Open Geospatial Consortium

Date: 2012-02-29

Reference number of this document:

Reference URL for this document: www.opengeospatial.net/doc/{short\_doc\_name}

Version: 1.0

Category: OpenGIS® Best Practices Document

Editor(s):

OGC® CUAHSI Hydrologic Information System Implementation Specification: Part 1 Concepts and Components

Migrating the CUAHSI HIS Service Oriented Architecture to OGC standard encodings and service interfaces

To obtain additional rights of use, visit <http://www.opengeospatial.org/legal/>.

**Warning**

This is an OGC® Best Practices document, which represents an official position of the OGC membership. This document is not an OGC Standard. Recipients of this document are invited to submit, with their comments, notification of any relevant patent rights of which they are aware and to provide supporting documentation..

Document type:  OpenGIS® Best Practices Document

Document subtype: NA

Document stage:  Draft

Document language:  English

Preface:

Funding for this document was provided by the Consortium of Universities for the Advancement of Hydrologic Science, Inc. (CUAHSI) under National Science Foundation Grant No. EAR-0622374. In addition, much input and feedback has been received from the CUAHSI Hydrologic Information System development team. Their contribution is acknowledged here. The authors also wish to acknowledge the significant contributions of David Arctur (Open Geospatial Consortium).

Contents

[Foreword 5](#_Toc312041471)

[1. Summary 6](#_Toc312041472)

[1.1 The mission and key challenges of CUAHSI HIS 6](#_Toc312041473)

[1.2 Enterprise-level questions addressed by this report 7](#_Toc312041474)

[2 The migration path to OGC services stack and related services 9](#_Toc312041475)

[3. Step-by-step migration to the OGC service stack and the roadmap of this report 10](#_Toc312041476)

[3.1 Step 1: prototyping a new infrastructure and assimilating results of international validation of new OGC specifications 11](#_Toc312041477)

[3.2 Step 2: Settle on a time series catalog information model that can be relayed via common WFS or CSW implementations 12](#_Toc312041478)

[3.3 Step 3: Settle on a service interface for time series catalog; create WFS interfaces over observation networks in the HIS Central catalog 12](#_Toc312041479)

[3.4 Step 4: Settle on a CSW profile for presenting registries of water data services and make the observation networks registry in HISCentral CSW compatible. 13](#_Toc312041480)

[3.5 Step 5: Establish a distributed system of federated hydrologic catalogs, using the chosen CSW profile and a set of compatible implementations 13](#_Toc312041481)

[3.6 Step 6: Create WaterML2/SOS data access endpoints, initially for observation networks collocated with the HIS Central Metadata Catalog at SDSC 14](#_Toc312041482)

[3.7 Steps 7-10: Migrating HydroServer publishing, HIS Central harvesting and client applications 15](#_Toc312041483)

[4 Mapping CUAHSI HIS services to the OGC services stack 15](#_Toc312041484)

[5. Identifiers and respective requirement classes 17](#_Toc312041485)

[5.1 Identifiers for services: requirements 18](#_Toc312041486)

[5.2 Identifiers for observation data series 18](#_Toc312041487)

[5.3 Identifiers and references for features and base feature layers 19](#_Toc312041488)

[6. Implementation of an OGC-based Hydrologic Information System 20](#_Toc312041489)

[6.1 Components identified by the Water Information Services Concept Development Study 20](#_Toc312041490)

[6.2 Additional Components based on the HIS architecture 21](#_Toc312041491)

[6.2.1 Vocabularies and semantic management: requirements 21](#_Toc312041492)

[6.2.2 Ontology 24](#_Toc312041493)

[6.2.2.1 Variable to Ontology Mapping 25](#_Toc312041494)

[6.2.2.2 Tagging service 26](#_Toc312041495)

[6.2.3 Controlled Vocabularies 27](#_Toc312041496)

[6.2.4 Term Lists 27](#_Toc312041497)

[6.3 Water Quality Data Services 27](#_Toc312041498)

[6.4 Feature Models 28](#_Toc312041499)

[6.4 Data Distribution Formats: Hydrologic Themes 29](#_Toc312041500)

[6.5 Authorization 31](#_Toc312041501)

[6.6. Data Synchronization 32](#_Toc312041502)

[7. HIS Components 33](#_Toc312041503)

[7.1 Metadata Catalogs 34](#_Toc312041504)

[7.1.1 Service Catalog 35](#_Toc312041505)

[7.1.1.1 Federated Searching via GI-CAT 36](#_Toc312041506)

[7.1.1.2 CSW – ESRI GeoportalServer 36](#_Toc312041507)

[7.1.1.3 CSW – registration of THREDDS sources 36](#_Toc312041508)

[7.1.2 Observation Series Catalog 37](#_Toc312041509)

[7.1.2.1 WFS for Observation Data Series Profiles: Requirements 37](#_Toc312041510)

[7.1.2.2.1 WFS Implementation 38](#_Toc312041511)

[7.1.2.2 CSW for Observation Data Series Profiles: Requirements 38](#_Toc312041512)

[7.1.2.2.1 CSW Implementation 39](#_Toc312041513)

[7.1.3 How to embed Ontology in CSW 39](#_Toc312041514)

[7.2 Data Services 40](#_Toc312041515)

[7.2.1 CUAHSI Profile of OGC Services for Time Series Data 40](#_Toc312041516)

[7.2.1.1 CUAHSI Profile of SOS for time series from network of observation sites 41](#_Toc312041517)

[7.2.1.1.1 SOS Discussion 42](#_Toc312041518)

[Observation Offering: 42](#_Toc312041519)

[7.2.1.2 WFS 44](#_Toc312041520)

[7.2.1.3 WMS 44](#_Toc312041521)

[7.2.1.4 Implementation: 44](#_Toc312041522)

[7.2.2 CUAHSI profile for time series from grid cells 44](#_Toc312041523)

[7.2.2.1 SOS Discussion 45](#_Toc312041524)

[7.2.2.2.WMS 46](#_Toc312041525)

[7.2.2.3 Implementation 46](#_Toc312041526)

[7.2.3 Query Filters 47](#_Toc312041527)

[7.3 Map Services 48](#_Toc312041528)

[Base Reference layers 48](#_Toc312041529)

[7.4 Vocabulary 48](#_Toc312041530)

[7.4.1 Ontology 48](#_Toc312041531)

[7.4.1.1 Ontology Format 49](#_Toc312041532)

[7.4.1.2 Ontology Term Ranking Service Definition 50](#_Toc312041533)

[7.4.2 Vocabularies 50](#_Toc312041534)

[7.5 Water Quality Information 51](#_Toc312041535)

[7.7 Legacy service interfaces 52](#_Toc312041536)

[7.8 Functions to be fully specified at a future date 53](#_Toc312041537)

[7.8.1 Methods 53](#_Toc312041538)

[7.8.2 Themes 53](#_Toc312041539)

[7.8.1 Theme Services 54](#_Toc312041540)

[7.8.3 Accessing THREEDS Service 54](#_Toc312041541)

[8. Processing Services 55](#_Toc312041542)

[9. HIS Node Types 55](#_Toc312041543)

[10. Sustainability and governance: CUAHSI HIS in the system of earth science information 55](#_Toc312041544)

[References 59](#_Toc312041545)

[Appendix 1. HIS 1.0 to CUAHSI OGC Mapping 60](#_Toc312041546)

[Appendix 2. Observations Data Model to WaterML 2 Mapping 63](#_Toc312041547)

[Sampling Feature 63](#_Toc312041548)

[Observed Property 64](#_Toc312041549)

[Process 65](#_Toc312041550)

[Result – Timeseries 65](#_Toc312041551)

[Result – Data Values 66](#_Toc312041552)

[Result – Data Value Defaults. 67](#_Toc312041553)

[Appendix 3: Water Observations Metadata Specification 69](#_Toc312041554)

[Appendix 4: Feature Data Models 73](#_Toc312041555)

[Appendix 4.1. INSPIRE Hydrography Data Model 73](#_Toc312041556)

[Appendix 4.2. ArcHydro 75](#_Toc312041557)

[Appendix 5. Test Cases 76](#_Toc312041558)

[A 5.1 Catalog 76](#_Toc312041559)

[A 5.1.1 Harvesting 76](#_Toc312041560)

[A 5.1.1.1 Services Catalog 76](#_Toc312041561)

[A 5.1.1.1 Observational Catalog Harvesting 76](#_Toc312041562)

[A 5.1.2 Retrieval 76](#_Toc312041563)

[A 5.1.1.1 Services Catalog 76](#_Toc312041564)

[A 5.1.1.1 Observational Catalog Harvesting 76](#_Toc312041565)

[A 5.2 Observational Data for Point Services 76](#_Toc312041566)

[Sites 76](#_Toc312041567)

[Series 76](#_Toc312041568)

[Observations 76](#_Toc312041569)

[Controlled Terminology 76](#_Toc312041570)

[A 5.3 Vocabularies and Ontology 76](#_Toc312041571)

[Appendix 6: XML Examples of Filters 77](#_Toc312041572)

[CSW Filters 77](#_Toc312041573)

[WFS filters 78](#_Toc312041574)

[SOS filters 79](#_Toc312041575)

[Appendix SKOS 83](#_Toc312041576)

[Appendix: Using Azure for centralized community data hosting 84](#_Toc312041577)

[Appendix XML Examples WaterML 2: 85](#_Toc312041578)

[Appendix: WQX Example 89](#_Toc312041579)

[Appendix Examples of Inspire Metadata Records 89](#_Toc312041580)

[Service: 89](#_Toc312041581)

[Series 96](#_Toc312041582)

Figures

[Figure 1. The HIS design follows the standard SOA "publish-find-bind" pattern, and integrates providers of data, catalog and processing services, and data synthesis and research clients. 12](file:///C:\Users\valentin\Documents\cuahsi\Implementation\CuahsiImplementaton_17.docx#_Toc318972329)

[Figure 2. An HIS client application developed at UT-Austin to demonstarte CSW and WFS services for water data 17](file:///C:\Users\valentin\Documents\cuahsi\Implementation\CuahsiImplementaton_17.docx#_Toc318972330)

[Figure 3. Application developed by Kisters as part of SWIE, Use Case 3, to demonstrate access to water information via WaterML 2.0/SOS services, and WFS. 18](file:///C:\Users\valentin\Documents\cuahsi\Implementation\CuahsiImplementaton_17.docx#_Toc318972331)

[Figure 4. Adding a WFS interface to networks registered in the HISCentral catalog 19](file:///C:\Users\valentin\Documents\cuahsi\Implementation\CuahsiImplementaton_17.docx#_Toc318972332)

[Figure 5. Federation of CUAHSI HIS service catalogs 20](file:///C:\Users\valentin\Documents\cuahsi\Implementation\CuahsiImplementaton_17.docx#_Toc318972333)

[Figure 6. Interaction diagram for OGC service calls. 24](#_Toc318972334)

[Figure 7. Ontology Tables of HIS Central 34](#_Toc318972335)

[Figure 8. Layered organization of parameter ontology in CUAHSI HIS (source: M. Piasecki). The detail Layer is comprised of variable names used by different data publishers. These variable names are associated with concepts managed within the Leaf Layer. The Compound Layer organizes leaf and core concepts into a navigable hierarchy. Search keywords are compiled from concepts in the Navigation, Compound and Core/Leaf layers. 35](#_Toc318972336)

[Figure 9. UML of WaterML2 MonitoringPoint 38](#_Toc318972337)

[Figure 10. XML fragment of representing a WaterObservationPoint 39](#_Toc318972338)

[Figure 11. Aggregating data from multiple organizations into hydrologic themes 40](#_Toc318972339)

[Figure 12. Mapping of themes using WFS 40](#_Toc318972340)

[Figure 13. . Azure Access Control System 41](file:///C:\Users\valentin\Documents\cuahsi\Implementation\CuahsiImplementaton_17.docx#_Toc318972341)

[Figure 15. Synchronizing observation data series from multiple formats to a staging repository and then to an HIS central catalog 43](#_Toc318972342)

[Figure 16. Components of the HIS system 44](#_Toc318972343)

[Figure 17. Metadata components, and harvesting of records from a WFS and a SOS service. 45](#_Toc318972344)

[Figure 18. Federated searching using GI-CAT 46](#_Toc318972345)

[Figure 19. CUAHSI OGC Compliant Data Server 53](#_Toc318972346)

[Figure 20. XML fragment MeasurmentTVP (time Value Pair) with a related observation 66](#_Toc318972347)

[Figure 21. Converting variable description to variable properties 69](#_Toc318972348)

[Figure 22. Vision of a reference architecture for EarthCube as is an integrated information system that includes research observatories generating large volumes of observations, domain systems (such as HIS) that publish the data according to community conventions about data models, vocabularies and protocols, and cross-domain knowledge layer that includes federated catalogs, normalized and curated datasets integrating data from domain systems and observatories, vocabulary cross-walks, as well as social networking, governance and compute infrastructure. 74](file:///C:\Users\valentin\Documents\cuahsi\Implementation\CuahsiImplementaton_17.docx#_Toc318972349)

Tables

[Table 1. Mapping from WaterOneFlow to OGC standards 22](#_Toc318972361)

[Table 2. HIS Central method mappings to OGC services 23](#_Toc318972362)

[Table 3. Components identified by the Water Concept Study. 28](#_Toc318972363)

[Table 4. Additional Components from the HIS architecture 29](#_Toc318972364)

[Table 5. Examples and best practice for WaterML2 55](#_Toc318972365)

[Table 6. GetCapabilities document - SOS Server Type A (procedure==sensor-type) (from Fuest, 2011) 56](#_Toc318972366)

[Table 7. Filters and applicability 61](#_Toc318972367)

[Table 8. Uniform Resource Locators (URL) for the Water Quality web services based on the USGS standard 67](#_Toc318972368)

[Table 9. CUAHSI Service Supporting Legacy WaterOneFlow Interface 68](#_Toc318972369)

[Table 10 Observations Metadata: Core Fields 81](#_Toc318972370)

# Introduction

## Scope

This document presents our vision for the transition from the initial CUAHSI HIS standards to the OGC specifications, as well as specific step-wise plans for this transition. It provides a detailed specification of CUAHSI HIS cyberinfrastructure to be implemented, including its technical and governance aspects.

This document follows the earlier “Water Information Services Concept Development Study (2011),” which outlined the CUAHSI HIS architecture and general patterns of mapping this architecture to OGC standards. The present document focuses on how this architecture is being implemented as the system is migrating to OGC reference model and service stack. It is expected that this document will be used by implementers who have participated in the earlier phase of CUAHSI HIS development, including decision-makers and technical personnel from federal and state government agencies, academic research centers and universities, and commercial companies, assisting them with the transition to the new set of services and other middleware components.

## Document contributor contact points

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

|  |  |
| --- | --- |
| Name | Organization |
| David Valentine | University of California, San Diego |
| Ilya Zaslavsky | University of California, San Diego |
|  |  |

## Revision history

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Date | Release | Editor | Primary clauses modified | Description |
| 2011-09 | 0.9 | Valentine |  | Moved WaterML 2 details to separate document |
| 2012-03 | 0.91 | Valentine | All. | Formatting to OGC template |
|  |  |  |  |  |

## Future work

Improvements in this document are desirable to

* Gridded Data
* Implementation
* Future items
  + Method
  + Themes
* Developing test datasets and validation tools

## Forward

CUAHSI Hydrologic Information System Project has developed services-based infrastructure for publishing, cataloguing, discovering and accessing hydrologic observations from multiple distributed repositories. The backbone of the HIS service-oriented architecture design is a set of standard web service APIs that define interactions between hydrologic data publication platform (HydroServer), the data cataloguing and discovery system (HydroCatalog) and client applications, such as HydroDesktop. The key standards used in the current operational version of HIS are WaterML 1.x and WaterOneFlow services. These specifications have been designed to unify hydrologic data discovery and access heterogeneous sources of hydrologic observations: academic data sources that store data in CUAHSI Observations Data Model (ODM) and large federal and state water data repositories (e.g. maintained by USGS, EPA, NCDC) that follow their own storage, metadata and access conventions. To establish a higher level of compatibility between a wider group of water data sources, at the national and international scales, and to take advantage of multiple third-party software applications, the CUAHSI HIS Service Oriented Architecture (SOA) is now evolving to ensure that the key interfaces are compatible with OGC standards. Another advantage of this transition is that Open Geospatial Consortium provides transparent and community-accepted procedures and protocols for governing standards development. In particular, OGC has assembled an international group of experts in standards for water data and related fields (as the OGC/WMO Hydrology Domain Working Group) with the mission to examine existing standards, develop standardization priorities, coordinate development of specifications, organize their testing in a series of interoperability experiments, and lead the standards to community adoption.

The first part of the document focuses on the main concepts and components of CUAHSI HIS as a service-compliant hydrologic information system, and the transition from the current system to the new OGC-compliant interfaces. Part II of the document describes the main workflows needed to keep the new system operational.

Comments on this document are welcome (as track changes submitted to the authors).

# References

# Terms and Definitions

# Summary

## The mission and key challenges of CUAHSI HIS

The mission of the CUAHSI Hydrologic Information System Project is to create a scalable, flexible and evolving data integration infrastructure for hydrologic information, that would become a foundation for addressing hydrologic research and education challenges for the next decade. The key challenges are as follows:

* As massive volumes of hydrologic data become available it is important to develop sophisticated data integration strategies and architectures enabling re-purposing and re-using observations, as hydrologic science is embracing larger-scale analysis and modeling made over longer time series.
* Distributed hydrologic data should be easily discoverable; yet they are currently described using different metadata and semantic conventions, thus making hydrologic data discovery and interpretation a complicated and time-consuming process which is often not easy to replicate. Unified hydrologic data discovery interfaces with community-accepted semantics are needed to .address this challenge.
* Data are structurally and semantically heterogeneous, follow different spatial and temporal sampling patterns, and undergo different types of processing before they become available, which complicates information integration across hydrologic data sources.
* Hydrologic science is addressing problems that could not be addressed before: regional and continental scale modeling, global climate change effects, large-scale disaster response. Besides data integration enabling such larger scale research, this requires clarification of the roles and responsibilities of all system stakeholders, including government and academic monitoring and research activities.
* Making hydrologic data publishing easy with standards-compliant mainstream software – this is critical for engaging wider groups of hydrologists and data collectors in hydrologic data sharing, achieving higher levels of data availability and, therefore, higher accuracy, predictive and explanatory value, and impact of hydrologic analysis and modeling.
* Integrating government and academic observation sources is needed for creating a more comprehensive hydrologic portrait at different spatial scales, and for better understanding of hydrologic processes in the water cycle. These groups play different roles in hydrologic research and monitoring, collect observational data for different purposes, and adhere to different data retention strategies, QA/QC procedures, and other data management lifecycle components.
* Integrating data across earth science research domains is an important challenge as cross-domain distributed information systems are being developed - such as those envisioned in the new NSF EarthCube initiative. In such systems, data from different subdomains of the earth sciences, managed by independently managed and independently evolving information systems, need to be integrated in comprehensive analysis and modeling of physical processes under a variety of research scenarios.
* Integrating data across different national systems and languages: this issue becomes important as larger-scale cross-border data integration, analysis, modeling and water resource management projects become a reality.

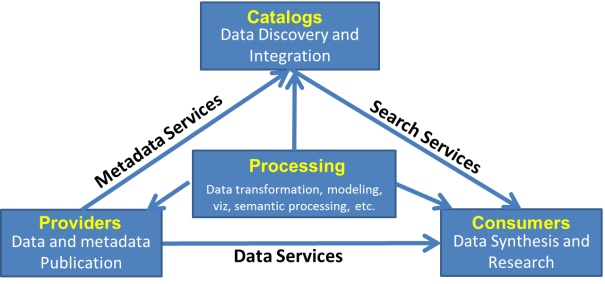


Figure 1. The HIS design follows the standard SOA "publish-find-bind" pattern, and integrates providers of data, catalog and processing services, and data synthesis and research clients.

Figure 1 describes the high-level logical organization of CUAHSI HIS, which provides the context for this study. The earlier “Water Information Services Concept Development Study” outlined the new architecture by following the collection of OGC reference model (ORM) viewpoints: enterprise, information model, computational and engineering. It also provided a description of key meta-use-cases supported by HIS: (1) Publication: publish data and their associated metadata using web services; (2) Cataloguing: aggregate metadata from published web services; (3) Discovery: search a catalog to identify data series or datasets of interest, (4) Access: acquire the selected data, and (5) Processing: generate aggregate measures or derived products, support cross-dataset query and mediation, and processing functions associated with other use cases.

This document follows a similar path, starting with enterprise-level questions that we identified based on the previous report. It then continues to explore information and services aspects of the new model, and describes a migration path from the current model to the new services stack.

## Enterprise-level questions addressed by this report

The experience operating CUAHSI HIS to date, and additional requirements expressed by our academic and government partners, lead to the following questions that drive this report:

1. What is the **migration path** from the current CUAHSI HIS service stack to OGC service stack? To answer this question, we focus on:
2. mapping CUAHSI HIS architecture to the OGC reference model;
3. mapping of information models (for features, time series, catalogs) to OGC encoding specifications described in the Information Viewpoint as presented in the previous report;
4. mapping of data access, metadata, catalog and processing services to OGC service interface specifications described in the Computational Viewpoint;
5. mapping of specific technologies used in CUAHSI HIS to technologies implementing OGC-RM as described in the Engineering Viewpoint;
6. exploring potential re-implementation of CUAHSI HIS components that are not explicitly covered by OGC reference model:
   1. managing semantic descriptions and controlled vocabularies;
   2. establishing and curating a uniform semantic framework based on community input;
   3. validating semantic compliance, and a related governance model;

We take the position that development of standards and best practices for managing semantic information is outside the OGC standardization process purview, as other standards bodies have been active in this area. However, management of controlled vocabularies and semantics is an inherent component of HIS, therefore migration of semantics management components to a standards-based encoding and services is within the scope of this document.

The CUAHSI HIS experience highlighted several key questions that have not been explicitly addressed by other studies. In particular, the interoperability experiments conducted by the OGC/WMO Hydrology Domain Working Group have focused on refining WaterML 2.0 encoding in the context of several data access scenarios, while not specifically focusing on the issues of data discovery and cataloguing, or governance aspects of information management and integration. Therefore, the additional specific questions we address include:

1. **Modeling of time series catalogs**, and their interaction with service registries. While OGC offers specifications for catalogs of services and datasets, these specifications may not be mature or specific for catalogs of observation time series. CUAHSI HIS experience points to modeling of time series catalogs at the information, computation and engineering levels as one of the key issues in the design of a hydrologic information system.
2. **Federation of catalogs and service registries** is an important component of the system as we’ve witnessed a growing demand from independent water data publishers to expose their catalogs in a compatible manner. This is complicated due to the multitude of profiles of catalog services (CSW, in particular) and their implementations. One of our aims is to define an appropriate CSW profiles and services to support federation of hydrologic catalogs for the use scenarios explored in CUAHSI HIS.
3. **Search patterns**, as enabled by the system of federated services and time series catalogs, specifically:
4. How OGC query interfaces express distributed discovery of hydrologic data, using semantic, spatial and temporal filters for services and time series?
5. What are pros and cons of different architecture patterns enabling data discovery and retrieval?
6. **Operational governance model,** which is required for maintaining the integrated water data system and the transition to new interfaces:
7. What are the key roles in the operational system, and how they change with the transition to the standards-based model?
8. What the operational management model for key HIS components, including: governance of encodings and service signatures; governance and curation of community vocabularies and ontologies; community management of mappings between semantic concepts and domain of values; procedures for catalog curation; system verification (against pre-defined verification use cases), validation (against science goals), and system monitoring?
9. What are the best practices for publishing hydrologic data and catalogs, and enabling community-supported and managed applications /implementations clearinghouse?

Governance model is a critical component for managing an operational hydrologic information system, yet it has received relatively little attention in our previous work, beyond setting data use and data publishing agreements, and requesting citations when HIS data are used. By now, the project has accumulated extensive experience collaborating with data and other service providers and with hydrologic data users, to start formulating governance principles. To be successful, the governance model must follow essential patterns of community organization with respect to:

* Data management responsibilities as established in the community and required by government regulations (for agencies) and academic rigor standards as expressed, for example, in NSF data management policies (for academic partners);
* Supporting local researchers in establishing and maintaining their observation networks, including identities of measurement sites, variables, methods, procedures, etc., while providing aggregation, validation and semantic mapping services to enable cross-network search and data integration;
* Enabling service providers and consumers to join the system with minimal initial investment and skill set, by using common standards-compliant software.

In addition, the scope of governance shall be clearly defined, with the expectation that governance of water data exchange standards shall belong to independent international standards bodies working in conjunction with projects like HIS to establish hydrology-related use cases and verify and validate proposed standards against such use cases.

This report is organized by the five groups of questions above. It starts with the step-by-step migration path (Question 1), and then focuses on several unique aspects of the transition as described in Questions 2-5.

# The migration path to OGC services stack and related services

The current organization of CUAHSI HIS has been described in several earlier reports (his.cuahsi.org/publications), so we are not repeating it here. At a high level, migrating these components to the new services stack, ensuring a smooth transition of the operational system, will include 10 steps:

Step 1: Prototype a new infrastructure and assimilate results of international validation of new OGC specifications.

Step 2: Settle on a time series catalog information model that can be relayed via common WFS or CSW implementations.

Step 3: Settle on a service interface for time series catalog (as described below, WFS has been chosen for this role); create WFS interfaces over observation networks in the HIS Central catalog, integrated with HIS Central administration interface.

Step 4: Settle on a CSW profile for presenting registries of water data services and make the observation networks registry in HISCentral CSW compatible.

Step 5: Establish a distributed system of federated hydrologic catalogs, using the chosen CSW profile and a set of compatible implementations.

Step 6: Create WaterML2/SOS data access endpoints, initially for observation networks collocated with the HIS Central Metadata Catalog at SDSC.

Step 7: Integrate the WaterML2/SOS endpoints in the HydroServer software stack, to provide an additional OGC-compatible data access interface to the HIS core data publication platform.

Step 8: Integrate WFS-based series catalog in HydroServer software stack, to enable data discovery at the HIS publication platform.

Step 9: Update HISCentral harvesting routines to utilize WFS Observation Data Series services, migrating it from relying on WaterOneFlow-based harvesting to this component of the OGC stack.

Step 10: Update the CUAHSI HydroDesktop client to interact with CSW and WFS services, as an example for other OGC-compatible client development for water observations.

In addition, this report addresses the emerging components of CUAHSI HIS, which include:

* Managing *themes*, as thematically organized named collections of time series (e.g. “salinity”, dissolved oxygen), which are curated by CUAHSI HIS, and contain sufficient provenance information for the included QA/QC-ed data to ensure trust: including theme generation and curation workflows, their registration, discovery and access;
* Workflows for monitoring and reporting usage and service performance, and alerting data administrators and the user community when data sources become available as a result of publication/approval workflows, or become unavailable as a result of data deprecation or violation of predefined service availability thresholds;
* Managing community process for vocabulary and ontology updates and curation;
* Migration to the cloud-based service provisioning model;
* Authentication and authorization for HIS services;
* Integration of CUAHSI HIS in a larger information system for the earth sciences such as EarthCube

# Step-by-step migration to the OGC service stack and the roadmap of this report

At the time of writing, we can already report on several migration steps outlined above, and present a plan for transitioning other components. The steps that have been accomplished, and the conclusions we reached in this process, are summarized below.

## Step 1: prototyping a new infrastructure and assimilating results of international validation of new OGC specifications

Figure 2 presents a snapshot of a client application built at UT-Austin to demonstrate how water information can be queried and accessed via OGC CSW and WFS interfaces. The application demonstrated how these OGC service interfaces can support hydrologic data search using the Who (data service providers) – What (variables) - Where (locations) search pattern. Here, the CSW interface enables federation of catalog services, while WFS is used to expose time series catalogs, as described below. Figure 3 shows a WISKI-based client developed by Kisters as part of the OGC Surface Water Interoperability Experiment. This client demonstrated access to WFS (for locations of sampling features) and SOS (for observational data encoded in WaterML 2.0), developed as part of global river discharge study (Use Case 3 of the SWIE.) The Groundwater (2009-2010) and Surface Water (2010-2011) Interoperability Experiments of the OGC/WMO Hydrology Domain Working Group have demonstrated serving water data encoded in WaterML2 using SOS1 and SOS2 services; their results are available in respective OGC reports.

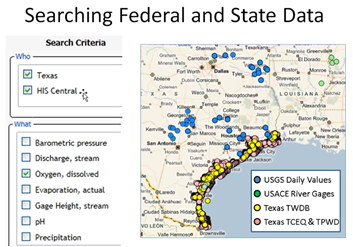


Figure 2. An HIS client application developed at UT-Austin to demonstarte CSW and WFS services for water data

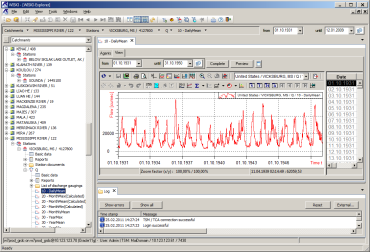


Figure 3. Application developed by Kisters as part of SWIE, Use Case 3, to demonstrate access to water information via WaterML 2.0/SOS services, and WFS.

The key conclusion from this analysis and prototyping step is that OGC specifications provide a sufficient foundation for implementing HIS functionality in a standards-compliant manner. Yet, additional work is needed to profile the respective service interfaces and information models to represent hydrologic data in unambiguous and easily interpretable manner, and meet requirements of hydrologic data discovery, access and processing use cases.

## Step 2: Settle on a time series catalog information model that can be relayed via common WFS or CSW implementations

This information model has been extensively discussed within the CUAHSI HIS team, and with USGS partners. It includes fields presented in Appendix 3, while a detailed discussion is in a section focused on the metadata catalogs. The key conclusion from this step of the study is that representing a time series catalog as a flat table is sufficient for a set of data discovery use cases supported by HIS.

## Step 3: Settle on a service interface for time series catalog; create WFS interfaces over observation networks in the HIS Central catalog

Based on the results of the previous step, the CUAHSI HIS project settled on using WFS as implemented in common COTS software to express the flat table time series metadata catalog structure. Initially, the additional WFS interfaces are developed alongside the HIS Central service interfaces for networks registered in the central metadata catalog. For example, a registered observation network with id = 52 (the Little Bear River water data service) is now additionally exposed as a WFS service (Figure 4). One of the key conclusions from this step is that the current system can be updated to include additional service interfaces without changes in the underlying data management infrastructure.

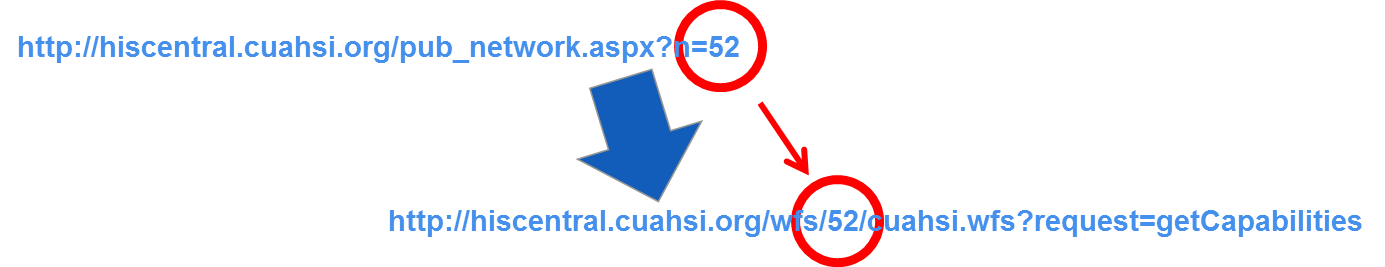


Figure 4. Adding a WFS interface to networks registered in the HISCentral catalog

Another key conclusion, with respect to exposing time series catalogs in a hydrologic information system, is to rely on simple WFS implementations, which are better supported in COTS software, while handling semantics via separate vocabulary services. Eventually, as the CSW/ebRIM model attains similarly mature implementations, the system may transition to additionally support this service interface.

## Step 4: Settle on a CSW profile for presenting registries of water data services and make the observation networks registry in HISCentral CSW compatible.

In the proposed system, we are supporting search and access of hydrologic data via a three-step process. At the first level, web services are discovered based on the spatial extent of the observation network they represent, and the names of variables they expose. At the second level, each of the found water data services is queried to discover time series that satisfy attribute, spatial and temporal constraints (which can be evaluated over the time series catalog information model mentioned above). The found data series can be used to formulate SOS GetObservation requests, in a way that is equivalent to WaterOneFlow GetValues calls. We concluded that, to support the first-level search against the service registry, it is sufficient to represent them as CSW/Dublin Core service profile.

## Step 5: Establish a distributed system of federated hydrologic catalogs, using the chosen CSW profile and a set of compatible implementations

Relying on the CSW/Dublin Core profile as a common denominator makes it easier to integrate different implementations of the service catalogs in a federated system of catalogs, relying on relatively mature service implementations. Our goal is to develop a hierarchical federated catalog system as shown in Figure 5 (more details and discussion in a section focused on federated catalogs later in the report.) In this model, the central services catalog (at the CUAHSI Office) integrates records from all subordinate catalogs by periodically harvesting them into the central registry. We use a combination of harvesting and distributed search to support efficient search patterns. Harvesting is a better model for situations when a quicker response to search requests is required, and results need to be globally ranked by relevance. At the same time, a harvesting strategy may result in metadata inconsistency if service metadata has changed between the harvests (a relatively low chance when the search is at the level of services). Distributed search is better in that latter respect but is lacking in response time and results presentation performance (e.g., all distributed search queries must complete before sorting by relevance can occur). In CUAHSI HIS, we evaluate the likelihood of service metadata changes for each catalog provider, to decide which strategy is appropriate for each local catalog reporting to the central CSW catalog. As a general rule, our experience points to distributed search as a better strategy for large federated catalogs exposed for federated repositories, while harvesting would be a more useful strategy for smaller academic services -but it shall be decided on a case by case basis. The implementation solution adopted by CUASHI HIS is take advantage of different capabilities of CSW implementations, to support indexing of both time series services and grid data services (e.g. THREDDS). As currently implemented, the CUAHSI Central Office maintains both Gi-CAT and ESRI HydroPortal CSW nodes, to federate catalogs from multiple organizations. The SDSC HydroPortal federates time series service catalogs and indexes academic WFS time series metadata services.

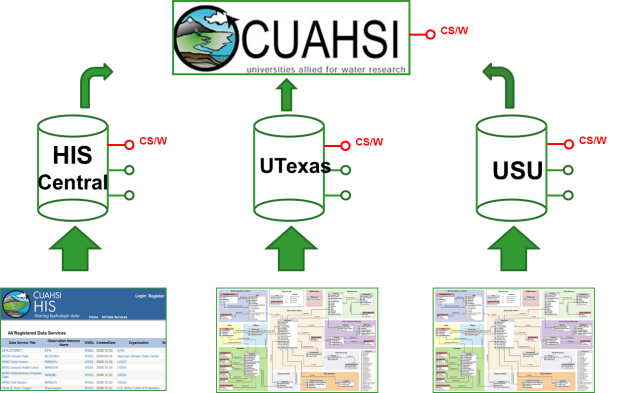


Figure 5. Federation of CUAHSI HIS service catalogs

Another key conclusion from steps 3.4-3.5 is that the 3-level data discovery and access (initial semantics-based and location-based discovery over integrated service registry at the CUAHSI Office, followed by time series discovery over the appropriate WFS services only, followed by data access and retrieval for the identified time series) is efficient over distributed water data services as it accommodates both frequently and less frequently updated catalogs, and provides enough flexibility to tune it catalog update patterns.

## Step 6: Create WaterML2/SOS data access endpoints, initially for observation networks collocated with the HIS Central Metadata Catalog at SDSC

This is a key migration step that involves mapping of WaterML 2/SOS requests to WaterOneFlow/WaterML 1 requests and software stack. This mapping is described in detail in the following sections, and a mapping table is provided in the Appendix.

## Steps 7-10: Migrating HydroServer publishing, HIS Central harvesting and client applications

Steps 7-10 of the migration are currently being implemented by the CUAHSI HIS project team. Once the new interfaces are implemented, approbated and tested at the central site at SDSC, they will be included in HydroServer software stack, so that HydroServers publish observation data in SOS, expose the hydrologic series catalog as WFS, and possibly expose water data service registries as CSW (along with other services published through HydroServers). The option of exposing HydroServer content via CSW rests with data publishers: typically CSW will be used if HydroServers provide access to large volumes of heterogeneous data, as opposed to relatively limited ODM-based deployments. As discussed in the previous Water Data Services report, we also expect the CSW publishing model to be attractive for agencies publishing hydrologic observations. The HIS Central harvesting will then use the standard service interfaces to retrieve data series metadata from the servers, compile a metadata catalog and make it available to applications such as HydroDesktop. The work on steps 7-10 is ongoing.

# Mapping CUAHSI HIS services to the OGC services stack

The WaterOneFlow SOAP API developed by the CUAHSI HIS project, has four basic methods: *GetSites*, *GetVariableInfo*, *GetSiteInfo* and *GetValues.*  Responses to WaterOneFlow requests follow WaterML 1.x, an XML schema based on the Observations Data Model (ODM) as well as on observation metadata published by several US federal agency repositories, in particular the USGS National Water Information System. As part of the migration strategy, we are proposing a mapping of WaterOneFlow methods to OGC standards (Table 1). To implement the HIS model using OGC encodings and service interfaces, several standards will be used, and the information will be split across several services. HIS organizes observational data by series, which is the central concept and the key managed object in the system. HIS uses a centralized catalog for discovery of observation data series (HISCentral), and a managed vocabulary system. Series information from registered WaterOneFlow services is harvested into the central catalog. The basic HIS Central API and HIS Controlled Vocabularies are mapped to OGC services as shown in Table 2. In addition to this mapping, semantic mapping occurs within the HIS Central System, where variable names are associated with concepts of the HIS parameter hierarchy. Semantic mapping and provision of ontology services is generally outside the scope of OGC, and thus can be placed inside/behind components of the OGC service stack.

Table 1. Mapping from WaterOneFlow to OGC standards

| Observations Data Model | WaterOneFlow | WFS | WFS | CSW | SOS |
| --- | --- | --- | --- | --- | --- |
| (All WaterML 1.x) | (DataCart) |
| Location/Site | GetSites (0..n) | GetFeature (0..n) | GetFeature (returns Simple GML) | GetRecords (returns Service Metadata Records) | (HDWG Best Practice: use WFS) |
| (returns GML /WaterObservationPoint) | GetFeatureOfInterest (optional ) |
|  | DescribeSensor(m) |
| Variable | GetVariables | GetCapabilities  (as keywords) | GetCapabilities (as keywords) |  | GetObservedProperty |
| Series | GetSiteInfo |  | GetFeature (returns Simple GML) | GetRecords | GetCababilities /offering |
| (returns Custom Record) | GetData Availability (optional extension) |
| DataValues/ Observations/ Results | GetValues |  | (records contain Pointer) | (records contain Pointer) | GetObservation |
| (returns WaterML 2) |
| (records contain pointer to Feature) |

Table 2. HIS Central method mappings to OGC services

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Method | WFS | SOS | CSW | OGC Common |
| Service Information | GetCapabilities (1) | GetCapabilties (1) | GetRecords (0-n) |  |
| GetSeries (with filters) | GetFeatures  (if feature == series) | GetCapabilities/Offerings (1-n) |  |  |
| GetDataAvailability (0-n) |
| HIS Controlled Vocabularies | GetCapbilities/Keywords |  |  | GetResourceByID (optional) (other resources) - |
| not an SOS or WFS service. Request defined in Web Services Common |

The interaction of clients with the OGC services is shown in Figure 6. A client queries the Catalog Services for the Web for service metadata (1). The client then parses the returned service records, and queries the Web Feature Services that contain information on Observation Data Series with a set of query filters (2). The client can then display the Observation Data Series on a map or in a tabular form. The user than selects the Observation Data Series of interest, and queries the services (4). This client will receive WaterML 2 timeseries (5). In these timeseries features are referenced (in WaterML2, water observation points are referred to as features of interest). Information on these features is retrieved from the Web Feature Service (5). The information from the WaterML 2 timeseries and features can be utilized by the client, and stored in the clients’ data model (6). The above represents a generic data discovery and access workflow in HIS as mapped to the OGC reference model and the emerging encodings in the hydrology domain.



Figure 6. Interaction diagram for OGC service calls.

# Identifiers and respective requirement classes

Managing persistent unique identifiers is essential for the distributed hydrologic information system. If information is to be harvested, synchronized and shared, then each system must maintain objects’ identity. We have seen issues with identifiers not being maintained or not being unique within a Hydroserver: this creates complications for harvesting, usage reporting and data integration. In addition, when CSW records managed by a Geoportal Server are harvested by a GI-Cat server, the identifiers were not maintained in some versions of GI-Cat, making us use titles and identifier codes combined as a proxy of unique persistent identifiers. Since it is simple to register a service into several Geoportal installations, multiple records for the same service can exist. For records in a CSW service, where harvesting is a component, service metadata record identifiers need to be maintained, to ensure that services are referenced in a distinct and persistent manner.

## Identifiers for services: requirements

1. Distinct identifiers must be generated for each service metadata record
2. Identifiers must be maintained when harvesting
3. Ideally, only one record must exist in the metadata catalog for a service. We should recognize that this may not always be the case, as the same service instance, or several versions of a given service, may be registered in one or more independently maintained but federated service catalogs.

|  |
| --- |
| Requirements Class |

|  |
| --- |
| his.cuahsi.org/spec/services/1.0/req/services/metadata |

|  |  |
| --- | --- |
| Target Type | Identifier |

|  |  |
| --- | --- |
| Name | Identifiers for services |
| Dependency |  |
| Requirement | his.cuahsi.org/spec/services/1.0/req/services/metadata/identifiers  Distinct identifiers must be generated for each service metadata. |
|  | his.cuahsi.org/spec/services/1.0/req/services/harvest/identifiers  Identifiers must be maintained when harvesting. |
|  | his.cuahsi.org/spec/services/1.0/req/services/catalog/records  Only one record must exist in the metadata catalog for a service. We should recognize that this may not always be the case |

## Identifiers for observation data series

The notion that observation data series have distinct IDs is new to HIS, which to date has relied on a standard concatenation of time series characteristics (site, variable, source, quality control level, method) as a proxy of persistent identifier. However, the experience shows that independently generated persistent series IDs are important, as we want to make the system flexible and extendable to scenarios when series are defined by their publishers in ways different than how it is prescribed in ODM. For example, within the CZO system data publishers can include additional attributes in defining a series (e.g. vertical offsets). For many federal data repositories a flexible definition of a series, and referencing it by a persistent identifies, is also a preferred approach, as series definitions may include additional site and variable qualifiers, while omitting attributes such as the QA/QC level. Therefore, we pose additional requirements for observation data series identifiers as follows:

|  |
| --- |
| Requirements Class |

|  |
| --- |
| his.cuahsi.org/spec/services/1.0/req/observations/metadata |

|  |  |
| --- | --- |
| Target Type | Observation data series |

|  |  |
| --- | --- |
| Name | Identifiers for observation data series |
| Dependency |  |
| Requirement | his.cuahsi.org/spec/services/1.0/req/services/observation/metadata/identifiers  Distinct identifiers must be generated for each observation data series record. |
|  | his.cuahsi.org/spec/services/1.0/req/services/observation/metadata/identifiers/stability  We need to recognize the organizations change identifiers for organizational codes, site codes, variable codes, and that the observation data series identifier should be independent of these codes. |
|  | his.cuahsi.org/spec/services/1.0/req/services/observation/metadata/identifiers/count  Only one record must exist in the catalog for an observation data series |
|  | his.cuahsi.org/spec/services/1.0/req/services/observation/harvest/identifiers  Identifiers must be maintained when harvesting or synchronizing data**.** |
|  | his.cuahsi.org/spec/services/1.0/req/services/observation/harvest/identifier/creations  Centralized harvesting services may create and manage identifiers for data sources that do not provide distinct identifiers for data series (e.g. ODM). |

## Identifiers and references for features and base feature layers

In the OGC standard, Observations and Measurements, a timeseries is related to a feature of interest. In present HIS functionality; this is a point location that is stored in an ODM database. In the future, this may become a reference to a standard hydrologic feature vocabulary, e.g. the WMO INSPIRE Hydrologic Features (Appendix 4). For the US, there are several such base datasets that can be referenced from feature descriptions, e.g. the EPA Reach files, and the Hydrologic Units identified by Hydrologic Unit Codes (HUC). Multiple copies of GIS based datasets will be available. The HIS System will need to create identifiers for common base datasets such as hydrologic units, river reaches and other datasets, and make persistent and curated feature dictionaries available via standard services

|  |
| --- |
| Requirements Class |

|  |
| --- |
| his.cuahsi.org/spec/services/1.0/req/features/metadata |

|  |  |
| --- | --- |
| Target Type | Identifier |

|  |  |
| --- | --- |
| Name | Identifiers for features |
| Dependency |  |
| Requirement | his.cuahsi.org/spec/services/1.0/req/services/features/metadata/identifiers  Distinct identifiers must be generated based on the original data source (us:hus:000000), and not the present service location/wfs/getfeature?id=0000. |
|  | his.cuahsi.org/spec/services/1.0/req/services/features/basemap/metadata/identifiers  Basemap feaures from different WFS servers should utilize a common identifitier schema. should be available via OGC WFS 1.1, |

# Implementation of an OGC-based Hydrologic Information System

## Components identified by the Water Information Services Concept Development Study

The Water Information Services Concept Development Study identified system components for data publication, cataloging, discovery and access, at the information, computational and engineering levels, which are summarized in Table 3. They will be discussed in details in the components section.

Table 3. Components identified by the Water Concept Study.

|  |  |  |  |
| --- | --- | --- | --- |
| **Component Type** | **Subcomponent Type** | **Computational Viewpoint Technology** | **Information Viewpoint Technology** |
| **(Data Model Encoding)** |  | **(Web Service)** |  |
| 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 | ISO 19139 |

## Additional Components based on the HIS architecture

There are additional components of HIS that are not detailed in the Water Information Services Concept Development Report, or do not have matching services developed as OGC specifications. These components are summarized below (Table 4). These additional components deal with managing hydrologic vocabularies, authentication/authorization, service approval/curation workflows, and service and usage monitoring. The major addition is a component for creating packaged datasets called “themes”, which represent named, thematically organized, normalized and curated collections of hydrologic data series compiled from multiple sources for a given area of interest (e.g. “salinity theme for Texas”.)

Table 4. Additional Components from the HIS architecture

|  |  |  |  |
| --- | --- | --- | --- |
| **Component Type** | **Subcomponent Type** | **Computational Viewpoint Technology (Web Service)** | **Information Viewpoint Technology** |
| Controlled terminology | For managing and exposing parameter ontology | Datastore | Table |
| Controlled terminology | For managing and exposing parameter ontology |  | SKOS |
| Controlled terminology | For exposing common vocabulary | SOAP/REST | SKOS |
| Data Server | Themes- Compiled datasets | Storage | Hydrodesktop |
| Data Server | Themes- Display on map | WMS/WFS | Image/GML |
| Data Server | Water quality | REST | WQX |
| Processing | For building themes | WPS |  |
| Authentication | Managing access to services | OAuth | OAuth |
| Workflow | Approval of new datasource |  |  |
| Workflow | Curating of a theme |  |  |
| Monitoring | Monitoring of the registered data services |  |  |

### Vocabularies and semantic management: requirements

Different types of vocabularies are used throughout the system. At present, the terminology is not exposed in a consistent manner. While the HIS stack has a controlled vocabulary (CV) system, the CV is underutilized, not compliant with terminology or ontology standards. Controlled vocabularies and parameter terminology are not exposed at the individual data services level. As a result, CV compliance enforcement requires that the centrally-managed CV system is updated before a new term can be used, which leads to publication delays. A centralized hierarchically organized vocabulary of hydrologic parameters (sometimes referred to as “CUAHSI HIS Ontology”, a term that would not satisfy ontology purists in this context) used in HIS Central queries. Our goal is to expose all vocabularies in a consistent manner, and – where possible - in standard or de-facto standard forms. Below are requirements for managing terms and vocabularies:

|  |
| --- |
| Requirements Class |

|  |
| --- |
| his.cuahsi.org/spec/services/1.0/req/terminology/ |

|  |  |
| --- | --- |
| Target Type | Terminology |

|  |  |
| --- | --- |
| Name | Contolled terminology |
| Dependency |  |
| Requirement | his.cuahsi.org/spec/services/1.0/req/terminology/organization  Terms should be grouped into term sets: eg variable names, sample medium. |
|  | his.cuahsi.org/spec/services/1.0/req/terminology/scope    Any listing of text values can, and should be a term list: organizations, qualifiers, quality control levels |
|  | his.cuahsi.org/spec/services/1.0/req/terminology/identifiers/  Terms should have a code that is unique in the system, or at least unique in a set of terms. |
|  | his.cuahsi.org/spec/services/1.0/req//terminology/name/  Terms must have a name**.** |
|  | his.cuahsi.org/spec/services/1.0/req//terminology/description/  Terms should have a brief description**.** |
|  | his.cuahsi.org/spec/services/1.0/req/terminology/provenance/  Terms should have provenance information: when added, by whom, when modified. |
|  | his.cuahsi.org/spec/services/1.0/req/terminology/distribution /  Terminology should be exposed by data services |
|  | his.cuahsi.org/spec/services/1.0/req/terminology/distribution / |

Term Ranking Service Requirements:

|  |
| --- |
| Requirements Class |

|  |
| --- |
| his.cuahsi.org/spec/services/1.0/req/terminology/services/ranking |

|  |  |
| --- | --- |
| Target Type | Terminology |

|  |  |
| --- | --- |
| Name | Contolled terminology services |
| Dependency |  |
| Requirement | his.cuahsi.org/spec/services/1.0/req/terminology/service/ranking/api  REST interface shall be used to expose terminology.  Pattern: http://host/{service}/terms/{term list abbreviation} |
|  | his.cuahsi.org/spec/services/1.0/req/terminology/scope    When requested, a list of terms shall be returned |
|  | his.cuahsi.org/spec/services/1.0/req//terminology/formats/controlledterms/defaults  Default format for Term lists shall be GML code lists. |
|  | his.cuahsi.org/spec/services/1.0/req//terminology/formats/controlledterms/skos/  For term lists that are hierarchical, SKOS will be a required format**.** |
|  | his.cuahsi.org/spec/services/1.0/req//terminology/service/formats/controlledterms/  SKOS shall be an additional format **f**or controlled vocabularies. |
|  | his.cuahsi.org/spec/services/1.0/req/terminology/service/formats/otology/skos  Support formats SKOS for ontologies. |
|  | his.cuahsi.org/spec/services/1.0/req/terminology/service/formats/otology/ owl  Support format OWL/RDF for ontologies. |
|  | his.cuahsi.org/spec/services/1.0/req/terminology/service/formats/otology/ html  Support human readable HTML output. |

Term Services Request

|  |
| --- |
| Requirements Class |

|  |
| --- |
| his.cuahsi.org/spec/services/1.0/req/terminology/services/query |

|  |  |
| --- | --- |
| Target Type | Terminology |

|  |  |
| --- | --- |
| Name | Contolled terminology services |
| Dependency |  |
| Requirement | his.cuahsi.org/spec/services/1.0/req/terminology/service/ query/api  REST interface shall be used to expose terminology.  Pattern: http://host/{service}/terms/query} |
|  | his.cuahsi.org/spec/services/1.0/req/terminology/ service/ query/keyword |
|  | his.cuahsi.org/spec/services/1.0/req/terminology/ service/ query/pattern |
|  | his.cuahsi.org/spec/services/1.0/req/terminology/ service/ query/relationship |

Term Management

|  |
| --- |
| Requirements Class |

|  |
| --- |
| his.cuahsi.org/spec/services/1.0/req/terminology/services/management |

|  |  |
| --- | --- |
| Target Type | Terminology |

|  |  |
| --- | --- |
| Name | Contolled terminology services |
| Dependency |  |
| Requirement | his.cuahsi.org/spec/services/1.0/req/terminology/service/ query/management  REST interface shall be used to expose terminology.  Pattern: http://host/{service}/terms/{term list abbreviation} |
|  | his.cuahsi.org/spec/services/1.0/req/terminology/service/management/addterm |
|  | his.cuahsi.org/spec/services/1.0/req/terminology/service/management/verisonterm |
|  | his.cuahsi.org/spec/services/1.0/req/terminology/service/management/addrelationship |
|  | his.cuahsi.org/spec/services/1.0/req/terminology/service/management/annotateterm |

### Ontology

The HIS Central catalog uses a vocabulary of hierarchically organized hydrologic parameters, which is referred to as CUAHSI HIS ontology. The HIS Central manages mappings between variable names harvested from individual observation networks (services) and leaf concepts in the ontology. While most relationships in the hierarchy are subClassOf and equality relationships, at the moment, the nature of hierarchy is not precisely defined nor it is consistent throughout the over 4100 terms in this vocabulary. Internally the vocabulary is represented as a set of tables with information about hierarchy and synonyms. This enables SQL querying of the concepts (Figure 7). Using SOAP API a hierarchy of terms or a term list can be returned.

While internally the ontology may be represented as tables, externally it should use standard formats for ontologies: SKOS and OWL (OWL can be derived from SKOS vocabularies). This will allow for a standard exchange mechanism to interact with other communities. SKOS is described in Appendix SKOS.

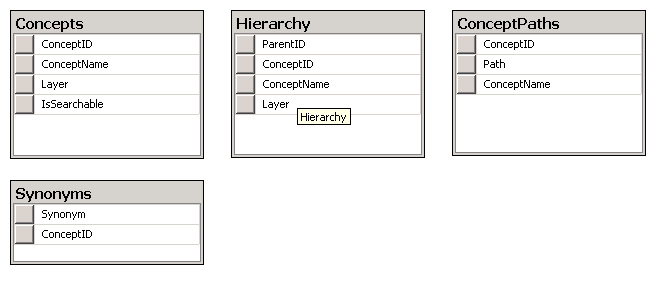


Figure 7. Ontology Tables of HIS Central

#### Variable to Ontology Mapping

In order to support concept-based search for variables and data series, variables must be linked with concepts. These linkages are presently created in an application named HydroTagger (manually, one by one), or by bulk-loading ontology tables with pre-defined tables of associations. HydroTagger provides a graphical interface which data publishers can use to list variables harvested from their observation network services, and see which variables are mapped to concepts in the ontology and which are not. For each unmapped variable, data managers can search for an appropriate concept associate it with the variable, and commit the mapping to central database. Non-tagged variables are “discovered” by a crawler that, currently once a week, trawls through all registered services published off ODM-based sources to find out what has been added in the last week. Metadata for services providing access to large nationwide and regional data sources is updated less frequently because the services do not support a GetSiteInfo or similar requests returning data series, making catalog updates very work intensive.

|  |
| --- |
| Requirements Class |

|  |
| --- |
| his.cuahsi.org/spec/services/1.0/req/ontology/mapping/ |

|  |  |
| --- | --- |
| Target Type | Variable |

|  |  |
| --- | --- |
| Name | Variable to ontolgy mapping |
| Dependency |  |
| Requirement | his.cuahsi.org/spec/services/1.0/req/ontology/mapping/variable  A variable shall be mapped to a concept in the variable ontology |

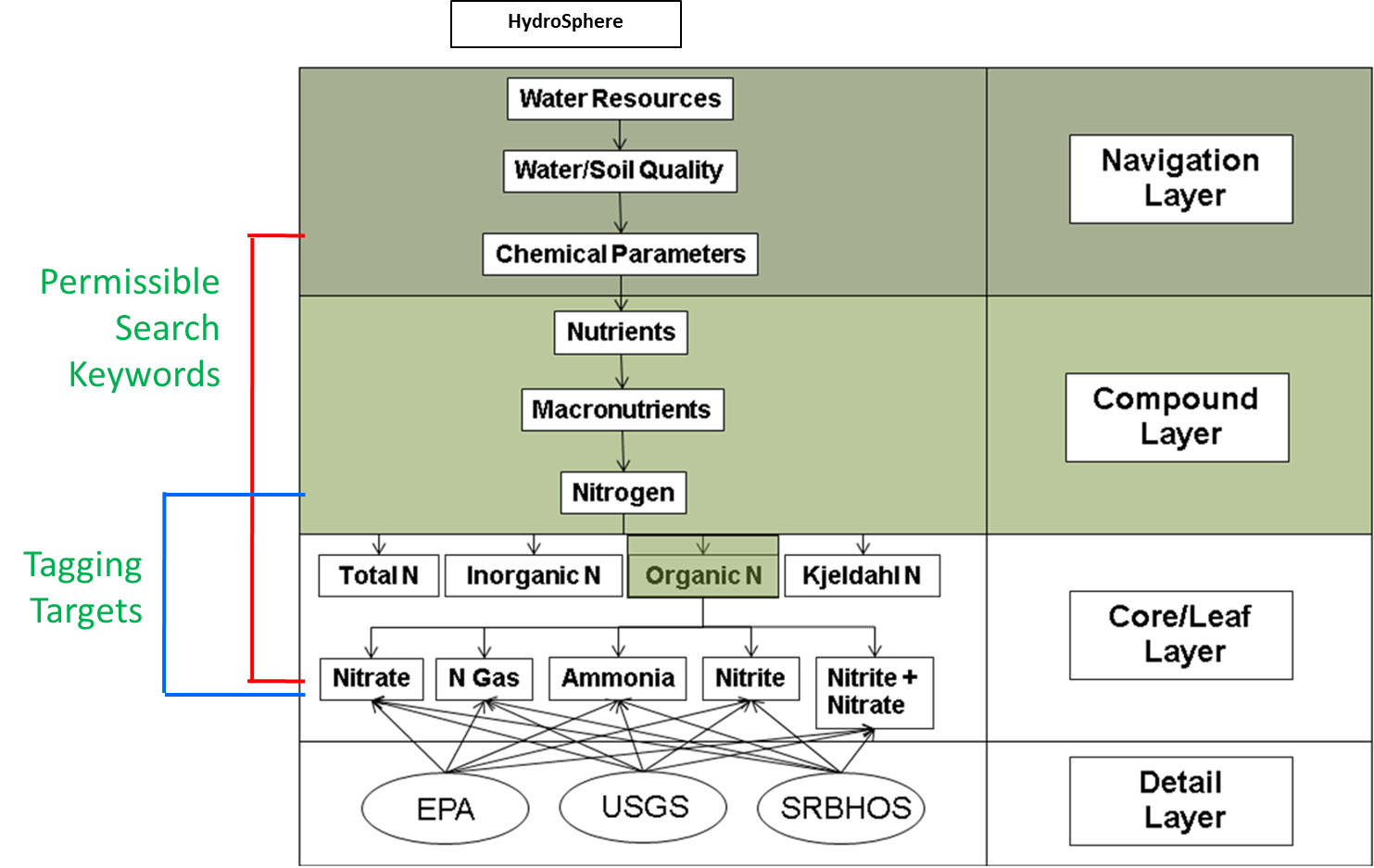


Figure 8. Layered organization of parameter ontology in CUAHSI HIS (source: M. Piasecki). The detail Layer is comprised of variable names used by different data publishers. These variable names are associated with concepts managed within the Leaf Layer. The Compound Layer organizes leaf and core concepts into a navigable hierarchy. Search keywords are compiled from concepts in the Navigation, Compound and Core/Leaf layers.

#### Tagging service

To alleviate the burden of manual variable-concept association, a service that can automatically tag variable names with concepts from one or more vocabularies, should be developed. Based on the SKOS names, and mapped variable names, an attempt to match should be made to assist the user in mapping. Potential simple rules to be used in automatic mapping are:

* exact match to variable name
* exact match to a synonym for a variable
* partial match to variable name
* exact match to a mapped name

More generally, semantic similarity measures may be used to associate variables with concept (e.g. Schwering 2008.)

Suggested approach is that a table of concepts and mapped variable names (with full text indexing) be created periodically in the HIS central, and used to help automate the process. The other would be a service that uses the Lucene full text indexing engine. Alternately, if variable names comply with a controlled vocabulary of parameter names, search services should be able to query the catalog without the necessary tagging step (the latter option would remove HIS reliance on tagging for series discovery, which is currently a bottleneck)

### Controlled Vocabularies

The HIS controlled vocabulary system is used to manage and curate terms in the observation data model. CUAHSI Vocabularies are expandable via community input. The following vocabularies are being used:

|  |  |
| --- | --- |
| Vocabulary | Source/Notes |
| Units | [Unitsof Measure](http://www.opengis.org/docs/02-007r4.pdf) |
| Censor Code | In WaterML spec |
| Dataset Keywords | WMO |
| Data Type | CUAHSI |
| Value Type | CUAHSI |
| Sample Media | CUAHSI |
| Speciation | CUAHSI |
| Spatial Reference | EPSG |
| Vertical Datum | EPSG |
| Methods | Possibly NEMI.gov |

HIS is setup to enforce compliance with these controlled vocabularies at the database level.

### Term Lists

Uncontrolled terminology used within a data source will be exposed by services. Harvesting and indexing this information at the catalog level will enable additional dimensions of data discovery in a data source. Currently, the term lists (which are not enforced, but are catalogued) include:

|  |
| --- |
| Term List |
| Organization |
| Qualifiers |
| Method |
| Quality Control Level |
| Variable Name |

The term lists here can be organized in a concept hierarchy, similarly to how this is done for variable names to enable concept-based search,

## Water Quality Data Services

Initially, efforts of the OGC Hydrology Domain Working Group have focused on standard encoding of time series, which are typically repeated in situ measurements not accompanied by detailed chemical analysis information. The WaterML 2.0 effort has produced a relevant profile of O&M as WaterML 2.0 – Part 1 – Time Series. The work has not focused yet on a water observation encoding profile of the “Samples and Measures” component of O&M, which provides the basis for OGC-compatible encoding analytical ex situ measurements. Absent a standard profile, USGS and EPA deliver analytical sample data via Water Quality Exchange (WQX) schema, using SOAP and REST APIs. The plan for extending HIS to handling analytical chemistry information for ex situ measurements in a common WQX encoding includes:

1. Analytical results will be made available in WQX via a REST endpoint
2. Time series containing analytical results shall reference a WQX record
3. When a specification from the HDWG creates a profile of the O&M schema for samples, sample information should be made available in that format.

## Feature Models

The existing HIS infrastructure relies on the notion of “sites” as points where observations are made. The WaterML 2.0 specification introduces the notion of MonitoringPoint which has similar characteristics. Initially, we will utilize the MonitoringPoint as defined by WaterML 2 – Part 1 (Figures 9-10) as the feature of interest for encoding time series information. Hydrologic Feature representation and encoding are under development as part of the INSPIRE Hydrography Data Model, which is summarized in Appendix 4. The OGC Hydrology Domain Working Group is reviewing this model and testing portions in the Surface Water Interoperability Experiment (SWIE). In HIS, we are planning to use this standard hydrologic feature encoding when it becomes available as a standard specification. In addition, HIS infrastructure may rely on ArcHydro geodatabase schema, which may be suitable as a feature model once it is mapped to OGC GML framework.

MonitoringPoint

Figure 9. UML of WaterML2 MonitoringPoint



Figure 10. XML fragment of representing a WaterObservationPoint

The following sections considers several new aspects of HIS related to data products derived from original observation series, maintaining access to such products, and maintaining synchronization across different distributed components of the system.

.

## Data Distribution Formats: Hydrologic Themes

Hydrologic themes are packaged datasets that represent thematically-organized named collections of observational series with associated metadata. Themes aggregate thematic data from multiple sources (organizations) into a single package, where data may be normalized (i.e. have consistent units and uniform metadata descriptions) and are curated to provide consistent hydrologic themes for multiple users. Themes shall be curated to support downloading, archival storage, or online distribution via WFS services. Some examples of common hydrologic themes assembled from multiple sources are shown graphically in Figure 11. Themes may be static, for example data published in a paper, or dynamic, where data are regularly updated (e.g. streamflow data series for a watershed updated monthly). A related concept is data cart. Data carts are collections of time series definitions typically assembled in a client application when a user creates or updates a selection set from a series catalog. Themes differ from datacarts in that they have data, plus metadata about the purpose of the theme, who created it, how often is it updated, and any additional provenance information describing the curation process. Themes may also be presented as WFS (Figure 12)

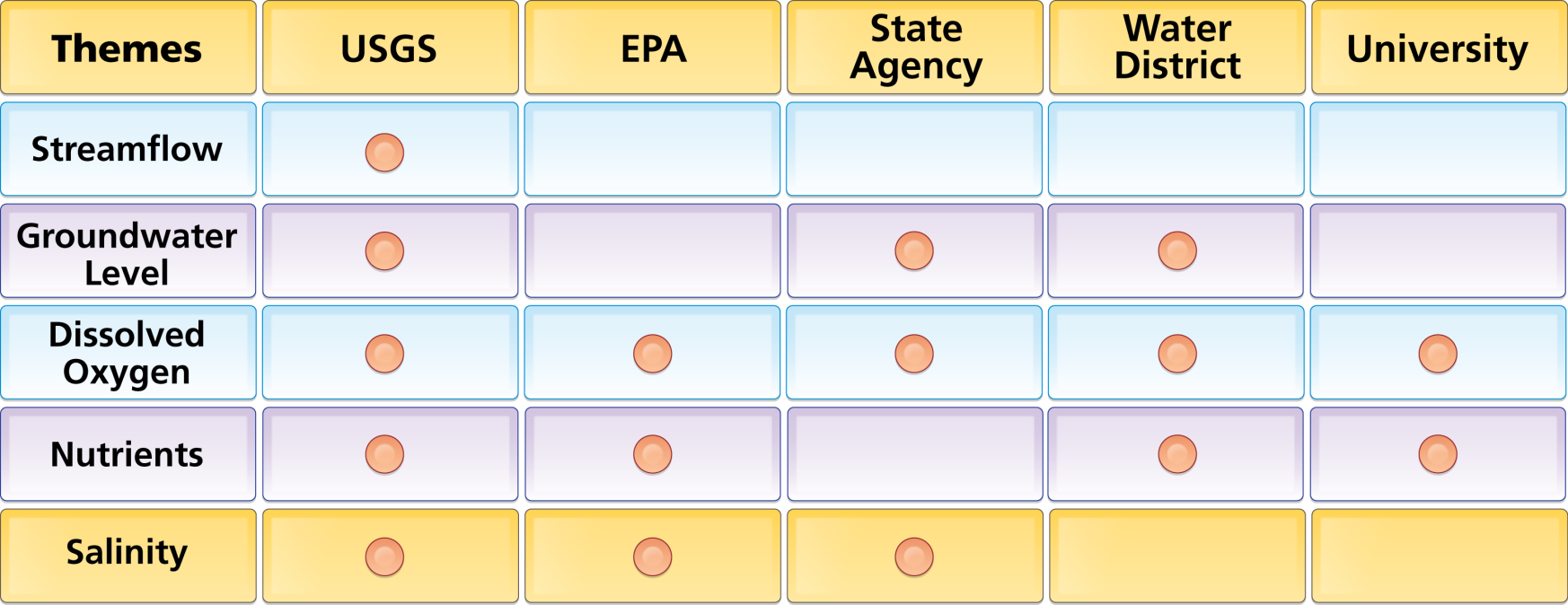


Figure 11. Aggregating data from multiple organizations into hydrologic themes

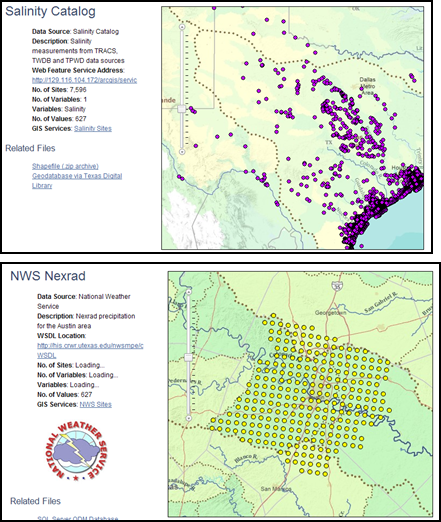


Figure 12. Mapping of themes using WFS

Curated themes and can be used in several use cases, including publishing data in OGC-compatible formats to be used in various analytical applications.in particular situational awareness and large-scale modeling.

To take advantage of theme-based representation of hydrologic data, HIS will need to develop several additional components:

1. define a theme information model, including sufficient metadata reflecting theme intent, update mechanism, provenance and curation procedures, and describing how data from different sources are combined into a theme (e.g., cover the same area or the same time period, semantically-related, or functionally-related). The current specification of a theme is included in the appendix.
2. develop consistent ways to store and represent packaged hydrologic themes in a system-independent manner, and materialize theme specification into a variety of common formats,
3. develop software for generating hydrologic themes based on series catalogs and OGC-compatible service interfaces,
4. create normalization routines to ensure that data within a theme are homogeneous with respect to units and other characteristics where automatic translations are possible,
5. develop theme validation routines, including ability to analyze theme metadata for consistency, completeness and gaps without retrieving all the data
6. create theme curation workflows and APIs,
7. in the future, shall extend theme specification to include model output,
8. specify how themes can incorporate semantic information
9. Develop routines to automatically register created themes in hydrologic catalogs

## Authorization

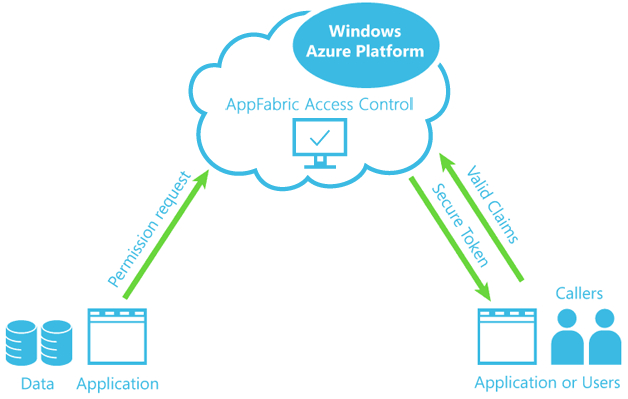
A capability commonly requested by HIS users is user authentication access control to hydrologic data. Several implementation options have been considered, in particular implementing granular (at the level of data series) access control at HydroServer level, and reliance on either internal to CUAHSI HIS or external identity providers. A centralized and distributed authentication and authorization system is required to manage data publication and access across distributed publishers and users. Microsoft Windows Azure Access Control (<http://www.microsoft.com/windowsazure/features/accesscontrol/>) would provide a sustainable authorization solution that meets the requirements, as it provides identity and access control to web applications and services, while integrating with standards-based identity providers (Figure 13). Azure Access control can integrate with a variety of authentication services (Windows Directory, Yahoo ID, Google, Facebook), and provides a centralize service that can be utilized by others who may wish to maintain alternative data systems. The latter is important as HIS should be designed in a flexible manner so as to be able to easily integrate with future data systems, and support API’s for .Net, Java, Ruby and Python.

Figure 13. . Azure Access Control System

## Data Synchronization

Data synchronization is another essential component of HIS. It would ensure consistency across distributed catalogs and, eventually, across distributed data as they are cached or incorporated in hydrologic themes or other derived data products. Identifiers (section 2) are essential to synchronization support. Persistent and managed identifiers enable duplicate information to be identified, and provide for new information to be distributed in a manner that does not violate integrity of the system. Synchronization of information occurs at several levels:

* Between data sources and the Central Catalog (metadata harvesting; eventually data harvesting and mirroring);
* Between distributed metadata catalogs;
* Between data sources and separately curated derived products such as hydrologic themes (data harvesting and normalization);
* Between a metadata catalog or a data source and an occasionally connected client.

Harvesting is equivalent to one-way synchronization. While we expect that observation series will be harvested from WFS observation series, we should not limit the harvest to only a single source type.

At the implementation level, data synchronization could utilize Microsoft synchronization framework. Source providers for different data sources will provide information for synchronization to a staging database (Figure 14). This staging database can be utilized to populate the HIS central metadata catalog. Each source provider may have a set of rules about its identifier generation: for example, series identifiers would be generated based the data provider’s notion of a series (e.g. including or not including QA/QC level, vertical offset, or other characteristics). Another important provider’s capability is returning “recent changes”, so that harvesting can focus on the metadata changes only. The Synchronization framework isolates changes in the destination from the source, and vice versa. Within HIS, we expect to need to develop providers for legacy WaterOneFlow sources, and Sensor Observations Services. The synchronization framework will be also used to synchronize the HIS Central Catalog with the staging database.

The synchronization framework will be explored as a way to subset observation data series catalogs and distribute them to an occasionally connected client, which would cache and maintain a local catalog copy (e.g., Hyrodesktop, which maintains a local metadata catalog fragment). By applying filters/queries, records from the staging database associated with the central metadata catalog can be returned to a client that is used for field work or in other situations where persistent connection to the central catalog is not guaranteed. When the client reconnects, new information can be uploaded, and old information synchronized using the sync framework.

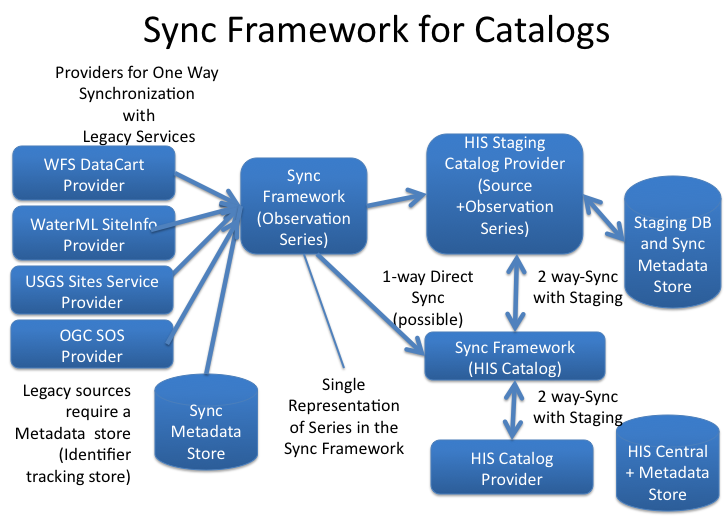


Figure 15. Synchronizing observation data series from multiple formats to a staging repository and then to an HIS central catalog

# HIS Components

The HIS system has many components. An overview was discussed in multiple earlier reports available online from the CUAHSI HIS web site. This section begins by discussing the metadata components, focusing of service catalogs and observation series catalogs. It then discusses the data access components, and two profiles of the OGC Sensor Observation Service that are used to carry hydrologic data in WaterML 2.0 or another OGC standard encoding. The discovery and access components for observation time series are shown on Figure 15.



Figure 16. Components of the HIS system

## Metadata Catalogs

HIS stores metadata for services and observation series; a service registry and observation series catalog are the two key metadata components in HIS (Figure 16). If a service registered in the service registry, is an Observation Series WFS (also referred to as “data cart WFS”), or SOS with the Data Availability extension, then the observation series metadata from such a service can be harvested into an observation series catalog. Registering and discovering services in a CSW catalog is well documented, although as a record is harvested into different CSW providers (e.g. ESRI GeoPortal Server, GI-CAT, GeoNetwork Open Source), translation errors can, and do, occur. Catalogs will need to be limited to specific tested implementations to insure accuracy and fidelity of the information as it is exchanged between different providers. Harvesting, storing and retrieving observation series metadata records using CSW is a fundamental change for the HIS system. The present HIS central catalog is not exposed via a standard CSW interface. The Water Data Services Concept Development Study recommends the use of the CSW-EbRim profile. While flexible and extendable, the CSW-EbRim profile is not yet being implemented in HIS because it is not currently supported by most CSW clients. Also, we do not know the scalability of CSW servers serving ISO metadata records, or nuances of ISO metadata record exchanges between CSW providers. HIS team needs to test how a CSW-ISO implementation would scale to millions of records so that it could be utilized as observational series catalogs. Given the scalability consideration, the preferred method of accessing observation series will be WFS observation data series profiles. Service metadata over WFS is a tested approach that has been utilized by many communities (Rob Atkinson and Steve Richards, Personal Communications).



Figure 17. Metadata components, and harvesting of records from a WFS and a SOS service.

### Service Catalog

The service catalog (registry) will represent a federation of CSW services, each of which indexes observation or other standard services. A single front end federation service will be utilized to access the multiple CSW catalogs, mostly implemented using the ESRI Geoportal Server, or some CSW proxy for other types of data. For example, THREDDS data server (TDS) records may be currently registered in a GI-CAT implementation of CSW, as described below. Accessing distributed CSW sources, or CSW proxy sources, via a single federation service will enable a community catalog of hydrologic data services of different types. One challenge to resolve is potential incompatibility between different CSW implementations. Catalog compatibility studies will need to be conducted, and run in order to ensure the fidelity of the information retrieved from CSW endpoints. While aggregating and querying into a single CSW can function, the fidelity of transmitted information becomes problematic when records are transmitted through more than one system, where information may be added, lost or transformed by different systems. In order to address the issue, the HIS project will utilizing the ESRI Geoportal Server, and the GI-Cat metadata servers.

#### Federated Searching via GI-CAT

To simplify discovery of services, we utilize a single search service that federates multiple CSW services (Figure 17). In a federated system of catalogs, multiple services can be searched simultaneously. In the HIS model, GI-CAT CSE implementation is utilized as the federation host. OGC-compatible services for observation series catalogs and features, map portrayal and gridded data (WFS, WMS, WCS) are expected to be registered to a CSW service, preferably implemented as ESRI Geoportal Server. The GI-CAT can also proxy services, such as a THREEDS catalog or a WFS server. In addition, it provides multiple search API, include CSW-ISO, CSW-EbRim, and OpenSearch.

Figure 18. Federated searching using GI-CAT

#### CSW – ESRI GeoportalServer

Community resources will be registered with a CSW service. In HIS, we selected ESRI Geoportal Server as a CSW service provider for distributed catalogs, as well as the web user interface that supports discovery over multiple CSW sources. It does not support federated searching, so only a single CSW can be searched at a time. The usage of Geoportal Server will follow these principal requirements:

* Each resource (service) is registered to a single CSW instance;
* A CSW instance translates the OGC GetCapabilities request, for each registered resource (service) into a service metadata record;
* Thus created metadata records have a distinct identifier, so that the identity of resources (services) is maintained when they are harvested by other CSW instances including the host CSW instance.

#### CSW – registration of THREDDS sources

GI-CAT can proxy a THREDDS catalog, allowing for CSW searching over THREDDS records. It does this by harvesting information from the THREEDS catalog into an XML catalog. The complication, however, is that THREDDS records do not follow a consistent metadata model, and that the rapid addition of records to the NCAR THREDDS servers means that the catalog is out of date when harvested. Some records in a THREDDS catalog are detailed, while others require a user to make assumptions or know the data model and implicit characteristics or defaults for a particular catalogued dataset. We have concluded that, within HIS, we should not proxy the entire THREDDS ecosystem, but should instead proxy subsets of information that we can retrieve. This will allow for specific testing of the keywords needed to discover and retrieve data accessible via THREEDS services. If testing does not meet expectations of hydrologists trying to discover relevant records, HIS catalog curators will either need to improve detail of the records in the THREEDS services, or improve the proxy capabilities by adding assumptions, defaults and heuristics based on expert knowledge of a group of datasets. This will ensure a better data discovery experience for the hydrologic community. The suggested sequence of steps for integrating THREDDS catalog search in HIS is summarized as follows:

* Develop a specific use case for a fragment of a THREEDS catalog.
* Register the fragment of the THREEDS catalog into a GI-CAT CSW proxy.
* Conduct metadata record retrieval tests.
* Once the tests are satisfactory (or THREDDS/proxy are improved), integrate the proxy into federated services.

### Observation Series Catalog

An observation series catalog is a list of observational data series that follows the Water Observations Metadata Specification (Appendix 2). A catalog reflects the content of one or more water data services. The WFS and CSW service options for series catalog implementation are described in this section. As was mentioned earlier, until the scalability of CSW-ISO implementations to millions of records is demonstrated, the preferred method of accessing observation series will be via WFS observation data series profiles.

#### WFS for Observation Data Series Profiles: Requirements

.

|  |
| --- |
| Requirements Class |

|  |
| --- |
| his.cuahsi.org/spec/observation/dataseries/services |

|  |  |
| --- | --- |
| Target Type | WFS |

|  |  |
| --- | --- |
| Name | WFS service for Observation Data Series |
| Dependency |  |
| Requirement | his.cuahsi.org/spec/observation/dataseries/services/wfs/version  WFS 1.1 |
| Requirement | his.cuahsi.org/spec/observation/dataseries/services/wfs/format  •Support the Datacart specification (Appendix 3 |
| Requirement | his.cuahsi.org/spec/observation/dataseries/services/wfs/filters  Supports filters listed in Table 7. Filters and applicability |
| Requirement | his.cuahsi.org/spec/observation/dataseries/services/wfs/keywords  Capabilities must include keywords to identify the type (datacart, CUAHSI), concepts in series, and variable names 1 |
| Requirement | his.cuahsi.org/spec/observation/dataseries/services/wfs/keywords/cardinality  Ideally multiple keyword elements to support multiple term lists will be utilized (not all servers support this)   * Keywords/type for Datacart: CUAHSI * Keywords/type for Concepts: Concepts * Keywords/type for Variable Names: VariableNames |
|  | his.cuahsi.org/spec/observation/dataseries/services/wfs/keywords/datacart  Keywords for the datacart should be in the datacart term list |
|  | his.cuahsi.org/spec/observation/dataseries/services/wfs/keywords/ Concepts  keywords for the concepts should be in the concepts keyword list |
|  | his.cuahsi.org/spec/observation/dataseries/services/wfs/keywords/ VariableNames  keywords for variable names should be in the variableNames kewords list |

##### WFS Implementation

There are many WFS implementations that can be utilized for the observations catalog as specified in Appendix 2, since this is essentially a flat file representation of a catalog. For example, a source for such a catalog service could be a simple shapefile with column names as specified in the observation catalog specification. ArcGIS Server would be a platform of choice for serving WFS observation catalogs to be consumed by client software based on ESRI products. For other client environments, Geoserver or Deegree are suggested.

#### CSW for Observation Data Series Profiles: Requirements

|  |
| --- |
| Requirements Class |

|  |
| --- |
| his.cuahsi.org/spec/observation/dataseries/services/catalog |

|  |  |
| --- | --- |
| Target Type | Catalog |

|  |  |
| --- | --- |
| Name | CSW Catalog services for Observation Data Series |
| Dependency |  |
| Requirement | his.cuahsi.org/spec/ observation/dataseries/services/catalog version  WFS 1.1 |
| Requirement | his.cuahsi.org/spec/ observation/dataseries/services/catalog format  • Support the INSPIRE data series specification |
| Requirement | his.cuahsi.org/spec/ observation/dataseries/services/catalog filters  Supports filters listed in Table 7. Filters and applicability |
| Requirement | his.cuahsi.org/spec/observation/dataseries/services/catalog/keywords  Capabilities must include keywords to identify the type (datacart, CUAHSI), concepts in series, and variable names 1 |
| Requirement | his.cuahsi.org/spec/ observation/dataseries/services/catalog cardinality  Ideally multiple keywords elements to support multiple term lists will be utilized (not all servers support this)   * Keywords/type for Datacart: CUAHSI * Keywords/type for Concepts: Concepts * Keywords/type for Variable Names: VariableNames |
|  | his.cuahsi.org/spec/observation/dataseries/services/catalog keywords/datacart  Keywords for the datacart should be in the datacart term list |
|  | his.cuahsi.org/spec/observation/dataseries/services/catalogkeywords/ Concepts  keywords for the concepts should be in the concepts keyword list |
|  | his.cuahsi.org/spec/ observation/dataseries/services/catalog/keywords/ VariableNames  keywords for variable names should be in the variableNames kewords list |

##### CSW Implementation

There are a limited number of CSW implementations that are compatible with the harvesting of data, and the searching of records across catalogs.

We expect that Gi-Cat proxy will be implemented to expose the observation series catalog. The HIS project has been also investigating the PyCSW project as a possible CSW implementation.

### How to embed Ontology in CSW

In order to query CSW catalogs by concepts, existing CSW clients will need to perform term expansion before submitting a series of requests to a CSW server. Currently, only ESRI Geoportal Server can be enabled to utilize a separate ontology service to expand the search using a query term prefixed by ‘like’. When ‘like’ is used, then the term is submitted to the ontology service, and a list of related terms based on an OWL ontology (GEMET, at present) will be returned.

With GI-CAT used as the catalog federation host, external ontology will be utilized via a proxy service. Integrating this capability in GI-CAT catalog service is one of the development goals as presented in the requirement class below.

|  |
| --- |
| Requirements Class |

|  |
| --- |
| his.cuahsi.org/spec/ontology/services/csw |

|  |  |
| --- | --- |
| Target Type | Catalog |

|  |  |
| --- | --- |
| Name | CSW Catalog services ontology |
| Dependency |  |
| Requirement | his.cuahsi.org/spec/ontology/services/csw/translation  ontology translation should happen in the CSW server |

## Data Services

Water Data Services is a set of services that return information about data values with the hydrologic context/semantics according to a standardized set of schemas. This section presents CUAHSI profiles of OGC services used or being developed by the HIS project.

### CUAHSI Profile of OGC Services for Time Series Data

Following recommendations of the earlier Water Data Services Concept Development Study, the CUAHSI profile for time series data uses a coupled SOS and WFS. WFS is used to serve hydrologic features of interest, as well as observation data series records (the data series catalog, as described earlier) encoded in a simple GML. An SOS 2 Data Availability Extension (the GetDataAvailability request) is also utilized to expose observation data series in a standard manner.

|  |
| --- |
| Requirements Class |

|  |
| --- |
| his.cuahsi.org/spec/services/1.0/req/services/ |

|  |  |
| --- | --- |
| Target Type | Services |

|  |  |
| --- | --- |
| Name | CUAHSI Profile for services |
| Dependency |  |
| Requirement | his.cuahsi.org/spec/services/1.0/req/services/ Observations/SOS/  Shall use SOS 2 to deliver data |
|  | his.cuahsi.org/spec/services/1.0/req/services/WFS  A WFS service shall serve the wml2 MonitoringPoint (Sampling Point) and Hydrologic Features (Sampled Point) of interest for hydrologic information systems. |
|  |  |



Figure 19. CUAHSI OGC Compliant Data Server

#### CUAHSI Profile of SOS for time series from network of observation sites

|  |
| --- |
| Requirements Class |

|  |
| --- |
| his.cuahsi.org/spec/services/1.0/req/services/Observations/SOS/ |

|  |  |
| --- | --- |
| Target Type | Services |

|  |  |
| --- | --- |
| Name | CUAHSI Profile for SOS |
| Dependency |  |
| Requirement | his.cuahsi.org/spec/services/1.0/req/services/ Observations/SOS  SOS, v2. Shall be used to deliver WaterML 2.. |
|  | his.cuahsi.org/spec/services/1.0/req/services/Observations/SOS/gda  SOS shall support the GetDataAvailability.extension |
|  | his.cuahsi.org/spec/services/1.0/req/services/SOS/Observations/profile  SOS defines Offering as SOS Type A |
|  | his.cuahsi.org/spec/services/1.0/req/services/SOS/Observations/FOI  SamplingPoint and Sampled Feature FOI should be reference a WFS service |
|  | his.cuahsi.org/spec/services/1.0/req/services/SOS/Observations/getobservaton/foi  GetObservation response FOI shall always be serialised by reference (xlink:href) with a proper xlink:title with a descriptive label (eg: Medon Road, Little Bear River, Utah |
|  | his.cuahsi.org/spec/services/1.0/req/services/SOS/Observations/getresult/lastvalue  Support a retrieval of the last value by using a TimeInstant/timePosition=latest |

##### SOS Discussion

The “SOS Usage Profile for the Hydrology Domain” (Fuest, 2011) discusses an SOS Usage profile for the Hydrology Domain. Fuest discusses four type of real world SOS implementations, and suggests that users go with SOS Type A. SOS type C would be simpler to implement, since the observation data series are discoverable with the WFS datacart. The CUAHSI OGC stack will adopt a model where an ObservationOffering is a procedure and features of interest (SOS Type A) and not an individual time series. This will require some aggregation of the of observation series catalog. This section uses information from Fuest.

###### Definitions

General definitions taken from the SOS 2.0 and O&M 2.0 specification documents:

##### Observation Offering:

An observation offering is a logical grouping of observations offered by a service that are related in some way. The parameters that constrain the offering should be defined in such as way that the offering is 'dense' in the sense that requests for observations that are within the specified parameters should be unlikely to result in an empty set.

###### Procedure:

Method, algorithm or instrument. (O&M: ...which is often an instrument or sensor but may be a process chain, human observer, an algorithm, a computation or simulator.)

###### ObservedProperty:

The observedProperty identifies or describes the phenomenon for which the observation result provides an estimate of its value. the minimum requirement is that every property of interest is identified by a URI. This allows the "by-reference" mode of the standard GML property pattern to be used, regardless of whether the value is available in a GML-encoded form or not

###### FeatureOfInterest:

The featureOfInterest is a feature of any type (ISO 19109, ISO 19101), which is a representation of the observation target, being the real-world object regarding which the observation is made.

Table 5. Examples and best practice for WaterML2

|  |  |  |
| --- | --- | --- |
|  | **Best Practice** | **Improper Practice** |
| Offering | Any logical grouping | With the Site as logical group because SITE is already encoded in the FOI |
| Procedure | RawData, DailyMean,  5minMean, … | Depends on type  SiteA~WL~RawData |
| ObservedProperty | WaterLevel  Discharge  Any Concept from the Ontology | WaterLevelDailyMean  DischargeXYZ  CologneWaterLevel |
| FeatureOfinterest | SiteA, StationB, MyLocationC | Cologne\_WaterLevel |

Table 6. GetCapabilities document - SOS Server Type A (procedure==sensor-type) (from Fuest, 2011)

|  |
| --- |
| SOS Server Type A (procedure==sensor-type)  A SOS service which serves more than raw data or medium/larger networks should use the following structure for the getCapabilities response. Here a procedure is seen as a sensor-type (==time-series type) and NOT a sensor instance (==time-series instance). This structure should be homogenously used and not be used with other encodings. (refer to http://kiwis.kisters.de for example) |
| SOS2 GetCapabilties Fragment Type A |
| (syntax has been reduced to get a better overview)  <contents> <sos:observableProperty xlink:href="WaterLevel" /> <sos:observableProperty xlink:href="Discharge" /> <sos:relatedFeature xlink:href="aSite1" /> <sos:relatedFeature xlink:href="aSite2" /> <sos:relatedFeature xlink:href="aSite3" /> <!--(... round about XXX other stations/sites…) --> </contents> <observationOffering>  <sos:procedure xlink:href="RawData" />  <sos:observableProperty xlink:href="WaterLevel" />  <sos:observableProperty xlink:href="Discharge" />  <sos:observedProperty xlink:href="Precipitation" /> </observationOffering> <observationOffering>  <sos:procedure xlink:href="DailyMean“ />  <sos:observableProperty xlink:href="WaterLevel" />  <sos:observableProperty xlink:href="Discharge" /> </observationOffering> <observationOffering>  <sos:procedure xlink:href="DailyTotal“ />  <sos:observableProperty xlink:href=" Precipitation " /> </observationOffering> |

#### WFS

|  |
| --- |
| Requirements Class |

|  |
| --- |
| his.cuahsi.org/spec/services/1.0/req/services/observations/WFS |

|  |  |
| --- | --- |
| Target Type | Observations Services |

|  |  |
| --- | --- |
| Name | CUAHSI Profile for observation services WFS |
| Dependency |  |
| Requirement | his.cuahsi.org/spec/services/1.0/req/services/observations/wfs/version  shall be wfs 1.1 |
|  | his.cuahsi.org/spec/services/1.0/req/services/observations/wfs/foi  WFS service return FOI for the observations SOS service. |
|  | his.cuahsi.org/spec/services/1.0/req/services/observations/wfs/formats/hydrographic features  WFS should return hydrographic features |

#### WMS

It is desirable for a data source to represent its locations in a web map server. For small sources, this may not be needed. For services with large numbers of data

#### Implementation:

Possible software for the HIS data server.

|  |  |  |
| --- | --- | --- |
| **Service** |  |  |
| SOS | Returns observations data | PySOS, 52North SOS. Deegree |
| WFS | Returns locations and hydrographic features | ArcGIS, GeoServer, Deegree |
| WMS | Returns images | ArcGIS, GeoServer, Deegree |

### CUAHSI profile for time series from grid cells

A time series for single grid cell, or an aggregate of cells over a region can be returned from a sensor observation service. Since there are many locations, the service should use a single feature of interest, the area that the grid covers, and a web map service to display the locations.

|  |
| --- |
| Requirements Class |

|  |
| --- |
| his.cuahsi.org/spec/services/1.0/req/services/grid/SOS/ |

|  |  |
| --- | --- |
| Target Type | Services |

|  |  |
| --- | --- |
| Name | CUAHSI Profile for SOS gird data |
| Dependency |  |
| Requirement | his.cuahsi.org/spec/services/1.0/req/services/grid/SOS  SOS, v2. Shall be used to deliver WaterML 2.. |
|  | his.cuahsi.org/spec/services/1.0/req/services/grid/SOS/gda  SOS shall support the GetDataAvailability.extension |
|  | his.cuahsi.org/spec/services/1.0/req/services/SOS/grid/profile  SOS defines Offering as SOS Type C |
|  | his.cuahsi.org/spec/services/1.0/req/services/SOS/grid/FOI  FOI should be a single feature of inerest |
|  | his.cuahsi.org/spec/services/1.0/req/services/SOS/grid/GetFeatureOfInterest  shall implement GetFeatureOfInterest with filter for bounding box.This will return locaitons within the filtered region |
|  | his.cuahsi.org/spec/services/1.0/req/services/SOS/grid/GetObservation  Implement GetObservation   * If a feature identifier of the SamplingFeatureCollection is requested, reject the request * If a feature identifier sampling feature is requested, then return the result |

#### SOS Discussion

This SOS would best be served by a using an SOS Type C, where a single feature of interest has one or more observed properties. This simplifies the implementation on the server side, while requiring the client to make an additional query to filter the result set to the region of interest. A grid is a set of point locations, and there can be several tens of thousands of locations.

In this case the client reads the getCapabilities document, and sees a single feature of interest that is a SamplingFeatureCollection. The client will do a DescribeFeatureOfInterest, and return the schema of the SamplingFeatureCollection. The client will then call GetFeatureOfInterest with a spatial filter, and features will be returned. The client can then call GetObservation with a feature identifier, and an observed property to get the timeseries.

|  |
| --- |
| SOS Server Type C (procedure==sensor-type/system)  A SOS with many sites, and a few properties should the following structure for the getCapabilities response. Here a procedure is seen as a sensor-type or system. This structure requires additional requests or knowledge to “drill” into the data if you want to do it by sensor instance (also refer to <http://external.opengis.org/twiki_public/bin/view/HydrologyDWG/GwIeGetCapabilitiesBestPractices>, <http://external>.opengis.org/twiki\_public/bin/view/HydrologyDWG/SOSLargeCollectionSensorDiscussion) |
| SOS2 GetCapabilties Fragment Type C |
| (syntax has been reduced to get a better overview)  <contents> </contents> <observationOffering>  <sos:procedure xlink:href="DailyMean"/>  <sos:observableProperty xlink:href="urn:ogc:def:property:OGC:GroundWaterLevel"/>  <sos:featureOfInterest xlink:href="urn:xngwd:  feature:OntarioWellNetwork"/> </observationOffering> <observationOffering>  <sos:procedure xlink:href="RawData"/>  <sos:observableProperty xlink:href="urn:ogc:def:property:OGC:GroundWaterLevel"/>  <sos:featureOfInterest  xlink:href="urn:xngwd:  feature:OntarioWellNetwork"/> </observationOffering> |

#### WMS

Purpose: In order to simply visualize the information from a gridded dataset, a web map service should be utilized.

|  |
| --- |
| Requirements Class |

|  |
| --- |
| his.cuahsi.org/spec/services/1.0/req/services/grid/WMS/ |

|  |  |
| --- | --- |
| Target Type | Services |

|  |  |
| --- | --- |
| Name | CUAHSI Profile for SOS grid data |
| Dependency |  |
| Requirement | his.cuahsi.org/spec/services/1.0/req/services/ grid /WMS  location information about the grid should be returned in a web map service |

#### Implementation

|  |  |  |
| --- | --- | --- |
| Service |  |  |
| SOS | Returns observations data | PySOS, 52North SOS. Deegree |
| WMS | Returns images | ArcGIS, GeoServer, Deegree |
| WCS | Returns coverage | THREEDS |

### Query Filters

In order to query the various services, OGC filters are used. Different standards utilize different filters. OpenGIS Filter Encoding 2.0 Encoding Standard (2009). Table 7 lists the information that should be supported by filtering, and the applicability to the different services. XML examples of filters are included in Appendix 6.

Table 7. Filters and applicability

|  | **Metadata Services** |  |  | **Data Services** |  |
| --- | --- | --- | --- | --- | --- |
|  | **Service metadata** | **Data Series** |  | **Time Series** | **Grid** |
| OGC Interface | CSW | WFS | CSW | SOS | WCS |
| Dataset name | X | X | X | N/A |  |
| Service provider name | X | X | X | N/A |  |
| Geographic bounding box | X | X | X | X | X |
| Hydrologic concepts | X | X | X | X |  |
| Hydrologic concepts with mapping | X | X | X |  |  |
| Temporal window |  | X | X | X | X |
| Named location |  | X | X |  |  |
| ObservedProperty Name |  | X | X | X |  |
| ObservedProperty Identifier |  | X | X | X |  |
| Boolean operators | X | Optional1 | X | X |  |
| Data Value | N/A |  |  | Optional2 |  |

1 ESRI WFS does not support Boolean operators, at least according to their documentation

2 at present, it is difficult to develop queries to the data value level based on the SOS specifications. The HDWG will need to develop such as specification.

## Map Services

### Base Reference layers

A hydro base map set of layers should be utilized, and exposed via WMS and WFS. ESRI has developed a US Hydrologic base map, which provides a common background map. These need to be exposed an OGC Map service (WMS), and an OGC Web Feature Service (WFS). There will be multiple servers and organizations hosting and utilizing reference datasets, either as OGC services, or as transformed data to be utilized in modeling and other applications.

One obstacle to be overcome is the sharing of reference datasets. Sharing of such information is difficult because no common identification system exists for registering or identifying features. Without a common reference, we cannot easily identify features between distributed systems, and many features may end up duplicated. .

|  |  |
| --- | --- |
| **Base Dataset** | **Original Source** |
| River network |  |
| Watershed |  |
| Place Names |  |

TDB: Base Layers

## Vocabulary

Controlled vocabularies and ontologies are needed to communicate the contents and concepts of a data source of hydrologic information.

### Ontology

An ontology is used to map a variable in a dataset to a set of common hydrologic variables. At present we only have a single ontology, variable names. The HIS master vocabulary controls the ODM concept of variable: Sample Media, Unit, Value Type, and Data Type, and these will be exposed as Ontologies. Time Periods is another ontology that needs to be encoded; we need to describe day, month, year, and other common periods. Events, such as hurricane or flood, also should be developed into community ontologies..

|  |
| --- |
| Requirements Class |

|  |
| --- |
| his.cuahsi.org/spec/information/1.0/req/ontology/ |

|  |  |
| --- | --- |
| Target Type | Variable |

|  |  |
| --- | --- |
| Name | Types of ontologies |
| Dependency |  |
| Requirement | his.cuahsi.org/spec/ information/1.0/req/ontology/ontologies  There should be more than one ontology that a variable can be mapped to.   * Sample Media, * Unit, * Value Type, * Data Type, * Named Time Period. |
|  | his.cuahsi.org/spec/ information/1.0/req/ontology/ontologies/mapping  A variable may be attached to more than one ontology |

|  |
| --- |
| Requirements Class |

|  |
| --- |
| his.cuahsi.org/spec/ information/1.0/req/ontology/ontologies/variable |

|  |  |
| --- | --- |
| Target Type | Variable |

|  |  |
| --- | --- |
| Name | Variable Mapping to the CUAHSI Vocabulary Ontology |
| Dependency |  |
| Requirement | his.cuahsi.org/spec/information/1.0/req/ontology/ontologies/variable/variableCode  A variable should be Mapped at least one concept in the CUAHSI Vocabulary Ontology. |
|  | his.cuahsi.org/spec/information/1.0/req/ontology/ontologies/variable/variableCode/Count  A variable may be mapped to more than one concept |

#### Ontology Format

The ontology and controlled terms should be made available in SKOS. Like the ontology tables, separate files containing the concepts/terms, relationships, and synonyms should be exposed.

|  |
| --- |
| Requirements Class |

|  |
| --- |
| his.cuahsi.org/spec/services/1.0/req/ontology/format |

|  |  |
| --- | --- |
| Target Type | Ontology |

|  |  |
| --- | --- |
| Name | Ontology format |
| Dependency | his.cuahsi.org/spec/information/1.0/req/ontology |
| Requirement | his.cuahsi.org/spec/services/1.0/req/ontology/format/skos  The ontology will be made available as SKOS |
|  | his.cuahsi.org/spec/services/1.0/req/ontology/format/skos/concepts  concepts as one file |
|  | his.cuahsi.org/spec/services/1.0/req/ontology/format/skos/relationships  relationships as one file |
|  | his.cuahsi.org/spec/services/1.0/req/ontology/format/skos/matches  matches/synoyms as one file |

#### Ontology Term Ranking Service Definition

ESRI Geoportal uses an ontology service. When a term is submitted to the ontology service, and a list of related terms and ranking will be returned in a JSON string. This is based on an OWL ontology (GEMET, at present), but it is possible to use any properly formatted ontology.. The ontology service utilizes Lucene full text indexing engine for its text matching, and ranking. The ESRI ontology service should be used to suggest terminology based on concepts served from an OWL-based ontology source.

|  |
| --- |
| Requirements Class |

|  |
| --- |
| his.cuahsi.org/spec/services/1.0/req/ontology/service |

|  |  |
| --- | --- |
| Target Type | Ontology |

|  |  |
| --- | --- |
| Name | Ontology Service |
| Dependency | his.cuahsi.org/spec/information/1.0/req/ontology |
| Requirement | his.cuahsi.org/spec/services/1.0/req/ontology/service/api  An ontology service should provide an API compatible with ESRI geoportal ontology service |

### Vocabularies

HIS controlled vocabulary services should expose terminology should use a REST API to expose terms from this system as fixed URLs, with distinct identifiers. These should be available as HTML, SKOS and GML Code lists, even if they are non-hierarchical controlled lists. SKOS is compatible with RDF and ontology logic processing applications. Term lists contained in a data sources, such as ODM’s should also be exposed via REST endpoints. Organizations, sources, quality control levels, data types, value types, methods, lab names should all be exposed, with identifiers based on local codes.

|  |
| --- |
| Requirements Class |

|  |
| --- |
| his.cuahsi.org/spec/services/1.0/req/vocabularies/api/services |

|  |  |
| --- | --- |
| Target Type |  |

|  |  |
| --- | --- |
| Name | Vocabulary service |
| Dependency |  |
| Requirement | his.cuahsi.org/spec/services/1.0/req/vocabularies/services/api/rest  Services should utilize a REST api |
|  | his.cuahsi.org/spec/services/1.0/req/vocabularies/services/identifiers  distinct identifiers need to be utilized |

|  |
| --- |
| Requirements Class |

|  |
| --- |
| his.cuahsi.org/spec/services/1.0/req/vocabularies/apiformat |

|  |  |
| --- | --- |
| Target Type | format |

|  |  |
| --- | --- |
| Name | vocabularies |
| Dependency |  |
| Requirement | his.cuahsi.org/spec/services/1.0/req/vocabularies/format/SKOS  SKOS shall be one format for vocabularies and term lists |
|  | his.cuahsi.org/spec/services/1.0/req/vocabularies/format/GmlCodelist  GML code lists shall be one format for vocabularies and term lists |

## Water Quality Information

There is lack of a standard for communicating Water Quality Information, and for Water Quality Services. The USGS and the US Environmental Protection Agency are working to provide interoperable services based on a standard called Water Quality Exchange (WQX). The services are interoperable in the sense that the outputs are consistent in format and nomenclature, and compliant with the WQX-Outbound 2.0 schema.

The USGS WQX web services provide a REST interface. Either method will retrieve the same data when the same retrieval parameters are specified. The basic building blocks for querying the services are similar (Table 8). The technical details are listed at: <http://qwwebservices.usgs.gov/technical-documentation.html>.

HIS clients should understand how to access the details of WXQ formatted information in order to determine details of the analytical analysis. WaterML 2.0 will provide links out to WQX services as relatedObservation (Figure 19)



Figure 20. XML fragment MeasurmentTVP (time Value Pair) with a related observation

|  |
| --- |
| Requirements Class |

|  |
| --- |
| his.cuahsi.org/spec/services/1.0/req/waterquality/ |

|  |  |
| --- | --- |
| Target Type | Water quality |

|  |  |
| --- | --- |
| Name | Analytical data |
| Dependency | his.cuahsi.org/spec/information/1.0/req/waterquality/service |
| Requirement | his.cuahsi.org/spec/services/1.0/req/waterquaiity/service/api  water quality data should be made available over a REST API as defined by the USGS and the EPA |
|  | his.cuahsi.org/spec/services/1.0/req/waterquaiity/format/wqx  water quality data should be delivered in WQX 2.0 outbound format as defiend by the EPA and the USGS |

Table 8. Uniform Resource Locators (URL) for the Water Quality web services based on the USGS standard

|  |  |  |
| --- | --- | --- |
| Web Service | Base URL [parameters must be appended to the REST URLs] | Purpose |
| Station | **REST** http://host/Station/search? | Retrieves data-collection station data |
| Result | **REST** http:// host /Result/search? | Retrieves water-quality result data |
| Domain values | **REST** http:// host /Codes/ | Retrieves choice lists for web-service arguments |

## Legacy service interfaces

For the near future, WaterOneFlow and OGC standards will be provided for HIS Water Web Services. The legacy service are defined as SOAP interfaces.

|  |  |
| --- | --- |
| Method | Parameters |
| GetVariableInfo | "VarVocab:code", "AuthToken" |
| GetSites | ["SiteVocab:Code","SiteVocab:Code2",...], "AuthToken" |
| GetSiteInfo | "SiteVocab:Code", "AuthToken" |
| GetValues | "SiteVocab:Code","VarVocab:code","YYYY-MM-DD","YYYY-MM-DD", "AuthToken" |

Table 9. CUAHSI Service Supporting Legacy WaterOneFlow Interface



## Functions to be fully specified at a future date

### Methods

HIS shall include a system to register methods, like NEMI.org. The proposal is to utilize methods documented at NEMI and other endpoints, and cross-reference them from the HIS stack.

### Themes

Themes shall use a REST API.

1. Themes shall be cataloged using INSPIRE Metadata for data series in Geoportal or other CSW
2. Themes should only be registered at one CWS server
3. Themes shall consist of a list of observational data series using the Water Observations Metadata Specification.
4. Timeseries shall be downloaded, and store with the theme.
5. Depending on the theme type, the theme may be static, or periodically updated
   1. A workflow will periodically update a non-static themes.
6. Themes will require curation, and a curation workflow will be developed.
7. Themes can be versioned. A new identifier shall be generated when a new version is generated.

#### Theme Services

Theme data should be exposed via standard OGC interfaces. Themes should be exposed via a WFS/WMS server, and styled to produce maps styled to emphasize the theme content. Managing the updating themes is in the workflow section.7.8.1 Ontology Mapping Automated

Based on the concepts, mapped variable names, and controlled vocabularies and attempt to match should be made to assist the user in mapping, and should be exposed as a service. We can extract information from the descriptive names. Rather than forcing users to comply to a single concept as the name rule, it might be better to allow for more descriptive names to be utilized. Figure 20 shows an example. Given a text variable description, ‘streamflow ft^3/sec daily’ we should be able convert this to a concept, a controlled variable name, and properties of a variable/series.

Rules:

* exact match to variable name
* exact match to a synonym for a variable
* partial match to variable name
* exact match to a mapped name

Figure 21. Converting variable description to variable properties

### Accessing THREEDS Service

Proxying THREEDS data services, to generate a catalog and provide access to these services is a future topic. Sections will need to include:

* Proxy Catalog via GI-Cat
  + Datasets to proxy
  + Retrieval
* Data access via WCS
* Data access via NetCDF

# HIS Node Types

The HIS system has several node types. A centralized catalog node is used to hold harvested metadata for discovery queries. A local catalog node is used by individual organizations to expose the metadata for their services

A Hydroserver node is the code data distribution component, and a cloud node, is a cloud hosted equivalent of a Hydroserver.

|  |  |  |
| --- | --- | --- |
| Type | Components | Software |
| Central Catalog Node | Database  HIS Central Services  CSW Services  WFS Services  Harvest  Authentication | MsSql  Custom  Geoportal  Custom  Ms Synchronization Framework  SingleProviderFramework |
| Local Catalog Node | Database  CSW Services  WFS Services | MsSql  Geoportal  ESRI |
| Hydroserver | Database  WFS Services  SOS  Leagcy WaterOneFlow  Authentication  Authorization | MsSql  Geoprotal  Custom  Custom  ?  ? |
| Cloud Data Node | Database  SOS Service  WFS Services | SQL Azure  Custom Code, or PySOS  ESRI? |

# Sustainability and governance: CUAHSI HIS in the system of earth science information

Governance model is a critical component in the management of an operational hydrologic information system. The model would define relationships between system partners, agreements, contracts and enforcement mechanisms, determine lifecycle management of system components (data, services, catalogs, identifiers, vocabularies, application clients, etc.), and specify authority, performance expectations, and risk management strategies. To date, governance issues have received relatively little attention in our previous work in the HIS project, beyond setting data use and data publishing agreements, and requesting citations when HIS data are used. By now, the project has accumulated extensive experience collaborating with data and service providers and with hydrologic data users, to start formulating governance principles. Outlining these principles and the main components of the governance model is the goal of this section.

To be successful, the governance model for a hydrologic information system must follow essential patterns of community organization with respect to:

* Data management responsibilities as established in the community and required by government regulations (for agencies) or academic rigor standards as expressed, for example, in NSF data management policies (for academic partners);
* Supporting local researchers in establishing and maintaining their observation networks, including identities of measurement sites, variables, methods, procedures, etc., while providing aggregation, validation and semantic mapping services to enable cross-network search and data integration;
* Enabling service providers and consumers to join the system with minimal initial investment and skill set, by using common standards-compliant software.
* Ensuring that hydrologic data and services maintained by the system can be reliably and easily consumed and integrated in applications developed in neighboring domains.

The scope of governance shall be clearly defined. Some components of the standards-based hydrologic information system will be governed by independent international standards bodies, such as the Open Geospatial Consortium. It is expected that the governance of water data exchange standards shall belong to such standards groups working in conjunction with projects like HIS to establish hydrology-related use cases and verify and validate proposed standards against such use cases.

In addition, governance arrangements extend beyond management of hydrologic data. Cross-domain research scenarios and designs are rapidly moving to scientific mainstream, as new research findings happen across traditional domain boundaries. Enabling cross-disciplinary and cross-jurisdictional research designs is within the purview of higher-level systems, such as the one envisioned in the NSF EarthCube initiative.

To understand the requirements of this higher-level infrastructure on hydrologic information system governance, we present a draft reference model of EarthCube. At the conceptual level, this model recognizes three main layers of cyberinfrastructure for the earth sciences: research observatory sites, domain infrastructure, and cross-domain information integration and knowledge management layer, or the EarthCube proper (Figure 21).

* Research observatory sites collect and analyze large volumes of observations of different types, both in situ and ex situ. In case of HIS, the research sites maintain HydroServers or otherwise publish their data as standards-compliant services, as described in this report and in the earlier Water Information Services Concept Development Study. In the CZO case, data are published as standard ASCII files, and harvested into the central repository. Governance arrangements for such sites are defined by a PI or a group of PIs, following their contracts with granting agencies. In the NSF model, these are separately funded and fairly independent efforts, with different science goals and scope. In order to participate in the larger EarthCube framework, the published data must comply with standard formats defined for this data type in domain information systems. Yet, new types of measurements are created at such sites, for which standard formats and data sharing arrangements are not yet defined: for such situations, the governance model shall specify how the new measurements are integrated in a domain information system such as HIS.
* Domain information systems. These systems manage common types of data generated by research sites, and support publication, discovery and access to the data across multiple sources, for multiple clients. Conceptually, they include four main infrastructure components for each data type: (a) community information model, (b) vocabularies, (c) uniform data access services, and (d) catalogs. These components enable interoperability within a domain system at several levels, needed for different research designs. To support simple dataset discovery, data of different types are registered in domain catalogs, with minimal metadata. At the next level, having the registered data conform with formal vocabularies enables data validation, semantics-based querying, and unambiguous interpretation of retrieved catalog and data records. Further, different data may be available via common service interfaces (e.g. SOS, as described earlier in this report), and conform with compatible information models (e.g. formal profiles based on OGC Observations and Measurements specification, such as WaterML 2.0). As long as each of the four components of the domain system is compliant with externally managed standards (e.g., in case of HIS: **CSW for catalogs; SKOS for vocabularies; SOS for service interfaces; and O&M profiles for information models**), these components can be made interoperable in a larger system such as EarthCube. From the governance model perspective, domain CI may present a fairly complex structure. In case of CUAHSI HIS, it includes many elements: a consortium of universities with a governing board; development teams; a user committee; operational and curation support of HydroServers and the central hydrologic metadata catalog; community consensus process about agreeing on domain models, vocabularies, data access protocols/services, and catalogs, which is leveraging the OGC framework for standardization and consensus building; agreements defining data publishing and data use rights and responsibilities, etc.
* Information integration and knowledge management layer is designed to support cross-cutting components of the earth systems research, taking advantage of the standards and best practices established at the domain level, in particular standard specifications for catalogs, vocabularies, services and information models. As long as these domain components can be accessed via standard interfaces described above, they can be integrated within the cross-domain knowledge integration layer. This layer would include: vocabulary cross-walks that establish correspondences between terms in domain vocabularies and support attribute-based data discovery and interpretation across disciplines; federated metadata catalogs that organize domain catalogs and enable discovery of resources across domains; integrated datasets that are compiled (and, ideally, curated) drawing data from several domains; persistent identifier management for cross-domain resources; high-performance and cloud-based compute facilities to manage and analyze large composite datasets; collaborative code development ecosystem; and a system of policies and governance ensuring that components work together and can be configured to address different research issues. Data provenance information available at this level, as well as social networking tools that link researchers from neighboring domains, are also critical in maintaining a level of confidence in data obtained from datasets that were created outside a researcher’s immediate area of expertise.

In a complex governance structure like this, policies need to be developed to orchestrate consensus-building and regulate potential conflicts at the boundaries of EarthCube subsystems. They would define, for example, what should happen when new or enhanced scientific data types and formats are developed and propagated to domain systems and the cross-domain knowledge integration layer; what should happen if a sensor network managed by an observatory needs to be reconfigured by another group (perhaps from another domain) to address their research problems; how data life cycle arrangements are coordinated across domains (since data collected in one domain often provide context for data from another domain).

The experience of OGC points to a successful governance model, where communities of practice (organized as OGC domain working groups, with members representing different organizations) develop and present specifications for data description and exchange protocols. Development of a standard specification is preceded by carefully defining the scope and use cases it would address, describing how it will relate to other existing and proposed standards, and how it can be extended. The proposed specifications are shared with larger OGC membership, who are requested to comment on them, and go through a series of approvals before coming to a final vote. In this way, OGC maintains an open, transparent and formal process for bringing standards to the community and orchestrating their discussion, refinement, compliance testing, approbation in various projects, and eventual community adoption.

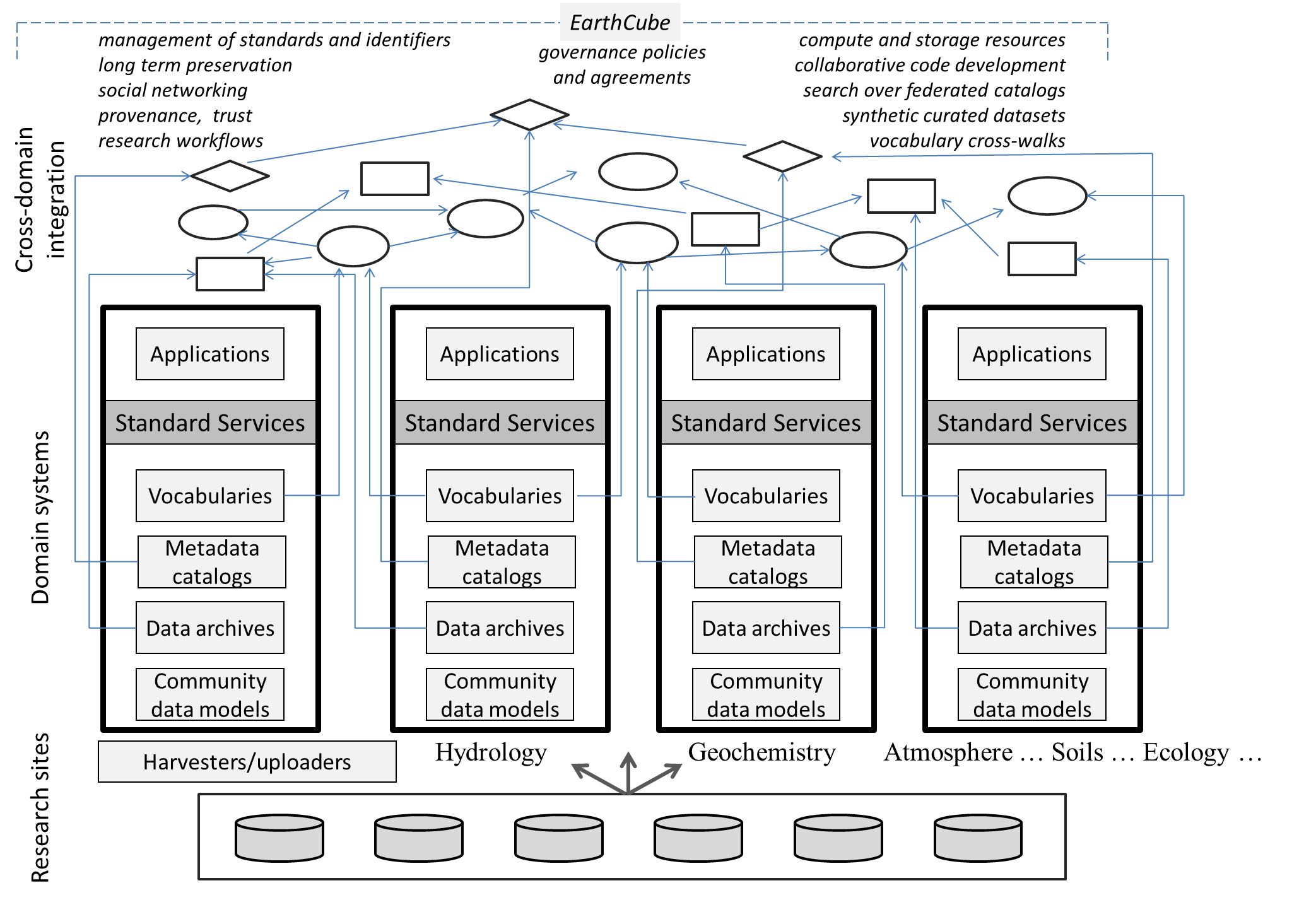


Figure 22. Vision of a reference architecture for EarthCube as is an integrated information system that includes research observatories generating large volumes of observations, domain systems (such as HIS) that publish the data according to community conventions about data models, vocabularies and protocols, and cross-domain knowledge layer that includes federated catalogs, normalized and curated datasets integrating data from domain systems and observatories, vocabulary cross-walks, as well as social networking, governance and compute infrastructure.

Bibliography

, David. 2011. Water Information Services Concept Development Study. OGC® Engineering Report . 83pp

Fuest, Stefan. 2011. SOS Usage Profile for the Hydrology Domain. Proposed OGC Discussion Paper.

INSPIRE Thematic Working Group Hydrography. 2009. D2.8.I.8 INSPIRE Data Specification on Hydrography – Guidelines <http://external.opengis.org/twiki_public/pub/HydrologyDWG/FeatureModels/INSPIRE_DataSpecification_HY_v3.0.pdf>

Isaac, Antoine and Summers, Ed. 2009. SKOS Simple Knowledge Organization System Primer. <http://www.w3.org/TR/skos-primer>

Jupp, Simon, Bechhofer, Sean, and Stevens, Robert. 2008. SKOS with OWL: Don’t be Full-ish! <http://www.webont.org/owled/2008/papers/owled2008eu_submission_22.pdf>

Looser, Ulrich. 2010. **Extracts from the reports of the CBS expert teams related to the use of keywords in the implementation of the WMO core profile of the ISO metadata standard.** <http://external.opengis.org/twiki_public/pub/HydrologyDWG/VocaBularies/keywords.doc>

Microsoft. 2011. Access Control. <http://www.microsoft.com/windowsazure/features/accesscontrol/>

National Environmental Information Exchange Network. 2011. Water Quality Exchange (WQX) <http://www.exchangenetwork.net/exchanges/water/wqx.htm>

Schwering, Angela 2008. Approaches to Semantic Similarity Measurement for Geo-Spatial Data: A Survey.Transactions in GIS, 2008, 12(1): 5–29.

US Geological Survey, 2011. National Water Information System (NWIS) Water-Quality Web Services. <http://qwwebservices.usgs.gov/technical-documentation.html>

Appendix 1. HIS 1.0 to CUAHSI OGC Mapping

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | HIS 1.0 |  | CUAHSI OGC |  |  |
| Component |  | Component | Format | Service | DataFormat | Software |
| Federated Metadata Search |  |  |  | CSW |  | Gi-Cat |
| Metadata Server | For service endpoints | HIS Central | Custom | CSW | Dublin Core | Geoportal |
|  |  |  |  | CSW | Inspire Service Metadata | Geoportal |
| Metadata Server | For time series data services | HIS Central | Custom | WFS | Simple GML Datacart | ESRI/Geonetwork/HIS |
|  |  | WFS | Simple GML Datacart | CSW |  |  |
| Ontology |  | HIS Central | Custom | ESRI Ontology Ranking | Owl |  |
|  |  |  |  | TBD | External- SKOS. Internal-Table | |
| Vocabulary Services |  | HIS Vocabulary | Custom | TBD | External- SKOS |  |
| Data Server | For time series from network of observation sites | WaterOneFlow | WaterML | SOS | WaterML 2-Timeseries |  |
| Data Server | For time series from grid cells |  |  | SOS |  |  |
|  |  |  |  |  |  |  |
| Data Server | For Reference Features |  |  | WFS | WaterML 2- Hydrologic Features | |
|  |  |  |  | WMS |  |  |
|  |  |  |  |  |  |  |
| Metadata Server | For dynamic & static maps. For map-visualization of all data services, even WFS. |  |  | CSW | Dublin Core | ESRI Geoportal |
|  |  |  |  | CSW | Inspire Service Metadata |  |
|  |  |  |  | THREDDS CSW Proxy |  | Gi-Cat |
|  |  |  |  |  |  |  |
| Data Server | For multidimensional arrays |  |  | WCS |  |  |
|  |  |  |  | WFS |  |  |
|  |  |  |  | THREDDS | NetCDF | THREEDS |
| Data Server | For static (GIS) data sets, not supported by open standards-based exchange | FTP |  | FTP |  |  |
|  |  |  |  |  |  |  |
| Authorization |  |  |  |  |  | Azure Access Control |
| Authentication |  |  |  |  |  | Oauth-wrap |
|  |  |  |  |  |  |  |
| Processing |  |  |  | WPS |  | Deegree |
|  |  |  |  | ESRI REST Geoprocessing Proposal | | ESRI ArcGIS |
|  |  |  |  |  |  |  |
| Data Synchronization | For time series metadata services |  |  | MS Sync Framework |  | MS Sync Framework |
|  | For Time Series Data to Azure Cloud |  |  | MS Sync Framework |  | MS Sync Framework |
|  |  |  |  |  |  |  |
| Public Exposure |  |  |  | ArcGIS Online | Package |  |
|  |  |  |  | Azure Marketplace | Data Offering |  |
|  |  |  |  | DataOne |  | DataOne |

Appendix 3: Water Observations Metadata Specification

The specification that follows is for the fields of a feature class that contains one record for each time series. Required fields are written in **bold text.** The remaining fields are there to help you assess the time series before actually requesting the data. For example, you might only want to download time series for observations recorded using a specific measurement method. In order to get a better sense of the data, you may also one to perform some quick summaries on the fields in the observations metadata table to identify all of the unique methods used or to determine all of the unique site locations (handy when a site measures more than one variable and thus appears in the table more than once).

Table 10 Observations Metadata: Core Fields

|  |  |  |  |
| --- | --- | --- | --- |
| Field Name  (Field Type) | Definition | Examples | USGS |
| RecordType | The type of information record. There are presently three, ObservationsCore, and two extensions for Generic ODM services and DataSet Services (Modis and DayMet) | ObservationsBrief  ObservationsODM  ObservationsWOFDataSet |  |
| OrgHier | Organizational Hierarchy. To be used for defining a hierarchy of services. Use the Java Inverted domain. | Gov.USGS  EDU.Texas | Gov.USGS |
| ServCode  (Text - 50) | Network prefix for site codes used by the WaterOneFlow service, giving the context within which the site code applies | CCBay  NWISDV | NWISDV |
| SiteCode  (Text - 50) | Unique text identifier for a site within a given WaterOneFlow service. For the USGS and EPA, an agency is bonded with the site number | H1 | USGS:02289050  FL005:02289050  (Different sites) |
| SiteName  (Text - 255) | Name of a site | Hypoxia\_1 |  |
| SiteType | Site Type. Defined by both the USGS, and the EPA. |  | Surface Water, Ground |
| Latitude  (Double) | Latitude of the site location in decimal degrees (WGS\_1984); for polygons can be *NULL* | 27.814 |  |
| Longitude  (Double) | Longitude of the site location in decimal degrees (WGS\_1984); for polygons can be *NULL* | -97.141 |  |
| Elevation | Elevation of site with units. This is needed for variables where parameter is observed referenced to the ground surface. | 1500 m | 37.9 ft |
| VarCode  (Text - 50) | Unique text identifier for a variable within a given WaterOneFlow service. DataProviders should create distinct codes. | DOC | 00065  00060:00003 |
| VarName  (Text - 255) | Name of a variable | Dissolved Oxygen Concentration | Stream water level elevation above NAVD 1988, in feet,Upstream |
| VarUnits  (Text -– 50) | Units of measure for the variable. Encouraged, but optional, since some providers for analytical chemistry may wish for all information | milligrams per liter | ft |
| DataType  (Text – 50) | Type of data. See the [CUAHSI Controlled Vocabulary](http://his.cuahsi.org/mastercvreg/edit_cv11.aspx?tbl=DataTypeCV&id=485576768). If needed request an additional DataType | Value, Average, Maximum, Minimum, StandardDeviation |  |
| Medium  (Text – 50) | Medium in which the variable applies. See [CUAHSI Controlled Vocabulary](http://his.cuahsi.org/mastercvreg/edit_cv11.aspx?tbl=SampleMediumCV&id=533576939) . If needed request an additional DataType | Surface Water |  |
| Vocabulary  (Text – 50) | Vocabulary prefix for variable codes giving the context within which the code applies | CCBay | NWISDV  NWISIID |
| Ontology  (Text – 50) | 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 | CUAHSI Variable Ontology v1.26 | SRS |
| Concept  (Text – 50) | Leaf concept keyword from the ontology to which this variable applies | dissolvedOxygen | Stream Stage |
| SerStatus | Series Status. Active, Inactive, Sporadic.  If Inactive, StartDate and EndDate are populated  If Active, and all data is available, EndDate should be null  If Active, and Data is available for a limited time, both StartDate and End Date are Null. | Inactive | Active |
| DataAvail | Limited Data Availability. If the series information is only available for a period of time, eg 120Days |  | P120D |
| IsRegular  (ShortInt) | 1 (TRUE) if variable is measured/calculated regularly in time; 0 (FALSE) otherwise | 0 | 1 |
| TimeStep | For regular data, the time step and time units give the length of time between measurements, e.g., 1 day, 6.5 hrs, 1 month. Represent as ISO Duration (P1D) Estimated, it ok. USGS does not know what the normal sampling interval. | P1D (one Day)  P1M (one Month)  PT12H (Time One Hour)  PT15M (Time 15 Minutes) | P1D  PT15M |
| StartDate  (Date) | 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 null or empty, and the value should be calculated as Now minus the DataAvail. | 5/3/94 8:40 AM |  |
| EndDate  (Date) | 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. | 8/31/06 11:26 AM |  |
| ValueCount  (LongInt) | Number of time series values for the variable at the site for the given time period | 270 |  |
| Last Updated | DataCart was Last Updated. This will allow for querying the WFS for when a record was last update | 2010-11-03 | 2010-11-03 |
| MaxRecords | 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 empty/null = unlimited. Optional | Null | null |
| ServType | Service Protocol Type. Rest or Soap. | SOAP | REST |
| Location  (Text – 255) | Properly formatted location parameter to pass to WaterOneFlow.GetValues SOAP request | CCBay:Hypoxia\_1  GEOM:BOX(-97.141 27.814,-93.5 30.2)  GEOM:POINT(-97.141 27.814) | NWISDV: FL005:02289050  NWISDV: USGS:02289050  (future) |
| Variable  (Text – 255) | Properly formatted variable parameter to pass to WaterOneFlow.GetValues SOAP Request | CCBay:DOC  NWISDV:00060/DataType=Maximum | NWISDV:00060:0003  NWISUV:00060 |
| ReqsAuth  (ShortInt) | Request authorization. 1 (TRUE) if authorization for download is required; 0 (FALSE) otherwise | 0 | 0 |
| WaterMLURI  (Text – 255) | URI of WaterOneFlow service WSDL or REST.  For a REST Service this will be a complete URL, with tokens to indicate where the start and end times can be substituted. {time:start} {time:end} | http://data.com/WoF/ /cuahsi\_1\_0.asmx?WSDL | http://waterservices.usgs.gov/nwis/iv?sites=01646500& startDT={time:start}&endDT={time:end}&parameterCd=00060 |
| WofVersion  (Text – 15) | Version of the WaterOneFlow service | 1.0 | 1.1 |
| SitesWFSURI  (Text – 255) | GetCapabilities URI of web feature service showing site locations | http://data.com/WFSServer&service=WFS&request=GetCapabilities |  |
| SiteWMSURI  (Text – 255) | URI of web mapping service related to the data | http://data.com/WMSServer&service=WMS&request=GetCapabilities |  |
| DCrtWFSUri (Text – 255) | GetCapabilities URI of the WFS Data Cart Service that contains the series | http://data.com/DataService&service=WFS&request=GetCapabilities |  |

Table 2. Observations Metadata: Additional Fields for Services from CUAHSI ODM

|  |  |  |
| --- | --- | --- |
| Field Name  (Field Type) | Definition | Examples |
| MethodID | Unique ID within a WaterOneFlow service for the method used to measure the variable | 1 |
| Method | Description of the method used to measure the variable | Multiprobe measurement |
| QCLevelID | Unique ID within a WaterOneFlow service for the quality control level of the time series | 0 |
| QCLevel | Description of the quality control level of the time series | Raw Data |
| SourceID | Unique ID within a WaterOneFlow service for the original source of the data | 9 |
| SourceName  (Text – 255) | Name of the original source of the data | Texas A&M University Corpus Christi |

Table 3. Additional Fields for DataCartWOFDataset Record Type

|  |  |  |
| --- | --- | --- |
| LocType  (Text – 25) | Type of service – indicates how the Location parameter of a WaterOneFlow.GetValues call should be formatted | LatLongBox  LatLongPoint |
| XLL  (Double) | For point data, Longitude of the point. For data defined by a lat/lon box, western longitude of the box | -97.141 |
| YLL  (Double) | For point data, Latitude of the point. For data defined by a lat/lon box, southern latitude of the box | 27.814 |
| XUR  (Double) | For data defined by a lat/lon box, eastern longitude of the box; otherwise can be *NULL* | -93.5 |
| YUR  (Double) | For data defined by a lat/lon box, northern latitude of the box; otherwise can be *NULL* | 30.2 |

Appendix 4: Feature Data Models

## Appendix 4.1. INSPIRE Hydrography Data Model

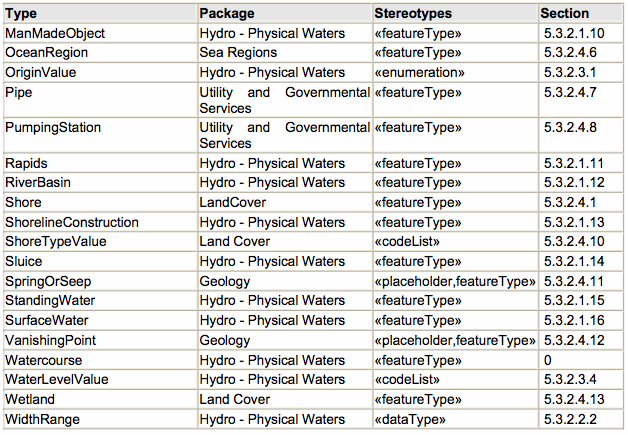
The INSPIRE hydrography data model is an application schema conformant to ISO UML modeling rules to be used as part of INSPIRE. The model consists of three main sub models:

Hydro - Networks: to be used in GIS and modeling

Hydro - Physical Waters: to be used for mapping

Hydro - Reporting: to be used for reporting

As part of the model a number of hydro features are defined. The features are copied from the specification below.

For interoperability, we should utilized the D2.8.I.8 INSPIRE Data Specification on Hydrography – Guidelines (detailed above).

Appendix 5. Test Cases

Test compliance procedures need to be detailed by developing a subset of the information expected to be in the HIS infrastructure. This will be utilized to insure that the interaction of the systems components is correct. For search and discovery this need to include both retrieval (you find the records desired), but also that non-applicable records are not returned.

We suggest that test cases for the following components be developed.

## A 5.1 Catalog

### A 5.1.1 Harvesting

#### A 5.1.1.1 Services Catalog

#### A 5.1.1.1 Observational Catalog Harvesting

### A 5.1.2 Retrieval

#### A 5.1.1.1 Services Catalog

#### A 5.1.1.1 Observational Catalog Harvesting

## A 5.2 Observational Data for Point Services

#### Sites

#### Series

#### Observations

#### Controlled Terminology

## A 5.3 Vocabularies and Ontology

* Support and enforce compliance with HIS controlled terminology

Appendix 6: XML Examples of Filters

### CSW Filters

Use case: Temporal Filter

<csw:GetRecords xmlns:csw="http://www.opengis.net/cat/csw/2.0.2" version="2.0.2" service="CSW"  
 resultType="results" startPosition="1" maxRecords="10">  
 <csw:Query typeNames="csw:Record" xmlns:ogc="http://www.opengis.net/ogc"  
 xmlns:gml="http://www.opengis.net/gml">  
 <csw:ElementSetName>full</csw:ElementSetName>  
 <csw:Constraint version="1.1.0">  
 <ogc:Filter>  
 <ogc:AND>  
 <ogc:PropertyIsLessThan>  
 <ogc:PropertyName>EndDate</ogc:PropertyName>  
 <ogc:Literal>2010-01-01</ogc:Literal>  
 </ogc:PropertyIsLessThan>  
 <ogc:PropertyIsGreaterThan>  
 <ogc:PropertyName>StartDate</ogc:PropertyName>  
 <ogc:Literal>2009-01-01</ogc:Literal>  
 </ogc:PropertyIsGreaterThan>  
 </ogc:AND>  
 </ogc:Filter>  
 </csw:Constraint>  
 <ogc:SortBy>  
 <ogc:SortProperty>  
 <ogc:PropertyName>dc:title</ogc:PropertyName>  
 <ogc:SortOrder>ASC</ogc:SortOrder>  
 </ogc:SortProperty>  
 </ogc:SortBy>  
 </csw:Query>  
</csw:GetRecords>

Use case: Recent Records

<csw:GetRecords xmlns:csw="http://www.opengis.net/cat/csw/2.0.2" version="2.0.2" service="CSW"  
 resultType="results" startPosition="1" maxRecords="10">  
 <csw:Query typeNames="csw:Record" xmlns:ogc="http://www.opengis.net/ogc"  
 xmlns:gml="http://www.opengis.net/gml">  
 <csw:ElementSetName>full</csw:ElementSetName>  
 <csw:Constraint version="1.1.0">  
 <ogc:Filter>  
 <ogc:PropertyIsGreaterThan>  
 <ogc:PropertyName>dct:modified</ogc:PropertyName>  
 <ogc:Literal>2000-01-01</ogc:Literal>  
 </ogc:PropertyIsGreaterThan>  
 </ogc:Filter>  
 </csw:Constraint>  
 <ogc:SortBy>  
 <ogc:SortProperty>  
 <ogc:PropertyName>dc:title</ogc:PropertyName>  
 <ogc:SortOrder>ASC</ogc:SortOrder>  
 </ogc:SortProperty>  
 </ogc:SortBy>  
 </csw:Query>  
</csw:GetRecords>

### WFS filters

Use Case: find Feature with concept = ‘nitrogen’

<Filter>  
 <AND>  
 <BBOX>  
 <PropertyName>NAME</PropertyName>  
 <Box srsName="EPSG:4327">  
 <coordinates>-115.2239,44.4879 115.8578,44.8471</coordinates>  
 </Box>  
 </BBOX>  
 <PropertyIsEqualTo>  
 <PropertyName>Concept</PropertyName>  
 <Literal>nitrogen</Literal>  
 </PropertyIsEqualTo>  
 </AND>  
</Filter>

Temporal Filter

<ogc:Filter>  
 <ogc:AND>  
 <ogc:PropertyIsLessThan>  
 <ogc:PropertyName>EndDate</ogc:PropertyName>  
 <ogc:Literal>2010-01-01</ogc:Literal>  
 </ogc:PropertyIsLessThan>  
 <ogc:PropertyIsGreaterThan>  
  
 <ogc:PropertyName>StartDate</ogc:PropertyName>  
 <ogc:Literal>2009-01-01</ogc:Literal>  
 </ogc:PropertyIsGreaterThan>  
 </ogc:AND>  
</ogc:Filter>

#### SOS filters

Use Case 1: Retrieve data for a variable over an area

<sos:GetObservation xmlns="http://www.opengis.net/sos/2.0" service="SOS" version="2.0.0"  
 xmlns:sos="http://www.opengis.net/sos/2.0" xmlns:fes="http://www.opengis.net/fes/2.0"  
 xmlns:gml="http://www.opengis.net/gml/3.2" xmlns:swe="http://www.opengis.net/swe/2.0"  
 xmlns:swes="http://www.opengis.net/swes/2.0" xmlns:xlink="http://www.w3.org/1999/xlink"  
 xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"  
 xsi:schemaLocation="http://www.opengis.net/sos/2.0  
 http://schemas.opengis.net/sos/2.0.0/sos.xsd">  
  
 <!--identifier of an offering-->  
 <offering> precipitation </offering>  
  
 <!--identifier of an observed property-->  
 <observedProperty>urn:ogc:def:phenomenon:OGC: precipitation </observedProperty>  
  
 <!--identifier of a procedure-->  
 <procedure>http://www.my\_namespace.org/sensors/Water\_Gage\_1</procedure>  
  
 <!--the observations returned shall match the spatial filter defined in this request (the spatial property defined in the ValueReference element must be within the passed polygon)-->  
 <spatialFilter>  
 <fes:Within>  
 <fes:ValueReference>om:featureOfInterest/sams:SF\_SpatialSamplingFeature/sams:shape</fes:ValueReference>  
 <gml:Polygon gml:id="Muenster" srsName="urn:ogc:def:crs:EPSG:4326">  
 <gml:exterior>  
 <gml:LinearRing>  
 <gml:posList> 52.90 7.52 52.92 7.51 52.96 7.54 52.90 7.52 </gml:posList>  
 </gml:LinearRing>  
 </gml:exterior>  
 </gml:Polygon>  
 </fes:Within>  
 </spatialFilter>  
</sos:GetObservation>

Use Case 2: Get Last (N) observations

Not supported. Client will need to calculate a time period of the request

<GetObservation xmlns="http://www.opengis.net/sos/1.0"  
 xmlns:ows="http://www.opengis.net/ows/1.1"   
 xmlns:gml="http://www.opengis.net/gml"  
 xmlns:ogc="http://www.opengis.net/ogc"  
 xmlns:om="http://www.opengis.net/om/1.0"  
 xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"  
 xsi:schemaLocation="http://www.opengis.net/sos/1.0  
 http://schemas.opengis.net/sos/1.0.0/sosGetObservation.xsd"  
 service="SOS" version="1.0.0" srsName="urn:ogc:def:crs:EPSG:4326">  
   
 <offering>precipitation</offering>  
   
 <eventTime>  
 <ogc:TM\_During>  
 <ogc:PropertyName>om:samplingTime</ogc:PropertyName>  
 <gml:TimePeriod>  
 <gml:beginPosition>2011-01-01</gml:beginPosition>  
 <gml:endPosition>2011-02-01</gml:endPosition>  
 </gml:TimePeriod>  
 </ogc:TM\_During>  
 </eventTime>  
   
 <observedProperty> urn:OGC:def:property:OGC::Precipitation</observedProperty>  
 <responseFormat>text/xml;subtype=" waterml/2.0.0"</responseFormat>  
</GetObservation>

Use Case 3: Get Last Measurement

Request by <gml:timePosition>latest</gml:timePosition> not supported by all SOS servers..

<GetObservation xmlns="http://www.opengis.net/sos/1.0"  
 xmlns:ows="http://www.opengis.net/ows/1.1"   
 xmlns:gml="http://www.opengis.net/gml"  
 xmlns:ogc="http://www.opengis.net/ogc"  
 xmlns:om="http://www.opengis.net/om/1.0"  
 xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"  
 xsi:schemaLocation="http://www.opengis.net/sos/1.0  
 http://schemas.opengis.net/sos/1.0.0/sosGetObservation.xsd"  
 service="SOS" version="1.0.0" srsName="urn:ogc:def:crs:EPSG:4326">  
   
 <offering>precipitation</offering>  
   
 <eventTime>  
 <ogc:TM\_Equals>  
 <ogc:PropertyName>om:samplingTime</ogc:PropertyName>  
 <gml:TimeInstant>  
 <gml:timePosition>latest</gml:timePosition>  
 </gml:TimeInstant>  
 </ogc:TM\_Equals>  
 </eventTime>  
   
 <observedProperty> urn:OGC:def:property:OGC::Precipitation</observedProperty>  
 <responseFormat>text/xml;subtype=" waterml/2.0.0"</responseFormat>  
</GetObservation>

Use case 4: retrieve by offering identifier

<GetObservation xmlns="http://www.opengis.net/sos/1.0"  
 xmlns:ows="http://www.opengis.net/ows/1.1"   
 xmlns:gml="http://www.opengis.net/gml"  
 xmlns:ogc="http://www.opengis.net/ogc"  
 xmlns:om="http://www.opengis.net/om/1.0"  
 xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"  
 xsi:schemaLocation="http://www.opengis.net/sos/1.0  
 http://schemas.opengis.net/sos/1.0.0/sosGetObservation.xsd"  
 service="SOS" version="1.0.0" srsName="urn:ogc:def:crs:EPSG:4326">  
   
 <offering> precipitation </offering>  
   
 <observedProperty> urn:OGC:def:property:OGC::Precipitation</observedProperty>  
 <observedProperty> urn:OGC:def:property:OGC::Precipitation</observedProperty>  
 <responseFormat>text/xml;subtype="waterml/2.0.0"</responseFormat>  
</GetObservation>

Use Case 5: get by Feature of Interest:

<sos:GetObservation xmlns="http://www.opengis.net/sos/2.0" service="SOS" version="2.0.0" xmlns:sos="http://www.opengis.net/sos/2.0" xmlns:fes="http://www.opengis.net/fes/2.0" xmlns:gml="http://www.opengis.net/gml/3.2" xmlns:swe="http://www.opengis.net/swe/2.0" xmlns:swes="http://www.opengis.net/swes/2.0" xmlns:xlink="http://www.w3.org/1999/xlink" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://www.opengis.net/sos/2.0  
 http://schemas.opengis.net/sos/2.0.0/sos.xsd">  
   
 <!--identifier of an offering-->  
 <offering>http://www.my\_namespace.org/water\_gage\_1\_observations</offering>  
   
 <!--identifier of an observed property-->  
 <observedProperty>urn:ogc:def:phenomenon:OGC:water\_level</observedProperty>  
   
 <!--identifier of a procedure-->  
 <procedure>http://www.my\_namespace.org/sensors/Water\_Gage\_1</procedure>  
   
 <featureOfInterest>http://wfs.example.org?request=getFeature&amp;featureid="river1"</featureOfInterest>  
   
</sos:GetObservation>

Appendix 7 SKOS

The Simple Knowledge Organization System (SKOS) is a specification for representing semi-formal knowledge organization systems (KOSs), such as thesauri, taxonomies, classification schemes and subject heading lists (Isaac, and Ed Summers, 2011). The aim of SKOS is not to replace original conceptual vocabularies in their initial context of use, but to allow them to be ported to a shared space, based on a simplified model, enabling wider re-use and better interoperability (Isaac, and Ed Summers, 2011). SKOS uses thesauri like structure where terms can have multiple names, or labels, and a broader of narrow reference to another term. In addition SKOS supports the definition of exact matches, and close matches (synonyms). Exact Match is used to link two concepts, indicating a high degree of confidence that the concepts can be used interchangeably across a wide range of information retrieval applications. In order to avoid governance issues, best practice recommend to maintain the mapping relationships between two SKOS concepts in a different file from the concepts themselves (CSIRO wiki, 2011).

<skos:Concept rdf:about="urn:cuahsi.org/vocabulary/samplemedia#water">  
 <skos:prefLabel xml:lang="en">water</skos:prefLabel>  
 <skos:altLabel xml:lang="en">water</skos:AltLabel>   
 <skos:altLabel xml:lang="en">waters</skos: AltLabel >  
 <skos:prefLabel xml:lang="es">agua</skos:prefLabel>  
 <skos:altLabel xml:lang="es"> agua </skos: AltLabel >  
</skos:Concept>

Best practice says that links should be maintained separately from the concepts (CSIRO,2011).

<rdf:Description rdf:about="urn:Cuahsi.org/ontology#nitrogen">  
 <skos:broader rdf:resource=" urn:Cuahsi.org/ontology#Chemistry"/>  
</rdf:Description>  
  
<rdf:Description rdf:about=" urn:Cuahsi.org/ontology#nitrogen ">  
 <skos:broader rdf:resource=" urn:user1.Cuahsi.org/ontology#nutrients "/>  
</rdf:Description>  
  
<rdf:Description rdf:about=" urn:Cuahsi.org/ontology#nitrogen ">  
 <skos:narrower rdf:resource=" urn:Cuahsi.org/ontology#nitrogen+NH4 "/>  
 <skos:narrower rdf:resource=" urn:Cuahsi.org/ontology#nitrogen+NH3"/>  
</rdf:Description>

Exact match and Close matches (aka Synonyms)

<rdf:Description rdf:about=" urn:Cuahsi.org/ontology#discharge ">  
 <skos:exactMatch rdf:resource=" urn:GEMET/TERM/Streamflow "/>  
</rdf:Description>

# Appendix: Using Azure for centralized community data hosting

We could use the capabilities of Microsoft Azure to host data from observations data model databases in the cloud. The capabilities of Azure include capabilities to sync and host data and services, and a set of authentication services. Having such a system would mean that we could keep all services up to date with the latest software.

This would require cloud hosting of a revised version of the web services that would work over Sql Azure databases. A data source would be synchronized to Sql Azure via data sync. This would require a revised version of the web services that would key off the Network ID. A brief outline of the w

Azure Capabilities

* IIS
* SQL Azure
* Authentication
* Data Sync

Workflow

* On a Hydroserver
  + Download sync agent
* On Azure,
  + Create New database (best practice, use NetworkCode)
  + Configure sync to access hydroserver database
* Register network code and database into services listing
* On Azure Generic WS,
  + When request comes in use connection for that service
  + If no known service throw error
* Future
  + Recode to use some form of authorization to get access to a particular service/series
  + Service to register which would create the DB, and setup the sync

Tasks/Actions

* Test azure sync
* Add code to generic WS to use different connections
* Add code to read services, and redirect

Appendix Examples of Inspire Metadata Records

## Service:

Mapping

<?xml version="1.0" encoding="UTF-8"?>  
<gmd:MD\_Metadata  
 xsi:schemaLocation="http://www.isotc211.org/2005/srv http://schemas.opengis.net/iso/19139/20060504/srv/srv.xsd"  
 xmlns:srv="http://www.isotc211.org/2005/srv" xmlns:gmd="http://www.isotc211.org/2005/gmd"  
 xmlns:gco="http://www.isotc211.org/2005/gco"  
 xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns:gml="http://www.opengis.net/gml"  
 xmlns:xlink="http://www.w3.org/1999/xlink">  
 <gmd:fileIdentifier>  
 <gco:CharacterString>6bbf8bf9-dbba-47b8-baa0-e64e78e38d29</gco:CharacterString>  
 </gmd:fileIdentifier>  
 <gmd:language>  
 <gmd:LanguageCode  
 codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO\_19139\_Schemas/resources/Codelist/ML\_gmxCodelists.xml#LanguageCode"  
 codeListValue="eng">eng</gmd:LanguageCode>  
 </gmd:language>  
 <gmd:hierarchyLevel>  
 <gmd:MD\_ScopeCode  
 codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO\_19139\_Schemas/resources/Codelist/ML\_gmxCodelists.xml#MD\_ScopeCode"  
 codeListValue="service">service</gmd:MD\_ScopeCode>  
 </gmd:hierarchyLevel>  
 <gmd:contact>  
 <gmd:CI\_ResponsibleParty>  
 <gmd:organisationName>  
 <gco:CharacterString>CUAHSI</gco:CharacterString>  
 </gmd:organisationName>  
 <gmd:contactInfo>  
 <gmd:CI\_Contact>  
 <gmd:address>  
 <gmd:CI\_Address>  
 <gmd:electronicMailAddress>  
 <gco:CharacterString>his@sdsc.edu</gco:CharacterString>  
 </gmd:electronicMailAddress>  
 </gmd:CI\_Address>  
 </gmd:address>  
 </gmd:CI\_Contact>  
 </gmd:contactInfo>  
 <gmd:role>  
 <gmd:CI\_RoleCode  
 codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO\_19139\_Schemas/resources/Codelist/ML\_gmxCodelists.xml#CI\_RoleCode"  
 codeListValue="pointOfContact">pointOfContact</gmd:CI\_RoleCode>  
 </gmd:role>  
 </gmd:CI\_ResponsibleParty>  
 </gmd:contact>  
 <gmd:dateStamp>  
 <gco:Date>2010-11-10</gco:Date>  
 </gmd:dateStamp>  
 <gmd:metadataStandardName>  
 <gco:CharacterString>ISO19119</gco:CharacterString>  
 </gmd:metadataStandardName>  
 <gmd:metadataStandardVersion>  
 <gco:CharacterString>2005/PDAM 1</gco:CharacterString>  
 </gmd:metadataStandardVersion>  
 <gmd:identificationInfo>  
 <srv:SV\_ServiceIdentification>  
 <gmd:citation>  
 <gmd:CI\_Citation>  
 <gmd:title>  
 <gco:CharacterString>WaterOneFlow Web Service</gco:CharacterString>  
 </gmd:title>  
 <gmd:date>  
 <gmd:CI\_Date>  
 <gmd:date>  
 <gco:DateTime>2010-11-10T09:08:01</gco:DateTime>  
 </gmd:date>  
 <gmd:dateType>  
 <gmd:CI\_DateTypeCode  
 codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO\_19139\_Schemas/resources/Codelist/ML\_gmxCodelists.xml#CI\_DateTypeCode"  
 codeListValue="creation">creation</gmd:CI\_DateTypeCode>  
 </gmd:dateType>  
 </gmd:CI\_Date>  
 </gmd:date>  
 <gmd:date>  
 <gmd:CI\_Date>  
 <gmd:date>  
 <gco:Date>2010-11-03</gco:Date>  
 </gmd:date>  
 <gmd:dateType>  
 <gmd:CI\_DateTypeCode  
 codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO\_19139\_Schemas/resources/Codelist/ML\_gmxCodelists.xml#CI\_DateTypeCode"  
 codeListValue="publication">publication</gmd:CI\_DateTypeCode>  
 </gmd:dateType>  
 </gmd:CI\_Date>  
 </gmd:date>  
 <gmd:date>  
 <gmd:CI\_Date>  
 <gmd:date>  
 <gco:DateTime>2010-11-10T09:07:56</gco:DateTime>  
 </gmd:date>  
 <gmd:dateType>  
 <gmd:CI\_DateTypeCode  
 codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO\_19139\_Schemas/resources/Codelist/ML\_gmxCodelists.xml#CI\_DateTypeCode"  
 codeListValue="revision">revision</gmd:CI\_DateTypeCode>  
 </gmd:dateType>  
 </gmd:CI\_Date>  
 </gmd:date>  
 <gmd:identifier>  
 <gmd:RS\_Identifier>  
 <gmd:code>  
 <gco:CharacterString>WOF\_11</gco:CharacterString>  
 </gmd:code>  
 <gmd:codeSpace>  
 <gco:CharacterString>his.cuahsi.org/waterml11</gco:CharacterString>  
 </gmd:codeSpace>  
 </gmd:RS\_Identifier>  
 </gmd:identifier>  
 </gmd:CI\_Citation>  
 </gmd:citation>  
 <gmd:abstract>  
 <gco:CharacterString>test</gco:CharacterString>  
 </gmd:abstract>  
 <gmd:pointOfContact>  
 <gmd:CI\_ResponsibleParty>  
 <gmd:organisationName>  
 <gco:CharacterString>SDSC</gco:CharacterString>  
 </gmd:organisationName>  
 <gmd:contactInfo>  
 <gmd:CI\_Contact>  
 <gmd:address>  
 <gmd:CI\_Address>  
 <gmd:electronicMailAddress>  
 <gco:CharacterString>valentin@sdsc.edu</gco:CharacterString>  
 </gmd:electronicMailAddress>  
 </gmd:CI\_Address>  
 </gmd:address>  
 </gmd:CI\_Contact>  
 </gmd:contactInfo>  
 <gmd:role>  
 <gmd:CI\_RoleCode  
 codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO\_19139\_Schemas/resources/Codelist/ML\_gmxCodelists.xml#CI\_RoleCode"  
 codeListValue="pointOfContact">pointOfContact</gmd:CI\_RoleCode>  
 </gmd:role>  
 </gmd:CI\_ResponsibleParty>  
 </gmd:pointOfContact>  
 <gmd:descriptiveKeywords>  
 <gmd:MD\_Keywords>  
 <gmd:keyword>  
 <gco:CharacterString>infoProductAccessService</gco:CharacterString>  
 </gmd:keyword>  
 </gmd:MD\_Keywords>  
 </gmd:descriptiveKeywords>  
 <gmd:descriptiveKeywords>  
 <gmd:MD\_Keywords>  
 <gmd:keyword>  
 <gco:CharacterString>WaterOneFlow</gco:CharacterString>  
 </gmd:keyword>  
 <gmd:thesaurusName>  
 <gmd:CI\_Citation>  
 <gmd:title>  
 <gco:CharacterString>CUAHSI</gco:CharacterString>  
 </gmd:title>  
 <gmd:date>  
 <gmd:CI\_Date>  
 <gmd:date>  
 <gco:Date>2010-11-10</gco:Date>  
 </gmd:date>  
 <gmd:dateType>  
 <gmd:CI\_DateTypeCode  
 codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO\_19139\_Schemas/resources/Codelist/ML\_gmxCodelists.xml#CI\_DateTypeCode"  
 codeListValue="creation">creation</gmd:CI\_DateTypeCode>  
 </gmd:dateType>  
 </gmd:CI\_Date>  
 </gmd:date>  
 </gmd:CI\_Citation>  
 </gmd:thesaurusName>  
 </gmd:MD\_Keywords>  
 </gmd:descriptiveKeywords>  
 <gmd:resourceConstraints>  
 <gmd:MD\_Constraints>  
 <gmd:useLimitation>  
 <gco:CharacterString>no conditions apply</gco:CharacterString>  
 </gmd:useLimitation>  
 </gmd:MD\_Constraints>  
 </gmd:resourceConstraints>  
 <gmd:resourceConstraints>  
 <gmd:MD\_LegalConstraints>  
 <gmd:accessConstraints>  
 <gmd:MD\_RestrictionCode  
 codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO\_19139\_Schemas/resources/Codelist/ML\_gmxCodelists.xml#MD\_RestrictionCode"  
 codeListValue="otherRestrictions"  
 >otherRestrictions</gmd:MD\_RestrictionCode>  
 </gmd:accessConstraints>  
 <gmd:otherConstraints>  
 <gco:CharacterString>(e) intellectual property rights;</gco:CharacterString>  
 </gmd:otherConstraints>  
 </gmd:MD\_LegalConstraints>  
 </gmd:resourceConstraints>  
 <srv:serviceType>  
 <gco:LocalName>other</gco:LocalName>  
 </srv:serviceType>  
 <srv:extent>  
 <gmd:EX\_Extent>  
 <gmd:geographicElement>  
 <gmd:EX\_GeographicBoundingBox>  
 <gmd:westBoundLongitude>  
 <gco:Decimal>-21.595547</gco:Decimal>  
 </gmd:westBoundLongitude>  
 <gmd:eastBoundLongitude>  
 <gco:Decimal>33.951328</gco:Decimal>  
 </gmd:eastBoundLongitude>  
 <gmd:southBoundLatitude>  
 <gco:Decimal>31.694376</gco:Decimal>  
 </gmd:southBoundLatitude>  
 <gmd:northBoundLatitude>  
 <gco:Decimal>72.299844</gco:Decimal>  
 </gmd:northBoundLatitude>  
 </gmd:EX\_GeographicBoundingBox>  
 </gmd:geographicElement>  
 </gmd:EX\_Extent>  
 </srv:extent>  
 <srv:extent>  
 <gmd:EX\_Extent>  
 <gmd:temporalElement>  
 <gmd:EX\_TemporalExtent>  
 <gmd:extent>  
 <gml:TimePeriod gml:id="ID6952ab3b-4fdb-433c-b139-2cab910146cc"  
 xsi:type="gml:TimePeriodType">  
 <gml:beginPosition>2010-11-03</gml:beginPosition>  
 <gml:endPosition>2010-11-17</gml:endPosition>  
 </gml:TimePeriod>  
 </gmd:extent>  
 </gmd:EX\_TemporalExtent>  
 </gmd:temporalElement>  
 </gmd:EX\_Extent>  
 </srv:extent>  
 <srv:couplingType gco:nilReason="missing"/>  
 <srv:containsOperations gco:nilReason="missing"/>  
 <srv:operatesOn xlink:href="http://hiscentral.cuahsi.org/"/>  
 </srv:SV\_ServiceIdentification>  
 </gmd:identificationInfo>  
 <gmd:distributionInfo>  
 <gmd:MD\_Distribution>  
 <gmd:distributionFormat>  
 <gmd:MD\_Format>  
 <gmd:name gco:nilReason="inapplicable"/>  
 <gmd:version gco:nilReason="inapplicable"/>  
 </gmd:MD\_Format>  
 </gmd:distributionFormat>  
 <gmd:transferOptions>  
 <gmd:MD\_DigitalTransferOptions>  
 <gmd:onLine>  
 <gmd:CI\_OnlineResource>  
 <gmd:linkage>  
 <gmd:URL>http://example.com/abec?wsdl</gmd:URL>  
 </gmd:linkage>  
 </gmd:CI\_OnlineResource>  
 </gmd:onLine>  
 </gmd:MD\_DigitalTransferOptions>  
 </gmd:transferOptions>  
 </gmd:MD\_Distribution>  
 </gmd:distributionInfo>  
 <gmd:dataQualityInfo>  
 <gmd:DQ\_DataQuality>  
 <gmd:scope>  
 <gmd:DQ\_Scope>  
 <gmd:level>  
 <gmd:MD\_ScopeCode codeListValue="service"  
 codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO\_19139\_Schemas/resources/Codelist/ML\_gmxCodelists.xml#MD\_ScopeCode"  
 >service</gmd:MD\_ScopeCode>  
 </gmd:level>  
 </gmd:DQ\_Scope>  
 </gmd:scope>  
 <gmd:report>  
 <gmd:DQ\_DomainConsistency xsi:type="gmd:DQ\_DomainConsistency\_Type">  
 <gmd:measureIdentification>  
 <gmd:RS\_Identifier>  
 <gmd:code>  
 <gco:CharacterString>Conformity\_001</gco:CharacterString>  
 </gmd:code>  
 <gmd:codeSpace>  
 <gco:CharacterString>INSPIRE</gco:CharacterString>  
 </gmd:codeSpace>  
 </gmd:RS\_Identifier>  
 </gmd:measureIdentification>  
 <gmd:result>  
 <gmd:DQ\_ConformanceResult xsi:type="gmd:DQ\_ConformanceResult\_Type">  
 <gmd:specification>  
 <gmd:CI\_Citation>  
 <gmd:title>  
 <gco:CharacterString>WaterML</gco:CharacterString>  
 </gmd:title>  
 <gmd:date>  
 <gmd:CI\_Date>  
 <gmd:date>  
 <gco:Date>2010-11-02</gco:Date>  
 </gmd:date>  
 <gmd:dateType>  
 <gmd:CI\_DateTypeCode  
 codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO\_19139\_Schemas/resources/Codelist/ML\_gmxCodelists.xml#CI\_DateTypeCode"  
 codeListValue="revision"  
 >revision</gmd:CI\_DateTypeCode>  
 </gmd:dateType>  
 </gmd:CI\_Date>  
 </gmd:date>  
 </gmd:CI\_Citation>  
 </gmd:specification>  
 <gmd:explanation>  
 <gco:CharacterString>See the referenced  
 specification</gco:CharacterString>  
 </gmd:explanation>  
 <gmd:pass gco:nilReason="template"/>  
 </gmd:DQ\_ConformanceResult>  
 </gmd:result>  
 </gmd:DQ\_DomainConsistency>  
 </gmd:report>  
 </gmd:DQ\_DataQuality>  
 </gmd:dataQualityInfo>  
</gmd:MD\_Metadata>

## Series

<?xml version="1.0" encoding="UTF-8"?>   
<gmd:MD\_Metadata  
 xsi:schemaLocation="http://www.isotc211.org/2005/gmd http://schemas.opengis.net/iso/19139/20060504/gmd/gmd.xsd"  
 xmlns:gmd="http://www.isotc211.org/2005/gmd" xmlns:gco="http://www.isotc211.org/2005/gco"  
 xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns:gml="http://www.opengis.net/gml"  
 xmlns:xlink="http://www.w3.org/1999/xlink">  
 <gmd:fileIdentifier>  
 <gco:CharacterString>9ebab0a7-4df1-42ac-89e7-a9f7a0610b25</gco:CharacterString>  
 </gmd:fileIdentifier>  
 <gmd:language>  
 <gmd:LanguageCode  
 codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO\_19139\_Schemas/resources/Codelist/ML\_gmxCodelists.xml#LanguageCode"  
 codeListValue="eng">eng</gmd:LanguageCode>  
 </gmd:language>  
 <gmd:hierarchyLevel>  
 <gmd:MD\_ScopeCode codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO\_19139\_Schemas/resources/Codelist/ML\_gmxCodelists.xml#MD\_ScopeCode" codeListValue="series">series</gmd:MD\_ScopeCode>  
 </gmd:hierarchyLevel>  
 <gmd:contact>  
 <gmd:CI\_ResponsibleParty>  
 <gmd:organisationName>  
 <gco:CharacterString>HIS Team</gco:CharacterString>  
 </gmd:organisationName>  
 <gmd:contactInfo>  
 <gmd:CI\_Contact>  
 <gmd:address>  
 <gmd:CI\_Address>  
 <gmd:electronicMailAddress>  
 <gco:CharacterString>his@sdsc.edu</gco:CharacterString>  
 </gmd:electronicMailAddress>  
 </gmd:CI\_Address>  
 </gmd:address>  
 </gmd:CI\_Contact>  
 </gmd:contactInfo>  
 <gmd:role>  
 <gmd:CI\_RoleCode codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO\_19139\_Schemas/resources/Codelist/ML\_gmxCodelists.xml#CI\_RoleCode" codeListValue="pointOfContact">pointOfContact</gmd:CI\_RoleCode>  
 </gmd:role>  
 </gmd:CI\_ResponsibleParty>  
 </gmd:contact>  
 <gmd:dateStamp>  
 <gco:Date>2010-11-10</gco:Date>  
 </gmd:dateStamp>  
 <gmd:metadataStandardName>  
 <gco:CharacterString>ISO19115</gco:CharacterString>  
 </gmd:metadataStandardName>  
 <gmd:metadataStandardVersion>  
 <gco:CharacterString>2003/Cor.1:2006</gco:CharacterString>  
 </gmd:metadataStandardVersion>  
 <gmd:identificationInfo>  
 <gmd:MD\_DataIdentification>  
 <gmd:citation>  
 <gmd:CI\_Citation>  
 <gmd:title>  
 <gco:CharacterString>CUAHSI DataSet</gco:CharacterString>  
 </gmd:title>  
 <gmd:date>  
 <gmd:CI\_Date>  
 <gmd:date>  
 <gco:DateTime>2010-11-10T09:08:01</gco:DateTime>  
 </gmd:date>  
 <gmd:dateType>  
 <gmd:CI\_DateTypeCode codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO\_19139\_Schemas/resources/Codelist/ML\_gmxCodelists.xml#CI\_DateTypeCode" codeListValue="creation">creation</gmd:CI\_DateTypeCode>  
 </gmd:dateType>  
 </gmd:CI\_Date>  
 </gmd:date>  
 <gmd:date>  
 <gmd:CI\_Date>  
 <gmd:date>  
 <gco:Date>2010-11-14</gco:Date>  
 </gmd:date>  
 <gmd:dateType>  
 <gmd:CI\_DateTypeCode codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO\_19139\_Schemas/resources/Codelist/ML\_gmxCodelists.xml#CI\_DateTypeCode" codeListValue="publication">publication</gmd:CI\_DateTypeCode>  
 </gmd:dateType>  
 </gmd:CI\_Date>  
 </gmd:date>  
 <gmd:date>  
 <gmd:CI\_Date>  
 <gmd:date>  
 <gco:DateTime>2010-11-10T09:07:56</gco:DateTime>  
 </gmd:date>  
 <gmd:dateType>  
 <gmd:CI\_DateTypeCode codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO\_19139\_Schemas/resources/Codelist/ML\_gmxCodelists.xml#CI\_DateTypeCode" codeListValue="revision">revision</gmd:CI\_DateTypeCode>  
 </gmd:dateType>  
 </gmd:CI\_Date>  
 </gmd:date>  
 <gmd:identifier>  
 <gmd:RS\_Identifier>  
 <gmd:code>  
 <gco:CharacterString>hydrology</gco:CharacterString>  
 </gmd:code>  
 <gmd:codeSpace>  
 <gco:CharacterString>cuahsi</gco:CharacterString>  
 </gmd:codeSpace>  
 </gmd:RS\_Identifier>  
 </gmd:identifier>  
 </gmd:CI\_Citation>  
 </gmd:citation>  
 <gmd:abstract>  
 <gco:CharacterString>test</gco:CharacterString>  
 </gmd:abstract>  
 <gmd:pointOfContact>  
 <gmd:CI\_ResponsibleParty>  
 <gmd:organisationName>  
 <gco:CharacterString>San Diego Supercomputer Center</gco:CharacterString>  
 </gmd:organisationName>  
 <gmd:contactInfo>  
 <gmd:CI\_Contact>  
 <gmd:address>  
 <gmd:CI\_Address>  
 <gmd:electronicMailAddress>  
 <gco:CharacterString>valentin@sdsc.ed</gco:CharacterString>  
 </gmd:electronicMailAddress>  
 </gmd:CI\_Address>  
 </gmd:address>  
 </gmd:CI\_Contact>  
 </gmd:contactInfo>  
 <gmd:role>  
 <gmd:CI\_RoleCode codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO\_19139\_Schemas/resources/Codelist/ML\_gmxCodelists.xml#CI\_RoleCode" codeListValue="pointOfContact">pointOfContact</gmd:CI\_RoleCode>  
 </gmd:role>  
 </gmd:CI\_ResponsibleParty>  
 </gmd:pointOfContact>  
 <gmd:descriptiveKeywords>  
 <gmd:MD\_Keywords>  
 <gmd:keyword>  
 <gco:CharacterString>Hydrography</gco:CharacterString>  
 </gmd:keyword>  
 <gmd:thesaurusName>  
 <gmd:CI\_Citation>  
 <gmd:title>  
 <gco:CharacterString>GEMET - INSPIRE themes, version 1.0</gco:CharacterString>  
 </gmd:title>  
 <gmd:date>  
 <gmd:CI\_Date>  
 <gmd:date>  
 <gco:Date>2008-06-01</gco:Date>  
 </gmd:date>  
 <gmd:dateType>  
 <gmd:CI\_DateTypeCode codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO\_19139\_Schemas/resources/Codelist/ML\_gmxCodelists.xml#CI\_DateTypeCode" codeListValue="publication">publication</gmd:CI\_DateTypeCode>  
 </gmd:dateType>  
 </gmd:CI\_Date>  
 </gmd:date>  
 </gmd:CI\_Citation>  
 </gmd:thesaurusName>  
 </gmd:MD\_Keywords>  
 </gmd:descriptiveKeywords>  
 <gmd:resourceConstraints>  
 <gmd:MD\_Constraints>  
 <gmd:useLimitation>  
 <gco:CharacterString>no conditions apply</gco:CharacterString>  
 </gmd:useLimitation>  
 </gmd:MD\_Constraints>  
 </gmd:resourceConstraints>  
 <gmd:resourceConstraints>  
 <gmd:MD\_LegalConstraints>  
 <gmd:accessConstraints>  
 <gmd:MD\_RestrictionCode codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO\_19139\_Schemas/resources/Codelist/ML\_gmxCodelists.xml#MD\_RestrictionCode" codeListValue="otherRestrictions">otherRestrictions</gmd:MD\_RestrictionCode>  
 </gmd:accessConstraints>  
 <gmd:otherConstraints>  
 <gco:CharacterString>no limitations</gco:CharacterString>  
 </gmd:otherConstraints>  
 </gmd:MD\_LegalConstraints>  
 </gmd:resourceConstraints>  
 <gmd:language>  
 <gmd:LanguageCode  
 codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO\_19139\_Schemas/resources/Codelist/ML\_gmxCodelists.xml#LanguageCode"  
 codeListValue="eng">eng</gmd:LanguageCode>  
 </gmd:language>  
 <gmd:topicCategory>  
 <gmd:MD\_TopicCategoryCode>inlandWaters</gmd:MD\_TopicCategoryCode>  
 </gmd:topicCategory>  
 <gmd:extent>  
 <gmd:EX\_Extent>  
 <gmd:geographicElement>  
 <gmd:EX\_GeographicBoundingBox>  
 <gmd:westBoundLongitude>  
 <gco:Decimal>20.152499</gco:Decimal>  
 </gmd:westBoundLongitude>  
 <gmd:eastBoundLongitude>  
 <gco:Decimal>34.039218</gco:Decimal>  
 </gmd:eastBoundLongitude>  
 <gmd:southBoundLatitude>  
 <gco:Decimal>49.733926</gco:Decimal>  
 </gmd:southBoundLatitude>  
 <gmd:northBoundLatitude>  
 <gco:Decimal>59.885294</gco:Decimal>  
 </gmd:northBoundLatitude>  
 </gmd:EX\_GeographicBoundingBox>  
 </gmd:geographicElement>  
 </gmd:EX\_Extent>  
 </gmd:extent>  
 <gmd:extent>  
 <gmd:EX\_Extent>  
 <gmd:temporalElement>  
 <gmd:EX\_TemporalExtent>  
 <gmd:extent>  
 <gml:TimePeriod gml:id="ID220dbbfd-d88a-4c6f-aa9a-41091d59f20a"  
 xsi:type="gml:TimePeriodType">  
 <gml:beginPosition>2010-11-10</gml:beginPosition>  
 <gml:endPosition>2010-11-22</gml:endPosition>  
 </gml:TimePeriod>  
 </gmd:extent>  
 </gmd:EX\_TemporalExtent>  
 </gmd:temporalElement>  
 </gmd:EX\_Extent>  
 </gmd:extent>  
 </gmd:MD\_DataIdentification>  
 </gmd:identificationInfo>  
 <gmd:distributionInfo>  
 <gmd:MD\_Distribution>  
 <gmd:distributionFormat>  
 <gmd:MD\_Format>  
 <gmd:name gco:nilReason="inapplicable"/>  
 <gmd:version gco:nilReason="inapplicable"/>  
 </gmd:MD\_Format>  
 </gmd:distributionFormat>  
 <gmd:transferOptions>  
 <gmd:MD\_DigitalTransferOptions>  
 <gmd:onLine>  
 <gmd:CI\_OnlineResource>  
 <gmd:linkage>  
 <gmd:URL>his.cuahsi.org</gmd:URL>  
 </gmd:linkage>  
 </gmd:CI\_OnlineResource>  
 </gmd:onLine>  
 </gmd:MD\_DigitalTransferOptions>  
 </gmd:transferOptions>  
 </gmd:MD\_Distribution>  
 </gmd:distributionInfo>  
 <gmd:dataQualityInfo>  
 <gmd:DQ\_DataQuality>  
 <gmd:scope>  
 <gmd:DQ\_Scope>  
 <gmd:level>  
 <gmd:MD\_ScopeCode codeListValue="series" codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO\_19139\_Schemas/resources/Codelist/ML\_gmxCodelists.xml#MD\_ScopeCode">series</gmd:MD\_ScopeCode>  
 </gmd:level>  
 </gmd:DQ\_Scope>  
 </gmd:scope>  
 <gmd:report>  
 <gmd:DQ\_DomainConsistency xsi:type="gmd:DQ\_DomainConsistency\_Type">  
 <gmd:measureIdentification>  
 <gmd:RS\_Identifier>  
 <gmd:code>  
 <gco:CharacterString>Conformity\_001</gco:CharacterString>  
 </gmd:code>  
 <gmd:codeSpace>  
 <gco:CharacterString>INSPIRE</gco:CharacterString>  
 </gmd:codeSpace>  
 </gmd:RS\_Identifier>  
 </gmd:measureIdentification>  
 <gmd:result>  
 <gmd:DQ\_ConformanceResult xsi:type="gmd:DQ\_ConformanceResult\_Type">  
 <gmd:specification>  
 <gmd:CI\_Citation>  
 <gmd:title>  
 <gco:CharacterString>CUAHSI</gco:CharacterString>  
 </gmd:title>  
 <gmd:date>  
 <gmd:CI\_Date>  
 <gmd:date>  
 <gco:Date>2010-11-10</gco:Date>  
 </gmd:date>  
 <gmd:dateType>  
 <gmd:CI\_DateTypeCode codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO\_19139\_Schemas/resources/Codelist/ML\_gmxCodelists.xml#CI\_DateTypeCode" codeListValue="creation">creation</gmd:CI\_DateTypeCode>  
 </gmd:dateType>  
 </gmd:CI\_Date>  
 </gmd:date>  
 </gmd:CI\_Citation>  
 </gmd:specification>  
 <gmd:explanation>  
 <gco:CharacterString>See the referenced specification</gco:CharacterString>  
 </gmd:explanation>  
 <gmd:pass gco:nilReason="template"/>  
 </gmd:DQ\_ConformanceResult>  
 </gmd:result>  
 </gmd:DQ\_DomainConsistency>  
 </gmd:report>  
 <gmd:lineage>  
 <gmd:LI\_Lineage>  
 <gmd:statement>  
 <gco:CharacterString>Raw Data</gco:CharacterString>  
 </gmd:statement>  
 </gmd:LI\_Lineage>  
 </gmd:lineage>  
 </gmd:DQ\_DataQuality>  
 </gmd:dataQualityInfo>  
</gmd:MD\_Metadata>