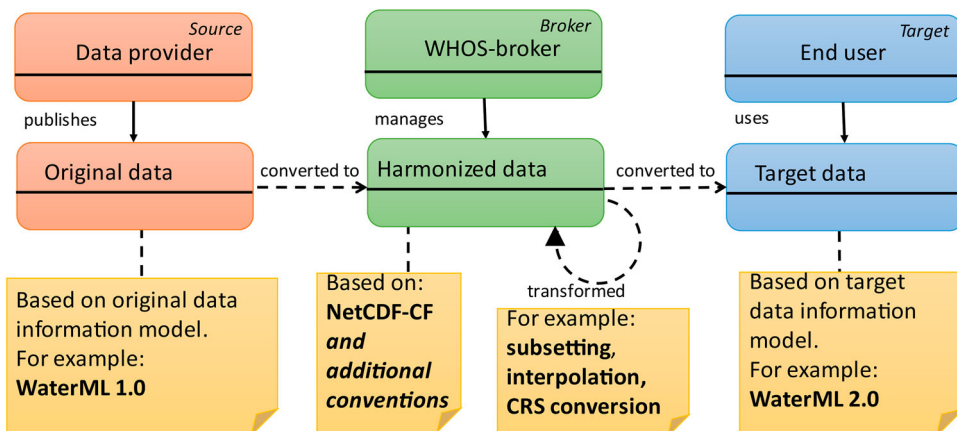

Figure 13. Data processing model.

harmonized one, using as its basis the NetCDF standard with Climate and Forecast (CF) conventions plus others.

Simple data transformations (e.g. subsetting, interpolation, and change of coordinate reference system) can be executed by the WHOS broker; they are especially needed whenever the data provider doesn't offer such functionalities, but these are requested by the end user. The self-relationship characterizing the 'Harmonized data' class indicates that simple transformations are applied to a Harmonized data instance and produce another Harmonized data object, as well, maintaining the same data encoding. Finally, data is converted from the harmonized data model to the target data model that is required by the end user.

For instance, a data provider might publish its data according to the 'WaterML 1.0' data model, WHOS broker can access it and convert it to 'NetCDF-CF' during the harmonization procedure, then it can optionally perform additional simple transformations (such as subsetting and change of CRS) and finally convert it to the 'WaterML 2.0' data model, required by the end-user application.

3.3. Computational service view

The computational viewpoint presents the functional decomposition of the WHOS system focusing on the WHOS broker components. The main components are described, along with their interactions, regardless of their distribution over networks and nodes. The macro components of the WHOS complex system (i.e. a System-of-Systems) are depicted in the diagram of Figure 14, along with the macro interfaces used to connect them:

- **WHOS broker**: the core component of the WHOS infrastructure, which interconnects all the other components.
- **User tool**: a user instrument that interacts with the WHOS broker through its discovery and/or access interfaces. It could be either a web portal, a desktop application, a model, or an alerting system.
- **Data provider publication system**: a software system for data publication, which is accessed by the WHOS broker through its discovery and/or access interfaces.
- **Semantics service**: a network service, accessed by the WHOS broker through its semantics interface.
- **Web configurator**: this application is used to configure the WHOS broker through its configuration interface.
- **Metadata DB**: this component is used by the WHOS broker as a metadata cache, to optimize discovery requests.
- **Data DB**: this component is utilized by the WHOS broker as a data cache, to optimize access requests.
- The macro components interact with each other at the computational interfaces. Depending on the implemented computation, it is possible to distinguish three abstract macro interfaces – or stereotypes:
- **Discovery interface**: to enable the discovery of resources of interest.
- **Access interface**: to enable access to selected resources.
- **Configuration interface**: to enable the configuration of the WHOS broker.

Each implemented service interface may belong to one or more macro interfaces (as represented in Figure 15) and is characterized by a set of operations depending on the available functionalities. Each operation is characterized by input and output parameters. An updated list of implemented interfaces is available on the WHOS official site.[8]

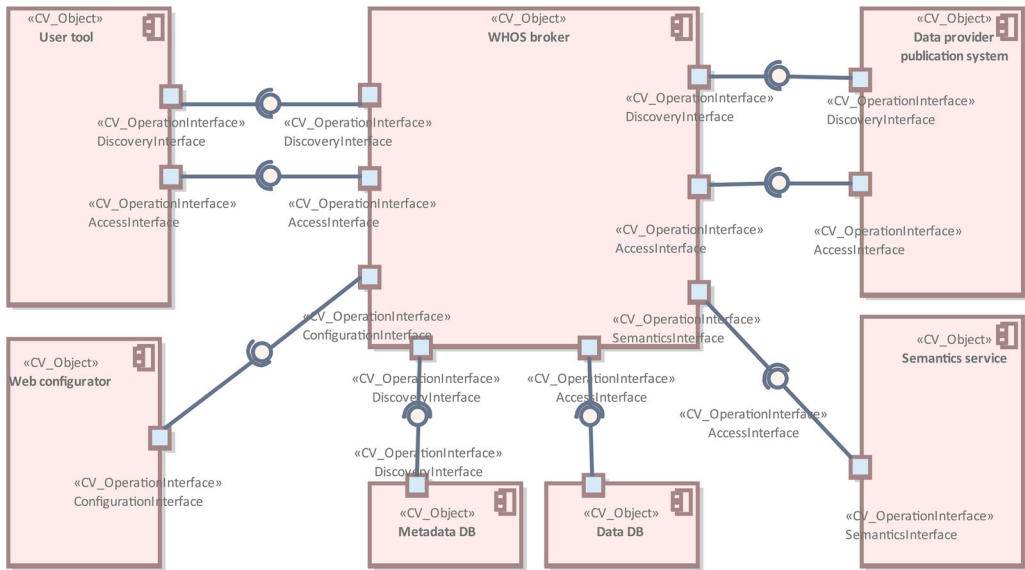The implemented discovery interfaces include:

Figure 14. Component diagram highlighting macro components of the WHOS System of Systems (SoS) and the main interfaces.

- **OAI-PMH** (The OAI Executive 2002): The Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) interface provides metadata harvesting functionalities. WHOS OAI-PMH support metadata encoding in Dublin Core, ISO 19115, ISO 19115-2 and WIGOS metadata models.
- **OGC CSW** (OGC 2007): The Open Geospatial Consortium (OGC) Catalogue Service for the Web (CSW) interface provides metadata search functionalities. WHOS OGC CSW support metadata encoding in Dublin Core, ISO 19115, ISO 19115-2 and ISO 19115 community profiles.
- **OpenSearch** (DeWitt 2018): Amazon A9 OpenSearch interface provides custom metadata search functionality. The actual available interface is described within the search engine description document, including parameterized URL templates that indicate how the search clients should make search requests. WHOS OpenSearch supports the GeoRSS/Atom metadata model and JSON.
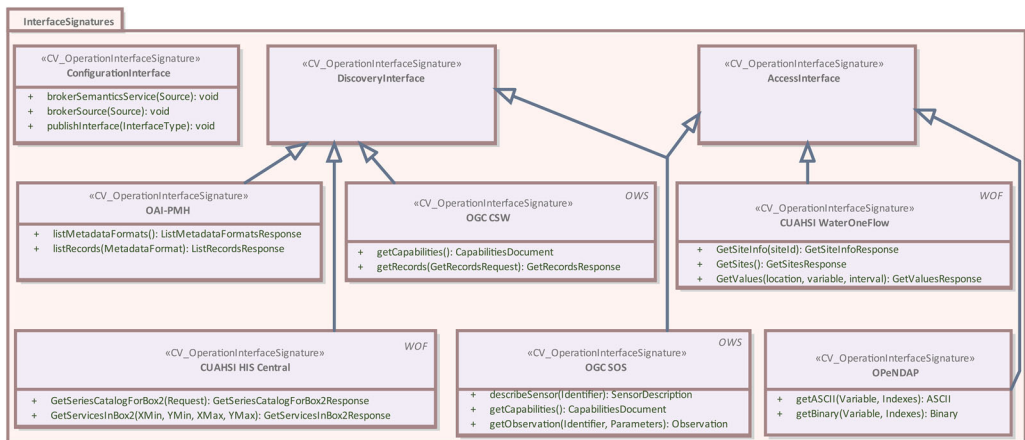

Figure 15. A partial set of interfaces implemented by WHOS broker.

- **CUAHSI HIS Central** (Whitenack, Zaslavsky, and Valentine 2008): The Consortium of Universities for the Advancement of Hydrologic Science, Inc. (CUAHSI) HIS Central interface provides service and time series discovery functionalities.

The implemented access interfaces include:

- **OPeNDAP** (OPeNDAP 2017): the Open-source Project for a Network Data Access Protocol (OPeNDAP) Data Access Protocol (DAP) interface provides scientific data access functionalities, including download and subset. The supported data encoding includes gridded and time series NetCDF as well as ASCII.
- **OGC WCS** (OGC 2018): the Web Coverage Service (WCS) interface provides multi-dimensional coverage data access (and processing) functionality, including download, subset, interpolation.
- **CUAHSI WOF** (OGC 2017): CUAHSI WaterOneFlow (WOF) provides a data access interface to download a specific subset of a time series, selected by its variable code and temporal period.
- **USGS RDB** (Engel 2003): an access interface enabling download and subset of time series and formatting the result according to the United States Geological Survey (USGS) RDB format.
- **ESRI Feature Service** (ESRI 2021): it provides a data access interface to discover and query for geographic features (for example, station locations).

Interfaces providing both discovery and access functionalities:

- **OGC SOS** (OGC 2012): The OGC Sensor Observing Service (SOS) interface provides sensor data discovery and access functionalities, including querying observations, sensor metadata, as well as representations of observed features. The data model supported by WHOS SOS are OGC Observations & Measurements (O&M) and OGC WaterML 2 (WML).
- **DAB API** (ESSI-Lab 2016): this JSON-based REST API interface provides discovery and access functionalities easing the development of web portals and custom user tools. It includes specific operations for timeseries, based on a JSON encoding of O&M.

Figure 16 shows the main subcomponents of WHOS broker, along with their interactions. The functionalities of each of them are here described:

- **Dispatcher**: it dispatches incoming Internet requests to either the Administration component or to one of the available Profiler depending on their nature. The Dispatcher uses a path-based strategy to select the component to forward the request to.
- **Administration**: it executes authorized administration requests, by reading and modifying the WHOS broker system options, making use of the Configuration manager.
- **Configuration manager**: the configuration holds all the WHOS broker options, this component is in charge of reading/writing it to the database; it periodically synchronizes the local configuration with the remote one (that could be changed by other instances); an update event is fired to the local subscribed components when an updated configuration is found on the database.
- **Profiler**: each of these components publishes a specific Internet service interface (e.g. OGC CSW, OGC WCS, OPeNDAP, etc.). In order to execute the incoming Internet requests, each Profiler performs a set of actions in a given order, delegating to internal subcomponents (for example the Request Transformer and the Result Set Mapper).
- **Semantics engine**: in charge of communicating with remote semantics services, in order to execute semantics queries (for example, to retrieve related terms in an ontology).
- **Discovery executor**: it executes the discovery of the resources matching the user queries (both count and retrieval) from the configured sources (both distributed and harvested), depending on user authorizations.
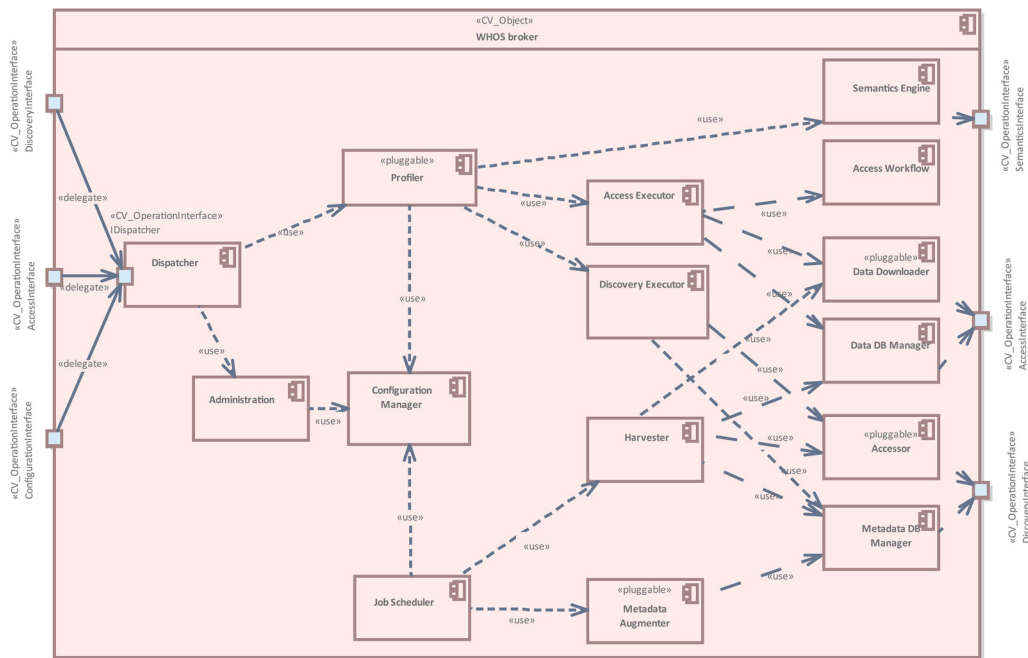
Figure 16. Main internal components of WHOS broker and their relations.

- **Accessor**: each of these components interacts with a publication system supporting a specific communication protocol and provides different functionalities: (1) discovery of resources from a distributed source; (2) download of original metadata records; (3) mapping the Original metadata to Harmonized metadata.
- **Metadata DB Manager**: it provides read, write, and especially discovery functionalities over the database holding the metadata records of all the harvestable configured sources, for caching purposes.
- **Access executor**: it executes access requests, by orchestration of the Data Downloader and the Access Workflow components responsible for data transformation. The Data DB Manager is used to optimize data requests that can be addressed using the cached data.
- **Data DB Manager**: it provides read/write functionalities over the database holding the harvested datasets, for caching purposes.
- **Data downloader**: each of these components is in charge of retrieving data from a specific data provider publication system.
- **Access workflow**: in charge of executing simple transformations (such as format conversion, CRS reprojection, subset, interpolation) to transform the downloaded data according to the user access request
- **Job Scheduler**: in charge of scheduling and launching WHOS broker recurrent jobs (for example, harvesting, metadata augmenters and access tests).
- **Harvester**: implements harvesting functionality which means collecting all the available metadata (and optionally data) from a Source and store them into the database for optimization of subsequent queries (and optionally downloads).
- **Metadata augmenter**: processes selected metadata records from the database with the aim of improving their content according to the specified algorithm.

Notably, some components (i.e. accessors, profilers and metadata augmenters) are tagged as , meaning that multiple instances are available and new instances of these components can be added to WHOS at operational stage to further enhance its capabilities (for example, to provide support for new publication systems or user tools).

Modularity and the mechanism are leveraged as well at lower levels to maximize reuse of code and extensibility. Each high-level component (for example, the **Profiler**) is indeed composed by different sub-components not detailed in Figure 16 and responsible for simpler computational tasks.

As an example, the inner pluggable components of the **Profiler** are:

- **Web Request Transformer**: in charge of validating and transforming an incoming web Request compliant with a specific request model to an internal harmonized Request;
- **Result Set Mapper**: in charge of mapping a Harmonized metadata instance from the Result Set to a specific metadata model;
- **Result Set Formatter**: in charge of joining and formatting the harmonized Result Set to be presented to the client.

Appendix provides an example of the implemented components to provide support for user tools compliant with the CUAHSI HIS-Central communication protocol.

Further examples of inner pluggable components of the **Accessor** are:

- **Connector**: in charge of iteratively retrieving metadata records from a remote service according to a specific communication protocol;
- **Metadata Mapper**: in charge of mapping from a metadata record from a specific remote model to a Harmonized metadata instance.

3.4. Engineering view

In the distributed geospatial data systems world, the interoperability issue can be seen as an MxN complexity problem, to connect M different user tools to N diverse data publication systems. From an architectural point-of-view, distributed and federated architectures can be implemented in a pure two-tier environment: the well-known Client-Server approach. In a traditional federation, the M clients interact with the N servers because only one type of interaction is allowed by defining the common federated protocol – i.e. services interface and content data model. In this way, the MxN complexity must be solved at the client and/or server level, making their protocols compliant with the common federation protocol. The brokered architectures introduced a middle-tier (i.e. the brokering tier) between the clients and servers, reducing the number of needed interactions to connect each client to every server from MxN to M + N, which are the necessary links to connect every client and server to the broker – see Section 1.2. WHOS is based on a three-tier architectural style (as shown in Figure 17) consisting of the following nodes:

- **ClientTier**: End user applications are executed on the end user machines (e.g. desktops, notebooks in case of desktop apps or web portals or servers in case of scientific models) representing the client tier. User interactions generate discovery and access requests sent to the broker tier.
- **BrokerTier**: WHOS broker middleware service executes on a cloud infrastructure, representing the broker tier. WHOS broker accepts incoming requests from the client tier and connects to the server tier to implement distributed discovery and access of resources.
- **ServerTier**: Multiple metadata and data publication systems execute on the data centers of the data providers, representing the server tier.

To realize a System-of-Systems, the three tiers brokering architecture has many advantages with respect to a traditional Client-Server architecture as discussed in Section 1.1. A possible configuration of WHOS is shown in the diagram of Figure 18, where two different clients are connected, through the WHOS broker services, to the resources offered by different data provider systems.

The deployment diagram, depicted in Figure 19, provides further details about the deployment of the WHOS broker in a cloud infrastructure environment. In the instance, represented by Figure 19, two Virtual Machines (VMs) are dedicated to the deployment with auto-scaling rules that allow to increase or decrease the number of dedicated VMs, according to resources needed. Each VM hosts a WHOS broker service, which is composed by an **auto-scaling** set of containers that provide duplicated brokering services (because instantiated by identical container images). An **Application Load Balancer** distributes incoming requests amongst the available **WHOS broker containers**. The **auto-scaling** feature is managed by a set of upscaling and downscaling rules, triggered by the request execution times. A **Healthcheck Monitor** constantly checks the health status of each container and remove containers that may start exhibiting a malfunctioning behavior. The container-based architecture addresses a set of important system requirements, including portability, reproducibility, and production level Quality of Service (QoS) – as to availability, reliability, and performance.

3.5. Technological view

Many different technologies contribute to the implementation of the WHOS ecosystem. The WHOS data providers can utilize different types of technologies for the implementation of their **data publication systems**, including:

- Technologies to create custom solutions – e.g. OpenAPI, Apache CXF.
- Installable solutions – e.g. CUAHSI HydroServer, Unidata TDS, OSGeo GeoNetwork, OSGeo GeoServer, 52North SWE.
- Cloud -based solutions – e.g. ESRI ArcGIS Online.

For the **end users tools and applications**, there are diverse technological solutions, including:

- Technologies to develop custom tools – e.g. BYU PyWaterML Library (Bustamante et al. 2021), R WaterML library, OpenAPI, Apache CXF, WCF Data Service Template plugin, Node.js WaterML client, USGS GWIS, DAB API.
- Installable tools – e.g. CUAHSI HydroDesktop, OSGeo GeoNetwork, OSGeo QGIS.
- Cloud -based tools – e.g. BYU Water Data Explorer Web Application (Bustamante et al. 2021), 52North Helgoland, Jupyter Notebook, ESRI ArcGIS Online.

The **WHOS broker** leverages and expands the DAB (Discovery and Access Broker) community edition technology, which is available as a source code project on its GitHub repository (ESSI-Lab
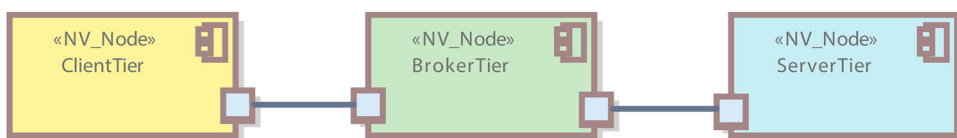

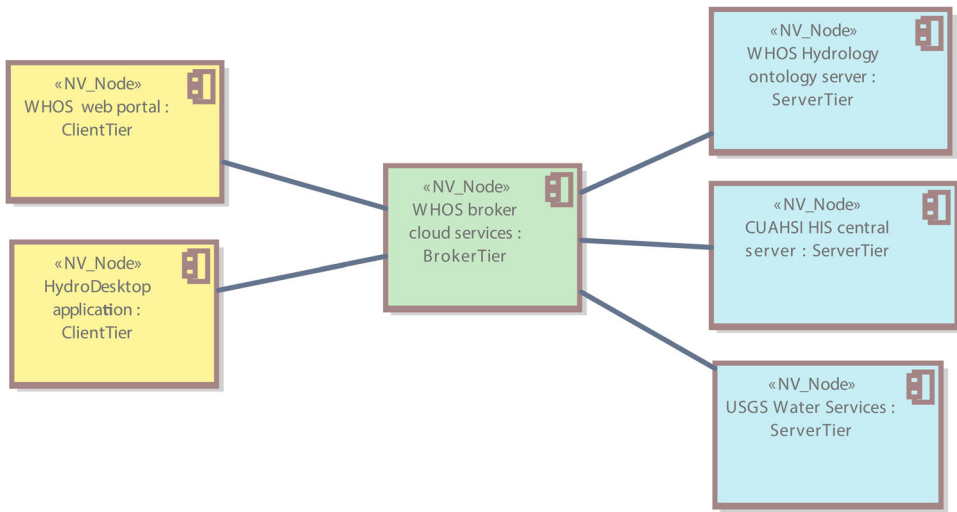**Figure 17.** Three-tier WHOS architectural style.

Figure 18. Example of WHOS configuration involving different nodes from the three-tier architecture.

2021). The DAB technology was first designed and is still maintained and advanced by the ESSI-Lab team of the Florence division of CNR-IIA.[9] Over the last ten years, the DAB software framework has been developed in the context of several National, European, and international projects funded and/or operated by different organizations including:
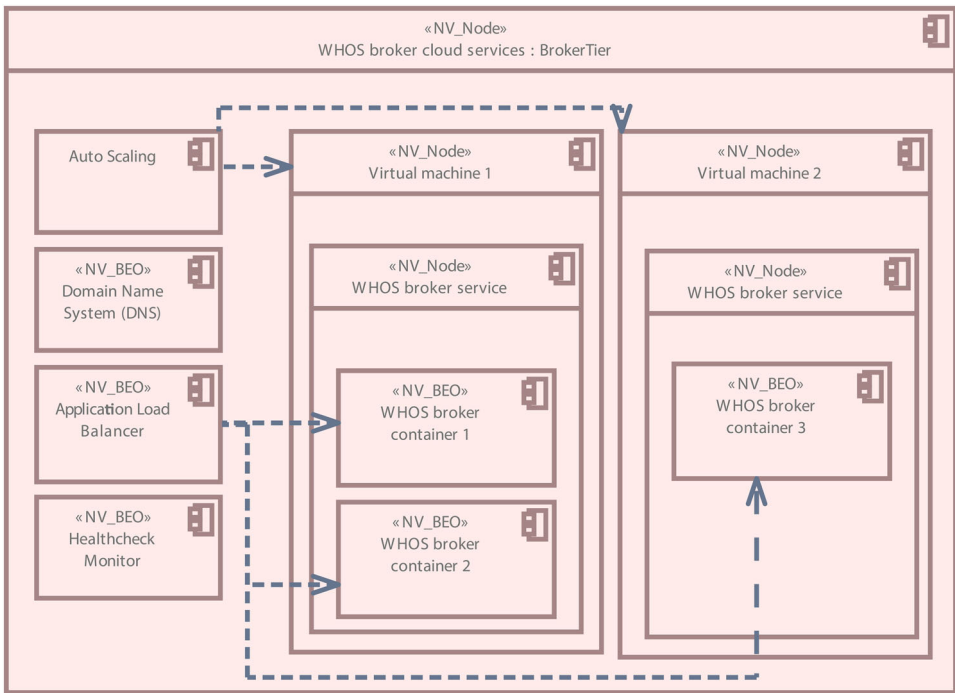

Figure 19. Details of WHOS broker deployment on a cloud infrastructure supporting virtualization, containerization and orchestration.

- Intergovernmental initiatives (GEO, WMO).
- European Union-funded projects (e.g. FP7, Horizon 2020, EASME).
- International Agencies (e.g. ESA).
- National research programmes (e.g. USA NSF, Italian Ministry of University and Research).

The WHOS broker builds on top of the DAB framework, by adding specific components to support those systems that are well-used in the hydrological and meteorological context. The WHOS broker realizes a Java-based software framework supporting a multiplatform deployment at its core. The support of the (Docker-based) containerization technology provides a normalized platform for both development, tests, and deployment. To enable optimized searches, the MarkLogic Server (MarkLogic 2022) technology was adopted as XML database for local cache of metadata content. To enable download optimization of gridded and time series data, the THREDDS Data Server[10] and ElasticSearch/OpenSearch[11] technologies were used, respectively. Presently, the WHOS broker services are deployed on AWS cloud infrastructure; nevertheless, there will soon be other deployments on different cloud infrastructures, by utilizing the Docker and Kubernetes technologies.

3.6. The implemented regional pilots

In the WHOS Phase II, three regional prototypes are currently being implemented:

1. the WHOS-Plata, in the La Plata Basin in South America;
2. the WHOS-Arctic, in the Arctic region;
3. the WHOS-DR, in the Dominican Republic.

The prototypes implementation is driven and carried out by the respective participating countries, in accordance with their local and regional requirements. The implementations of the three prototypes have reached their final stage and already engage 14 countries that freely exchange and reuse hydrological metadata and data, in an interoperable way across organizational and national boundaries.

3.6.1. La Plata River basin pilot

The WHOS-Plata implementation facilitates the exchange of meteorological and hydrological data collected in the La Plata River basin by Argentina, Bolivia, Brazil, Paraguay and Uruguay in order to strengthen national and basin capacities and develop more accurate hydrometeorological products and services. The WHOS-Plata data can be visualized, downloaded, analyzed, and modeled by means of various supported tools and applications. Based on the user's needs, any eventual new tools can be supported in the future. To easily leverage common WHOS functionalities such as data discovery and data access on the web by means of common web browsers, the WHOS-Plata web-portal is available online. Also, the Hydrometeorological Forecasting and Early Warning System called PROHMSAT-Plata, which is being developed for the La Plata River basin, ingests the data shared through WHOS. It is foreseen that the forecast results will also be shared through WHOS, allowing users to further use them in different tools and applications. The WHOS-Plata web portal is implemented using the Water Data Explorer application and is online available.[12]

3.6.2. The Arctic pilot

WHOS-Arctic implementation aims at freely exchanging hydrological data for the Arctic basin to enable better climate research and predictions in the Northern Hemisphere. In the frame of the Arctic-HYCOS project, the Basic Network of Hydrological Stations (BNHS) was created by selecting key existing observation stations of the national hydrological networks within the Arctic basin.

For the selected stations, historical and/or real-time data are collected and shared through WHOS by the participating countries. As with the WHOS-Plata implementation, the set of

supported tools are available to users allowing them to visualize, download, analyze and model the Arctic-HYCOS data. Based on user needs, any new tools can be supported in the future. Also, a WHOS-Arctic web-portal is available online.[13] It is foreseen that the Arctic-HYPE hydrological model, which is a Swedish contribution to the Arctic-HYCOS project, will ingest the data shared through WHOS. The Arctic-HYPE's goal is to increase the understanding of climate impact on fine-scale hydrology in the entire drainage basin of the Arctic Ocean, with the aim to improve predictions of river discharge into the ocean in the present and future climate. The Arctic-HYCOS model results will also be shared through WHOS allowing users to use them by means of various tools and applications.

WHOS-Arctic web portal provides hydrometeorological data shared by Canada, Finland, Denmark (for Greenland), Iceland, Norway, Russian Federation and United States of America for the Arctic basin. WHOS-Arctic web portal is implemented using ESRI ArcGIS Online[14] for the map interface and USGS GWIS[15] (Graphing Water Information System) for the time-series plots.

3.6.3. The Dominican Republic pilot

The main objective of WHOS-DR is to ease the sharing of hydrological and meteorological data acquired by different Dominican Republic organizations (i.e. ONAMET and INDRHI). One interesting feature of this pilot is the composition of the shared information, made both by point data (i.e. observations from monitoring points) and gridded data (i.e. results of modeling forecasts). Specific communication protocols have been implemented to assure the support of both data types (i.e. the THREDDDS data server protocol). As a result of the WHOS-DR implementation, all the tools already supported by WHOS can be used by the end users to discover and access the Dominican Republic data. Notably, the Water Data Explorer has been demonstrated to address point data and Met Data Explorer has been demonstrated to address gridded data. In the next future, additional tools will be readily available to the users, with no additional efforts required by the Dominican Republic data providers – as soon as they will be supported by the WHOS ecosystem. The WMO Meteorological, Climatological and Hydrological database management system (known as MCH[16]) was added as one of the data sources supported by the WHOS ecosystem (the connection was enabled through its web API), making it a convenient option for both hydrology data storing and sharing.

4. Conclusions and future work

The development process of WHOS and, more generally, of WIGOS-WIS has now been taking place for a few years. It must be viable with respect to the continuous advancement of IT instruments and the constant evolution of community needs. The adoption of a brokering architecture shifts this significant burden from the data systems and user tools (which contribute to the WHOS ecosystem) to the WHOS middleware solution, which is specifically designed and devoted to such a task.

In the regional pilots, the brokering approach has proved to be inclusive and sustainable. The data provider organizations were able to readily join the ecosystem with the System of Systems by using their already available data publication services. The providers were not required to undergo any further development, which would have been difficult to sustain, since this required, for example, the adoption of new technology and/or additional human resources. During the interoperability tests, the mutual feedbacks between the WHOS broker support center and the data providers, proved to be profitable on both sides. For instance, the WHOS broker technology improved its functionalities according to the received Community requirements. On the other hand, the data provider services were improved by enriching the quality and the content of their published metadata. Once trained, the regional Community offered to contribute to the advancement and evolution of the WHOS brokering framework. For this reason, a community edition of the WHOS brokering technology was created and published as a GitHub project.

The development of the WHOS digital ecosystem could not be possible without a clear, formal, scalable, and flexible administrative approach. As discussed, WHOS Community decided to implement a regional-based approach realizing a set of sub-ecosystems that contribute to the larger WHOS one. This approach ensures the necessary scalability, while the brokering architectural style guarantees the required political, technological, and administrative flexibility. WMO secretariat provides the necessary coordination for the WHOS implementation. This has allowed a homogeneous data sharing, across the data providers, and an overall development that is consistent with the objectives shared by the WHOS Community. The WMO secretariat has monitored and steered the developments to assure a high quality of the results.

Big Data and AI play an essential role in the present era of digital transformation of the entire society: in fact, they enable the development and use of innovative scientific and engineering models and instruments, identified as Digital Twins of the Earth (Nativi and Craglia 2021b). The WHOS digital ecosystem provides the required platform to share the key components of sound Digital Twins of the Earth: heterogenous and continuous observations (i.e. data), self-learning models (i.e. Machine Learning and Deep Learning models), and the interpretation instruments to generate actionable intelligence. These twins are used to monitor, simulate, and predict the behavior of Earth systems and phenomena – making our society smart and more sustainable.

In the next future, WHOS will continue the evolution of the existing regional ecosystems. In addition, it will pursue the development of new regional pilots. These tasks will be accompanied by a constant support to the Community, including their necessary related training. The WHOS operational plan will secure an allocation of infrastructural resources, which are needed for the WHOS functioning maintenance, to the technical support centers.

Notes

1. WMO (2021c).
2. WMO (2020).
3. WMO (2021b).
4. ISO (2009).
5. ESSI-Lab (2021).
6. WMO (2021a)
7. INMET (2020).
8. WMO (2021d).
9. CNR-IIA (2020).
10. Unidata (2019).
11. OpenSearch (2021).
12. BYU (2021).
13. WMO HydroHub (2020).
14. ESRI (2018).
15. Texas Water Science Center (2017).
16. WMO Secretariat (2022).
17. Freepik Company (2022).

Acknowledgments

Data availability statement

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

Disclosure statement

Funding

ORCID

Enrico Boldrini 🔟 http://orcid.org/0000-0003-2075-4742
Stefano Nativi 🔟 http://orcid.org/0000-0003-3185-8539
Paolo Mazzetti 🔟 http://orcid.org/0000-0002-8291-1128

References

Bustamante, Giovanni R., Everett J. Nelson, Daniel P. Ames, Gustavious P. Williams, Norman L. Jones, Enrico Boldrini, Igor Chernov, and Jorge L. Sanchez Lozano. 2021. "Water Data Explorer: An Open-Source Web Application and Python Library for Water Resources Data Discovery." *Water* 13: 1850.

BYU. 2021. "Water Data Explorer" INMET. https://tethys.inmet.gov.br/apps/water-data-explorer/.

CNR-IIA. 2020. "About IIA-Florence" CNR-IIA. https://www.uos-firenze.iia.cnr.it/about.

DeWitt, Clinton. 2018. "OpenSearch 1.1 Draft 6." https://github.com/dewitt/opensearch/blob/master/opensearch-1-1-draft-6.md.

Domenico, B., J. Caron, E. Davis, S. Nativi, and L. Bigagli. 2006. "GALEON: Standards-based Web Services for Interoperability among Earth Sciences Data Systems." In *IEEE International Symposium on Geoscience and Remote Sensing*. Denver (CO).

Engel, Glenn B. 2003. "RDB Format." In User's manual for the National Water Information System of the U.S. Geological Survey: Automated Data Processing System (ADAPS), by U.S. Geological Survey. https://pubs.usgs.gov/of/2003/ofr03123/6.4rdb_format.pdf.

ESRI. 2018. "ArcGIS Online" ESRI https://www.esri.com/en-us/arcgis/products/arcgis-online/overview.

ESRI. 2021. "Feature Service." https://developers.arcgis.com/rest/services-reference/enterprise/feature-service.htm.

ESSI-Lab. 2016. The GEOSS Discovery and Access Broker APIs. Accessed 2021. http://api.eurogeoss-broker.eu/docs/index.html.

ESSI-Lab. 2021. *DAB Community Edition GItHub repository*. https://github.com/ESSI-Lab/DAB.

Fielding, Roy Thomas. 2000. "Architectural Styles and the Design of Network-based Software Architectures.".

Freepik Company. 2022. "FlatIcon" Freepik Company. https://www.flaticon.com/.

GEO. 2016. *GEO Strategic Plan 2016-2025: Implementing GEOSS*. Geneva: GEO. http://www.earthobservations.org/documents/GEO_Strategic_Plan_2016_2025_Implementing_GEOSS.pdf.

Guo, H., S. Nativi, D. Liang, M. Craglia, L. Wang, S. Schade, C. Corban, et al. 2020. "Big Earth Data Science: An Information Framework for a Sustainable Planet." *International Journal of Digital Earth* 13 (7): 743–767.

INMET. 2020. "Instituto Nacional de Meteorologia" INMET. https://portal.inmet.gov.br/

ISO. 2009. "ISO/IEC 10746-3:2009" ISO. https://www.iso.org/standard/55724.html.

ISO/IEC. 1998. "Information technology — Open Distributed Processing — Reference model: Overview." 12 15. http://standards.iso.org/ittf/PubliclyAvailableStandards/c020696_ISO_IEC_10746-1_1998(E).zip.

ITU-T. 2014. "Information technology - Open distributed processing - Use of UML for ODP system specifications." 10. http://www.lcc.uma.es/%7Eav/download/UML4ODP_IS_V2.pdf.

MarkLogic. 2022. *MarkLogic Server homepage*. https://www.marklogic.com/.

Mazzetti, P., S. Nativi, M. Santoro, G. Giuliani, D. Rodila, A. Folino, S. Caruso, G. Aracri, and A. Lehmann. 2022. "Knowledge Formalization for Earth Science Informed Decision-Making: The GEOEssential Knowledge Base." *Environmental Science & Policy* 131: 93–104.

Nativi, Stefano, and Max Craglia. 2021a. "*Destination Earth: Ecosystem Architecture Description*. JRC, European Commission." *Luxembourg: European Union* EUR 30646 EN : 1–63 10.2760/08093.

Nativi, S., and M. Craglia. 2021b. "Digital Twins of the Earth." In *Encyclopedia of Mathematical Geosciences. Encyclopedia of Earth Sciences Series*, edited by B. Daya Sagar, Q. Cheng, J. McKinley, and F. Agterberg, 1–4. Cham: Springer.

Nativi, S., and P. Mazzetti. 2021. "Geosciences Digital Ecosystems." In *Encyclopedia of Mathematical Geosciences. Encyclopedia of Earth Sciences Series*, edited by B. Daya Sagar, Q. Cheng, J. McKinley, and F. Agterberg, 1–6. Cham: Springer.

Nativi, S., P. Mazzetti, and M. Craglia. 2021. "Digital Ecosystems for Developing Digital Twins of the Earth: The Destination Earth Case." *Remote Sensing* 13 (11): 2119–2144.

Nativi, S., M. Santoro, G. Giuliani, and P. Mazzetti. 2020. "Towards a Knowledge Base to Support Global Change Policy Goals." *International Journal of Digital Earth* 13 (2): 188–216.

The OAI Executive. 2002. *The Open Archives Initiative Protocol for Metadata Harvesting.* http://www.openarchives.org/OAI/openarchivesprotocol.html.

OGC. 2007. *OpenGIS Catalogue Services Specification.* https://portal.ogc.org/files/?artifact_id=20555.

OGC. 2012. "OGC Sensor Observation Service Interface Standard." https://portal.ogc.org/files/?artifact_id=47599.

OGC. 2017. "CUAHSI WaterML." https://portal.ogc.org/files/?artifact_id=21743.

OGC. 2018. OGC Web Coverage Service (WCS) 2.1 Interface Standard - Core. http://docs.opengeospatial.org/is/17-089r1/17-089r1.html.

OPeNDAP. 2017. "User Guide – OPeNDAP Documentation." https://opendap.github.io/documentation/UserGuideComprehensive.html.

OpenSearch. 2021. "About OpenSearch" OpenSearch. https://opensearch.org/.

Texas Water Science Center. 2017. "GWIS JavaScript API" USGS. https://txpub.usgs.gov/dss/gwis/0.0/doc/.

Unidata. 2019. "THREDDS Data Server" UCAR. https://www.unidata.ucar.edu/software/tds/current/.

United Nations. 2015. "Resolution adopted by the General Assembly on 25 September 2015." October 21. Accessed October 02, 2016. http://www.un.org/ga/search/view_doc.asp?symbol=A/RES/70/1&Lang=E.

United Nations. 2016a. *Sustainable Development Knowledge Platform.* Accessed 2016. https://sustainabledevelopment.un.org/index.html.

United Nations. 2016b. "Sustainable Development Goals." August 21. http://www.un.org/sustainabledevelopment/sustainable-development-goals/.

Whitenack, T, I Zaslavsky, and D W Valentine. 2008. "HIS Central and the hydrologic metadata catalog." *AGU Fall Meeting Abstracts* 1142. https://hiscentral.cuahsi.org/webservices/hiscentral.asmx.

WMO. 2020. "WIGOS" WMO. https://community.wmo.int/activity-areas/wigos.

WMO. 2021a. "Watch the recording of the WHOS webinar" WMO. https://community.wmo.int/news/watch-recording-whos-webinar.

WMO. 2021b. "WIS 2.0" WMO. https://community.wmo.int/activity-areas/wis/wis2-implementation.

WMO. 2021c. "WMO Hydrological Observing System (WHOS)" WMO. https://community.wmo.int/activity-areas/wmo-hydrological-observing-system-whos.

WMO. 2021d. "WHOS web services and supported tools" WMO. https://community.wmo.int/whos-web-services-and-supported-tools.

WMO. 2022. "WMO Hydrological Observing System (WHOS)." Accessed February 25, 2022. https://public.wmo.int/en/our-mandate/water/whos.

WMO HydroHub. 2020. "Arctic-HYCOS" WMO. https://hydrohub.wmo.int/en/projects/Arctic-HYCOS.

WMO Secretariat. 2022. 'MCH Database Management System' WMO. https://public.wmo.int/en/resources/bulletin/mch-database-management-system.

Appendix. Implemented components for CUAHSI HIS-Central support

Figure A1 shows an example of implemented components, concurring to 'construct' the **CUAHSI HIS-Central Profiler** component. This profiler is responsible for implementing the CUAHSI HIS-Central service interface, composed of different operations (Whitenack, Zaslavsky, and Valentine 2008) (e.g. GetSeriesCatalogForBox, GetSites, GetWaterOneFlowServiceInfo, etc.). Each operation is implemented by a different handler (four of them are shown in the diagram).

The **Handler Selector** chooses the correct handler to use, based on the user request. Two types of handlers are depicted in the figure:

- **WebRequest Handler:** this general handler type is used for maximum flexibility (e.g. when customized functionalities are requested to the handler)
- **Discovery Handler**: this handler is used when a typical discovery of dataset functionality is expected to be implemented, following predetermined steps: (1) transformation of the user request, (2) mapping of results, (3) formatting of results

In the CUAHSI HIS-Central case, the **WebRequet Handler** is used for three different operations: (1) to return the web service description (i.e. the WSDL document), (2) to retrieve the available sites and (3) to retrieve the statistics per data provider.

The **Discovery Handler** is instead used by the **GetSeriesCatalogForBox Handler** to implement the GetSeriesCatalogForBox operation. The intended aim of this operation is to discover the available time series matching a set of user constraints (i.e. a keyword, and spatial-temporal extent). The handler pluggable sub-components are:

- **GetSeriesCatalogForBox Transformer**: in charge of transforming web requests valid according to the HIS Central GetSeriesCatalogForBox operation (i.e. a HTTP GET request in the form: *hiscentral.asmx/GetSeriesCatalogForBox2?xmin = string&xmax = string&ymin = string&ymax = string&conceptKeyword = string&networkIDs = string&beginDate = string&endDate = string*) to the internal harmonized encoding of the query.
- **GetSeriesCatalogForBox ResultSet Mapper**: in charge of mapping the matching records of the **Resultset** from the **Harmonized metadata** model to the SeriesRecord object in the output data model of the GetSeriesCatalogForBox operation
- **GetSeriesCatalogForBox ResultSet Formatter**: in charge of creating a valid web response according to the GetSeriesCatalogForBox output data model (i.e. ArrayOfSeriesRecord object) to be filled with mapped records

The interaction diagram in Figure A2 shows the interactions between the described components during a CUAHSI HIS-Central GetSeriesCatalogForBox operation execution.
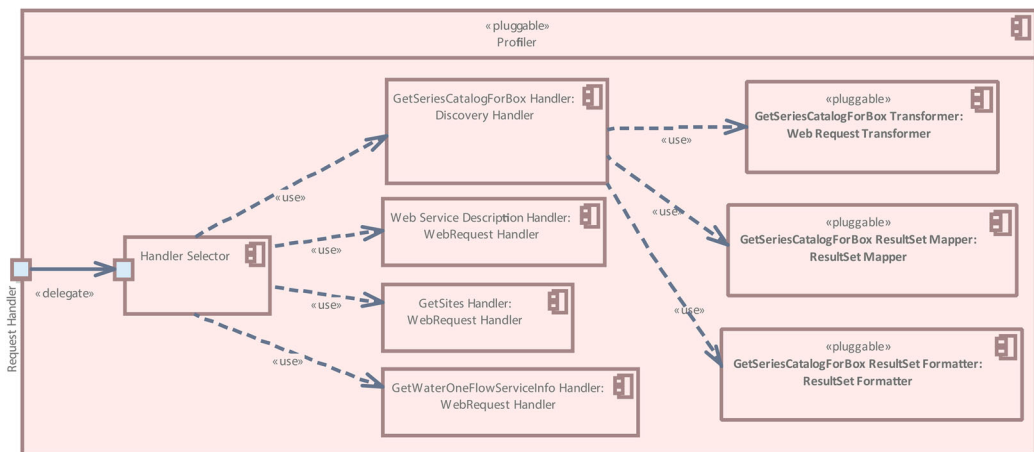

Figure A1. Internal structure of the CUAHSI HIS-Central Profiler component, showing the implemented subcomponents.
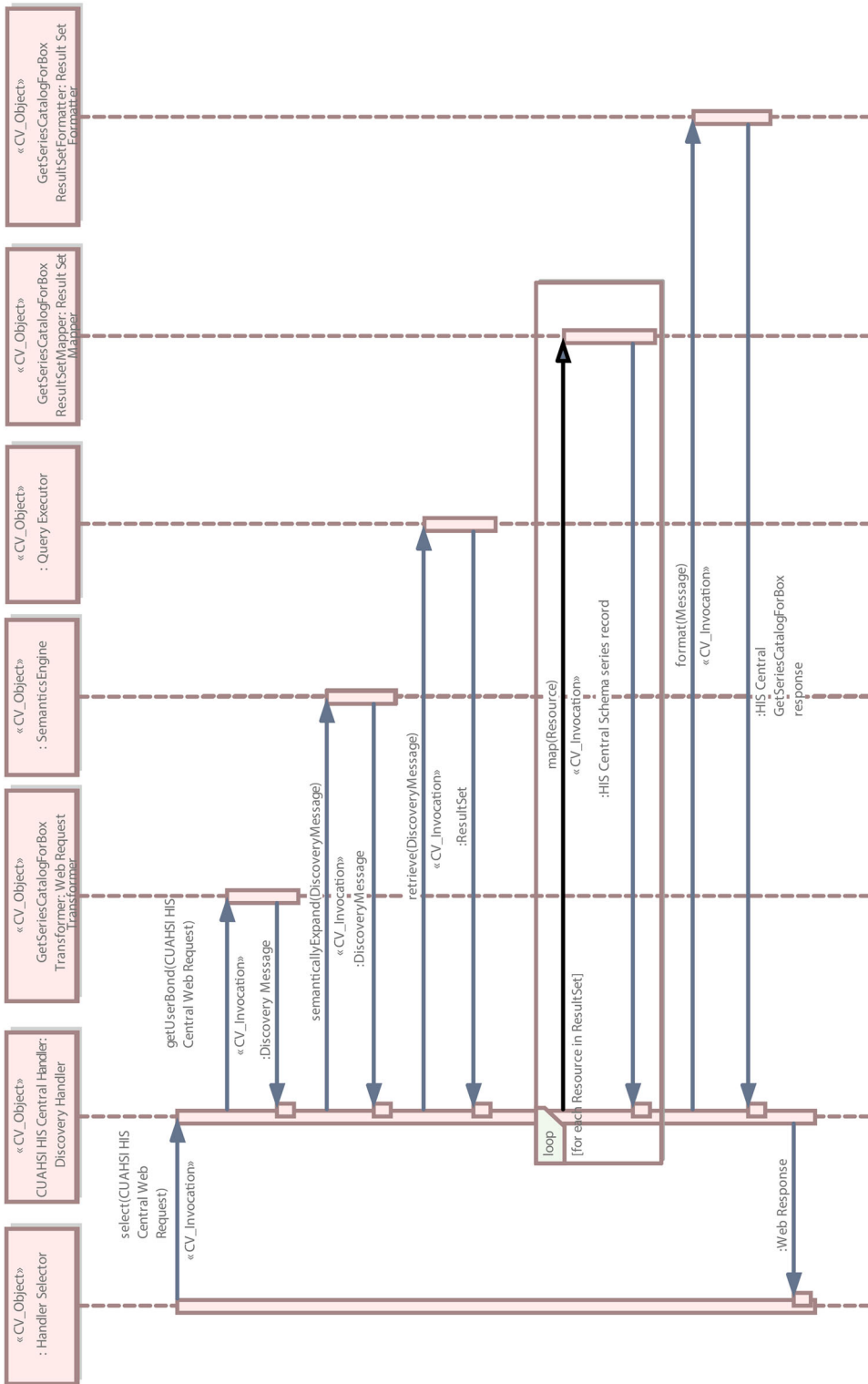
Figure A2. Interaction diagram for the GetSeriesCatalogForBox operation focusing on the CUAHSI HIS-Central Profiler subcomponents.