Open Geospatial Consortium

Date: 2011-07-12

Reference number of this document: OGC 11-013r6

Category: Engineering Report

Editors: Luis Bermudez, David Arctur

OGC® Engineering Report: Water Information Services Concept Development Study

Warning

This document is not an OGC Standard. This document is an OGC Public Engineering Report created as a deliverable in an OGC Interoperability Initiative and is not an official position of the OGC membership. It is distributed for review and comment. It is subject to change without notice and may not be referred to as an OGC Standard. Further, any OGC Engineering Report should not be referenced as required or mandatory technology in procurements.

Document type:	OGC Engineering Report
Document subtype:	NA
Document stage:	Approved for public release
Document language:	English

Preface

Water resources, weather, and natural disasters are not constrained by local, regional or national boundaries. Effective research, planning, and response to major events call for increasing coordination and data sharing among many organizations, with the potential for these important economic benefits:

- Improved coordination for disaster response;
- Reduced cost of software development, training, maintenance;
- Improved scalability, flexibility, and security from using modern architectures; and
- Access to extensive industry expertise, mentoring, tools & practices.

The necessary coordination would include agreement on conventions and policies for interoperable data exchange that work across jurisdictions, not just national-state-regional-local boundaries and agencies, but transcontinental. Trans-jurisdictional data sharing raises issues around handling multiple languages, as well as differing laws and customs for privacy, security, and rights management. The more technical challenges for achieving effective software and data interoperability include the design and maintenance of an adaptive, scalable, dependable architecture of distributed web services, information models and encodings, that support discovery and access for very large and very different types of data collections, even having different classification systems and meanings. All this is quite demanding, but increasingly urgent for managing water and other critical environmental resources.

The Consortium of Universities for the Advancement of Hydrologic Science, Inc. (CUAHSI) has conducted significant, extensive research and development addressing these challenges within the United States since 2004, with support from the National Science Foundation. The CUAHSI Hydrologic Information System (HIS) project has greatest focus on time-series observations of water resource variables at point locations throughout the U.S., as well as gridded precipitation and other related weather data to support hydrologic and other analytical modeling. HIS began with a focus on harmonizing data across U.S. agencies, such as the U.S. Geological Survey's (USGS) National Water Information System (NWIS), the Environmental Protection Agency's (EPA) Water Quality Exchange (WQX) and STOrage and RETrieval (STORET) data warehouse, and the National Oceanic and Atmospheric Administration's (NOAA) National Climate Data Center (NCDC). Even as the initial implementations of CUAHSI HIS were gaining acceptance and support from the U.S. agencies, it became evident that CUAHSI needed to align itself with the broader international geospatial data exchange standards community.

Based on sponsorship from CUAHSI, the OGC Interoperability Program conducted a Water Information Services Concept Development Study (Water IS CDS) from January through April 2011. The study was presented to the OGC HDWG. This report presents the findings of this study, and recommendations for further work. This document is anticipated to serve as the starting point to define future OGC Interoperability Program initiatives.

Forward

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. The Open Geospatial Consortium Inc. shall not be held responsible for identifying any or all such patent rights.

Recipients of this document are requested to submit, with their comments, notification of any relevant patent claims or other intellectual property rights of which they may be aware that might be infringed by any implementation of the standard set forth in this document, and to provide supporting documentation.

Document Contributor Contact Points

All questions regarding this document should be directed to the editor or the contributors:

Name	Email	Organization
Arctur, David (Editor)	Darctur<at>opengeospatial.org	Open Geospatial Consortium, Inc.
Bermudez, Luis (Editor)	Lbermudez<at>opengeospatial.org	Open Geospatial Consortium, Inc.
Ames, Dan	dan.ames<at>isu.edu	Idaho State University
Booth, Nathaniel	Nlbooth<at>usgs.gov	U.S. Geological Survey (USGS)
Choi, Yoori	Ychoi<at>cuahsi.org	CUAHSI
Goodall, Jon	Goodall<at>engr.sc.edu	University of S. Carolina - Columbia
Hooper, Richard	Rhooper<at>cuahsi.org	CUAHSI
Horsburgh, Jeffery	jeff.horsburgh<at>usu.edu	Utah State University
Jirka, Simon	Jirka<at>52north.org	52°North GmbH
Maidment, David	Maidment<at>mail.utexas.edu	The University of Texas at Austin
Percivall, George	Gpercivall<at>opengeospatial.org	Open Geospatial Consortium, Inc.
Salas, Fernando	fernando.r.salas<at>gmail.com	The University of Texas at Austin
Schreuders, Kim	kim.schreuders<at>usu.edu	Utah State University
Seppi, James	james.seppi<at>gmail.com	The University of Texas at Austin
Tarboton, David	Dtarb<at>usu.edu	Utah State University
Valentine, David	Valentin<at>sdsc.edu	University of California, San Diego Supercomputer Center
Veluppillai, Tevaganthan	Veluteva<at>isu.edu	Idaho State University
Vretanos, Panagiotis (Peter) A.	Pvretano<at>cubewerx.com	CubeWerx
Whiteaker, Tim	Twhit<at>mail.utexas.edu	The University of Texas at Austin
Zaslavsky, Ilya	Zaslavsk<at>sdsc.edu	University of California, San Diego Supercomputer Center

Contents

Page

Figures

Page

Tables

Page

OGC® Engineering Report: Water Information Services Concept Development Study

1 Introduction

1.1 Scope

The purpose of this report is to recommend appropriate architectures and procedures for migrating the CUAHSI HIS to the OGC-based WaterML 2.0 encoding (profile of OGC O&M standard) and OGC web services such as Sensor Observation Service (SOS), Web Feature Service (WFS), Web Mapping Service (WMS), Web Coverage Service (WCS), and Catalogue Service for the Web (CSW). This report may be used as the basis for future OGC Interoperability Program initiatives.

The report includes draft guidelines for five general use cases: publishing, cataloguing, discovering, accessing, and processing time series and point observations of hydrologic data using OGC encodings and web service standards. These guidelines reflect experience gained from operational prototypes handling large, federated catalogues and datasets of U.S. state and national agencies' current holdings, primarily for point observations (continuous time series) of surface water discharge.

This report is intended primarily for providers of water resources data in the United States, such as the Environmental Protection Agency (EPA), the U.S. Geological Survey (USGS) National Water Information System (NWIS), the National Oceanic and Atmospheric Administration (NOAA) National Weather Service (NWS), corresponding state level agencies and university research facilities (e.g. NSF supported observatories. Support for multiple languages and other international considerations are not discussed here, however these are being considered in other OGC HDWG activities, such as the Groundwater Interoperability Experiment (GWIE) and the Surface Water Interoperability Experiment (SWIE). It is hoped this Concept Development Study and report will be of interest outside the U.S. for international organizations, agencies, universities and research organizations who collect water data and who wish to make the data broadly available. The intended audience also includes data consumers who need to discover, access and integrate data from multiple sources in studies related to hydrologic science and water resources management, and developers building applications to support these functions.

1.2 Study Approach

Given this background of experience, limitations, and intended audience, the current study and report seek to address the following set of requirements, as expressed by the key stakeholders among government agencies and users.

1. Transition to OGC model for better interoperability, including international: what is the suggested path, what are new service interfaces, and what may be missing from this proposed reference model? The study provides a crosswalk between water data

encodings and services implemented in CUAHSI HIS (including WaterOneFlow services and HIS Central services) and OGC standard service interfaces.

2. Federation of catalogues for scalability, and since many data providers stand up catalogs: what is the suggested combination of catalogue technologies and interfaces? The study identifies certain interoperability issues between key commercial and open source catalogue services. We also suggest appropriate conditions for using WFS as a metadata server, and example portal user interfaces to support federated queries.

3. Efficient data discovery, recognizing that we don't need to search over all services: what are the most appropriate search patterns? This study identifies a 3-step approach: identify services, perhaps by thematic content or provider; then extract time series metadata; and then request data content for the time series.

4. Cultural and institutional integration and governance, recognizing that we can (and need to) rely on common implementations of mature, modular standard specifications: what is an appropriate operational governance model for distribution of roles and responsibilities within such a modular system? This includes supporting mandated data management roles and policies; maintaining data integrity, quality, and provenance throughout aggregation and semantic mapping processes; and enabling new data and service providers and consumers to join the system as simply as possible.

Two phases of the initiative are envisaged. Phase 1 is the Concept Development Study reported here. Phase 2 is anticipated to be an OGC Interoperability Program pilot initiative to implement, test and refine the architecture and design choices described here, possibly resulting in profiles and/or change requests to one or more OGC standards. The study and the pilot will reduce risks and help engage federal agencies and others towards the implementation of standard web services for hydrologic data.

2 Introducing CUAHSI HIS and RM-ODP

This report primarily relies on the experience developing the CUAHSI Hydrologic Information System (HIS). At the same time, it takes into account other related developments of service architecture for water data, in particular those undertaken in the course of OGC Interoperability initiatives, to present alternative solutions in information models, service interfaces, and implementation choices.

The CUAHSI HIS is an Internet-based system for sharing hydrologic data. It is comprised of databases and servers connected through web services to client applications, allowing for the publication, discovery, and access of data. CUAHSI HIS is based on a Service Oriented Architecture (SOA) where catalogues help mediate across data providers and data consumers (Figure 1).

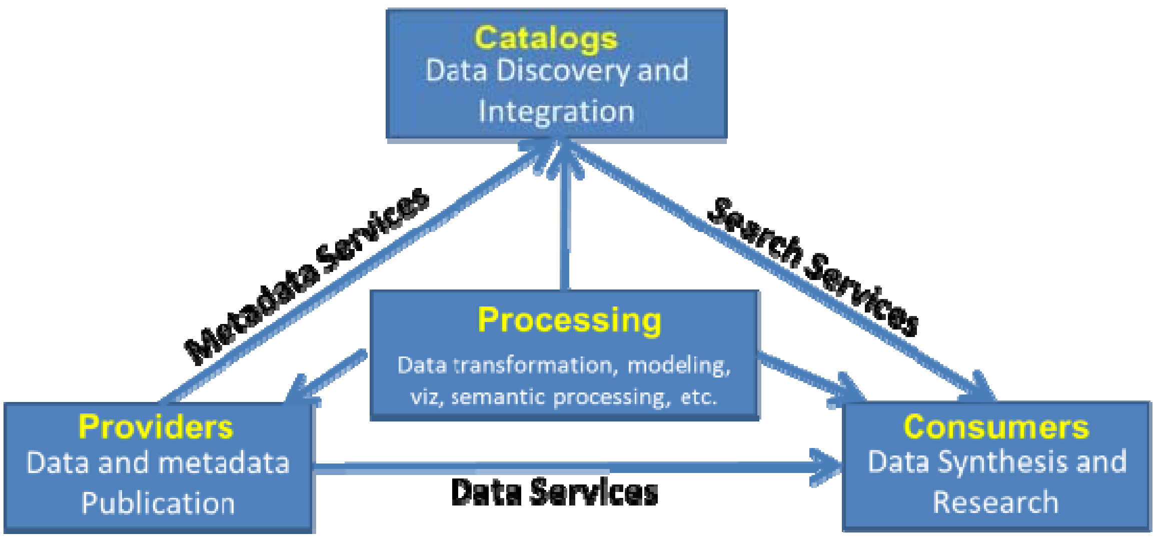


Figure 1 – Water Data Services Architecture

The functionality illustrated in Figure 1 follows the general publish-find-bind paradigm of the Internet: data providers publish metadata about data content and services to web-based catalogs; data consumers can then discover these datasets and services by querying the catalogs, and receive enough information to access (bind) the desired data content. Additional processing support may be provided to providers, catalogs and consumers as needed.

Data providers include federal, regional, state, local, international, and non-governmental organizations; universities; research teams; and volunteers. Consumers of hydrologic data include hydrologists, scientists, engineers, researchers, planners, decision-makers, students, and any other parties interested in retrieving data about the water environment. Consumers rely on desktop or web-based client applications to facilitate searching in catalogues accessing data services, analyzing and using the data.

Catalogues support data discovery based on indexed metadata, similar to the way search engines support the discovery of Internet content. A catalogue provides a centralized registry of services and descriptive metadata about the data that is published through those services. Catalogues also provide a searchable interface so that data consumers can discover the data in which they are interested.

Any one of the three corner nodes in Figure 1 is a simplification, and could involve multiple components & processes for transformations, mediation, brokering, and other functions, as indicated by the processing node shown in the model.

To support this architecture there are four main categories of web services in this services-oriented architecture (SOA):

- Data Services – provide the interface to access the data from the data provider
- Metadata Services – provide the interface to convey the metadata about a web service to a catalogue.
- Search Services – provide the interface to allow consumers to search, discover and assess applicability of services of interest. The catalogue provides the metadata for accessing the data using data services.
- Processing Services – provide specific functionality for operations on the data, such as for semantic mediation, transformations, language support, etc., assuming that such operations and their execution can be reasonably factored and distributed over the Internet.

Note that there may be overlap between data services and metadata services, in that what could be considered metadata in one context could also be seen as data in another context. We will try to distinguish between these as needed for clarity.

Fundamental to the architecture in Figure 1 is the information model and community support infrastructure upon which the system is founded. The information model comprises the conceptual and logical models used to organize and define sufficient metadata about hydrologic observations for them to be unambiguously interpreted and used. The information model is based on the CUAHSI Observations Data Model (ODM) (Horsburgh et al., 2008; Tarboton et al., 2008) that was developed through a comprehensive survey and review process (Bandaragoda et al., 2005; Tarboton, 2005) to identify the data and metadata required to provide sufficient and unambiguous description of point water observations in the hydrology domain.

A widely used approach for building a SOA for geospatial information is based on the Reference Model for Open Distributed Processing (RM-ODP) [ISO/IEC 10746-1:1998], illustrated in Figure 2. We have already mentioned concepts such as information models, web services, and service architectures. The RM-ODP provides a well-structured approach for relating these concepts in a given system. The first step is to define the community objectives in an *Enterprise Viewpoint,* to describe in a broad way why the effort is being undertaken, and its scope and objectives. This is aided by defining a set of *use cases* that describe how producers and consumers of data services perform their required functions. The second step involves a pairing of an information viewpoint and a computational viewpoint, which are complimentary to one another in the sense that the *Information Viewpoint* presents a set of conceptual models that describe the information (metadata and data) elements and how they will be encoded, while the *Computational Viewpoint* describes the interfaces by which service consumers interact with service providers. These viewpoints provide an inventory of concepts to be used in the system architecture. The *Engineering Viewpoint* then describes the architecture as a whole and how the information models and services relate to each other. The *Technology Viewpoint* completes the description with the physical implementation of specific software, network and hardware components.

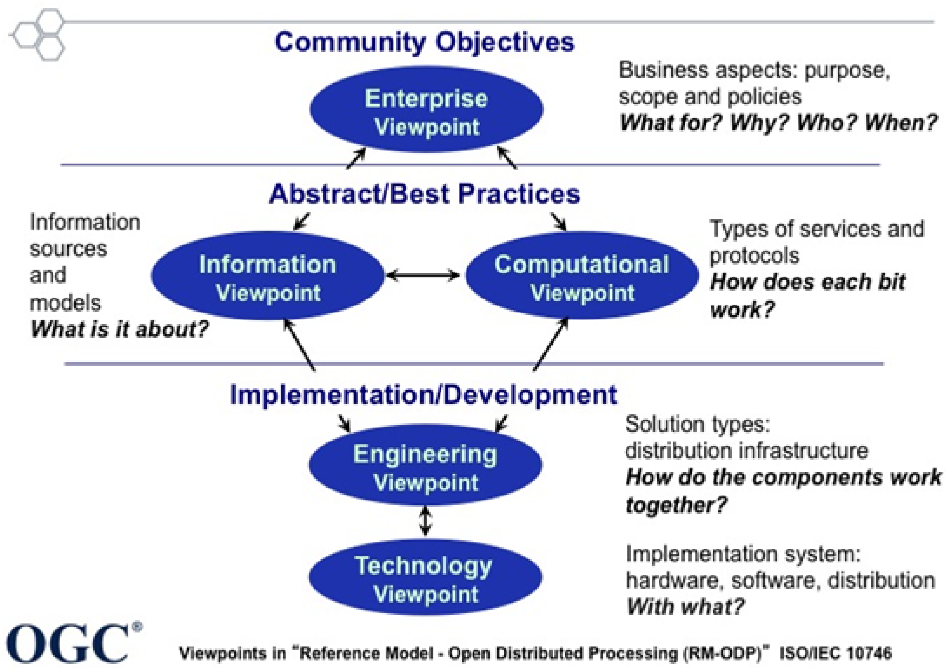


Figure 2 – Geospatial Service Architecture Viewpoints

This Concept Development Study focuses on the Enterprise, Information, Computational and Engineering Viewpoints. The study will not cover the Technology viewpoint, which would typically be based on final implementation of the complete architecture.

3 Enterprise Viewpoint: Water Information Community

3.1 Overview

As scientists and water resource managers begin to investigate complex hydrologic processes at expanding spatial and temporal scales, integration of data from multiple agencies, sources, projects, and research efforts is becoming more important. At the same time, data volumes are rapidly growing and changing in structure with advancements in sensor and other data collection and storage technologies. In a world of increasing availability and desire to use massive amounts of data, people need sophisticated data integration strategies to address the following key issues:

- Difficulty in discovering data: Most data published on the Internet are not inherently discoverable using traditional web search capabilities because they are generally encapsulated within files or databases, the contents of which cannot easily be discovered or catalogued by web crawler technologies employed by major web search engines.
- Data heterogeneity: Syntactic and semantic heterogeneity in data from different sources make data integration and synthesis difficult, and data are rarely annotated with sufficient attribute information, or metadata, to make their interpretation unambiguous by investigators other than those who collected the data.

The challenges of data organization and publication are different for government agencies and research organizations such as universities. In the U.S., major government agencies that hold hydrologic data are the U.S. Geological Survey (USGS), Environmental Protection Agency (EPA), National Climatic Data Center (NCDC), National Weather Service (NWS), Natural Resource Conservation Service (NRCS), U.S. Army Corps of Engineers (USACE), and U.S. Bureau of Reclamation. Each of these typically has one or more of its own (often-complex) internal data management systems. These agencies are mandated to make their data available to consumers, under specific scope (e.g., excluding sensitive areas), procedural (e.g., data collected according to approved set of protocols and meeting certain accuracy or detection standards) and/or currency/timeliness constraints.

The Open Government Initiative and the creation of the Federal Geographic Data Committee are good exemplars of mandates related to data publishing in the U.S. The Open Government Initiative (Orszag, 2009) mandates agencies in the U.S. to make available information online, improve the quality of information and foster the culture of open government. Previously an executive order (Clinton, 1994) mandated the use of standards and created the Federal Geographic Data Committee (FGDC). Agency requirements include:

- Making data available in open and standard formats;
- Making data available at useful and meaningful granularities or scales;
- Making data available with consideration for privacy, confidentiality and other restrictions.

Research organizations, on the other hand, are less likely to have established data management systems. University research often results in the accumulation of valuable data that may languish in the files of the investigators. Recognizing this problem, the U.S. National Science Foundation (NSF) now mandates that NSF-funded studies have a data management plan. This is critical for large interdisciplinary projects that are increasing in the hydrology domain. These include Critical Zone Observatories and Water Sustainability and Climate Observatories where NSF has recently (last four years) funded establishment of six and four of these integrated study observatories, respectively. With a mandate for data collected in these studies to be published soon (usually within at most two years) after collection, there is a need for effective, standards-based data publication capability that universities and other research organizations can use, including ability to:

- Make data available promptly;

- Make data available following standards;
- Make it easier for other scientists to discover and use data.

3.2 Use Cases

This Concept Development Study considers both government agency and research organization cases. It takes into account data consumer and data provider needs. An example of an application that provides integration of data server and data consumer needs is presented in Figure 3. This shows a user's client software application displaying possible data providers (Who), variables (What), and locations (Where) of hydrologic observations series in a map. In this instance, there are two federated catalogues of information that are being searched – one at the CUAHSI HIS Central located at the San Diego Supercomputer Center (SDSC) which identifies federal water data (e.g., USGS, USACE), and the other at The University of Texas at Austin, which identifies data published by Texas state water agencies. The application searches across both catalogues to identify dissolved oxygen data within an area of interest. This represents a simple case of federated catalogue search and access, in which the Texas state and U.S. Federal (HIS Central) catalogues are known to the user's desktop client application. In a more general application, the client software may need to discover new data providers and associated content.

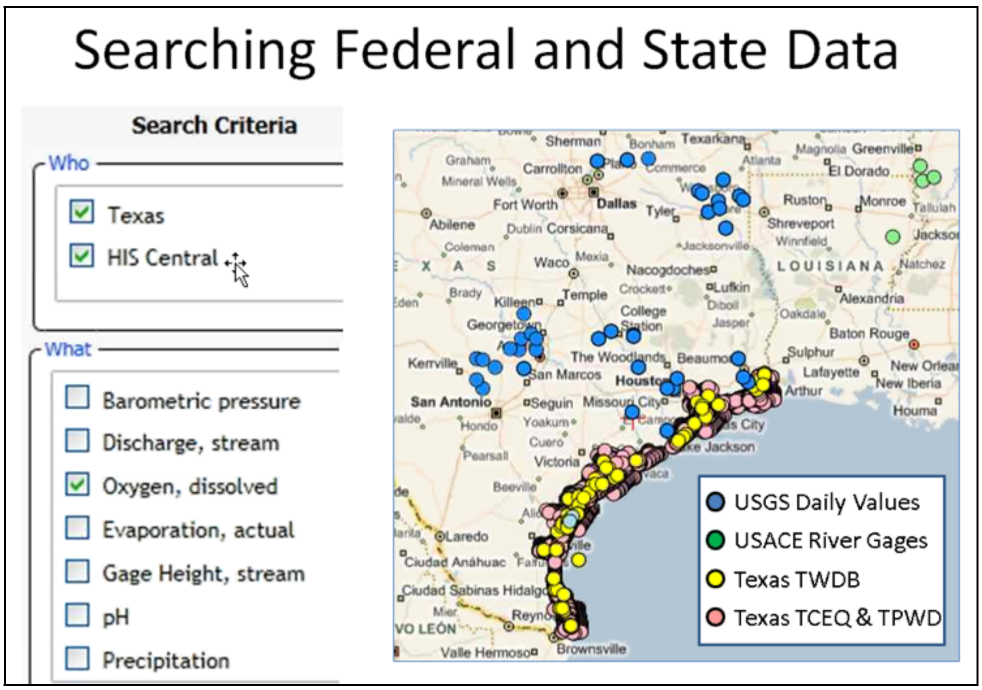


Figure 3 – Hydrologic Data Client Software Application

The following primary use cases have been chosen for discussion. It's worth considering that these all reflect a fundamental desire or need to "re-purpose data." This brings into focus the need to identify and describe original data well enough to support each and all of these operations without loss of information or quality.

1. **Publication** – make data and their associated metadata available using web services.
2. **Catalogue** – aggregate metadata from published web services.
3. **Discovery** – search a catalogue to identify data series or datasets of interest.
4. **Access** – acquire the selected data from the web service.
5. **Processing** – apply statistical or other processing routines to a dataset, such as aggregating time series data to a common time step.

The Publication use case describes the situation where a water agency or researcher wishes to make water and hydrologic data along with descriptive metadata available online through a service. The number of publishers is unlimited and ultimately in the United States may extend to thousands of state and local water agencies, utilities, research groups, and even public volunteers (consider the Open Street Map project).

The Cataloguing use case describes registration of published water data services (from Use Case 1) in a service registry, accompanied by harvesting of service or dataset metadata to enable subsequent data search and access. The catalogues may be organized by geographic and/or subject areas of interest, and reference other catalogues. Catalogue information shall be accessible via web services from different applications, which may include an online discovery portal. There will most likely be national aggregation catalogues, such as that provided by CUAHSI HIS Central, as well as regional and thematic catalogues hosted by individual state and local agencies and other organizations.

The Discovery use case refers to a consumer of the data who searches the metadata in a particular catalogue or across catalogues to identify data series or datasets of interest. The Access use case describes how the consumer downloads data for use on his/her local machine. An example of a specific discovery/access request might be to find and download spatially and temporally coincident observations of surface water nitrogen and stream flow with at least 100 days of values at a given location.

Finally, the Processing use case involves chaining of data publication, cataloguing, discovery or access services into workflows, to carry out analytical, statistical, and other kinds of operations on the data or metadata at varying granularities.

The scope of the use cases presented here are based primarily on common uses of the CUAHSI HIS, with published hydrologic data comprising time series of water measurements made at a network of observation sites. Other organizing principles or constraints on the use cases include the following:

- The published hydrologic data values or datasets may be grouped in regular or irregular time series, which may be further grouped into higher-level constructs such as hydrologic themes (time series collections).
- The data publication case covers publication of both primary (raw) measurements and derived or aggregated products (e.g., themes).
- Both real-time measurements and historical time series (organized and stored on the server) can be made accessible via a standard set of services.

- Besides the core metadata elements specifying where and when the measurements are taken, what is measured, by whom, and how; additional metadata elements may be included, depending on the publisher's metadata model and publication mandate (e.g., rating curves).

Caveats: These use cases are general enough to apply to other types of data actively being used in hydrologic research: (1) time series of gridded radar rainfall values; (2) multi-dimensional arrays in space and time, such as the output of climate models, or a set of products derived from remote sensing; (3) static geospatial datasets, such as LIDAR data for an experimental watershed. However, these types of datasets would involve additional infrastructure components (for example query tools for grids, or publication tools for multi-dimensional arrays) that are not considered in the proposed architecture.

This study and report do not address certain known issues that will need to be considered for any subsequent implementation project. One of these issues is to handle semantic differences among the data sources. The major data providers, such as the USGS, EPA, and NCDC, have many observational variables in common, but use different classifications and coding schemes for the definitions and data values. It is impractical to expect or require all agencies to agree on a single, universal data model. A solution to this was built into the initial design of CUAHSI HIS Central by establishing a common ontology for water resources, and mapping the various data providers' unique ontologies to this in advance. This enables users to search for data content based on the common ontology, and rely on HIS Central's mappings to find the correct records across all the data providers, without the user having to know the differences in ontologies and query syntax required among the various data sources. This greatly simplifies the user's experience. The addition of a new data source and/or provider would require advance registration of any new ontology, with guidelines to enable consistent query capability across all supported data sources. However, the best way to migrate this capability to the OGC standards framework will require further investigation that could not be accomplished during the current study. This kind of capability is a matter of current research and development in OWS-8, which should provide useful guidelines in time for a future Water Information System implementation project.

Other issues to be considered are to handle metadata and data content having multiple different languages, and with potentially different symbology styles, which would be required for international data sharing applications. These are being considered in other OGC interoperability projects as well, and should be feasible to factor into a subsequent Water Information System implementation project.

Another capability not considered in this study is to support forecasting applications. This would involve additional temporal fields in the information model, and patterns for usage. This is a subject being considered for a future interoperability experiment as a joint project between the OGC Hydrology Domain Working Group (HDWG) and Met-Ocean DWG.

The following tables organize and describe the five key use cases considered in this study.

Table 1 – Use Case 1: Publication

Overview	
Title	Hydrologic Data Service Publication
Description	A provider of hydrologic data publishes data and corresponding descriptive metadata using standards-based web services. Metadata should conform to a standard metadata specification. The data publisher may wish to control access to both data and metadata.
Actors	• Hydrologic data service provider
Initial Status and Preconditions	• Hydrologic data service provider stores hydrologic data and metadata, such as a database, structured data files, or a small memory device in a sensor. • Hydrologic data service provider has web services for data and metadata access, or can make a data resource available on the web. • A hydrology metadata model(s) comprising structural metadata elements, shared vocabulary, and discovery ontology concepts is in place. The metadata models include description for services, description for granule (time series, layers, etc.) and for variable types and value ranges. • A hydrology data model (schema) to convey the data is established. • A hydrologic data service provider has a policy and procedures for making data available, including appropriate workflows for quality assurance, access control rules, and guidelines for post-processing of data that may be required.
Basic Flow	
1. The data provider selects a standard web service interface, a standard data encoding, a standard metadata model, and relevant control vocabularies. 2. The data provider establishes a system to associate its data store into corresponding structural elements of the standard data and metadata models, including using appropriate shared vocabularies and discovery ontology concepts. 3. The data provider sets up standards-based web services on a host web server for accessing its hydrologic data and metadata which complies with its own policies and procedures. 4. Data access constraints are established and appropriate rules for controlling access are set up on the host web server.	
Post Condition	
A data provider's data and descriptive metadata are available on a web server. The data and metadata are accessible using standard web service interfaces. The metadata	

content includes shared vocabulary and discovery ontology concepts to ensure that the data can be catalogued and discovered. Appropriate access control rules have been established for the data and metadata and are enforced by the publication web services.

Table 2 – Use Case 2: Cataloguing

Overview	
Title	Hydrologic Data Service Cataloguing
Description	A data provider registers data web services to a service catalogue by providing the data service end point or link to the data. Descriptive metadata about the data location or data service and information about how to access it is made available to the catalogue. The catalogue has automated harvesting capability to periodically refresh and update the content to ensure that it maintains a current description of the registered services and the data that each service provides.
Actors	• Hydrologic data service provider • Catalogue service provider
Initial Status and Preconditions	• One or more web services that provide water related data and metadata have been published on a web server by a data provider (as in the Publication use case).
Basic Flow	
1. The data provider identifies a catalogue for registering the provider's data service. 2. The data provider registers its data and metadata service in the catalogue by providing a link to the web service method that describes the web service's capabilities. The metadata required to describe a web service are defined by an agreed-upon metadata standard for the hydrologic community and are retrieved automatically by the catalogue from the service's capabilities web service method. 3. The catalogue harvests metadata from the registered service to support subsequent queries. 4. The catalogue exposes all of the metadata that it has harvested and compiled about registered web services through a web service interface that can be searched by client software applications.	
Post Condition	
A data provider's services are registered in a catalogue. The catalogue has harvested the metadata for a data provider's service. The catalogue is queryable and accessible by data consumers over the Internet. The catalogue is searchable using several criteria, or facets, that enable data consumers to narrow search results. The catalogue can be	

indexed by other catalogues or meta-catalogues.

Table 3 – Use Case 3: Discovery

Overview	
Title	Hydrologic Data Discovery
Description	A data consumer wishes to find hydrologic data of any of the types listed in Section 4.2.1 “Hydrologic Data” that meet a number of search criteria. Search criteria will generally follow a who-what-when-where pattern, and the following are specific search criteria that will be supported: ***Who***: Find data from a particular data source (e.g., originating agency), project (e.g., scientific experiment in a watershed), or service (e.g., a particular data service instance). ***What***: Find data for one or more particular variables or concepts from the hydrologic ontology. This may include derived data, such as mean daily values, where data providers offer results of server-based operations. ***When***: Find data that fall within a specified temporal window with a specified frequency. ***Where***: Find data within a particular geographic bounding box, named region, or region type. The data consumer may specify one or more of these criteria together within a single search. The data consumer may also wish to filter and subset the results of a query to further refine search results.
Actors	• Hydrologic data consumer • Catalogue service provider
Initial Status and Preconditions	Water related data services from multiple data providers have been published online using standard service interfaces and standard data transmission languages. The data published by these services have been semantically annotated using the shared vocabularies and discovery ontology. The published services have been registered with a catalogue, and the catalogue has compiled metadata describing the services that is searchable by the data consumer.
Basic Flow	
1. A data consumer uses a client application to build a query that includes one or more	

of the search criteria described above. 2. The client application executes the query against the search web services provided by the catalogue. 3. The catalogue handles the query request and assembles a set of metadata records matching the query criteria that will be sent back to the data consumer. The published ontology is used to mediate the client application's vocabulary in the query to match that of the data service. 4. Metadata records for services that contain datasets that meet the criteria of the data consumer's query are returned and displayed in the client application. 5. The data consumer using the client application may then further filter and refine the query results. The client application will refine the results by further interacting with the catalogue or directly with the data provider web services.
Post Condition
Search results containing metadata about the hydrologic data that meet a data consumer's search criteria are returned. Search results returned to the data consumer contain references to the data access services that provide the actual data of interest and contain enough information that the data consumer can retrieve the data using their client application. Search results also contain enough information to enable data consumers to refine search results.

Note: In handling the discovery use case, it is necessary to consider the potential for "huge" responses in terms of the number of datasets in answer to a given query. With millions of water resource gauges and sampling sites in the U.S., each collecting numerous variables and covering numerous time series intervals, the number of total datasets is currently on the order of 23 million. Just keeping the metadata harvesting up to date presents a large task. Performance and scalability must be considered in addressing this use case. Approaches to this implementation are presented in detail later in this report.

Table 4 – Use Case 4: Access

Overview	
Title	Hydrologic Data Access
Description	A data consumer retrieves a hydrologic dataset from a specific data provider's web service, subject to filter criteria (what-when-where). This may include raw or derived data, with historical or real-time values.
Actors	• Hydrologic data consumer • Hydrologic data service provider
Initial Status and	Hydrologic data have been previously published online in a consistent format through a standard service protocol. A data

Preconditions	consumer has discovered a particular dataset that (s)he wants to download. The data consumer has obtained metadata describing how and where to get the dataset. If the data require authorization for access, the data consumer has obtained necessary permission to retrieve the data.
Basic Flow	
1. A data consumer uses a client application to connect to a data provider's web service. 2. Access and authorization to the data and data service is negotiated between data consumer client and data provider service. 3. The web service address and parameters used to retrieve the data have been previously obtained through the data discovery process (as in Use Case 3), or mediated through introspection of the data service, or the researcher may manually enter the required information. 4. The data consumer may specify filter criteria (what-when-where parameters) for which data from the service are desired. If processing steps are needed to generate the desired data, these are carried out. 5. Data that meet the requested parameters are used by the client application for use on the data consumer's local machine. The client can download the data in a standard format or use it in a workflow.	
Post Condition	
Hydrologic data are downloaded to a data consumer's local machine in standard format(s) for further study or analysis. The user receives the output of the workflow.	

Table 5 – Use Case 5: Processing

Overview	
Title	Hydrologic Data Processing
Description	Any actor in the system initiating processing of hydrologic data/metadata of any of the four types, or any related type of processing (e.g semantic mediation). The type of processing ranges from transformation operations on individual data sets to data fusion merging several data sources; from simple operations to complex workflows. The processing generates aggregate measures or derived products. Processing is typically characterized by formally defined input and output parameters, including their semantic types. Processing functions have metadata sufficient to ensure that operations are traceable and transparent so that users have

	confidence in the results.
Actors	• Hydrologic data consumer • Hydrologic catalogue provider • Hydrologic data provider • Hydrologic processing service provider (note: the Processing provider role may be associated with other roles in the system)
Initial Status and Preconditions	Processing services (workflows) have been published online and registered in a service registry, which may be integrated into a hydrologic data catalog. Processing services have sufficient standard metadata to be discoverable. The exposed processing capability may be applicable to specific datasets (e.g. when the processing service is published along with a data service by data provider), or may accept references to datasets available elsewhere via published data services. Invoking a processing service may require authentication/authorization; it is assumed that the processing service user has obtained the necessary authorization. It is also assumed that a provider of hydrologic processing service has a mechanism for service provisioning at certain level of availability (i.e. there are enough resources to support processing, and consumers are notified when it is not the case; the resource usage and service provisioning are monitored and recorded).
Basic Flow	
1. A processing service is discovered by a consumer application, and its input and output parameters are retrieved from the service description. As part of the same negotiation process, processing service provider determines if the consumer is authorized to use the service, and if resources are available to satisfy the processing request. 2. Alternately, a processing service may be coupled with other services as part of a workflow, such as data publication, cataloguing, discovery or consumption, and not invoked independently of the workflow. For example, data and metadata may require processing before they are delivered to data consumer or before they can be indexed by the catalog: in this case processing services may appear hidden behind data access or search services. 3. If multi-step processing is required, the client application may assemble a chain of services and execute them at once, either returning results to the caller after each step, or keeping them at the processing service provider (if the provider allows this and has sufficient resources). Such application chains may be preserved and appear as single step processing functions.	
Post Condition	
Processing results and status information are returned to the calling client. Information about the service use is recorded.	

We make a distinction here between provision of processing functionality via services, and provision of applications. Hydrologic models run from CUAHSI HydroDesktop plugins, or the HydroDesktop application itself, are examples of the latter, and are not included in this processing use case. A processing use case is only created when the model becomes wrapped as a standard processing service and registered, with all appropriate metadata, in a service registry. At the same time, we allow for more than one processing service API – as long as the functionality exposed in this way is reusable and can be provisioned on demand. Also, the functionality would be normally viewed as stateless from the perspective external to the processing service or service workflow.

A processing function is exposed by a processing service provider, and the function may be either coupled with other roles in the system, or be exposed independently. In the latter case, it would be registered in a service catalog, and semantically annotated to enable discovery. One of the key features of processing functionality compared with data access, discovery, and cataloguing functions is that processing requires compute resources which may be limited. Hence their use requires recording/monitoring/reporting. In many cases, however, this difference between processing and other services (characterized rather as query services) is blurred, since queries may require significant resources to execute.

4 Information Viewpoint: Water Information

4.1 Introduction

The Information Viewpoint describes the information models of the data and metadata that might used when exchanging information over the web. This is to organize and define the information needed to support all of the use cases above (e.g., publication, cataloguing, discovery, access, and processing). This section first provides a description of the high-level information model for the different types of hydrologic data that must be supported. Following that, we present requirements based on this information model and the use cases in Section 3. Finally, an overview of existing technologies to be considered for encoding hydrologic data and metadata for the purposes of this study is presented.

4.2 Information Model for Hydrologic Data

The information model defined by this study supports the use cases defined in Section 3 and provides the framework for describing hydrologic data and metadata (i.e., characteristics of data) required to represent each data value unambiguously. In what follows we distinguish between the following three groups:

- **Hydrologic Data –** The complete information needed for unambiguous interpretation of hydrologic observations.

- **Discovery metadata –** The information needed for data discovery that is the basis for the content of a Catalogue.

- **Services metadata –** The information needed to describe web services that transmit hydrologic data.

Subsections below address each of these. Core metadata can be grouped into the following categories of content:

1. Who observed it
2. What was observed
3. When was it observed
4. Where was it observed
5. How was it observed

All of these categories are needed when providing metadata for the transmission of data values, and as such need to be persisted in data storage systems. However not all this information is required for data discovery or as part of the description of data services used to access the data.

Annex A lists the specific attributes that comprise the content of this information model. This is grouped into Attribute groups that provide a hierarchical organization of this metadata. There are also columns in Annex A that indicate which attributes are important for discovery, and as such are part of discovery metadata, and which attributes are used with different classes of hydrologic data and/or are part of services metadata.

4.2.1 Hydrologic Data

Hydrologic data are those that describe the water environment or the characteristics of weather and climate that affect the water environment. Observational hydrologic data may include measurements of the variable physical, chemical, and biological characteristics of the water environment or atmosphere and are made to characterize not only natural processes, but also the interaction of the hydrologic cycle with the human-built environment and water management infrastructure. Observational data can be made *in situ* (e.g., in the case of a continuous stream gauge) or *ex situ* (e.g., in the case of a water sample that is collected and then analyzed in a laboratory) and may be made at a point location or over wide areas of the land surface in the case of remotely-sensed data. Additionally, data with characteristics similar to those of hydrologic observations are created as the result of models, simulations, and forecasts, and are, in some cases, estimates of variables that cannot be measured directly.

Based on these characteristics, we define several classes of hydrologic data, irrespective of whether the data are from actual observations or arise as a result of a model or simulation: (1) time series of water observations at a point; (2) time series of water observations on a geographic feature (not limited to point); (3) multidimensional arrays of water observations in space and time; and (4) static geospatial datasets that describe the water environment. The distinction between types (1) and (2) here is described in Section 4.2.1.2 below. While not all of these data classes have been developed within CUAHSI HIS, these would be natural extensions from the current capability.

4.2.1.1 Water Observations at a Point

Water observation time series are sets of time-indexed values collected at gauges and sampling sites. Such data are collected by many water agencies and also by water and atmospheric scientists. Three main types of observational time series are involved:

1. Data collected continuously through time using gauges, sensors, or automated samplers (e.g., physical hydrology variables, groundwater levels, weather and climate variables, or water quality variables);
2. Data collected intermittently at sampling sites and analyzed later in a laboratory (e.g., water quality data samples for surface or groundwater);
3. Data collected intermittently in the field (e.g., groundwater levels collected intermittently at wells, fauna/flora/river behavior description at a given point).

At a high level, the information model for water observation time series is as follows:

- A data “Source” operates a “Network” containing one or more monitoring “Sites” attached to feature(s) from referential datasets (e.g : Mississippi river from point A to B bearing the code ‘XYZ’; cf feature model).
- At each “Site,” a number of “Variables” are measured or estimated over time using one or more “Methods,” resulting in “Data Values.” Methods may have sensor instance level descriptive metadata, such as calibration, sensor ID, and others.
- The set of “Data Values” measured or estimated over time at a particular “Site” by a particular “Source” for a particular “Variable” using a particular “Method” make up an individual “Time Series.”
- “Time Series” can be copied, processed, and versioned, leading to multiple copies with differing “Quality Control Levels.”
- “Time Series” can be grouped thematically into “Themes,” which define sets of “Time Series” that are logically grouped.

4.2.1.2 Time Series of Water Observations on a Geographic Feature

These are sets of time-indexed values collected over a geographic feature. Such data are similar in many respects to time series at a point, except that the “sampling location(s)” are not limited to points, and may differ from the geographic extent of the “domain feature” to which they apply.

4.2.1.3 Multidimensional Arrays in Space and Time

Multidimensional spatiotemporal arrays are generally collected or generated for an unchanging spatial domain. The spatial domain is usually modeled as a regularly-spaced grid, with one data value per grid cell at each time. This may be based on a grid of latitude/longitude values (for which distance and area measurements between coordinates gets smaller moving north), or transformed (projected) to a grid having well defined distance and area measures. For example, simulated weather and climate products generally have output organized for presentation on a regularly spaced grid.

At a high level, the information model for multidimensional arrays in space and time is as follows:

- A data "Source" defines one or more "Spatial Domains" for which they have assembled data.
- For each "Spatial Domain," a number of "Variables" are measured or estimated over time using one or more "Methods," resulting in a regularly spaced grid of "Data Values."
- The grid of "Data Values" measured or estimated over time for a particular "Spatial Domain" by a particular "Source" for a particular "Variable" using a particular "Method" make up an individual "Space-time Array."
- "Space-time Arrays" can be copied, processed, and versioned, leading to multiple copies with differing "Quality Control Levels."
- "Space-time Arrays" can be grouped thematically into "Themes," which define sets of "Space-time Arrays" that are logically grouped.

4.2.1.4 Static Geospatial Datasets

Static geospatial datasets contain data that describe characteristics of or features within a particular spatial domain. The context of the particular spatial domain is usually an experimental watershed, data collection area, or region, but can be any spatial domain. Examples of static geospatial datasets include watershed boundaries, stream hydrography, land cover, soil types, and digital elevation models. These datasets are static in that they either represent features that do not change or change very little over time, or they represent snapshots in time (e.g., satellite or aircraft based imagery taken at a particular time). Static geospatial datasets can contain either vector or raster data.

At a high level, the information model for static geospatial datasets is a follows:

- A data "Source" defines one or more "Spatial Domains" for which they have assembled geospatial data.
- For each "Spatial Domain," the "Source" creates one or more "Geospatial Datasets" that define one or more "Spatial Fields," which cover the "Spatial Domain," or sets of "Spatial Features" within the "Spatial Domain."

- For each “Spatial Field,” the “Source” defines an attribute that describes the “Spatial Domain” (e.g., elevation, land cover), and zero or more “Attributes” for the “Spatial Field.”
- For each set of “Spatial Features,” the “Source” defines an attribute that describes the “Spatial Domain” (e.g., elevation, land cover), and one or more “Attributes” for each of the “Features.”
- “Geospatial Datasets” can be copied, processed, and versioned, leading to multiple copies with differing “Versions.”
- “Geospatial Datasets” can be grouped thematically into “Themes,” which define sets of “Geospatial Datasets” that are logically grouped.

4.2.2 Discovery Metadata

Discovery metadata comprises the attributes that should be used when describing data series of water observations data indexed in a catalogue that supports identification and preliminary screening of data, prior to it being downloaded and used. The focus is thus on metadata that is descriptive of the content of data series (e.g., variables being measured and period of record) but not content that while required for unambiguous interpretation does not describe the content (e.g., units).

4.2.3 Hydrologic Data Services

This Concept Development Study considers the service interfaces and data encodings that are appropriate for publishing the four hydrologic data types described above and detailed in Annex A. In addition to the metadata listed in Annex A that describes particular attributes of the data, there are also metadata elements that describe a web service that offers one or more datasets. Service-level metadata elements are important in cataloguing the existence and contents of data services.

4.2.4 Semantic Annotations of Hydrologic Metadata

A major goal of this Concept Development Study is to examine service and data standards that will resolve much of the syntactic heterogeneity in data from different sources (e.g., the file format, encoding structure, etc.) However, there is still a high probability of semantic differences among the metadata provided by different data publishers within a federated system. Mediating across the vocabularies used by different data publishers to enhance the discoverability and integration of data requires the use of semantic technologies, such as specialized registries based on RDF (Resource Description Framework).

Many of the metadata elements described in the sections above, and particularly those that are important for data discovery, can potentially be the subjects of one or more controlled vocabularies (CVs) that define the set of allowable values for particular metadata elements. Rather than using their own terms within their metadata, data publishers choose one of the existing terms from a CV to describe their data, thus reducing the number of terms used across all sources to those in the agreed upon CV. CVs can be defined by a community process and are an encoding of the vocabulary used within the domain.

There are, however, instances where the use of CVs is impractical. An example of this is a large federal data provider like the USGS that has an existing data management system and an existing vocabulary that differs from the domain CVs. In these cases, a mechanism is needed to map from the vocabulary used by the data publisher to one that is common with other data sources within the domain. An ontology and semantic mappings between the ontology and the domain CVs and any other vocabulary can be used for this purpose.

4.3 Information Viewpoint Requirements

The following Information Viewpoint requirements are derived from the high level information model described above and from the use cases described in Section 6:

Table 6 – Information Viewpoint Requirements

Requirement	Use case(s)	Discussion
Data models for the exchange of hydrologic data, metadata, and semantic annotations.	Publication, Access, Processing	In order to publish data and metadata on the Internet, they must first be organized into some open/accessible/queryable data store.
Encoding of hydrologic data for exchange across the Internet.	Publication, Access, Processing	Data must be published using standard encodings to support interoperable access to data on the Internet.
Encoding of metadata describing hydrologic data for exchange across the Internet.	Publication, Cataloguing, Processing	A data provider must publish metadata about their data, which is harvested by a catalogue. Standard, dataset-level metadata encodings support interoperable access to the metadata across all sources of data on the Internet.
Encoding of metadata describing web services that provide hydrologic data and metadata for exchange across the Internet.	Publication, Cataloguing, Discovery, Processing	A data provider must publish metadata describing their web services. Service-level metadata is harvested by a catalogue. Standard service-level metadata encodings support interoperable access to metadata across all services published on the Internet.
Semantic annotation of hydrologic metadata and data.	Publication, Cataloguing, Discovery, Processing	A data publisher should tag metadata with concepts from standard ontologies or controlled vocabularies in order to support cataloguing and semantic mediation in discovery of the data and services that offer the data.
Registry metadata model for time series observations that conforms to standard metadata as agreed by the hydrologic community.	Cataloguing, Discovery, Processing	A standard catalogue information model is required in order to support the cataloguing and discovery of data and the services that offer that data.

4.4 Technologies to be Considered for the Information Viewpoint

4.4.1 Overview

The following table presents an overview of the standards to be considered as data models and exchange encodings for the information viewpoint.

Table 7 – Overview of Standards and Encodings for Information Viewpoint

Type of Information	Storage Models / Information Models	Exchange Encodings
Hydrologic Data	• Observations Data Model • O&M • ISO 19123 (OGC Abstract Spec Topic 6) • WaterML 2.0 Profile of O&M • INSPIRE Data Specifications on Hydrography, and on Enviromental Monitoring Facilities • Global Runoff Data Center Hydrology Feature Model	• Water ML v2.0 • OMXML • netCDF • GeoSciML/GWML
Hydrologic Metadata	• ISO 19115	• about Data o ISO 19139 o Hydrologic Observations Information Model (as GMLSF) o SensorML including • SensorML Profile for Discovery (OGC 09-033) • SensorML Extension Package for ebRIM Application Profile (OGC 09-163r2) • about Services o ISO 19119 o Capabilities Document as defined in the SOS specification
Semantic Annotations	• ebRIM v3.0 (SQL) • RDF, RDFa	• OGC GetCapabilities • ebRIM v3.0 (XML)
Catalogue	• ebRIM v3.0 (SQL) • OGC Cataloguing of ISO Metadata (CIM) using the ebRIM profile of CS-W (0.1.7) (OGC 07-038)	• ebRIM v3.0 (XML) • ISO 19139

4.4.2 Details

4.4.2.1 Introduction

The following sections present detailed information about the storage models and exchange encodings presented in Table 7.

4.4.2.2 Hydrologic Data

4.4.2.2.1 Encoding(s) of Hydrologic Data for Exchange over the Internet

The following standard(s) should be considered for encoding hydrologic data for exchange over the Internet.

Title:	WaterML: An O&M profile for water observation data
Version:	2.0 (to be published)
Description:	WaterML 2.0, an O&M profile for water observations data is designed as an extensible schema to allow encoding of data to be used in a variety of exchange scenarios. Example areas of usage are: exchange of data for operational hydrological monitoring programs; supporting operation of infrastructure (e.g., dams, supply systems); cross-border exchange of observational data; release of data for public dissemination; enhancing disaster management through data exchange; and exchange in support of national reporting. The core aspect of the model is in the correct, precise description of time series. Interpretation of time series relies on understanding the nature of the process that generated them. This standard provides the framework with which time series can be exchanged with appropriate metadata to allow correct machine interpretation and thus correct use for further analysis. Existing systems should be able to use this model as a conceptual bridge between existing schemas or systems, allowing consistency of the data to be maintained.
Doc. Number:	OGC 10-126 (Draft version of WaterML)
Link:	http://portal.opengeospatial.org/files/?artifact_id=41546&version=1
OGC Dependencies:	Observations and Measurements - XML Implementation (OGC 10-025)

Title:	Observations and Measurements - XML Implementation
Version:	2
Description:	This standard defines an XML implementation of schemas for observations, and for features involved in sampling when making observations. This provides document models for the exchange of information describing observation acts and their results, both within and between different scientific and technical communities. The implementation is derived from a conceptual model defined in ISO/DIS 19156, and follows the rules for GML Application Schemas described in ISO 19136:2007.
Doc. Number:	OGC 10-025r1
Link:	http://portal.opengeospatial.org/files/?artifact_id=41510&version=2
OGC Dependencies:	GML v3.2.1 (OGC 07-036) O&M v2.0 (OGC 10-004r1)

Title:	OGC Network Common Data Form (netCDF) Core Encoding Standard
Version	1.0
Description:	NetCDF (network Common Data Form) is a data model for array-oriented scientific data. A freely distributed collection of access libraries implementing support for that data model, and a machine-independent format are available. Together, the interfaces, libraries, and format support the creation, access, and sharing of multi-dimensional scientific data.
Doc. Number:	OGC 10-090r3
Link:	http://portal.opengeospatial.org/files/?artifact_id=43732

4.4.2.3 Hydrologic Metadata

4.4.2.3.1 Encoding(s) of Hydrologic Metadata about data for exchange over the Internet

The following standard(s) should be considered for encoding metadata about hydrologic data for exchange across the Internet:

Title:	Hydrologic Observations Information Model
Version:	1
Description:	The Hydrologic Observations Information Model, defined in Annex A gives the fields that should be used when describing hydrologic observations data. The fields can be thought of as fields in a table, where each row defines a single data series. The core set of fields is presented in Annex A, derived from ODM (Horsburgh et al., 2008) and WaterML, but not following either exactly.

Title:	Geography Markup Language Simple Features profile
Version:	2
Description:	The GML standard declares a large number of XML elements and attributes meant to support a wide variety of capabilities. For example, the GML standard can encode dynamic features, spatial and temporal topology, complex geometric property types and coverages. With such a wide scope, interoperability can only be achieved by defining profiles of GML that deal with a restricted subset of GML capabilities. Such profiles limit the number of GML object types that can appear in compliant schemas and consequently are easier to process. The generation and parsing of Geographic Markup Language (GML) [OGC 07-036] and XML Schema [W3C XML-1, W3C XML-2] code are required in the implementation of many components that deal with GML encoded content. The Simple Features profile defines a restricted but useful subset of XML-Schema and GML to lower the "implementation bar" of time and resources required for an organization to commit for developing software that supports GML. It is hoped that by lowering the effort required to manipulate XML encoded feature data, organizations will be encouraged to invest more time and effort to take greater advantage of GML's rich functionality. Development of this profile does not reduce the need for distinct communities of users to develop application schemas (data models) for information sharing. However, to the extent that users' application schemas fit within the scope of GML-SF capabilities, this profile facilitates the ability to use WFS for interoperable feature data exchange with much less software development investment.
Doc. Number:	OGC 10-100r2
Link:	http://www.opengeospatial.org/standards/gml
OGC Dependencies:	GML 3.2.1 (OGC 07-036 / ISO 19136:2007)

Title:	Geographic Information – Metadata – XML schema implementation
Version:	ISO 19139:2007
Description:	ISO/TS 19139:2007 defines Geographic MetaData XML (gmd) encoding, an XML Schema implementation derived from ISO 19115.

Title:	Sensor Model Language SensorML
Version:	1.00
Description:	The Sensor Model Language Encoding Standard (SensorML) specifies models and an XML encoding that provide a framework within which the geometric, dynamic, and observational characteristics of sensors and sensor systems can be defined. There are many different sensor types, from simple visual thermometers to complex electron microscopes and earth observing satellites. These can all be supported through the definition of atomic process models and process chains. Within SensorML, all processes and components are encoded as application schema of the Feature model in the Geographic Markup Language (GML) Version 3.1.1. SensorML is one of the elements of the OGC Sensor Web Enablement (SWE) suite of standards.
Doc. Number:	OGC 07-000

Link:	http://www.opengeospatial.org/standards/sensorML

Title:	OWS-6 SensorML Profile for Discovery Engineering Report
Version:	1.00
Description:	This document defines a basic SensorML profile for discovery purposes. Besides a minimum set of metadata also the structure of according SensorML documents is defined in order to ensure a consistent metadata description. This goal is achieved by a set of Schematron rules that can be used to validate if a given SensorML document complies with the profile described in this engineering report.
Doc. Number:	OGC 09-033
Link:	http://portal.opengeospatial.org/files/?artifact_id=33284

4.4.2.3.2 Encoding(s) of Hydrologic Metadata about services for exchange over the Internet

The following standard(s) should be considered for encoding metadata about services (that offer hydrologic data and metadata) for exchange across the Internet:

Title:	Geographic Information – Services
Version:	ISO 19119:2005
Description:	ISO 19119:2005 identifies and defines the architecture patterns for service interfaces used for geographic information, defines its relationship to the Open Systems Environment model, and presents a geographic services taxonomy and a list of example geographic services placed in the services taxonomy. It also prescribes how to create a platform-neutral service specification, how to derive conformant platform-specific service specifications, and provides guidelines for the selection and specification of geographic services from both platform-neutral and platform-specific perspectives.

4.4.2.4 Catalogue Information Models

The following standard(s) should be considered as storage and exchange models for catalogues that register hydrologic data and metadata for the purpose of discovery and access:

Title:	ebXML Registry Information Model (ebRIM)
Version:	Version 3.0
Description:	An ebXML Registry is an information system that securely manages any content type and the standardized metadata that describes it. The ebXML Registry provides a set of services that enable sharing of content and metadata between organizational entities in a federated environment. This document defines the types of metadata and content that can be stored in an ebXML Registry.

Link:	http://www.oasis-open.org/committees/download.php/13591/docs.oasis-open.orgregrepv3.0specsregrep-rim-3.0-os.pdf
Comments:	• The ebRIM standard defines both a storage model (as SQL) and an exchange mode (as XML) • ebRIM catalogues can be used to manage ISO19115 and ISO19119 metadata documents

The acronym ebRIM stands for electronic business Registry Information Model. It is a standard developed by OASIS (see http://www.oasis-open.org) and adopted as one of the application profiles of web protocol binding of the OGC Catalogue Service Specification (see 07-006r1).

Title:	SensorML Extension Package for ebRIM Application Profile
Version:	Discussion Paper
Description:	This document describes the mapping of description of sensors using SensorML specification 1.0 [OGC 07-000] to an ebRIM structure within an OGC Catalogue 2.0.2 (Corrigendum 2 Release) [OGC 07-006r1] implementing the CSW-ebRIM Registry Service – part 1: ebRIM profile of CSW [OGC 07-110r4]. In addition this document contains the definition of a SensorML profile for Discovery which defines a minimum set of metadata to be provided within SensorML documents as well as the structure this data shall possess. This profile is based on the OGC OWS- 6 SensorML Profile for Discovery Engineering Report [OGC 09-033]. It defines the way sensors metadata are organized and implemented in the Catalogue for discovery, retrieval and management.
Doc. Number:	OGC 09-163r2
Link:	http://portal.opengeospatial.org/files/?artifact_id=37944

5 Computational Viewpoint: Services

5.1 Introduction

The computational viewpoint is concerned with the functional decomposition of architecture into a set of services that interact at interfaces. It reflects the components, interfaces, interactions and constraints of the service architecture without regard to their distribution. This section provides an overview of the technologies to be considered in the computational viewpoint, the requirements extracted from the use cases in Section 6, and the proposed best practice.

5.2 Technologies to consider for the Computational Viewpoint

5.2.1 Requirements

The following computational viewpoint requirements are derived from the use cases described in Section 3:

Requirement	Use Case
Ontology service supports annotation of hydrologic data and metadata describing hydrologic data	Publication
Data web service supports user access control	Publication
Data web service is a standard web service accessible via a URI	Publication
Registry Service provides as part of the result the end point of the service	Cataloguing
Registry has a web service interface	Cataloguing
The catalogue can be indexed by other catalogues	Cataloguing
The catalogue can index other catalogues	Cataloguing
The catalogue provides automated harvesting capabilities for registered services	Cataloguing
Data web service provides information for automatic harvesting by catalogues	Cataloguing
Catalogue service supports querying by provider name	Discovery
Catalogue service supports querying by service provider name	Discovery
Catalogue service supports querying by named location	Discovery
Catalogue service supports querying by variable or hydrologic concepts	Discovery
Catalogue service supports querying by temporal window	Discovery
Catalogue service supports querying by geographic bounding box	Discovery
Catalogue service supports multiple facets in one query	Discovery
Catalogue service returns link to data or the data service end point	Discovery
Catalog service is able to query Data Web Service (data availability on specific what-when-where conditions)	Discovery
Data Web Service supports querying by provider name	Access
Data Web service supports querying by service provider name	Access
Data Web service supports querying by named location	Access
Data Web service supports querying by variable or hydrologic concepts	Access
Data Web service supports querying by temporal window	Access
Data Web service supports querying by geographic bounding box	Access
Data Web service supports multiple facets in one query	Access
Data Web service supports downloading data in standard formats in asynchronous mode	Access

5.2.2 Web Services to Consider for the Publication and Access Use Cases

The following standards(s) describe web service interfaces that support the publication of metadata about hydrologic data services:

Service:	WFS
Version:	2
Title:	OpenGIS Web Feature Service 2.0 Interface Standard
Description:	This International Standard specified the behaviour of a service that provides transactions on and access to geographic features in a manner independent of the underlying data store. It specifies discovery operations, query operations, locking operations, transaction operations and operations to manage stored parameterized query expressions.
Doc. Number:	OGC 09-025r1 / ISO 19142:2010
Link:	http://www.opengeospatial.org/standards/wfs

OGC Dependencies:	Definition identifier URNs in OGC namespace (OGC 07092r3) Filter Encoding 2.0 Encoding Standard (OGC 09-026r1) OpenGIS Geography Markup Language (GML) Encoding Standard v3.2 (OGC 07-036 / ISO19136:2007)

The following standards(s) describe web service interfaces that support the publication and access of hydrologic data on the web:

Service:	SOS
Version:	1 (with regard to subsequent implementation pilots, SOS 2.0 should be considered)
Title:	OpenGIS Sensor Observation Service
Description:	The OpenGIS® Sensor Observation Service Interface Standard (SOS) provides an API for managing deployed sensors and retrieving sensor data and specifically “observation” data. Whether from in-situ sensors (e.g., water monitoring) or dynamic sensors (e.g., satellite imaging), measurements made from sensor systems contribute most of the geospatial data by volume used in geospatial systems today. This is one standard within the OGC Sensor Web Enablement (SWE) [http://www.opengeospatial.org/ogc/markets-technologies/swe] suite of standards. Metadata about observation offerings are advertised in the GetCapabilities response and metadata about the procedure (e.g., network, site, platform and sensors) are provided via SensorML.
Doc. Number:	OGC 06-009r6
Link:	http://www.opengeospatial.org/standards/sos
OGC Dependencies:	OGC Web Service Common Implementation Specification v1.1 (OGC 06-121r3) Enhanced Filter Encoding (OGC 05-093) Geography Markup Language (GML) Implementation Specification, version 3.1.1 (OGC 03-105r1 / ISO 19136) Observations and Measurements v1.0 (OGC 07-023r3) Sensor Model Language for In-Situ Remote Sensors v1.0 (OGC 07-000) Sensor Planning Service (SPS) v1.0 (OGC 07-104r3) Transducer Markup Language (TML) v1.0 (OGC 06-010r6) NOTE: efforts should be made during implementation phase to use SOS 2.0 which is expected to be formally adopted by OGC during 2011, in order to advance the use of SOS 2.0 and improve coherence of the current Sensor Web Enablement (SWE) standards framework. This would change the dependencies for O&M to v2.0, Sensor Planning Service to v2.0.

Service:	WCS
Version:	2.0
Title:	Web Coverage Service (WCS) 2.0 Interface Standard

Description:	The OpenGIS® Web Coverage Service Interface Standard (WCS) defines a standard interface and operations that enables interoperable access to geospatial "coverages" [http://www.opengeospatial.org/ogc/glossary/c]. The term "grid coverages" typically refers to content such as satellite images, digital aerial photos, digital elevation data, and other phenomena represented by values at each measurement point.
Doc. Numbers:	WCS 2.0 Core: OGC 09-110r3 WCS 2.0 KVP Protocol Binding Extension (1.0): OGC 09-047r1 WCS 2.0 XML/SOAP Protocol Binding Extension (1.0): OGC 09-149r1 WCS 2.0 XML/POST Protocol Binding Extension (1.0): OGC 09-148r1.
Link:	http://www.opengeospatial.org/standards/wcs
OGC Dependencies:	OGC Web Services Common Implementation Specification 2.0, OGC 06-121r9.

Service:	WCS-netCDF
Version:	1.1
Title:	Web Coverage Service (WCS) 1.1 extension for CF-netCDF 3.0 encoding (0.2.2)
Description:	The OpenGIS® Web Coverage Service Interface Standard (WCS) defines a standard interface and operations that enable interoperable access to geospatial "coverages" [http://www.opengeospatial.org/ogc/glossary/c]. The term "grid coverages" typically refers to content such as satellite images, digital aerial photos, digital elevation data, and other phenomena represented by values at each measurement point. This extension of the WCS standard specifies an Information Community data model with the related encoding that may optionally be implemented by WCS servers. This extension specification allows clients to evaluate, request and use data encoded in CF-netCDF3 format from a WCS server.
Doc. Number:	OGC 09-018
Link:	http://www.opengeospatial.org/standards/wcs
OGC Dependencies:	Web Coverage Service (WCS) 1.1 Corrigendum 2 (version 1.1.2) Implementation Standard (OGC 07-067r5) NOTE: efforts should be made during implementation phase to use NetCDF with WCS 2.0, and submit change requests to the WCS-netCDF extension based on that experience, in order to advance the use of WCS 2.0 and improve coherence of the current standards framework.

The publication use case has a requirement for semantic annotations to be added to published metadata. The following services are presented here for consideration as ontology services to support this requirement:

Service:	CSW-ebRIM
Version:	1.0.1
Title:	CSW-ebRIM Registry Service - Part 1: ebRIM profile of CSW

Description:	The CSW-ebRIM Registry Service profile is based on the HTTP protocol binding (the CSW part) documented in Clause 10 of the OGC Catalogue Services Specification (version 2.0.2, OGC 07-006r1). The profile imposes some constraints on the use of the base specifications and introduces additional search, retrieval, and registry management capabilities. It provides facilities for advertising and discovering a wide variety of information resources. While such resources are often labeled as "metadata", it is rarely possible to maintain an absolute distinction since what is deemed data in one context may well be treated as metadata in another.
Doc. Number:	OGC 07-110r4
Link:	http://www.opengeospatial.org/standards/cat
OGC Dependencies:	OGC Catalogue Services Specification v2.0.2 (OGC 07-006r1) OGC 03-105r1, OpenGIS® Geography Markup Language (GML) Implementation Specification, Version 3.1.1 (ISO/CD 19136) NOTE: efforts should be made during implementation phase to investigate best ways to coordinate with Europe's INSPIRE requirements to use the ISO 19115/19119 Application Profile of CSW 2.0.2.

Service:	Ontology Services (CSW-OWL)
Version:	0.3.0
Title:	OGC Catalogue Services - OWL Application Profile of CSW
Description:	Catalogues based on the CSW standard [OGC 07-006r1] are used to store information to manage geospatial resources in an interoperable environment. This information may include metadata about resources, including data sets and web services, as well as other related information to manage the resources. The existing CSW application profiles based on ebRIM [OASIS ebRIM] and ISO 19115/19 [ISO 19115/19] do not address the systematic inclusion of semantic information about the resources in catalogues beyond the inclusion of basic metadata, so up until now, ad hoc approaches have been required in order to include such formalized semantics for geospatial resources. OWL is a language for expressing ontologies that store semantic information about concepts using a description logic language that provides opportunities for various kinds of reasoning. Furthermore, particular ontologies have been developed to allow web services to be semantically described in a way that can support automated orchestration in an interoperable environment (for example, OWL-S [W3C OWL-S]). The inclusion of OWL in a geospatial catalogue has a number of benefits, including the representation of richer semantic information to describe resources, to assist discovery and to provide opportunities for web services orchestration. The addition of the catalogue specification to the usual description logic of ontology standards provides for interoperability among semantic resource descriptions within a spatial data infrastructure.
Doc. Number:	OGC 09-010
Link:	http://www.opengeospatial.org/standards/catalog
OGC Dependencies:	OGC Web Service Common Implementation Specification v1.0 (OGC 05-008c1) OGC Catalogue Services Specification v2.0.2 (OGC 07-006r1)

5.2.3 Web Services to Consider for the Cataloguing and Discovery Use Cases

The following standard(s) should be considered to support the Cataloguing and Discovery use cases:

Service:	CSW-ebRIM
Version:	1.0.1
Title:	CSW-ebRIM Registry Service - Part 1: ebRIM profile of CSW
Description:	A service-oriented architecture must support some fundamental interactions: publishing resource descriptions so that they are accessible to prospective users (publish); discovering resources of interest according to some set of search criteria (discover); and then interacting with the resource provider to access the desired resources (bind). Within such an architecture a registry service plays the essential role of matchmaker by providing publication and search functionality, thereby enabling a requester to dynamically discover and communicate with a suitable resource provider without requiring the requester to have advanced knowledge about the provider.
	The CSW-ebRIM Registry Service profile is based on the HTTP protocol binding (the CSW part) documented in Clause 10 of the OGC Catalogue Services Specification (version 2.0.2, OGC 07-006r1). The profile imposes some constraints on the use of the base specifications and introduces additional search, retrieval, and registry management capabilities. It provides facilities for advertising and discovering a wide variety of information resources. While such resources are often labeled as “metadata”, it is rarely possible to maintain an absolute distinction since what is deemed data in one context may well be treated as metadata in another.
Doc. Number:	OGC 07-110r4
Link:	http://www.opengeospatial.org/standards/catalog
OGC Dependencies:	OGC Catalogue Services Specification v2.0.2 (OGC 07-006r1) OGC 03-105r1, OpenGIS® Geography Markup Language (GML) Implementation Specification, Version 3.1.1 (ISO/CD 19136)

Service:	Sensor Instance Registry (SIR)
Version:	
Title:	Sensor Instance Registry
	The Sensor Instance Registry (SIR), is a web service interface for managing the metadata and status information of sensors. Furthermore this service is capable of automatically harvesting sensor metadata, transforming the collected metadata sets into a data model compatible to OGC Catalogues and to push harvested metadata into OGC Catalogue instances.
Doc. Number:	OGC 10-171
Link:	http://portal.opengeospatial.org/files/?artifact_id=40609

The service below may help with semantic mediation (e.g., by handling the semantics of the observed properties of a sensor).

Service:	Sensor Observable Registry (SOR)
Version:	
Title:	Sensor Observable Registry (SOR) Discussion Paper
	The Sensor Observable Registry (SOR) is a web service interface for managing the definitions of phenomena measured by sensors as well as exploring semantic relationships between these phenomena.
Doc. Number:	OGC 09-112r1
Link:	http://portal.opengeospatial.org/files/?artifact_id=40571

5.2.4 Filtering

The following standard defines the predicate encoding used in WFS and CSW requests:

Service:	FES
Version:	2
Title:	OpenGIS Filter Encoding 2.0 Encoding Standard
Description:	This jointly developed OGC and ISO TC/211 International Standard describes an XML and Key-Value Pair encoding of a system neutral syntax for expressing projections, selection and sorting clauses collectively called a query expression. These components are modular and intended to be used together or individually by other standards which reference this International Standard.
Doc. Number:	OGC 09-026r1 / ISO 19143:2010
Link:	http://www.opengeospatial.org/standards/filter
OGC Dependencies:	OGC Web Service Common Implementation Specification v1.1 (OGC 06-121r3) OpenGIS Geography Markup Language (GML) Encoding Standard v3.2 (OGC 07-036 / ISO19136:2007)

5.2.5 Visualization

The WMS standard is presented here for consideration because it can work in concert with data services such as SOS, WFS, and WCS to provide efficient visualization capabilities:

Service:	WMS
Version:	1.3.0
Title:	OpenGIS Web Map Service (WMS) Implementation Specification
Description:	The OpenGIS® Web Map Service Interface Standard (WMS) provides a simple HTTP interface for requesting geo-registered map images from one or more distributed geospatial databases. A WMS request defines the geographic layer(s) and area of interest to be processed. The response to the request is one or more geo-registered map images (returned as JPEG, PNG, etc.) that can be displayed in a browser application. The interface also supports the ability to specify whether the returned images should be transparent so that layers from multiple servers can be combined or not.
Doc. Number:	OGC 06-042
Link:	http://www.opengeospatial.org/standards/wms

Service:	SLD
Version:	1.1.0
Title:	OpenGIS Styled Layer Descriptor Profile of the Web Map Service Implementation Specification

Description:	The OpenGIS® Styled Layer Descriptor (SLD) Profile of the OpenGIS® Web Map Service (WMS) Encoding Standard [http://www.opengeospatial.org/standards/wms] defines an encoding that extends the WMS standard to allow user-defined symbolization and coloring of geographic feature[http://www.opengeospatial.org/ogc/glossary/f] and coverage[http://www.opengeospatial.org/ogc/glossary/c] data. SLD addresses the need for users and software to be able to control the visual portrayal of the geospatial data. The ability to define styling rules requires a styling language that the client and server can both understand. The OpenGIS® Symbology Encoding Standard (SE) [http://www.opengeospatial.org/standards/symbol] provides this language, while the SLD profile of WMS enables application of SE to WMS layers using extensions of WMS operations. Additionally, SLD defines an operation for standardized access to legend symbols.
Doc. Number:	OGC 05-078r4
Link:	http://www.opengeospatial.org/standards/sld

5.3 Dependencies Analysis

A large number of OGC services have been presented in these sections that have potentially numerous dependencies upon each other and upon other OGC specifications. Figure 4 illustrates these dependencies. Understanding these dependencies is important when making decisions about the standards and versions to be chosen for implementation.

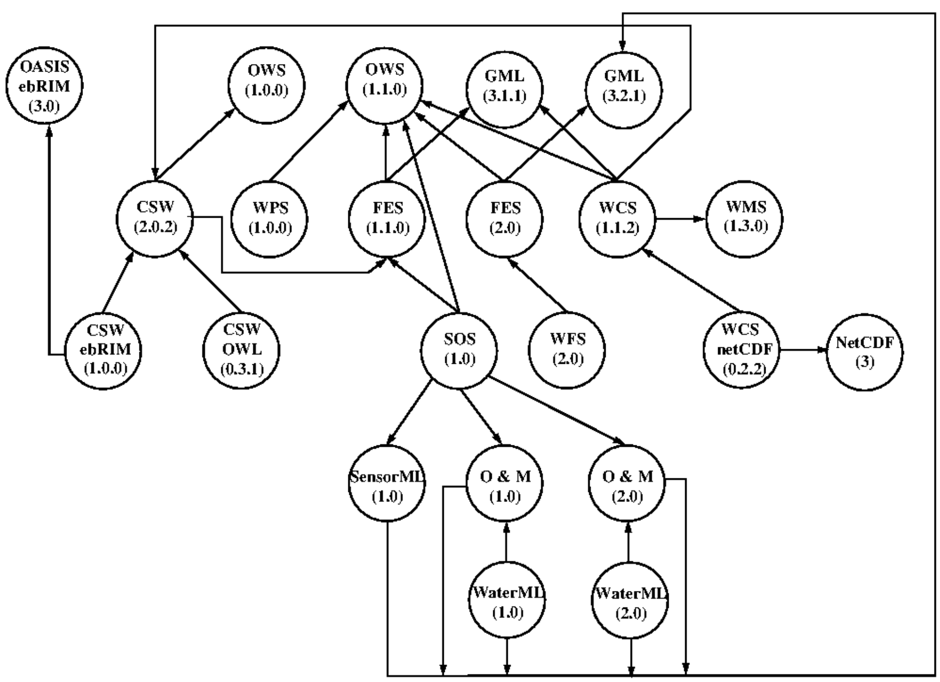


Figure 4 – Service Specification Dependencies

6 Engineering Viewpoint

6.1 Introduction

The engineering viewpoint defines a set of component types that provide the basis for deployment in a distributed environment. Initial consideration for identification of Engineering is to consider the components identified in the Enterprise viewpoint. The services and data that are used to define engineering components are defined in the informational (Section 4) and computational (Section 5) viewpoints.

In this clause we present a deployment architecture based on the CUAHSI prototype. The function of each component type is described and some specific topics, such as semantic annotation and query, are described in more detail. At the end of the clause we present a variation of the proposed architecture that reduces the number of services that a data publisher has to deploy but also provides more flexibility. It is anticipated that OGC prototype activities (pilot or testbed) will exercise both variations in order to compare and contrast them.

6.2 Deployment Architecture for WFS-Based Meta-information Servers

The UML component diagram (Figure 5) demonstrates the proposed deployment architecture for a water information system based on the components identified in the previous viewpoints. The components are laid out in a triangle to represent the familiar Publish-Find-Bind paradigm. Cataloguing components are found at the top of the diagram. Server components, provided by data publishers, are found on the left side of the diagram and client components are found on the right side of the diagram. In some cases a component (e.g. Data Server) may have one or more sub-components (e.g. Map Server, Coverage Server, etc.). Components are connected to each other by lines that represent interfaces that the component requires or provides. Lines with closed-circle terminators represent an interface that the connected component provides. Lines with open, half-circle terminators represent an interface that the connected component requires. At the point where a close-circle terminator meets an open half-circle terminator, a label is used to indicate the specific interface required. In some cases, the label includes a specific operation from the interface (e.g. GetCapabilities).

The red frame around certain components and interfaces in the lower-left corner represents one likely option for insertion of identity management (authentication) and access control (authorization) processes and data flows. However, security configurations are beyond the scope of the current study and will not be discussed.

The interactions among these components are detailed in the sections that follow.

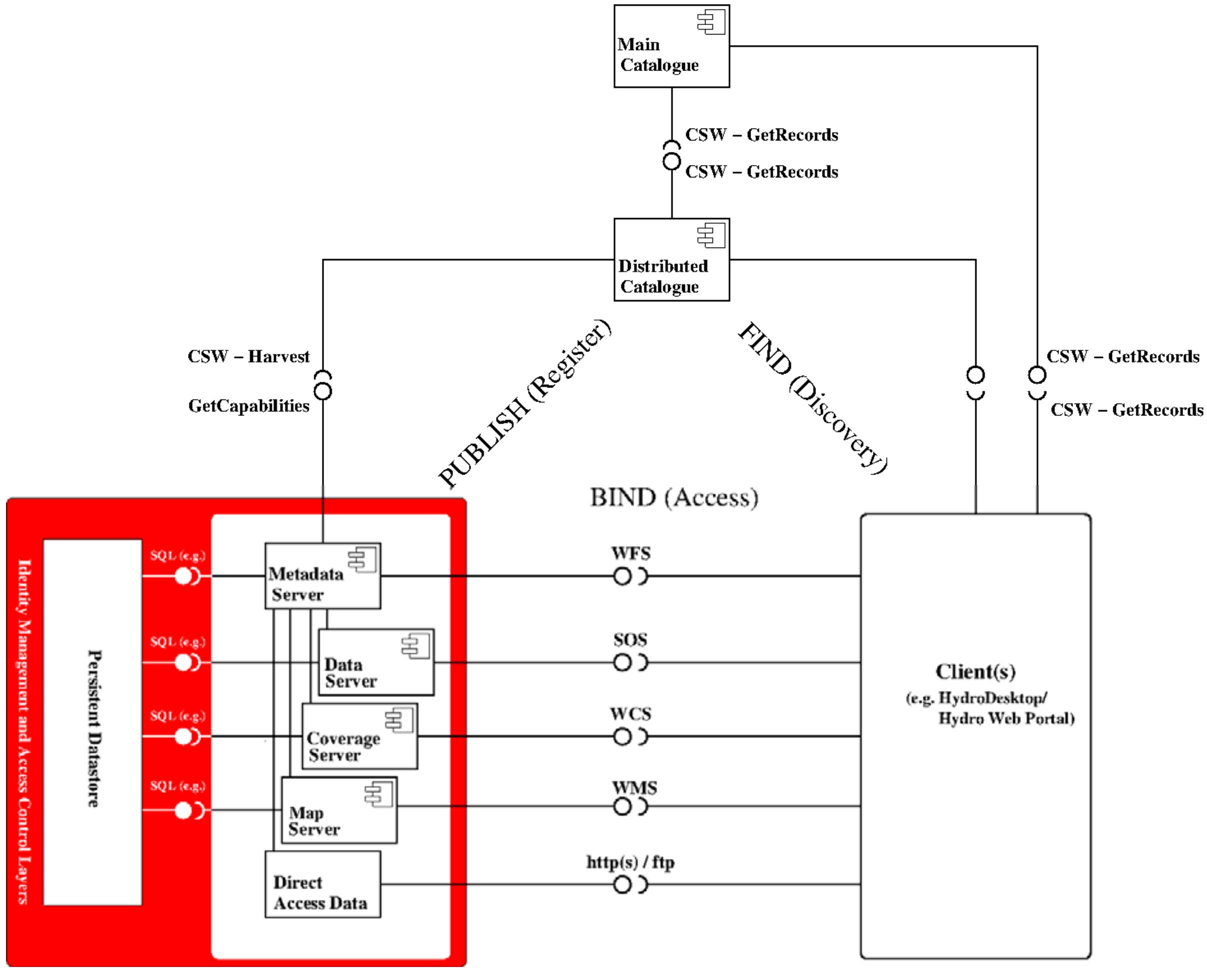


Figure 5 – Deployment Architecture

6.3 Component Types

6.3.1 List of Component types

The services and data types that were used in the deployment architecture are summarized in Table 8 and described in detailed in this section.

Table 8 – Component List for Proposed Architecture

Component Type	Subcomponent Type	Computational Viewpoint Technology (Web Service)	Information Viewpoint Technology (Data Model Encoding)
Metadata Server	For service endpoints	OWS GetCapabilities	OGC Capabilities document
Metadata Server	For data services	WFS	GMLSF encoding of Core Metadata Fields for Hydrologic Time Series Services

Data Server	For time series from network of observation sites	SOS	WaterML v2.0, O&M
Data Server	For time series from grid cells	SOS	O&M, SWE Common
Data Server	For multidimensional arrays	WCS	netCDF
Metadata Server	For dynamic & static maps. For map-visualization of all data services, even WFS.	WMS	Image: PNG, JPEG, etc.
Data Server	For static (GIS) data sets, not supported by open standards-based exchange	HTTP/FTP	Various
Data Server	For service endpoints	OWS GetCapabilities	OGC Capabilities Document
Metadata Server	For service and data discovery	CSW	ebRIM v3.0, ISO 19139

6.3.2 Publication and Access

6.3.2.1 Introduction

In general, a publisher of hydrologic data and metadata needs to be concerned about access control, persistent storage and web services. This sub-section describes the specific components, data models, exchange formats and their relationships as required by a data provider in order to publish hydrologic data and metadata onto the web.

6.3.2.2 Metadata server components

6.3.2.2.1 Introduction

The function of the metadata server components is to provide metadata about service endpoints and provide metadata about the hydrologic data that a data provider publishes (see Use Case 1, Table 1).

6.3.2.2.2 For Service Endpoints

Each deployed web service provides metadata about itself in the form of a capabilities document (see OGC 06-121r3). The GetCapabilities operation, which is mandatory for all OGC web services, is to be used to obtain the service metadata.

6.3.2.2.3 For Data Services

Detailed metadata about the hydrologic time series data (including sampling location(s) at which the time series was recorded, or for which it was generated) that a data provider publishes shall be made accessible to web clients via the Web Feature Service interface (see OGC 09-025r1). As more widespread GIS support for SOS becomes available, the DescribeSensor operation may be used, which provides SensorML encoded metadata.

The primary query method for the WFS is the GetFeature request. The canonical response encoding for a WFS GetFeature request shall be a GML document that validates against a GMLSF (see OGC 10-100r2) application schema that encodes the attributes of the Hydrologic Observations Information Model (see Annex A). In addition SOS 2.0 includes a significantly extended GetFeatureOfInterest operation, which might also be considered.

6.3.2.3 Data Server Components

6.3.2.3.1 Introduction

The function of data server components is to provide web access to a data providers'provider's hydrologic data (see Use Case 1, Table 1).

6.3.2.3.2 For Time Series from a Network of Observation Sites

Time series of hydrologic data from a network of observation sites can be made accessible to web clients via the Sensor Observation Service (see OGC 06-009r6) interface (see Use Case 4, Table 4).

The primary query method for the SOS is the GetObservation request. The canonical response encoding for an SOS GetObservation request is an XML document that conforms to the WaterML 2.0 standard (see OGC 10-126) which is a profile (see 9.2.2.5.2) of the Observations and Measurements standard (see OGC 10-125r1).

6.3.2.3.3 For Time Series from Grid Cells and other Multidimensional Arrays

Time series of hydrologic data from grid cells shall be made accessible to web clients via the Web Coverage Service (see OGC 09-110r4) interface.

The primary query method for the WCS is the GetCoverage request. The canonical response for a WCS GetCoverage request shall be a grid coverage encoded in CF-netCDF3 format (see OGC 09-018). As THREDDS implementations also serve GeoTIFF, this would be another preferred encoding for responses to WCS GetCoverage requests.

6.3.2.3.4 For Dynamic and Static Maps

Data publishers that wish to provide efficient visualization of data to web clients should incorporate a Web Map Service (OGC 06-042) interface on top of their hydrologic data, metadata, GIS data, and other static map content. For example a web map server may render onto a map the location of the sensors from which water observations are being published and overlay that with a hydrographic map. WMS is well suited to styling displays, and provides much better performance that WFS or WCS because it only has to move a JPEG, PNG or similar graphic image over the network, instead of potentially very large sets of geographic coordinates and related data.

The primary query method for WMS is the GetMap request. There is no canonical response format for a WMS GetMap request as the client and the server negotiate on the response image format to use.

6.3.2.3.4.1 Static Image Handling

Web map servers may generate their maps dynamically, by accessing the styling source data from a persistent data store, or they may offer static maps in the form of image files. In the latter case, static maps are materialized through the WMS interface as layers where the noSubset flag is set to "true" and the fixedWidth and fixedHeight parameters are set to the size of the fixed image (see OGC 06-042, 7.2.4.7.1). This indicates to clients that this particular layer is static.

6.3.2.3.5 For Static GIS Datasets

A data provider may also offer static GIS data sets, in various formats. These may include proprietary vendor-specific formats in order to include data content not currently handled by open standards, such as topology, network definitions, and coded-value domains. In such cases where no standard encodings or web services are suitable, these datasets can be offered for direct download via HTTP or FTP.

6.3.3 Cataloguing and Discovery

6.3.3.1 Introduction

Catalogues are used by data publishers to register their web services and hydrologic data. Client software uses the catalogue to find, discover, evaluate and help the end user understand services of interests (see Use Cases 2 and 3). Figure 5 shows two levels of catalogues; a main catalogue and a distributed catalogue. It is anticipated that the architecture shall contain only one main catalogue and any number of distributed catalogues (e.g., jurisdictional/regional catalogues).

6.3.3.2 Main Catalogue Component

The proposed main catalogue component is the CSW-ebRIM Registry Service (07-110r4) which is a profile of the HTTP binding defined in the OpenGIS Catalogue Service Implementation Specification (07-006r1). The reason for choosing the ebRIM profile of CSW is that it has an extensible information model, which enables support for user/client queries that filter on fields from the dataset information model, not just the usual metadata fields normally catalogued. This is an essential capability for the current project, to enable filtering on temporal and hydrologic variable values. This choice of main catalogue component may face difficulty in acceptance in Europe due to INSPIRE rules which favor CSW 2.0.2 with ISO Metadata Application Profile (OGC 07-045). However, the ISO metadata can be supported with CSW-ebRIM, and this situation may evolve with further implementation experience.

The purpose of the main catalogue component is to harvest and cache metadata from the distributed catalogues and act as the primary search node for the discovery use case (see Table 3). The main catalogue should support discovery based on locally cached copies of the metadata from distributed catalogues and also on distributed search in those cases where the distributed catalogue cannot be harvested (see Annex B). This approach is in current use by CUAHSI HIS Central with the U.S. national data sources.

The primary query method for the CSW-ebRIM is the GetRecords request. The canonical response encoding for a CSW-ebRIM GetRecords request is a catalogue record encoded in XML that validates against the ebRIM v3.0 schema. The anticipated query pattern is to first search for data using "who-what-where-when" predicates to discover data services of interest. Once the data are discovered, the GetRecords operation can also be used to discover which services offer the data of interest and the binding information for those services (see Use Case 4, Table 4). However, caching and performance scalability considerations may affect this approach, with further implementation experience.

6.3.3.3 Distributed Catalogue Component

The proposed distributed catalogue component is the CSW-ebRIM Registry Service (see OGC 07-110r4) which is a profile of the HTTP binding in the OpenGIS Catalogue Service Implementation Specification (see OGC 07-006r1).

The purpose of distributed catalogue components is to aggregate metadata about published web services by harvesting the capabilities documents of published metadata servers. The content of a distributed catalogue shall include queryables that allow data/metadata and service discovery by: data source, experimental site, variable or concept of interest from the hydrology ontology, temporal extent, geographic extent, etc. These queryables are referred to as the "who-what-where-when" criteria in Use Case 3.

The primary harvesting method (the means by which the catalogue is populated with metadata) for CSW-ebRIM catalogues is the Harvest request (see OGC 07-110r4, section 14). The Harvest request reads well-known metadata, such as the Capabilities document of an OGC service, and registers the information into the catalogue. The Harvest request can also be used to periodically refresh this information so that it remains current. Direct-access GIS data can be registered into the catalogue by having the Harvest request read ISO19139 metadata describing the direct-access data.

The primary query method for the CSW-ebRIM interface is the GetRecords request. The canonical response encoding for a CSW-ebRIM GetRecords request is an XML document that validates against the ebRIM v3.0 schema. The GetRecords operation can be used to search for data using a variety of criteria (i.e., the who-what-where-when queryables from Use Case 3) and then find the related service metadata that describes how to access data from the discovered service (i.e., bind). This can be done by adding classification schemes to support hydrologic searches (e.g., variable types, site types, and Earth realms).

6.3.3.4 ebRIM Information Model

ebRIM has been introduced as a profile of CSW in Section 4.4.2.3. The main features of ebRIM include:

1. The ability to register any type of information including all the information types used in hydrology.

2. The ability to create associations between information registered in the catalogue. This capability allows, for example, data to be associated with the service that offers it so that if either is discovered the other is immediately accessible. Another example

of the use of associations is the ability to organize data into themes; a theme object is registered into the catalogue and each data item that is part of the theme is associated with the theme object.

3. The ability to classify and search for information based on a taxonomy or ontology that is stored in the catalogue.

4. The ability to store binding information for services.

5. The ability to extend the model with domain specific metadata.

The ebRIM specification defines both a SQL encoding of that model, which can be implemented in any SQL-relational database, and an XML encoding for exchange over the web. A SQL database is not required to support ebRIM, though it is most often used in practice.

6.3.3.5 OpenSearch Interface

The CSW-ebRIM API is a full-featured API designed to satisfy the demanding catalogue needs of the hydro community. The CSW standard also provides a simpler, more lightweight interface to allow users outside the hydro community to query the available catalogues by using OpenSearch (see http://www.opensearch.org).

OpenSearch is a lightweight query interface designed to support cross-catalogue queries and will be a mandatory part of the upcoming CSW 3.0 specification. Supporting OpenSearch from a CSW catalogue does not require any interface changes but does require that the catalogue support ATOM (see http://datatracker.ietf.org/doc/rfc5023/) as a response encoding and that the catalogue provide a description document containing templates telling OpenSearch clients how to form valid query URLS.

A significant benefit of OpenSearch is that most available browsers can natively act as OpenSearch clients by simply providing an OpenSearch description document. Most users are familiar with the search box in major web browsers that provides a dropdown list of search engines that may be used to perform a search; the search engines in that list were added to the browser using OpenSearch description documents and the list can be extended by simply providing the OpenSearch description document of the search engine to be added. So, by supporting OpenSearch, hydro catalogues can now be integrated over the web at large and become simply another search engine that can be used to discover resources.

6.3.3.6 Semantic Annotations and Search

6.3.3.6.1 Introduction

As described in Use Case 1, a data provider shall semantically annotate (tag) its metadata using controlled vocabularies such as the hydrologic domain ontology. The purpose of these annotations is to support search based on ontology.

The OGC has done some work on semantic annotation and search but no standards have emerged thus far. This sub-clause presents some approaches that might be explored in a test bed to convey semantic annotations from a service to the catalogue and then perform semantic searches on the catalogue.

In order to enable semantic search in a catalogue the following conditions must be satisfied by the proposed architecture:

1. The controlled vocabularies or ontologies that are used for semantic annotation must be preloaded into the catalogue which means that the catalogue's information model must include the necessary structures to store this information.

2. The data provider must annotate its information and must make those annotations available in a form that a CSW-ebRIM catalogue can harvest.

3. The catalogue's predicate language must include an operator the supports query based on classification schemes or ontologies.

6.3.3.6.2 Storing Ontology in the Catalogue

The proposed CSW-ebRIM catalogue component includes a set of structures in its information model (see ebRIM, 7.2.6) that provide the ability to store controlled vocabularies such as the hydrologic ontology and to tag information stored in the catalogues with those annotations. Figure 6 illustrates the UML model for ebRIM. The area shaded in blue identifies the available model elements for storing ontology and annotating information with terms from that ontology.

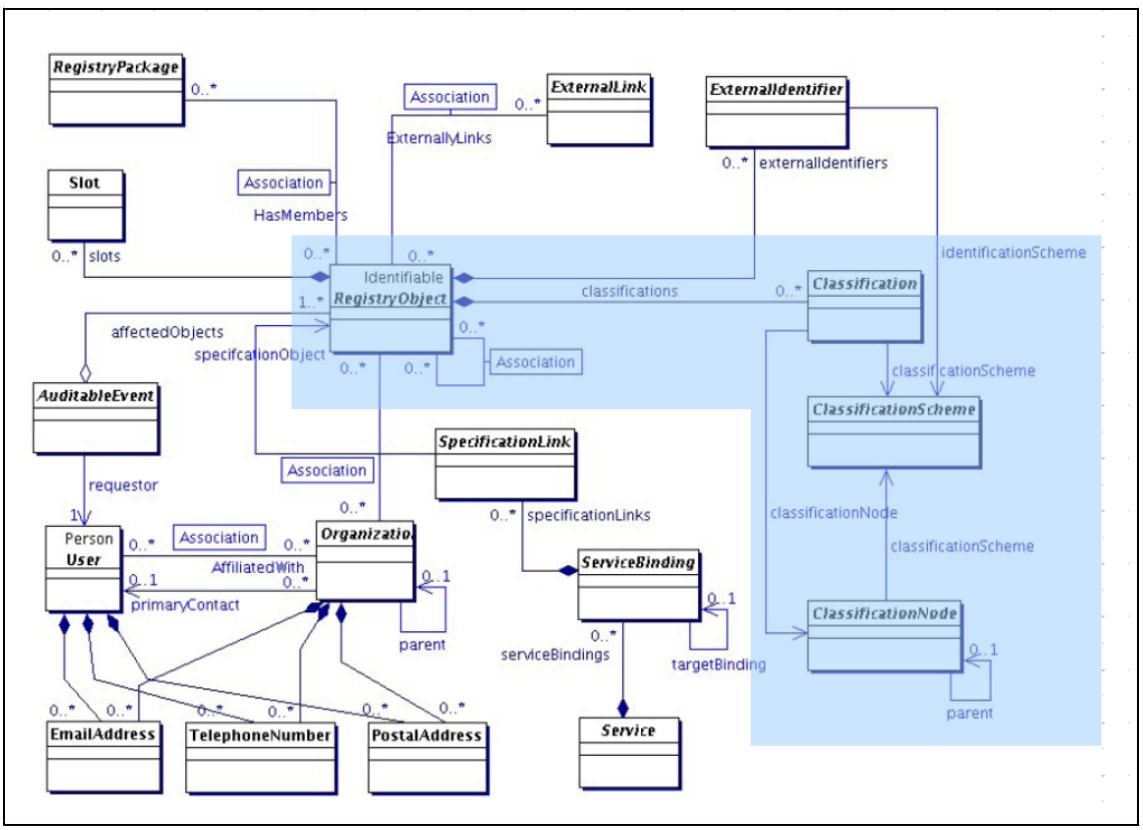


Figure 6 – ebRIM Model

6.3.3.6.3 Harvesting Semantic Annotations

A necessary requirement for supporting semantic search is that the semantic annotations be available in some form that a CSW-ebRIM catalogue can harvest. There is not yet a standard way to do this within OGC standards, but this is an area of current investigation (see OWS-8 discussion in Section 7.4 below).

One specific approach could be to harvest the semantic annotations using the OGC Capabilities document. The OGC Web Services Common Implementation Specification (see 06-121r9) offers a number of extension points in the capabilities document where such annotations can be included. These extension points include:

1. Using keywords with a special notation to indicate that specific keywords are in fact semantic annotations (see OGC 08-167, clause 3.1.2).

2. Using the ows:ExtendedCapabilities section to add the necessary XML structures to encode the semantic annotations.

3. Using the available ows:Metadata element to encode or reference the semantic annotations.

4. As part of the data content information of the service.

Example 1:

The following example (from OGC 08-167, clause 3.1.2) shows how keywords may be used to annotate information is OGC web services that use OWS common:

```
...
<ows:Keywords>
  <ows:Keyword>snow</ows:Keyword>
  <ows:Type codeSpace="http://www.cuahsi.com/ontology/hydrophere#">ontology</ows:Type>
</ows:Keywords>
...
```

Example 2:

The following example illustrates the use of the standard gml:metaDataProperty element to convey semantic annotation in the capabilities document of a service.

The capabilities document of a Sensor Observation Service (see OGC 06-009r6) includes a Content section that contains a list of observation offerings. The XML encoding of observation offerings is derived from gml:AbstractFeatureType which includes an element called gml:metaDataProperty that can contain metadata about the offering. The following XML fragment is an example illustrating how the gml:metaDataProperty element can be used to convey semantic annotations with an offering:

```
<sos:ObservationOffering>
   <gml:metaDataProperty>
      <cuahsi:annotation concept="http://www.cuahsi.org/ontology/hydropshere#snow"/>
   </gml:metaDataProperty>
   .
   .
   .
</sos:ObservationOffering>
```

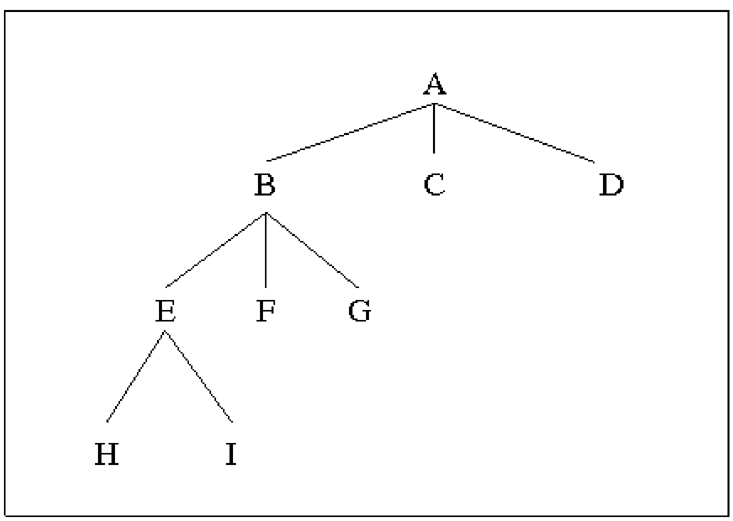
In this example the fictitious cuahsicuahsi:annotation element has been defined to encode the concept associated with the observation offering. In fact any element may be embedded within a gml:metaDataProperty element so we could have also reused the ows:Keyword element from Example 1.

In either case, Example 1 or Example 2, a catalogue registering the information would note the annotation and tag the incoming information appropriately.

6.3.3.6.4 ClassifiedAs Operator

In order to actually be able to query the catalogue using semantic annotations, the CSW must include an operator that is capable of traversing the ontology to identify records that have been tagged with the desired concepts. The proposed operator for this purpose is called ClassifiedAs (see OGC 08-167). The proposed prototype for this operation is: classifiedAs(concept URI, scope string) where the valid values for scope are narrow, exact, and full.

To understand the behavior of the classfiedAs() operator consider the following diagram that illustrates a fictitious ontology where the concepts are represented as letters:



In addition, imagine that the following information has been registered within a CSW-ebRIM catalogue:

Annotation	Record Id	Record Content
H	1	… content for record 1 …
H	2	… content for record 2 …
I	3	… content for record 3 …
F	4	… content for record 4 …
B	5	… content for record 5 …
C	6	… content for record 6 …
G	7	… content for record 7 …
I	8	… content for record 8 …
H	9	… content for record 9 …
H	10	… content for record 10 …

The predicate classfiedAs(C,"exact") would match exactly 1 record; record 6 which is tagged with concept C.

The predicate classifiedAs(E,"broad") would match records 1,2,3,8,9,10. The use of the scope "broad" indicates that the operator should match all catalogue records tagged with concept E and any child of that concept (H and I in this example).

The predicate classifiedAs(F,"narrow") would match records 4 and 5. The use of the scope "narrow" indicates that the operator should match all catalogue records annotated with the concept F and any ancestor of that concept (B or A in this example).

6.3.3.6.5 Summary

In summary, with all the components now defined, the process of harvesting and search based on semantic annotations is:

1. The catalogues are preloaded with the controlled vocabularies and the hydrologic ontology.

2. Publishers annotate their data/metadata and make those annotations available in the capabilities documents of their data/metadata servers.

3. Publishers register their data/metadata servers to the catalogues using the Harvest operation.

4. The Harvest operation reads the capabilities documents of the data/metadata services and registers the services in the catalogue AND tags each piece of information with the appropriate annotation based on the information in the capabilities document.

5. Using the ClassifiedAs() operator users perform queries for records that satisfy specified concepts or sets of concepts.

One objective of an OGC test bed would be to study the possible methods and recommend how semantic annotations should be included in an OGC capabilities document. Another objective of an OGC test bed would be to exercise the proposed ClassifiedAs() operator and determine its suitability in satisfying the use cases.

6.3.4 Client Component(s)

6.3.4.1 Introduction

The client component is the tool end users use to interact with the system. It helps the user discover what is available (Use Case 3), and download or use the data of interest (Use Case 4). The client application interacts with the catalogue and the data provider's service following the interfaces discussed in Section 6.3.2 "Publication and Access."

6.3.4.2 Example User – Client Interaction

Description of a concrete, though hypothetical, application could make the picture and likely alternatives more clear: consider a user who is querying for data using a web-based search tool. The search tool enables him/her to specify the who-what-where-when criteria, which might be displayed as categories (facets) in the search dialog. The results appear in some appropriate place in the browser window, and show the number of results for each category. The user continues to filter and query, until the user has found all the data of interest. There is a button for preview of the data that allows the user to get information about the extent and possible preview of some values. There might be an "add-to-cart" button beside each result which allows that data be added to the download cart. Once the client is satisfied that they have found all the data they require, they go to the download cart and download the data in a variety of formats (WaterML, GML, KML, ATOM, etc.). The available formats are determined by the output formats that the published services support.

6.4 Architecture Variations

6.4.1 Introduction

The currently proposed deployment architecture (Section 6.2), based on the CUAHSI prototype, employs one or more distributed catalogues to act as meta-catalogues of water services. That is, the distributed catalogues do not contain metadata about data services but rather hold meta-information for time series inventory WFS servers. A discovery client would query a distributed catalogue to find the set of available time series meta-information servers, typically using the OGC core set of queryable properties, and then queries the meta-information server(s) using WFS requests to examine the time series inventory for specific content such as the service provider name or the SOS endpoint that may be accessed to obtain data via the GetObservation operation. The GetObservation operation is then used to return WaterML to the client.

We now propose an alternative architecture that leverages already existing interfaces to handle discovery of services. This alternative is based on CSW-ebRIM. All metadata previously provided by the WFS meta-information server would instead be made available via distributed CSW-ebRIM catalogues.

The CSW-ebRIM specification, like the WFS specification, uses the OGC Filter Encoding specification and can thus support the same query types offered by the WFS-based meta-information servers.

The CSW-ebRIM would use the GetCapabilities response and other operations to get information about data services and populate the catalogue appropriately. For example, it can use the GetCapabilities response of SOS to harvest information about the service endpoint. The SOS DescribeSensor operation would be used to convey, via SensorML, the metadata about a site and process (e.g., site identifier, quality assurance, contact information). SWE common would be used to convey metadata about the data result (e.g., variable names, units).

The catalogue would be configured to periodically re-harvest all metadata from the source services in order to keep the information current. In fast-changing environments a data provider might consider deploying a Sensor Instance Registry Service (see OGC 10-171) and/or a Sensor Observable Registry Service (see OGC 09-112) which feed metadata changes directly to the catalogue.

In cases where the data provider has a large set of observations, services or site, the data provider may deploy their own CSW-ebRIM catalogue as a distributed node in a network of catalogues.

6.4.2 Deployment Architecture for CSW-Based Meta-information Servers

The following UML component diagram illustrates the proposed alternative deployment architecture for a water information system based on the components identified in the previous viewpoints with the exception of the WFS meta-information server whose role is now assumed by one or more of the distributed catalogues:

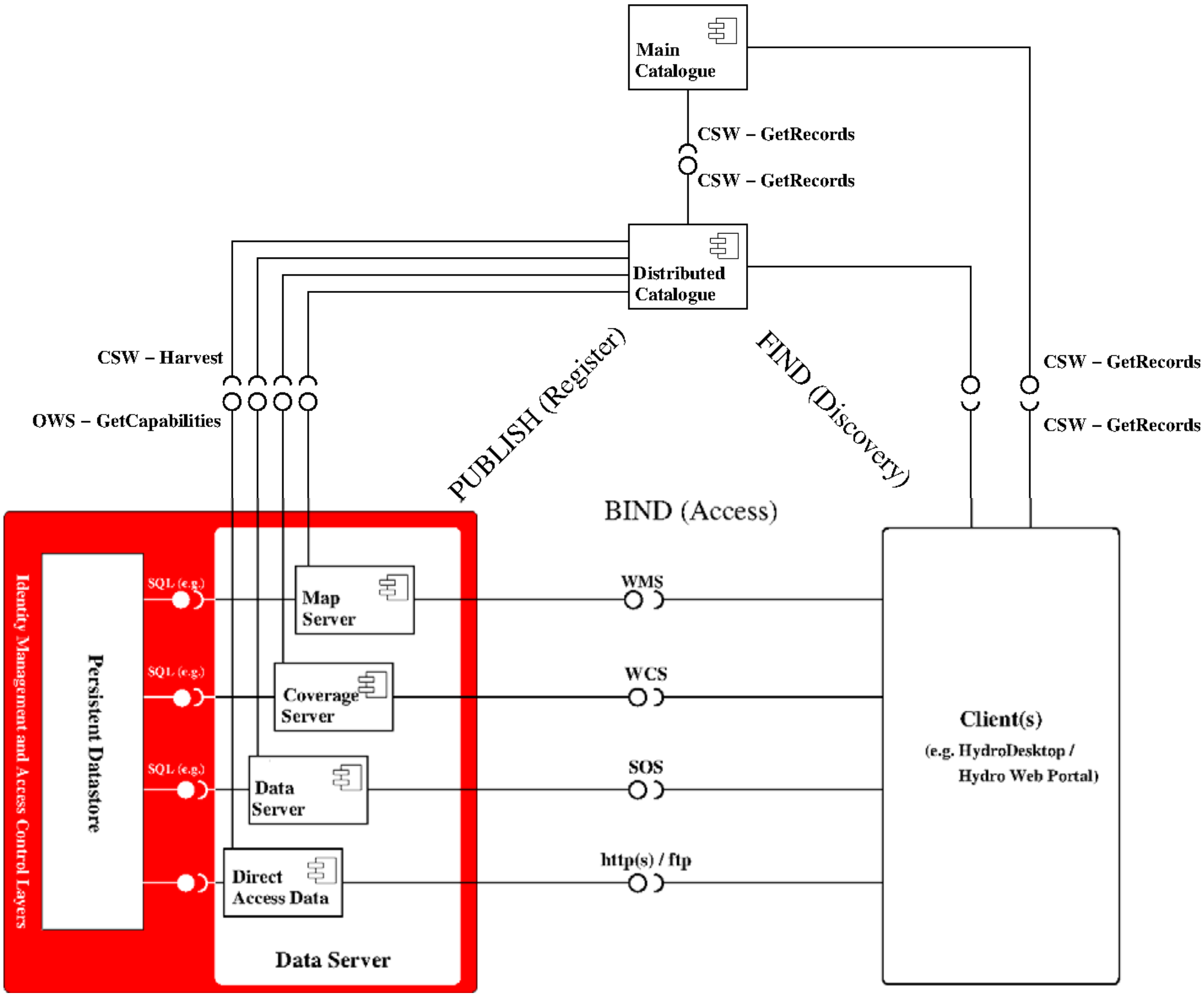


Figure 7 – Alternative (CSW-Based) Deployment Architecture

6.4.3 Components

The services and data types that are used in this alternate deployment architecture are summarized in Table 9. Unlike the previous architecture, which uses a WFS as a meta-information service providing metadata about service endpoints, sensors, observations, etc., in this alternative architecture one or more distributed CSW-ebRIM catalogues assume this role. Also, in the case of time series from sensors it is proposed to leverage other SWE standards, such as providing SensorML for site descriptions.

Table 9 – Component List for Alternative Architecture

Component Type	Subcomponent Type	Computational Viewpoint Technology (Web Service)	Information Viewpoint Technology (Data Model Encoding)
Metadata Server	For service endpoints	OWS GetCapabilities	OGC Capabilities document
Metadata Server	For sensors	SOS	SensorML

Metadata Server	For observations	SOS	O & M
Metadata Server	For multidimensional arrays	THREDDS metadata	netCDF/CF header
Data Server	For time series from network of observation sites	SOS	WaterML v2.0, O&M
Data Server	For time series from grid cells	SOS	O&M, SWE Common
Data Server	For multidimensional arrays	WCS	netCDF
Data Server	For dynamic & static maps	WMS	Image: PNG, JPEG, etc.
Data Server	For static (GIS) data sets	HTTP/FTP	Various
Data Server	For service endpoints	OWS GetCapabilities	OGC Capabilities Document
Metadata Server	For service and data discovery	CSW	ebRIM v3.0, ISO 19139

7 Related Work

There are additional prior and ongoing activities related to this concept report, including OGC Interoperability experiments lead by the Hydro Domain Working Group, portions of the current OGC Web Services testbed initiative, Phase 8 (OWS-8), and a study being developed by CSIRO.

Over the last several years OGC has run several interoperability experiments in the domain of water information systems that have exercised aspects of the deployment architecture presented in this report. The two most relevant previous interoperability experiments are the Surface Water Interoperability Experiment and the Ground Water Interoperability Experiment.

7.1 Ground Water Interoperability Experiment

The Ground Water Interoperability Experiment (GWIE) was the first IE to be initiated by members of the Hydro DWG, in this case to advance the development of WaterML 2.0. This involved the interaction of several existing representations of water information including WaterML (U.S.), Observations & Measurements (O&M from OGC) and GroundWaterML (GWML from Canada). The interoperability experiment tested the use of WaterML 2.0 with various services in the OGC stack; namely SOS, WFS, WMS and CSW. The participants and main components of the GWIE are shown in Figure 8; for the final report, see OGC 10-194r3.

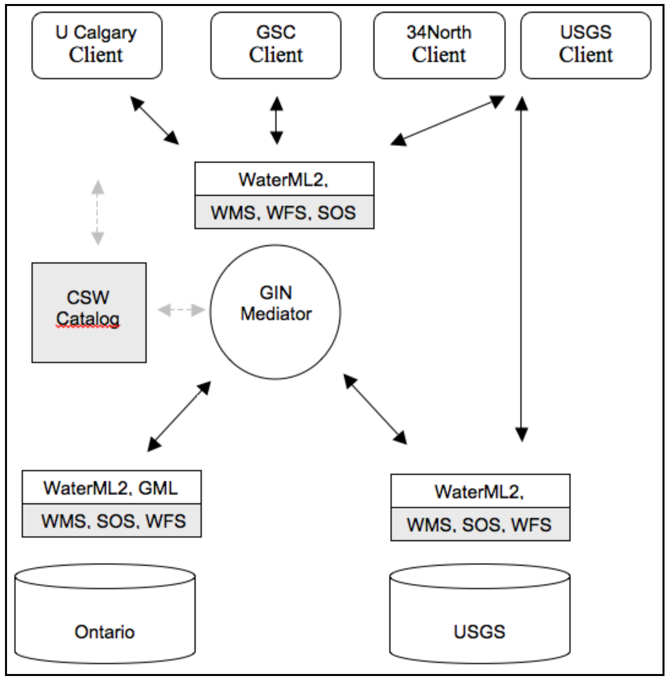


Figure 8 – Ground Water Interoperability Experiment

The architecture of the GWIE was mainly driven by a desire to optimize usability for clients, maximize interoperability, and re-use existing infrastructure. Important guiding principles include:

- One web-service access point for clients: this eliminates heterogeneity in service delivery that is often present in distributed systems;
- A single integrated result returned for clients: this reduces the burden on clients in integrating responses to queries from multiple clients;
- Semantic and schematic interoperability: this eliminates the need for schematic and semantic translation to be carried out at the server end.

The GWIE adopted a mediated architecture as shown in Figure 8. In this architecture a central mediator brokers requests by clients as well as responses by data providers: the mediator carries out query distribution, semantic and schematic translation, and response integration. It is accessed as a standard OGC WMS, WFS, or SOS. These services also wrap spatial databases at the data provider end. A catalogue provides discovery services for: the data servers, the mediator services, and potentially for metadata about each sensor (e.g., time and value of last reading) to facilitate execution of queries across data providers.

The final report from the GWIE presents a large number of recommendations and identifies (and proposes solutions to) a small number of issues, a sample of which is presented here:

1. Sensor observation metadata (solution: use CSW)
2. Relationships between services, e.g., between WMS and WFS and SOS (partial

solution: use WMC to encode relationship) (better solution from Editor: use CSW to maintain association)

3. SOS GetCapabilities and serialization of features-of-interest (solution: allow composite/nested features and feature collections in capabilities)
4. Complex thematic queries in SOS (solution: queries shall be written against SWE meta schema rather than a GML XSD schema)
5. Spatial queries in SOS (solution: use standard identified in filter predicate for the geometric property to which the spatial filter is applied)
6. Handling large volume of results from SOS operations (solution: introduce a resultType parameter with "metadata" and "results" values similar to WFS)
7. Handing large volume of results from SOS operations(solution: introduce maxObservations, maxResults, maxFeatures, sortBy and sortOrder parameters in SOS)
8. Handling of units of measure (solution: add ConversionCapabilities section to capabilities document; add uom attribute to all comparison operations in FES)

Several change requests to OGC SOS resulted from the GWIE, to better support WaterML 2. Many of these were incorporated into SOS 2.0, but some remain to be considered. See OGC 10-194r3 GWIE Final Report and the GWIE twiki for more details.

7.2 Surface Water Interoperability Experiment

The Surface Water Interoperability Experiment (SWIE) began in 2010, and is underway to test and refine WaterML 2.0 for encoding surface water data; test the comparability of exchanging surface water information encoded using Water ML 2.0 with SOS, WFS and WMS services; and advance the near-real-time exchange of surface water data between international participants (Germany & France). Figure 9 and the text following presents an overview of the experiment.

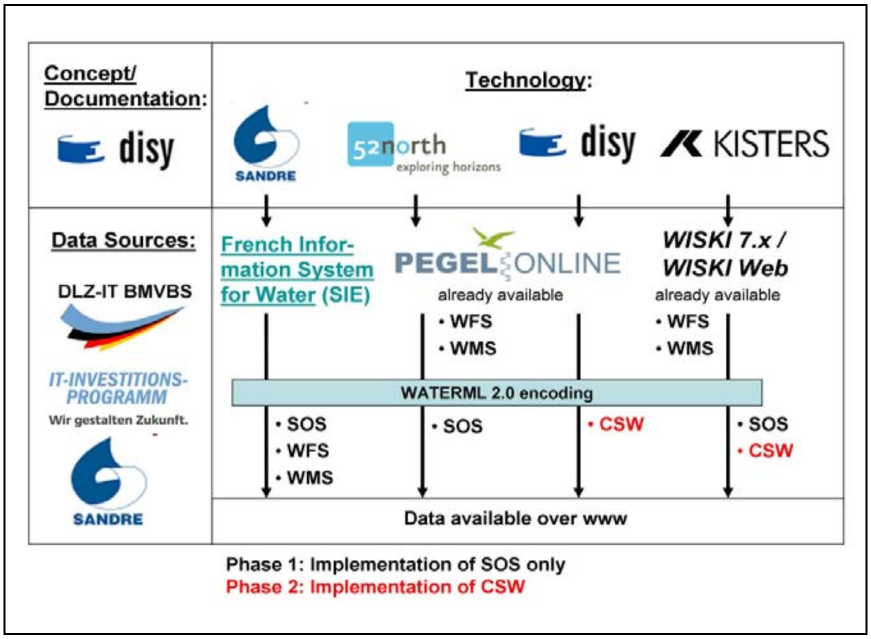


Figure 9 – Surface Water Interoperability Experiment

In this SWIE architecture WaterML 2.0 and OGC technologies are expected to be used in the following manner:

- WMS is used to display the location of available hydrometric stations on a map
- WFS is used to provide gauge locations
- SOS is used to access observations and time series from the hydrometric stations
- CSW is used to make available observations and time series discoverable
- A number of technical issues are discussed during the IE including:
- SOS services that service WaterML 2.0 are used
- Support for different RDBMS/storage layers is enabled
- Feature models to use for the IE (the INSPIRE Hydrography Data Model is discussed in detail) are defined.

7.3 Water Resources Observation Network Reference Model

The Australian Commonwealth Scientific and Industrial Research Organisation (CSIRO) began development of a Water Resources Observation Network Reference Model (WRON-RM), a framework for enabling interoperability of various different data sources (O'Hagan, et al., 2007). The study was halted before completion, but provides a service-oriented architecture with registries to publish and discover data services. The WRON-RM study included use of Data Warehouses, a useful and complementary addition to the architecture in the current report, as well as proposed operational governance policies and principles that should be considered if and when an implementation project can begin.

7.4 OWS-8 Testbed: Cross-Community Interoperability Development

The OGC Web Services Initiative, Phase 8 (OWS-8) has an activity thread now underway called Cross-Community Interoperability (CCI), which is making good progress in designing and implementing a web service architecture for semantic mediation across different data models having overlapping data content of interest to users. The basic issue is very similar to what CUAHSI has to cope with: USGS and Department of Defense (DoD) national-scale data overlaps in many areas of the U.S. where DoD may be concerned about anticipating and preventing a terrorist attack. The goal of this thread in OWS-8 is to develop an architecture based on OGC standards that enables users to query datasets having different semantics, and retrieve the correct features from both datasets. Good progress is being made, and some aspects of the work will be directly applicable to the current Water Information System Concept Development Study. The results of this work will be made public during or shortly after the final demonstration in September 2011. A public description of the initial requirements can be found in the Call for Participation here: http://www.opengeospatial.org/standards/requests/74

8 Interoperability Pilot

Water resources information is provided and collected by a large, heterogeneous ecosystem of individuals and organizations that is complicated even in a small country, becomes even more so when viewed from the scale of the United States, and still more complicated from the global scale. There seems little possibility that there could ever be a single catalogue for all water information in the world, and although CUAHSI maintains the largest water observations data catalogue comprising over 23 million time series of integrated information from several federal and state agencies and multiple research groups, there are thousands more local and state agencies in just the U.S. that are not yet involved. Maintenance and scalability issues associated with this catalogue prompted its current focus on federating different hydrologic data catalogues, and hence standard information models and web service stacks supporting interoperability across catalogues, with efficient interoperation between data services and catalogue services. This issue is especially relevant to large public agencies in the U.S. whose mandate is to collect and make available large volumes of diverse hydrologic data.

The review demonstrated that there are multiple implementation options for hydrologic information architecture, which could derive from the OGC services baseline. However, the services must work together in a coherent way, structurally and semantically, and conform to expectations and requirements of the hydrologic research and modeling community. This points to a series of issues that need to be resolved by an OGC interoperability pilot focused on a hydrologic data web services software stack. In addition to this study, a Hydrologic Web Services Pilot should take into account results of interoperability experiments, conducted within the Hydrology Domain Working Group, where similar architectures are being put forward. The role of the interoperability pilot will be to provide definitive recommendations to the implementers, given the alternative solutions presented in this report.

This concept development report presents a draft architecture for a national-scale Water Information System (WaterIS) based on a stack of data, metadata and service standards most of which have been developed within the OGC. Previous, current and future interoperability projects within OGC and CUAHSI, just described in Section 7, address many of the issues raised here, but further work with federating realistic scales of national and state data is needed to make sure the options and recommendations presented here are the right ones. Scalability and performance are key measures we need to test.

The main goal of an implementation pilot project based on this concept report should be to exercise and stress-test the two alternative system architecture patterns described here, taking into account emerging standards and practices where possible. Broadly, these were the WFS-based metadata server approach, and the CSW-ebRIM approach for metadata server.

Presented here are some specific aspects of the deployment architecture(s) to be investigated in a pilot. They are presented under the broad headings of Publish-Find-Bind, followed by a discussion of managing complexity.

Publish (i.e., data provider deploys services which are harvested by catalogues for discovery):

- Advance the use of SensorML to represent metadata of sites, sensors and observed data based on the Hydrologic Observations Information Model (map Annex A to SensorML). This has already been started in the SensorML profile for discovery. This could be extended to a Hydrology Profile of SensorML. (Currently especially the European FP7 projects GENESIS and EO2HEAVEN are addressing the question of a SensorML profile for discovery purposes.)
- Create a mapping of SensorML to ebRIM in order to support the discovery of sensors and time series data through CSW ebRIM. This would be the second step after a SensorML Hydrology Profile has been created. There are already ebRIM Catalogue implementations available that support the latest version of the SensorML profile for discovery.
- Even though not discussed in this report, future investigation can be performed on the use of ISO 19115 Part 2, which provides a metadata model that can be used to describe hydrologic networks, sites and time series.
- Advance the use of the methodologies and procedures for mapping data provider concepts with ontologies in the CSW.
- Advance the mapping of WQX to SOS/WaterML 2.0.
- Advance open source tools (e.g., GeoNetwork) to provide the CSW-ebRIM.
- Advance the use of CSW-ebRIM to harvest SOS service. This can replace the WFS metadata service to have a consistent interface to query metadata in a distributed hierarchical (system of systems) environment. This requires the creation of harvesting methods that utilize the SOS discovery methods (GetCapabilities, DescribeSensor, etc.). To support near-real-time updates of the catalogue, this investigation may also involve looking at the SIR/SOR services.

Find (i.e., query a catalogue to discover services that provide data of interest):

- Incorporate in the ebRIM model the classes required to provide faceted searches.
- Test distributed discovery using federated query of catalogues using the who-what-when-where pattern of Use Case 3.

Bind (i.e., access services that provide data of interest to download the data):

- Explore the use of open search based on the CSW-ebRIM model.
- Advance tools to visualize metadata records in a GUI.
- Develop a web based Water Information System portal that allows client to discover services that offer hydrologic data of interest to them and then bind to those services to download the desired data.

Pilot sponsors may include agencies, research institutions, and companies that want to advance interoperability among their systems and the whole hydrologic community. Ideas for the pilot also include implementation at the sponsor's organization of the components described in the report, for example setting an SOS that provides data encoded in WaterML 2.0.

Mitigating Complexity: Semantics and Complex Information Models

As the study demonstrated, as long as the information models of the various data sources being combined are reasonably consistent, there may be more than one service interface for hydrologic information system components. For example, time series catalogues may have both WFS and CSW-ebRIM endpoints, and implementing organizations should be able to choose which technology is more compatible with their publication mandates and internal procedures – as long as the service interfaces are formally defined and documented, and express information models that hydrologists can interpret.

It is important that patterns of hydrologic information infrastructure conform to how the hydrologic research community is organized. There are large and small data providers, which may choose to implement different components from the hydrologic services stack. For example, larger data publishers may implement CSW-ebRIM catalogues supporting semantics-based discovery and harvesting, which would index multiple data services within the organization. Smaller data publishers (typically, academic projects) may limit the software stack to data services and have them catalogued by community organizations such as CUAHSI. The interoperability pilot would address interoperation between hydrologic information nodes of different complexity. The approximate "implementation complexity" progression is shown in the figure below.

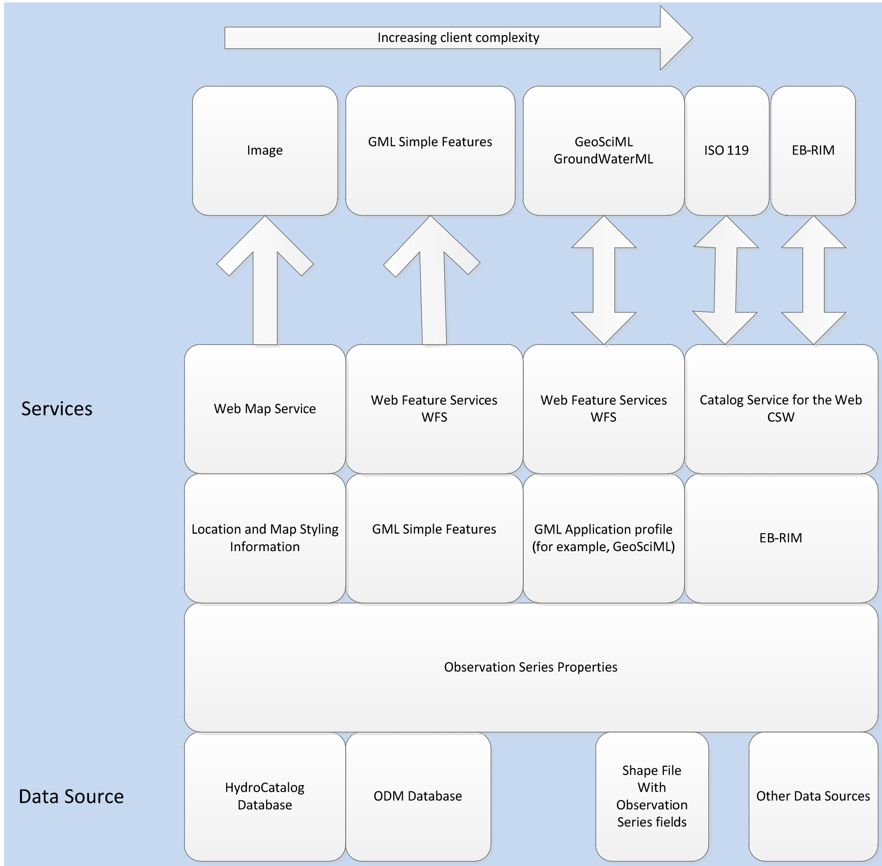


Figure 10 – Implementation Complexity

To be interoperable, services should support semantically consistent binding patterns. For example, the key notions of “feature”, “dataset”, etc. should have consistent semantics throughout the services stack. Currently, even within a single service this meaning may vary between requests (e.g., as demonstrated in HDWG Interoperability Experiments, which showed that the semantics of “feature” may be different in common usage in SOS). One of the tasks of an implementation pilot would be to address such inconsistencies within and between services in the hydrologic services stack.

9 Summary and Conclusions

This Concept Development Study was conducted by the Open Geospatial Consortium (OGC) Interoperability Program at the request of the Consortium of Universities for the Advancement of Hydrologic Science, Inc. (CUAHSI), with the goal of helping CUAHSI to understand how to best utilize and adapt standard encodings and web services developed by OGC to convey water resources data through the Internet.

This Concept Development Study has produced a framework for considering and testing relevant, alternative approaches. It has not arrived at complete, definitive solutions for all these questions. Further investigation in the various OGC programs is needed.

This report provides a snapshot of service standards and practices available and in use "now". As standards evolve and implementations mature, other implementation options should be considered. For example, our rendition of time series catalogue services is to a large extent influenced by wide availability of WFS implementations to relay GML Simple Features. These implementations are easy to use and incorporate in existing agency or research group workflows, thus providing a timely – though perhaps not perfect – solution for the key use cases put forward in this document. As service standards are evolving to handle more and better data collections, we envision that the framework created by this report will be continuously updated.

Annex A

Attributes of the Hydrologic Observations Information Model

This annex lists the attributes deemed required for unambiguous representation of the hydrologic data addressed in this study. This is derived and adapted from sources such as ODM and WaterML, drawing upon our experience working with these systems, but does not follow any one of these exactly. Rather the focus is on the content deemed most suitable for representation of hydrologic information to support the use cases of this study.

Attribute Name and Description describe the actual attributes. Attribute Group organizes these by what, where, when, who, how and by data value, dataset, web service and theme. The "Important for Discovery" column indicates the subset of attributes that comprise the authors' assessment of the subset most useful for discovery and thus the basis for the catalogue content. We recognize that this is a judgment or design trade off. Holding more data in a catalogue can support more extensive search, at the expense of greater complexity. There are five scope columns on the right that indicate the scope of each attribute as follows:

- PTS – Point time series
- FTS – Feature time series
- MA – Multidimensional array
- S – Static
- DS – Data Services

Attribute Group	**Attribute Name**	**Attribute Description**	**Important for Discovery**	**Scope**				
				PTS	**FTS**	**MA**	**S**	**DS**
Data - What	Accuracy	Quantification of the measurement accuracy associated with observation values	No	X	X	X	X	
Data - What	Censoring	An indication of whether a data value is above or below a detection limit	No	X	X	X		
Data - What	Qualifiers	Data qualifying comments that describe data values	No	X	X	X	X	
Data - When	Date and Time	The date and times at which the observations were made	Yes	X	X	X		
Data - Where	Offset and Offset Type	Distance from a reference point to the location at which the observation was made (e.g., 5 meters below the water surface)	No	X	X	X		
Dataset	Authorization Required	1 (TRUE) if authorization for download is required; 0	No	X	X	X	X	

Attribute Group	**Attribute Name**	**Attribute Description**	**Important for Discovery**	**Scope**				
				PTS	**FTS**	**MA**	**S**	**DS**
		(FALSE) otherwise. Used to manage data retrieval						
Dataset	Publication Date	The date on which the dataset was published	No	X	X	X	X	
Dataset	Spatial Type	Spatial field (raster) or vector feature type (point, line, polygon)	Yes		X		X	
Dataset	Value Count	Number of time instances that the feature point or multi array has been observed. Used as query filter.	Yes	X	X	X		
Dataset - What	Characteristic Categories	The definition of the categories that correspond to the values stored in the dataset (e.g., land cover classes)	No	X	X	X	X	
Dataset - What	General Category	An indication of the general category within which the variable falls (e.g., hydrology, water quality, weather and climate)	Yes	X	X	X	X	
Dataset - What	Medium	Medium in which the dataset applies. See CUAHSI Controlled Vocabulary for a recommended list of terms. Used as query filter	Yes	X	X	X	X	
Dataset - What	Ontology Concept	Leaf concept keyword from the ontology to which this variable applies. Used as query filter, e.g., "Discharge, stream".	Yes	X	X	X	X	
Dataset - What	Ontology Name	Unique name for the ontology containing the concept to which the given variable has been mapped. USGS and EPA are under mandate to create and use and SRS ontology. Used as query filter, e.g., "CUAHSI Variable Ontology v1.26".	Yes	X	X	X	X	
Dataset - What	Quality Control Level	An indication of the level of quality of the data or the level of quality control to which the data have been subjected	No	X	X	X	X	
Dataset - What	Speciation	For concentration measurements, the species in which the concentration	No	X	X	X	X	

Attribute Group	**Attribute Name**	**Attribute Description**	**Important for Discovery**	**Scope**				
				PTS	**FTS**	**MA**	**S**	**DS**
		is expressed (e.g., as N, or as NO3, or as NH4)						
Dataset - What	Units	Unabbreviated (e.g., cubic meters per second) and abbreviated (e.g., m3/s) units of measure along with the unit type (e.g., length, volume/time) for the variable. See the CUAHSI Controlled Vocabulary for a recommended list of terms. Used as query filter	No	X	X	X	X	
Dataset - What	Variable Name	The name of the variable. Used as query filter, e.g., "Streamflow". For spatial fields, the names of the characteristics that were measured or estimated (e.g., elevation)	Yes	X	X	X	X	
Dataset - What - How	Method Description	A description of the method or laboratory analytical procedure by which the observation was generated	No	X	X	X	X	
Dataset - What - How	Method Name	The name of the method or laboratory analytical procedure by which the observation was generated	Yes	X	X	X	X	
Dataset - What - How	Samples	A description of the samples from which data values were generated	No	X	X	X	X	
Dataset - What - How	Value Type	An indication of whether the data value represents an actual measurement, a calculated value, or is the result of a model simulation	Yes	X	X	X	X	
Dataset - What - When	Data Type	An indication of the kind of quantity being measured over the time interval (e.g., continuous, end of interval, cumulative, sporadic)	Yes	X	X	X		
Dataset - What - When	Is Regular	1 (TRUE) if variable is measured/calculated regularly in time; 0 (FALSE) otherwise. Used as query filter	Yes	X	X	X		
Dataset - What - When	Temporal Spacing	For regular data, the duration and units giving the length of time between measurements. Used as	Yes	X	X	X		

Attribute Group	**Attribute Name**	**Attribute Description**	**Important for Discovery**	**Scope**				
				PTS	**FTS**	**MA**	**S**	**DS**
		query filter, e.g., P1D (one Day), P1M (one Month), PT12H (Time One Hour), or PT15M (Time 15 Minutes)						
Dataset - What - When	Temporal Statistic Type	An indication of the aggregation statistic reported over the time interval of the measurement. See the CUAHSI Controlled Vocabulary for a recommended list of terms. Used as query filter; e.g., Value, Average, Maximum	Yes	X	X	X	X	
Dataset - What - When	Temporal Support	The time interval and unit over which each observation was collected or implicitly quantified by the measurement method.	Yes	X	X	X	X	
Dataset - What - Where	Network Description	A description of the network of monitoring sites	No	X	X	X	X	
Dataset - What - Where	Network Name	The name (Unique ID) of the network of monitoring sites	Yes	X	X	X	X	
Dataset - What - Where	Spatial Spacing	The grid spacing between individual measurements	Yes			X		
Dataset - What - Where	Spatial Statistic	An indication of the kind of quantity being measured over the spatial interval (e.g., minimum, maximum, average)	Yes	X	X	X	X	
Dataset - What - Where	Spatial Support	The spatial interval over which each observation was collected or implicitly quantified by the spatial statistic	Yes	X	X	X	X	
Dataset - When	Data Availability	Indicates a floating period of record by representing the time period for which data are available; otherwise *NULL*. Uses ISO 8601 duration, indicating duration from the beginning of the period of record up to the current date and time. Used for harvesting and maintaining a catalogue. E.g., P120D (data are available from	No	X	X	X		

Attribute Group	**Attribute Name**	**Attribute Description**	**Important for Discovery**	**Scope**				
				PTS	**FTS**	**MA**	**S**	**DS**
		120 days ago up to now)						
Dataset - When	Date and Time	A description of the time period of the content of the dataset	Yes				X	
Dataset - When	End Data and Time	End date and time for the time period of the variable at the site. If the site is active, then this will be null or empty. The value of Now or LastUpdated is appropriate. Used as query filter, e.g., "2006-08-31T11:26-06"	Yes	X	X	X		
Dataset - When	Is Active	1 (TRUE) if data provider may update series (e.g., real-time data); 0 (FALSE) otherwise. Used for harvesting and maintaining a catalogue	No	X	X	X		
Dataset - When	Last Update	Date when series was last updated. Used as query filter. E.g., "2010-11-03T16:00-06"	Yes	X	X	X		
Dataset - When	Start Date and Time	Start date and time for the time period of the variable at the site. If data is available for a limited time, StartDate will be calculated as Now minus the DataAvail. Used as query filter. E.g., 1994-05-03T08:40-06.	Yes	X	X	X		
Dataset - Where	Feature	A geographic feature, either point, line or polygon with its spatial reference. May be two or three dimensional			X		X	
Dataset - Where	Location Keywords	Keywords describing the spatial domain of the dataset	Yes	X	X	X	X	
Dataset - Where	Site Name	Name or Unique ID of the location at which data were collected. Used as query filter. E.g., "Colorado River at Austin"	Yes	X				

Attribute Group	**Attribute Name**	**Attribute Description**	**Important for Discovery**	**Scope**				
				PTS	**FTS**	**MA**	**S**	**DS**
Dataset – Where	Feature Name	Name or Unique ID of the feature being observed, e.g., “Colorado River”	Yes	X	X	X	X	
Dataset – Where	Feature Type	Type of the feature being observed, e.g., “River / Stream water column.” Terms are defined by a controlled vocabulary	Yes	X	X	X	X	
Dataset - Where - XYZ	Bounding Box	The coordinates and spatial reference system of the geographic bounding box containing the dataset	Yes		X	X	X	
Dataset - Where - XYZ	Elevation	The elevation of the location and its reference to a vertical datum	Yes	X				
Dataset - Where - XYZ	Latitude	The latitude and spatial reference system of the location	Yes	X				
Dataset - Where - XYZ	Longitude	The longitude and spatial reference system of the location	Yes	X				
Dataset - Who	Abstract	An abstract that describes the characteristics of observations from a particular source	Yes	X	X	X	X	
Dataset - Who	Citation	The citation for a particular set of observations	No	X	X	X	X	
Dataset - Who	Contact	The name and contact information for the person responsible for the dataset	Yes	X	X	X	X	
Dataset - Who	Organization	The name and description of the organization or agency that created the dataset	Yes	X	X	X	X	
Dataset - Who	Title	A title that indicates the characteristics of observations from a particular source	Yes	X	X	X	X	
Service	Max Records	Maximum number of records returned. If a service wished to limit the number of data values returned, it should indicate so by populating this value. Zero or *NULL* indicates no maximum. Used to manage data retrieval	No					X
Service	Publication	The date on which the	No					X

Attribute Group	Attribute Name	Attribute Description	Important for Discovery	Scope				
				PTS	FTS	MA	S	DS
	Date	service was published						
Service	Publication Place	The place where the service is published	No					X
Service	Service Location	The address of the service on the Internet	No					X
Service	Service Type	A description of the type of data service	No					X
Service - What	Ontology Concepts	Keywords or concept names describing the variables that make up the contents of the service	Yes					X
Service - What	Ontology Name	Unique name for the ontology containing the concept to which variables have been mapped. USGS and EPA are under mandate to create and use and SRS ontology. Used as query filter, e.g., "CUAHSI Variable Ontology v1.26"	Yes					X
Service - When	Date and Time, Begin and End	The time extent of the data contained within the service, or time period of content	Yes					X
Service - When	Last Update	The date on which the data within the service was last updated	Yes					X
Service - When	Update Frequency	The frequency with which the data are updated within the service	Yes					X
Service - Where	Bounding Box	The coordinates and spatial reference system of the geographic bounding box containing the contents of the service	Yes					X
Service - Where	Location Keywords	Keywords describing the spatial domain of the service	Yes					X
Service - Who	Abstract	A description of the data offered by the service	No					X
Service - Who	Contact	The name and contact information for an individual who is responsible for the data	No					X
Service - Who	Organization	The name and description of the service publisher	No					X
Service - Who	Purpose	The purpose for which the data were collected and or published	No					X
Service - Who	Title	The title or name of the service	No					X

Attribute Group	**Attribute Name**	**Attribute Description**	**Important for Discovery**	**Scope**				
				PTS	**FTS**	**MA**	**S**	**DS**
Theme	Theme Description	The description of a theme to which a particular dataset belongs	No	X	X	X	X	
Theme	Theme Name	The name of a theme to which a particular dataset belongs	Yes	X	X	X	X	

Annex B
Trade Study: Distributed Search vs. Harvest

B.1 Introduction

The main catalogue has two ways to interact with distributed catalogues: Harvest and Distributed Search. This clause discusses both of the alternatives. After defining the alternatives, a set of evaluation criteria is defined followed by an analysis of the alternatives using the criteria. Conclusions are presented at the end.

B.2 Trade Study Alternatives

The Harvest alternative retrieves metadata from Distributed Catalogues, often on a regular basis, and stores the retrieved metadata in the Main Catalogue cache. This alternative creates a copy of the Distributed Catalogue metadata in the Clearinghouse. Advantages of this approach are that searches of the Main Catalogue are completed quicker because the data is local and offers a fall back solution if the Distributed Catalogue is not online. Disadvantages include that the Main Catalogue copy of the metadata may be out of date and the required effort to initially harvested the metadata from the Distributed Catalogues.

The Distributed Search alternative is invoked when the Main Catalogue receives a search request from a client and propagates a secondary request to one or more Distributed Catalogues. An advantage of this approach is that the metadata is maintained and managed closer to the data provider. A disadvantage is that distributed searching takes longer to complete and has more chances for the search to not be completed.

B.3 Trade Study Evaluation Criteria

The following criteria are used to evaluate the options.

<u>User response time</u>:
The performance in terms of time to respond to a user search placed against the Main Catalogue. The User's search is assumed to trigger distributed searching. The duration is from the time the search is received by the Main Catalogue until the operation has fully completed and the Main Catalogue has replied to the user with results.

<u>Results ranking</u>:
Users desire that the results of a search be ranked in a fashion that provides the result of most interest at the top of the list. Ranking of results requires that the entire result set can be evaluated in a uniform fashion.

<u>Metadata accuracy</u>:
The consistency of the metadata received by the user in comparison to the most accurate metadata available at any location at the time of the search.

Metadata ownership:
Compatibility of the alternative to stewardship of metadata by the organization that is charged with maintaining the metadata.

Robustness
Ability of alternative to gracefully handle unanticipated changes of the distributed catalogues, e.g., catalogue off-line, catalogue schema changed.

Adaptive Modularity:
Ability to accommodate the addition and deletions of Distributed Catalogues from the architecture.

Clearinghouse Cost:
Cost for creating and maintaining the Main Catalogue – not including the costs of the Distributed Catalogues.

B.4 Trade Study Assessment Matrix

The following matrix presents the analysis of the Alternatives using the evaluation criteria defined in the previous clause.

	Alternative 1: Distributed Search	**Alternative 2: Harvesting**
1. User response time	Comparatively **Worse** Distributed searches require network and searches not required by harvest option.	Comparatively **Better** All searches are satisfied locally.
2. Results Ranking	Comparatively **Worse** Sorting of results from distributed catalogues can only be accomplished after all queries complete and are processed at the Clearinghouse.	Comparatively **Better** Sorting of a result set from the local cache is quick and accurate.
3. Metadata accuracy	Comparatively **Better** Metadata is retrieved from closer to the maintaining organization at the time of the user's search.	Comparatively **Worse** Metadata may be changed by the maintaining organization after the harvest.
4. Metadata ownership	Comparatively **Better** Metadata remains on the servers owned by the organization the metadata.	Comparatively **Worse** Bulk copy of metadata from a community catalogue may not be permitted by the maintaining organization.
5. Robustness	Comparatively **Worse** Changes are detected at time of search with limited response options.	Comparatively **Better** Addition of a new resource is accomplished offline from user queries.
6. Adaptive Modularity	Comparatively **Equivalent**	Comparatively **Equivalent**
7. Clearinghouse Cost	Comparatively **Better** Minimizes resources required at single Clearinghouse node.	Comparatively **Worse** Higher costs for Clearinghouse to maintain a larger store of metadata.

B.5 Trade Study Conclusions

The consensus is that the most important criteria are: 1) User Response Time and 2) Results Ranking. Therefore, the Harvest option should be selected for as many Distributed Catalogues as possible.

Given the nature of WIS's, i.e., a system of systems, there will be catalogues that cannot or will not be harvested. Where a Distributed Catalogue distinguishes between collection and granule metadata, only the collection metadata should be harvested. Some catalogues will object to being harvested, i.e., criteria 4) Metadata ownership. Therefore, the Main Catalogue should provide a distributed search functionality but its use should be minimized.

To meet this hybrid recommendation, further analysis of the Main Catalogue server interface is required. Currently there is not a widely implemented catalogue interface standard that blends a full response from the local harvested cache with stateful distributed queries. What is needed is a stateless catalogue interface for hybrid search: immediate response to user while distributed searches are continuing.

Annex C
Trade Study: Serving Metadata via WFS/CSW vs. CSW/CSW

C.1 Introduction

The current CUAHSI HIS system provides a meta-catalogue of water services. The meta-catalogue follows a CSW Dublin Core profile. The records in the CSW reference WFS services. The WFS services allow to query specifics about the time series (e.g., service via provider name, site name, variables, time frame), using OGC Filter Encoding (FE). WFS services point to SOS services which allow access to data via GetObservation. The GetObservation result returns O&M with WaterML 2.0 encoding.

This approach can be contrasted to only having an SOS as a data service endpoint for data, and a rich CSW-ebRIM for metadata cataloguing of data services which provides time series description details. The CSW-ebRIM can replace the functionality currently being provided by WFS/FE.

C.2 Trade Study Alternatives

Two alternatives are envisioned:

1) Provide a meta-catalogue CSW that contains records that describe WFS services that provide descriptions of time series metadata.
2) Provide CSW as the main metadata interface. Still data providers can provide a metadata catalogue of time series, but via CSW instead of the WFS. The catalogue metadata model may need to use an ebRIM model, following the time series meta model that the WFS is based on (Annex A).

C. 3 Trade Study Evaluation Criteria

The evaluation criteria of this trade study are shown below. The value of the criteria varies from 1 to 5, where 5 is the most important. The ease of implementation is the most important criteria with a value of 5.

Criteria	Description	Value (1-5)
Consistency accessing metadata	Ability to provide a common interface when accessing metadata. Includes metadata at different levels. For example for catalogues, for services, for data sets, for time series.	2
Querying/filtering of metadata on data provider service	Ability to query any metadata elements in the metadata server.	3
Ease of implementation	Existing tools that support the interface.	5

Criteria	Description	Value (1-5)
Support of semantic search	Ability to provide semantic searches using higher level concepts not provided by the data provider.	3

C.4 Trade Study Assessment Matrix

The table presents an analysis of using WFS and CSW vs. CSW only to provide metadata access and discovery. Approach 2 shows the CSW only approach and presents SOS as one possible service that can be registered at the CSW. Other registered services will follow the same pattern as in the case of SOS.

	Approach 1 (WFS/CSW)	Approach 2 (SOS/CSW)
1. Consistency when accessing metadata	Comparatively **Worse** Not consistent. Metadata about the time series (sites and observations) is accessed via WFS, but metadata about services is accessed via CSW. Current WFS service is tightly coupled to time series and currently supports a simple feature profile. Difficult to support more complex metadata models.	Comparatively **Better** Consistent. CSW can be used consistently at different levels (meta level, regional and data provider level) to search, discover and evaluate the different types of services.
2. Querying/filtering of metadata on data provider service	Comparatively **Equivalent** WFS via Filter Encoding (FE) can provide excellent support to filter the metadata of a data provider server.	Comparatively **Equivalent** CSW has the same support as WFS for FE.
3. Ease of implementation	Comparatively **Better** WFS is more mature than CSW-ebRIM, so more tools exist to deploy WFS servers. WFS is also much simpler than CSW-ebRIM. CUAHSI already has tools that support the WFS providing the metadata for time series services. However, the available tools support simple features which limits the expressiveness of more complex models. Another factor is that this usage of WFS is not fully consistent with established practice. That is not to say it should not be done, but it’s not the usual design pattern web developers would expect. Opinions vary on the importance of this (some developers are more resistant to this approach than others).	Comparatively **Worse** Even though CSW-ebRIM is a standard, few implementations and supporting tools exist. For example GeoNetwork, the best open source tool for catalogues available, only has an experimental ebRIM, whose code is not yet in a stable release. However, the GENESIS project in Europe has already implemented this approach. It relies on SOS instances providing SensorML encoded metadata, the SIR for converting these metadata to ebRIM, and CSW/ebRIM for making these data sets searchable. One Catalogue that is supported is the Buddata ebXML registry, and ERDAS has worked on this as well (and are editor of the related OGC Discussion Paper)

	Approach 1 (WFS/CSW)	Approach 2 (SOS/CSW)
4. Support of semantic search	Comparatively **Worse** WFS servers do not support search on higher concepts. This is currently done via keywords in the metadata model. (See Ontology Concept in Annex A.)	Comparatively **Better** CSW-ebRIM allows search on higher concepts that are bound to services registered. The higher concepts can be conceptualized in a model, which can provide support for faceted searching, instead of relying only a “keyword” field.

C.5 Trade Study Conclusions

The summary of the preceding comparisons is shown below. A maximum value for each criterion, defined in C.3, was given to the approach that was better or equivalent. The value given to the “worse” option for each criterion was a matter of judgment by the report authors. These rankings could change over the coming months as well, depending on standards up-take and further implementation experience. Given that these rankings are somewhat subjective, the two approaches seem fairly well matched at present. This is the reason that the report is presenting both architectures.

Criteria	Approach 1 (WFS/CSW)	Approach 2 (CSW/CSW)
1. Consistency when accessing metadata	0	2
2. Querying/filtering of metadata on data provider service	3	3
3. Ease of implementation	5	2
4. Support of semantic search	0	3
Total	**8**	**10**

Annex D
Data and Metadata Storage Recommendations

D.1 Persistent Data Store Component

D.1.2 Purpose

The function of the persistent data store is to store and manage hydrologic data and metadata as stated in Use Case 1 (see Table 1). Since the persistent data store is opaque to clients accessing the data from the web there is considerable latitude in how this component is implemented. Requirements for the persistent data store will likely be driven by the specific web service interface components chosen to deploy the data onto the web.

D.1.3 Content and Encoding

In some hydrologic information systems, the persistent data store for time series water observations and related metadata is done by using the Observations Data Model. Although it is represented by a single box in Figure 5, persistent storage and management of hydrologic data and metadata may be distributed across multiple data stores.

The persistent data store may be driven by temporal restrictions on the data or limitations of the procedure collecting the data. For example, data streaming from a sensor may only be persistently maintained for 24 hours, due to the sensor's limited storage capacity.

D.1.4 Interface

The component diagram in Figure 5 shows the interface between the persistent data store and the web services as being SQL. This is only an example based on the fact that most modern persistent data stores are implemented using SQL relational databases. Again the persistent store is opaque to the web and so implementers are free to choose whichever interfaces make sense for their deployment.

Revision History

Date	Release	Editor	Primary sections	Description
2011-02-10	11-013r1	L. Bermudez	Throughout	Pending doc posted for consideration at OGC TC, Bonn.
2011-03-31	11-013r3	L. Bermudez	Throughout	Last changes posted to Pending docs before OGC TC, Bonn.
2011-04-19	11-013r4	D. Arctur	Throughout	Entered comments from Hydro DWG on scope, technology
2011-07-07	11-013r5	D. Arctur	Throughout	Final report for public release.
2011-07-11	11-013r5	C. Reed	Various	Final scrub for publication

Bibliography

Bandaragoda, C. J., D. G. Tarboton and D. R. Maidment, (2005), "User Needs Assessment, Chapter 4," in Hydrologic Information System Status Report, Version 1, Edited by D. R. Maidment, p.48-87, http://www.cuahsi.org/docs/HISStatusSept15.pdf.

Clinton, William. 1994. Executive Order 12906 of April 11 Coordinating Geographic Data Acquisition and Access: The National Spatial Data Infrastructure. The White House.

Horsburgh, J. S., D. G. Tarboton, D. R. Maidment and I. Zaslavsky, (2008), "A Relational Model for Environmental and Water Resources Data," Water Resour. Res., 44: W05406, doi:10.1029/2007WR006392.

Horsburgh, J. S., D. G. Tarboton, D. R. Maidment and I. Zaslavsky, (2011), "Components of an environmental observatory information system," Computers & Geosciences, 37(2): 207-218, http://dx.doi.org/10.1016/j.cageo.2010.07.003.

ISO/IEC 10746-1 (1998), Information technology -- Open Distributed Processing -- Reference model: Overview.

ISO 19115, (2003) Geographic Information – Metadata.

ISO 19115-2 (2008) Geographic information – Metadata - Part 2: Extensions for imagery and gridded data.

ISO 19119, (2005) Geographic Information – Services.

ISO 19139, (2007) Geographic Information – Metadata – XML schema implementation.

ISO 19156, (2010) Geographic information – Observations and Measurements; cobranded in the OGC Abstract Specification [OGC doc 10-004r2].

Maidment, D. R., ed. (2005), Hydrologic Information System Status Report, Version 1, Consortium of Universities for the Advancement of Hydrologic Science, Inc, 224 p, http://his.cuahsi.org/documents/HISStatusSept15.pdf. [This includes results of the user surveys.]

O'Hagan, RG, Atkinson, R, Cox, S, Fitch, P, Lemon, D, and Walker, G (2007), "A reference model for a water resources observation network", in MODSIM 2007 International Congress on Modelling and Simulation , December 2007, Christchurch, New Zealand. Oxley, L. and Kulasiri, D., eds, Modelling and Simulation Society of Australia and New Zealand, pp. 1145-1151.

OASIS regrep-rim-3.0-os, (2005) ebXML Registry Information Model http://www.oasis-open.org/committees/download.php/13591/docs.oasis-open.orgregrepv3.0specsregrep-rim-3.0-os.pdf

OGC 03-038, (2003) Access Control System (DACS) DIPR

OGC 06-009r6, (2007) OpenGIS Sensor Observation Service http://www.opengeospatial.org/standards/sos

OGC 06-042, (2006) OpenGIS Web Map Service (WMS) Implementation Specification

OGC 06-121r3, (2007) Web Services Common Specification version 1.1.0

OGC 06-121r9, (2010) OGC Web Services Common Standard

OGC 07-006r1, (2007) OpenGIS Catalogue Services Specification V2.0.2

OGC 07-110r4, (2009) CSW-ebRIM Registry Service - Part 1: ebRIM profile of CSW

OGC 08-062r4, (2008) OGC Reference Model

OGC 08-167 , (2008) Semantic annotations in OGC standards

OGC 09-010, (2009) OGC Catalogue Services - OWL Application Profile of CSW

OGC 09-018, (2009) Web Coverage Service (WCS) 1.1 extenson for CF-netCDF 3.0 encoding (0.2.2) http://www.opengeospatial.org/standards/wcs

OGC 09-025r1 / ISO 19142:2010, (2010) OpenGIS Web Feature Service 2.0 Interface Standard http://www.opengeospatial.org/standards/wfs

OGC 09-026r1 / ISO 19143:2010 , (2010) OpenGIS Filter Encoding 2.0 Encoding Standard

OGC 09-110r3, (2010) OGC® WCS 2.0 Interface Standard - Core, http://portal.opengeospatial.org/files/?artifact_id=41437.

OGC 09-147r1, (2010) OGC® Web Coverage Service 2.0 Interface Standard - KVP Protocol Binding Extension (1.0), http://portal.opengeospatial.org/files/?artifact_id=41439.

OGC 09-148r1, (2010) OGC® Web Coverage Service 2.0 Interface Standard - XML/POST Protocol Binding Extension (1.0), http://portal.opengeospatial.org/files/?artifact_id=41440.

OGC 09-149r1, (2010) OGC® Web Coverage Service 2.0 Interface Standard - XML/SOAP Protocol Binding Extension (1.0), http://portal.opengeospatial.org/files/?artifact_id=41441.

OGC 10-025r1, (2011) Observations and Measurements XML (OMXML) Encoding Standard

OGC 10-037, (2010-08-27 draft) Sensor Observation Service (SOS) 2.0 (SOS 2.0 SWS working document, can be made available with license agreement)

OGC 10-090r2, (2011) OGC Network Common Data Form (NetCDF) Core Encoding Standard

OGC 10-100r2, (2010) Geography Markup Language simple features profile http://portal.opengeospatial.org/files/?artifact_id=39853

OGC 10-125r1, (2010) Observations and Measurements - XML Implementation http://portal.opengeospatial.org/files/?artifact_id=41510&version=1

OGC 10-126, (2010) WaterML: An O&M profile for water observation data http://portal.opengeospatial.org/files/?artifact_id=41546&version=1

OGC 10-194r3, (2011) OGC HDWG GWIE Final Report, B. Brodaric and N. Booth, editors, http://portal.opengeospatial.org/files/?artifact_id=43545&version=1

OGC 10-208, (2010, Pending Documents) SOS 2.0 RFC Comments, http://portal.opengeospatial.org/files/?artifact_id=41977&version=1

OGC 10-209, (2010, Pending Documents) SOS 2.0 - GetDataAvailability Extension RFC Comments, http://portal.opengeospatial.org/files/?artifact_id=42033&version=1

Orszag, P., (2009), "M-10-06 Open Government Directive." EXECUTIVE OFFICE OF THE PRESIDENT OFFICE OF MANAGEMENT AND BUDGET, December 8. http://www.whitehouse.gov/sites/default/files/omb/assets/memoranda_2010/m10-06.pdf.

Tarboton, D. G., J. S. Horsburgh, and D. R. Maidment (2008) CUAHSI Community Observations Data Model (ODM), Version 1.1, Design Specifications. http://his.cuahsi.org/documents/ODM1.1DesignSpecifications.pdf (This is the specification for the data model, including a relational database schema, that HIS uses to store point observations.)

Tarboton, D. G., (2005), "Review of Proposed CUAHSI Hydrologic Information System Hydrologic Observations Data Model." Utah State University. May 5, 2005. http://www.engineering.usu.edu/dtarb/HydroObsDataModelReview.pdf.