Open Geospatial Consortium Inc.

Date: 2009-12-04

Reference number of this OGC® project document: **OGC 09-124r1**

Version: 0.2.0

Category: OpenGIS© Discussion Paper

Editor: Peter Taylor

Harmonising Standards for Water Observation Data - Discussion Paper

Warning

This document is not an OGC Standard. This document is an OGC Discussion Paper and is therefore not an official position of the OGC membership. It is distributed for review and comment. It is subject to change without notice and may not be referred to as an OGC Standard. Further, an OGC Discussion Paper should not be referenced as required or mandatory technology in procurements..

Document type: OGC® Discussion Paper
Document stage: Draft
Document language: English

Contents

Figures

Tables

i. Preface

This is an OGC Discussion Paper for review by OGC members and other interested parties. It is a working draft document and may be updated, replaced by other documents at any time. It is inappropriate to use OGC Discussion Papers as reference material or to cite them as other than "work in progress." This is work in progress and does not imply endorsement by the OGC membership.

ii. Submitting organizations

The following organizations submitted this document to the Open Geospatial Consortium Inc.

a) CSIRO

b) CUAHSI

iii. Submission contact points

All questions regarding this submission should be directed to the editor or the submitters:

CONTACT	COMPANY
Peter Taylor	CSIRO
David Valentine	CUAHSI
Gavin Walker	CSIRO

iv. Revision history

Date	Release	Author	Paragraph modified	Description
2009-09-11	0.1.0	Peter Taylor	All	First version of Discussion Paper.
2009-12-04	0.1.1	Gavin Walker	Section 7 and general.	Added descriptions of time components for core standards. General edits and additions.
2009-12-04	0.1.2	Peter Taylor	Section 6 and general edits.	Added section 6 on hydrological observations and requirements. Added section on methodology and code lists.

v. Changes to the OGC® Abstract Specification

The OGC® Abstract Specification does not require changes to accommodate this OGC® standard.

Foreword

This work has, for the most part, been funded through a water information research and development alliance between CSIRO's Water for a Healthy Country Flagship and the Australian Bureau of Meteorology. The work has also been supported by The Consortium for the Advancement of Hydrological Sciences Inc. (CUAHSI).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. Open Geospatial Consortium Inc. shall not be held responsible for identifying any or all such patent rights. However, to date, no such rights have been claimed or identified.

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

Introduction

This discussion paper has two broad goals. Firstly, it will investigate the core requirements for an information model which describes the results of hydrological observations, focusing on time series. This will be done by analysing existing data standards for hydrology, or closely related domains. It will be shown that existing standards contain concepts that are sufficiently aligned that a harmonised view may be developed.

Secondly, the discussion paper will provide an approach for developing a harmonised core conceptual model for hydrological observations. It is proposed that such a model provides a basis for, in the first instance, generating an XML Schema and accompanying documentation.

The explored approach will focus on re-using existing open standards and information modelling best practices. Re-use allows development to focus on the domain specific problems rather than re-solving commonly addressed issues. Developing a standard that is usable to a wider audience improves the ability for communities to share tools that address common needs such as encoding and decoding of a standard schema.

Harmonising Standards for Water Observation Data

1 Scope

This document investigates the potential for harmonisation of water data standards, with the goal of developing an OGC compliant standard for the exchange of water observation data. It will be based on OGC's Observations and Measurements model, creating profile of it in the water domain. The profile will be developed by examining the content and structure of existing standards and suggesting future methodology for developing a harmonised model for observation data. This model will make use of existing standards where possible.

2 Normative references

The following normative documents contain provisions which, through reference in this text, constitute provisions of this document. For dated references, subsequent amendments to, or revisions of, any of these publications do not apply. For undated references, the latest edition of the normative document referred to applies.

ISO 19101:2003, *Geographic Information--ReferenceModel*

ISO 19109:2006, *Geographic Information — Rules for application schemas*

ISO 19123:2005, *Geographic Information — Coverages*

ISO DIS 19136:2006, *Geographic Information — Geography Markup Language*

ISO/FDTS 19139:2006, *Geographic Information — Metadata — XML schema implementation*

OpenGIS® Implementation Specification *Observations and Measurements – Part 1: Observation Schema, OGC document OGC 07-022r1.*

OpenGIS® Implementation Standard *Observations and Measurements – Part 2: Sampling Features, OGC document 07-002r3.*

OpenGIS® Implementation Specification *Sensor Model Language (SensorML), OGC Document OGC 07-000*

OpenGIS® Implementation Specification *Sensor Observation Service, OGC document OGC 06-009r6.*

W3C XLink, *XML Linking Language (XLink) Version 1.0. W3C Recommendation (27 June 2001)*

W3C XML, *Extensible Markup Language (XML) 1.0 (Second Edition), W3C Recommendation (6 October 2000)*

W3C XML Namespaces, *Namespaces in XML. W3C Recommendation (14 January 1999)*

W3C XML Schema Part 1, *XML Schema Part 1: Structures. W3C Recommendation (28th October 2004)*

W3C XML Schema Part 2, *XML Schema Part 2: Datatypes. W3C Recommendation (28th October 2004)*

3 Terms and definitions

For the purposes of this document, the following terms and definitions apply:

4.1
Discharge
The volume of fluid passing a point per unit time
4.2
Gauging station
Monitoring point for making observations of terrestrial water bodies
4.3
Phenomenon
Concept that is a characteristic of one or more feature types, the value for which may be estimated by application of some procedure in an observation.
4.4
Rating curve
A curve describing the relationship between river level and river flow (or discharge)

4 Conventions

4.1 Symbols (and abbreviated terms)

AWRIS	Australian Water Resources Information System
CF-netCDF	NetCDF Climate and Forecast Metadata Convention
CSIRO	Commonwealth Scientific and Industrial Research Organization
CSML	Climate Science Modelling Language
CSV	Comma Separated Values
CUAHSI	Consortium of Universities for Advancement of Hydrologic Science
DIF	Data Integration Framework

EPA	Environmental Protection Agency
ESAR	Environmental Sampling, Analysis, and Results
FTP	File Transfer Protocol
GeoSciML	Geological Sciences Markup Language
GEOSS	Global Earth Observation System of Systems
GIS	Geographic Information System
GML	Geography Markup Language
GRDC	Global Runoff Data Centre
GWML	Groundwater Markup Language
HDWG	Hydrology Domain Working Group
HIS	Hydrologic Information System
ISO	International Organisation for Standardization
ISO/TC 211	ISO/TC 211 standards catalogue
INSPIRE	Infrastructure for Spatial Information in the European Community
MDA	Model Driven Architecture
MMI	Marine Metadata Interoperability
NSF	National Science Foundation
netCDF	network Common Data Form
O&M	Observations and Measurements
ODM	Observation Data Model
OGC	Open Geospatial Consortium
OMG	Object Management Group
OpenGIS	Abstract Specification
REST	Representational State Transfer
SDI	Spatial Data Infrastructure
SensorML	Sensor Markup Language

SOS	Sensor Observation Service
SWE	Sensor Web Enablement
UML	Unified Modeling Language
URN	Uniform Resource Name
USGS	United States Geological Survey
W3C	World Wide Web Consortium
WaterML	Water Markup Language
WCS	Web Coverage Service
WISE	Water Information System for Europe
WMO	World Meteorological Organisation
WQX	Water Quality Exchange
XMI	XML Metadata Interchange
XML	eXtensible Markup Language

4.2 UML Notation

The diagrams that appear in this standard are presented using the Unified Modeling Language (UML) static structure diagram. The UML notations used in this standard are described in the diagram below.

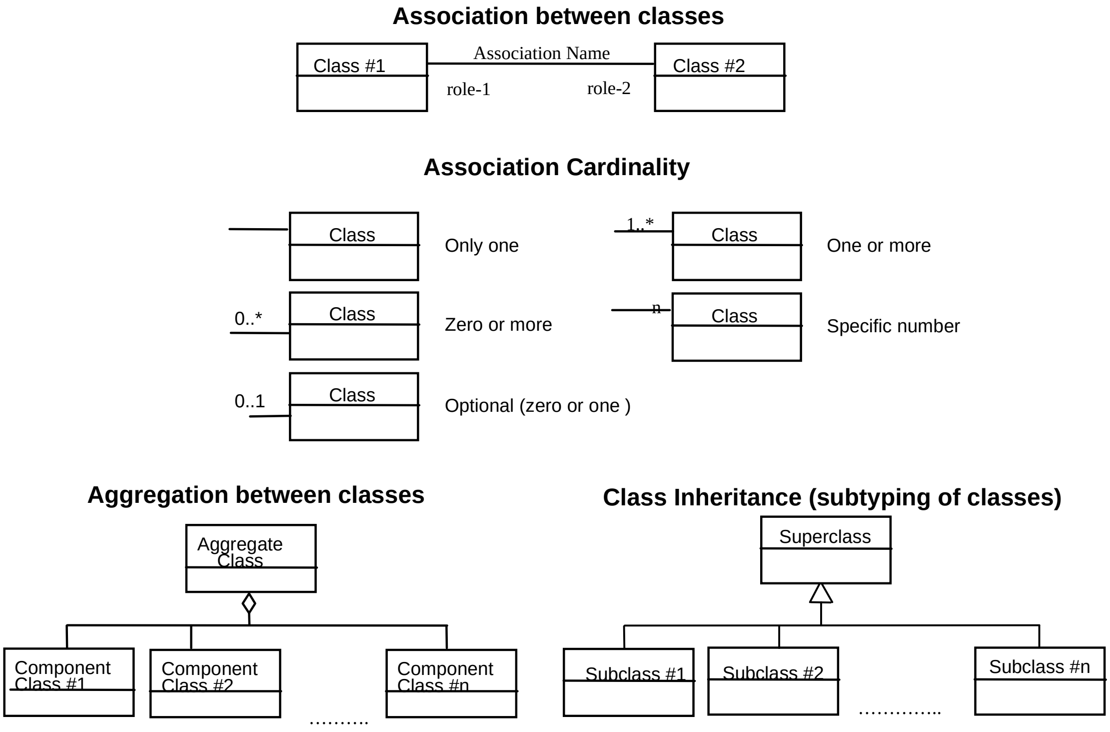


Figure 1 — UML notation

In this document, the following three stereotypes of UML classes are used:

a) <<Interface>> A definition of a set of operations that is supported by objects having this interface. An Interface class cannot contain any attributes.

b) <<DataType>> A descriptor of a set of values that lack identity (independent existence and the possibility of side effects). A DataType is a class with no operations whose primary purpose is to hold the information.

c) <<CodeList>> is a flexible enumeration that uses string values for expressing a list of potential values.

In this document, the following standard data types are used:

a) CharacterString – A sequence of characters

b) Integer – An integer number

c) Double – A double precision floating point number

d) Float – A single precision floating point number

5 Motivation

There is a global push within information communities for the development of consistent information models for the capture of spatial and temporal data and metadata. The current state of data existing in 'stove pipes' is seen as inconsistent, inefficient and a major barrier to improving interoperability of information systems.

A worldwide initiative, Global Earth Observation System of Systems (GEOSS), has the goal of developing a system to allow world-wide integration of observation data to improve our understanding of the global environment. Its 10-year plan outlines issues with data availability: "*…the current situation with respect to the availability of Earth observations is not optimal. This situation is particularly true with respect to coordination and data sharing among countries, organizations and disciplines, and meeting the needs of sustainable development.*" [GEO2005]

Within the US there are programmes and initiatives to promote data sharing and re-use through the use of standards and Web Services for information exchange. For example the Consortium for the Advancement of Hydrologic Sciences Inc. (CUAHSI) has developed a number of schemas and technologies to facilitate improved sharing of hydrological data sets.

Within Australia, the Bureau of Meteorology (the Bureau) is developing an Australian Water Resources Information System (AWRIS), with the goal of obtaining a deeper understanding of the current state of water resources across the country. This is resulting in developments that are addressing data standards within the hydrology community. The Australian Government's Water Act 2007 [AUSWA2007] empowers the Bureau to collect and set standards for water information across the country.

In Europe, the Water Information System for Europe (WISE) is developing a gateway for water information with the aim of providing data to the public collected by institutions across the member countries. More broadly in Europe, the Infrastructure for Spatial Information in the European Community (INSPIRE) initiative has the directive to develop an EU wide spatial data infrastructure for sharing spatial data sets. The directive states: "The loss of time and resources in searching for existing spatial data or establishing whether they may be used for a particular purpose is a key obstacle to the full exploitation of the data available" [EU2007].

These initiatives are dealing with the large scale complexity of disparate data sets and all are working on improved standards for water information. This type of information covers both spatial and temporal data sets, each of which has its own level of complexity. In order to avoid re-solving well understood issues with handling such data, most initiatives are looking to leverage existing standards and methodologies where possible [EU2007] [GEO2005]. The International Standards Organisation (ISO) and the Open Geospatial Consortium (OGC) are two bodies that define standards directly relevant to the issues being addressed.

The OGC and the World Meteorology Organisation (WMO) have recently formed the Hydrology Domain Working Group (HDWG) [LEM2008] which is a forum for the collaboration and development of standards for hydrological data. This group has members from countries dealing with similar issues of developing and reusing standards.

An international workshop held in Australia on Water Resources Information models in September 2007 [COX2007] indicated some of the benefits of developing shared models were: "…improved efficiency and quality of local information models and systems; wider use and re-use of information; new tool development, and new value from existing information via unexpected uses". This workshop suggested the development of a harmonised information model and transfer formats for water data.

5.1 Structure of this document

Section 6 gives an introduction into the type of observations that occur within the hydrology domain, continuing into an overview of the need for the exchange of such data sets.

Section 7 of this document will give an overview of existing relevant standards; firstly standards associated with methodologies for developing information models and secondly existing standards for hydrological information. Sections 7-10 examine these standards in each of four core areas: results, features and sampling, procedures and observed properties. From this analysis the discussion paper will outline a core set of requirements for an information model for hydrological observations. These are summarised in section 15.

Section 16 and onwards proposes an approach for developing a core conceptual model for hydrological observations. This approach will make use of existing best practice and standards by examining projects that have employed similar techniques. This model should be extensible to suit particular requirements for the exchange of hydrological observations. The potential uses of this model are considered in section 14 and a discussion on the repercussions of adoption is put forward in section 15.

6 Hydrological Observations

"Water is found on Earth in significant amounts in all three of its physical phases: liquid, solid, and gaseous. It is also found in all three of Earth's major environments that are readily accessible to humans: the atmosphere, the seas and oceans, and the land masses. Because water can readily move from one environment to another and can change from one phase to another in response to its environment, it is a dynamic medium in both space and time." [WMO1994].

The field of hydrology focuses on the water cycle as it interacts with land; hydrological observations are performed in order for us to increase our understanding of this interaction. Such observations can occur at any point within the hydrologic cycle, each employing different techniques for making measurement and estimates of water quantity and quality.

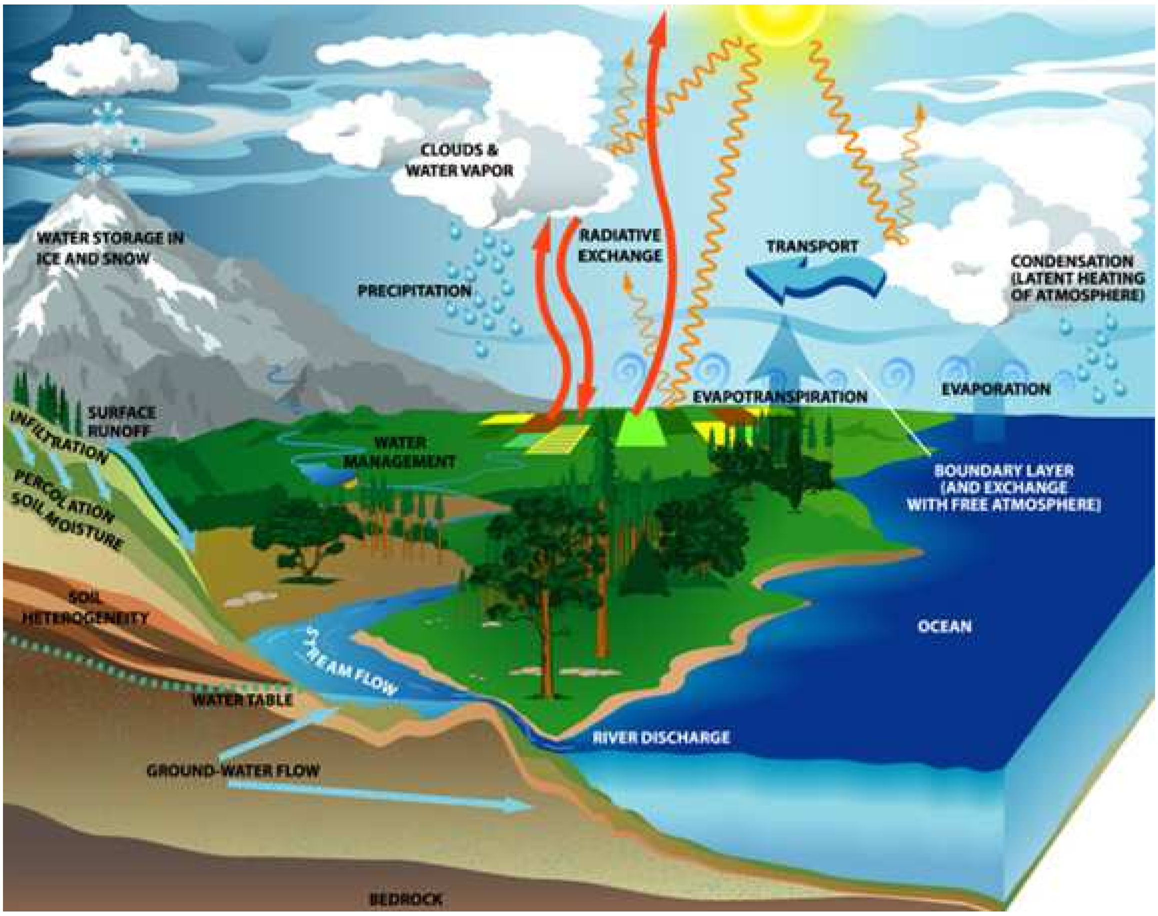


Figure 2 – The hydrologic cycle [ENE2009]

The WMO Technical Regulations [WMO2006] define concepts and standard types of observations typically made within the hydrology domain as well as relevant observations from other domains such as climatology and meteorology. The classifications defined are useful in separating the various categories of hydrological observations as they represent not only different observed phenomena but also different sampling techniques.

The regulations break hydrometric stations into the following categories:

(a) Hydrometric stations;

(b) Groundwater stations;

(c) Climatological stations and precipitation stations for hydrological purposes;

(d) Hydrological stations for specific purposes.

From these categories they define the types of phenomena that are recommended to be measured. The following lists show an adapted summary of these:

Hydrometric stations

- River, lake or reservoir stage;
- Stream flow;
- Sediment transport and/or deposition;

- Temperature and other physical properties of the water of a river, lake or reservoir;
- Characteristics and extent of ice cover on rivers, lakes and reservoirs;
- Chemical and biological properties of the water of a river, lake or reservoir.

Climatological stations

- Precipitation:

  (i) Amount;

  (ii) Time of occurrence;

  (iii) Form (e.g. rain, snow, sleet);

  (iv) Character (continuous, intermittent, scattered showers, etc.);

  (v) Intensity;
- Air temperature (including extreme temperatures);
- Air humidity
- Wind:

  (i) Speed and direction (10-minute wind average);

  (ii) Daily run
- Amount and type of cloud;
- Snow cover:

  (i) Snow depth;

  (ii) Density;

  (iii) Water equivalent;
- Evaporation (measured with evaporation pan);
- Solar radiation;
- Sunshine;
- Soil temperature;
- Atmospheric pressure;
- Soil moisture.

Groundwater stations

- Water level;
- Temperature and other physical properties of the water;
- Chemical properties;
- Rate and volume of abstraction or recharge.

The WMO specifications are not exhaustive but are written as a guide to standard observational practices. They give an indication of the types of observed phenomena that are crucial to the hydrology domain.

The Australian Water Regulations [BOM2008c] categorise water information into the following categories:

1. Surface water resource information
2. Ground water resource information
3. Information on major and minor water storages
4. Meteorological information
5. Water use information
6. 6257 Tf 974(a)250]TJ /R35 12 Tf 9.00531 0 Td ( )Tj /R8 12 Tf 9.00531 0 Td [(I)12.8115(n)-10.3

2. Ex-situ, complex processing observations (e.g. water quality): temporally sparse, spatially sparse, many observed phenomena. Examples: nutrients (nitrate, phosphorus etc.), pesticides (atrazine, glyphosate etc.), biologicals, pH, turbidity etc.
3. Complex data products. These consist of processed or synthesised observational data, mainly created to provide estimation of not directly measurable phenomena or predictions of future values. Examples: outputs from models or algorithms, water storage estimates.

These definitions are not clear-cut; it is possible to have water quality measurements that are made continuously by in-situ measurements (such as dissolved oxygen, turbidity etc.). Similarly, storage volume may be viewed as a complex data product as it often involves the integration of survey data and estimation algorithms. Exchange formats addressing category 1 may be capable of capturing data within category 3, but representation of the procedure used to generate the data set implies extra requirements on metadata (if it is to be supported through transfer). Generally, the more complex the process of making the measurement, the less likely it is to be available as a continuous observation.

6.1 Need for exchange of observational data

The driving need for the exchange of water observation data is varied and operates on different levels, from intra-agency sharing to sharing across international borders. Traditionally the impetus for the exchange of data has been for reporting requirements arising from across agency collection of data sets. More recently the development of enabling technologies, such as distributed computing and web services, have allowed for data to be shared with a broader audience. In addition to this is the increasing demand of cross-disciplinary research to access previously inaccessible data sets, such as climate science, where scientists attempt to merge data sets from a wide variety of influencing factors such as oceanography..

In 1999 the WMO adopted Resolution 25 which states:

"*a stand of committing to broadening and enhancing, whenever possible, the free and unrestricted exchange of hydrological data and products, in consonance with the requirements of WMO's scientific and technical programmes."*

This Resolution led to a report on the exchange of hydrological data and products [WMO2001] which explores the requirements for data exchange and defines three typical categories of data products:

(i) Data for the "*…the provision of services in support of the protection of life and property and for the well-being of all nations."*

(ii) "*additional hydrological data and products, where available, which are required to sustain programmes and projects of WMO, other UN agencies, ICSU and other organizations of equivalent status, related to operational hydrology and water resources research at the global, regional and national levels.*"

(iii) "*all hydrological data and products exchanged under the auspices of WMO, for the non-commercial activities of the research and education communities*"

These have a WMO perspective (i.e. international), but the categorisation goes further to describe principle uses of hydrological information globally as:

1. Real-time applications: forecasting and warning of floods, low flows and other extreme events;
2. Real-time applications: project operation;
3. Engineering design;
4. Hydrological and environmental science;
5. Monitoring trends in the global environment.

The report describes operational hydrology (primarily items 1, 2 & 3 above) as generally being performed on the national scale, where items 4 and 5 may require international exchange of information. It is noted that international exchange may need to occur in the first three cases where there are shared river basins across the borders of countries.

A report from the Global Terrestrial Observing System (GCOS) on the establishment of a Global Hydrological Observation Network for Climate [GCO2000] provided some synthesis of requirements for hydrological observation data and identified five major drivers for exchange:

1. Improved Climate and Weather Prediction
2. Characterising Hydrological Variability to Detect Climate Change
3. Developing the Ability to Predict the Impacts of Change
4. Assessing Water Sustainability as a Function of Water Use Versus Water Availability
5. Understanding the Global Water Cycle

7 Relevant standards

7.1 Standards and best practices for information modelling

In a Model Driven Architecture (MDA) [OMGa] approach, a domain modeller captures a conceptual model of an information system with a formal modelling language such as the Unified Modeling Language (UML) [OMGb]. UML allows construction of an abstract graphical representation of information artefacts and their relationships using diagrammatic elements that have well defined semantics. From this model it is possible to generate specific implementations of the model, such as the eXtensible Markup Language (XML) Schema [W3Ca] or database schema.

Combining MDA with existing standards, the ISO Technical Committee 211 has developed standards and methodologies to support the development of information models for Spatial Data Infrastructures (SDIs). The ISO19101 model is the reference model for the 19100 series on the development of geographic information standards; it, along with ISO19109 (rules for application schema), outline a methodology for developing conceptual models and application schema with a goal to improve interoperability. Note that geographic information is defined as “information concerning phenomena implicitly or explicitly associated with a location relative to the Earth" [ISO19101]; hydrological phenomena fit into this category. The INSPIRE project has taken the IS019101 approach in defining a methodology for developing its standards for data exchange.

The methodologies described in ISO19000 define rules for the use of tools such as UML for creating conformant conceptual models that have explicit relationships to other platforms such as XML. This allows construction, sharing and composition of standards in UML, removing the complexity of dealing with XML Schema or other implementations.

The INSPIRE project has developed broad methodologies for the development of common information models to promote data sharing and re-use across the EU:

"…a key step in the data harmonisation process is to achieve interoperability on the conceptual level (semantic interoperability) so that users and implementers of different information systems can understand the semantics of the relevant information provided by the other system(s). " [INS2007]

Whilst the INSPIRE project is focusing on spatial data sets, the general methodology for harmonisation is relevant to developing information models for observational data sets. The Observations & Measurements (O&M) model from OGC is within scope of INSPIRE and suggests that temporal data will play a role in the specification of INSPIRE's information models.

Model-driven approaches to information modelling are also being investigated within the Microsoft Active Data Objects (ADO) framework through its Entity Framework [MSEF] concept. This uses a conceptual model to define element structure, relationships and constraints. Using this framework a conceptual model is built using a tool called the Entity Data Model, which is similar to UML. Mappings can be created to underlying storage mechanisms (databases, schema etc.), allowing automatic generation from the conceptual model to a particular storage type. This generation of schema is still being developed. It will allow information to be modelled without artefacts that are specific to the underlying storage technology, creating a cleaner separation of concerns.

There are other model-driven frameworks emerging such as the two open source products, AndroMDA [AND2009] and Hibernate OOMEGA [OOM2009].

A large, actively developed standard that uses a model-driven approach is GeoSciML. The standard is used to describe geologic features with an emphasis on geological information for use in portraying geologic maps. The project has been a driving force in the development of tools to support MDA approaches to information modelling, such as the HollowWorld [HOL2009] and FullMoon [FUL2009] toolsets.

7.2 Observations and Measurements (O&M)

The OGC Observations and Measurements (O&M) standard is the point of convergence of a range of ISO TC 211 and OGC activities [BAC2007].

OGC has been developing a suite of specifications relating to observational data, known as Sensor Web Enablement (SWE)[BOT2006] that complements the generic access models provided by WMS, WFS and WCS. The distinguishing element of observational data is that the procedure used to obtain the data, and the resulting uncertainties, are of interest to data users.

O&M extends the existing ISO-specified models with the components related to detailed provenance and uncertainty issues, which are necessary to manage and make use of observations.

A growing list of application communities (including Geology, Climate Science and Water) have evaluated the formalization of observation and sampling information provided in O&M and have committed to implementing data-transfer and even database systems based on it. While the analysis required for this is often challenging initially, the rigorous and explicit model, and its integration in the ISO/TC 211 and OGC methodology is expected to provide significant benefits in interoperability and sustainability, with its modular rather than monolithic basis.

O&M's conceptual model defines an observation as "…an action whose result is an estimate of the value of some property of the feature-of-interest, obtained using a specified procedure." [COX2007b]

This model provides a separation of the elements involved in observations as well as defining the relationships between them. By separating the core elements of observation descriptions, we have a basis for exchanging, and discussing, observational data sets.

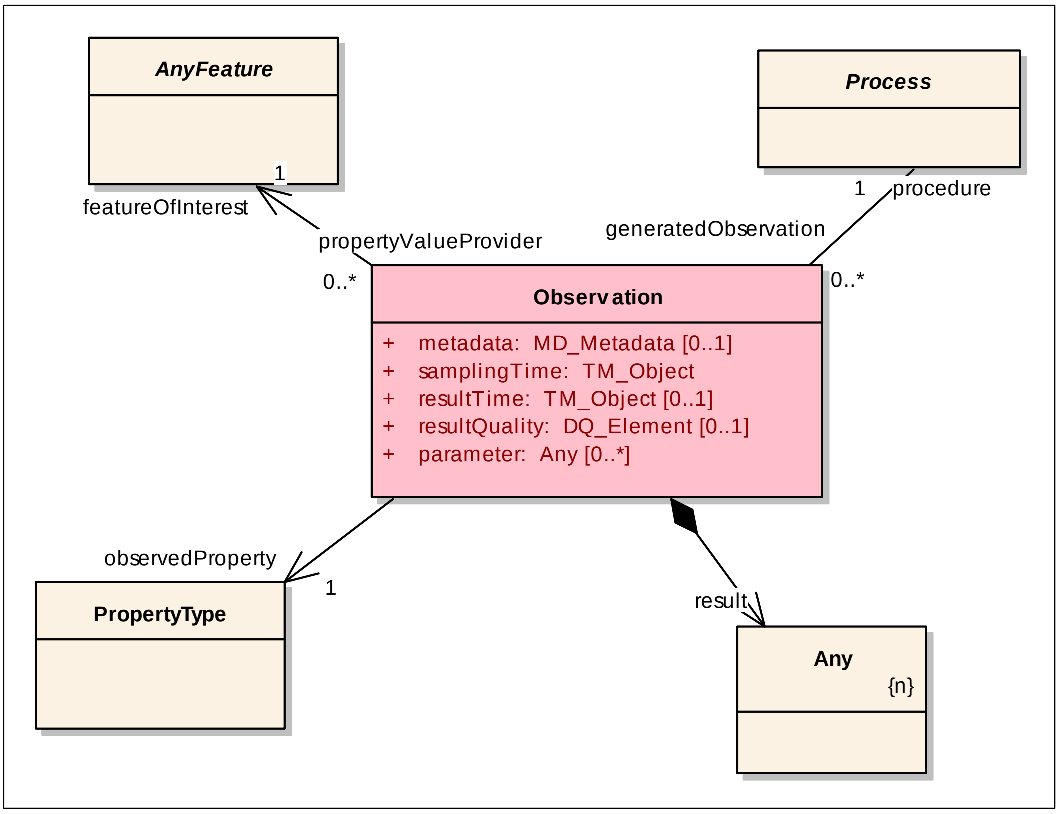


Figure 3 - Observations and Measurements UML model

Part 2 - Sampling Features of O&M [COX2007c] defines a sampling model for capturing cases where the actual target of an observation is not the ultimate feature but a proxy for measuring a property of the feature. It introduces the concept of a *sampling feature* that is the proxy for the measurement. This sampling feature has a relationship named *sampled feature*, which is the real world feature being observed.

The O&M model can be used to describe hydrological observations, and the sampling features section helps to describe common observation patterns in the domain. The sampling features concept does not describe 'domain features' such as lakes and rivers, but the intermediate process that occurs in observing them. Such intermediate concepts are often call stations but may include profiles or other sampling dimensions.

For example, a river level gauge is actually sampling the height of a river at a particular point (a gauging station); this observation is providing an estimate for the 'ultimate feature of interest' which is the actual river height. The sampling point here would be the gauging station.

An example of a hydrological observation using this model is shown using UML in Figure 4. This diagram shows three river flow observations, each at a different station, ultimately measuring two rivers. These use a calculation to estimate the river flow. There is another observation provided by a sensor measuring the temperature of the river. All have time series as their results, described here using discrete coverages [COX2008].

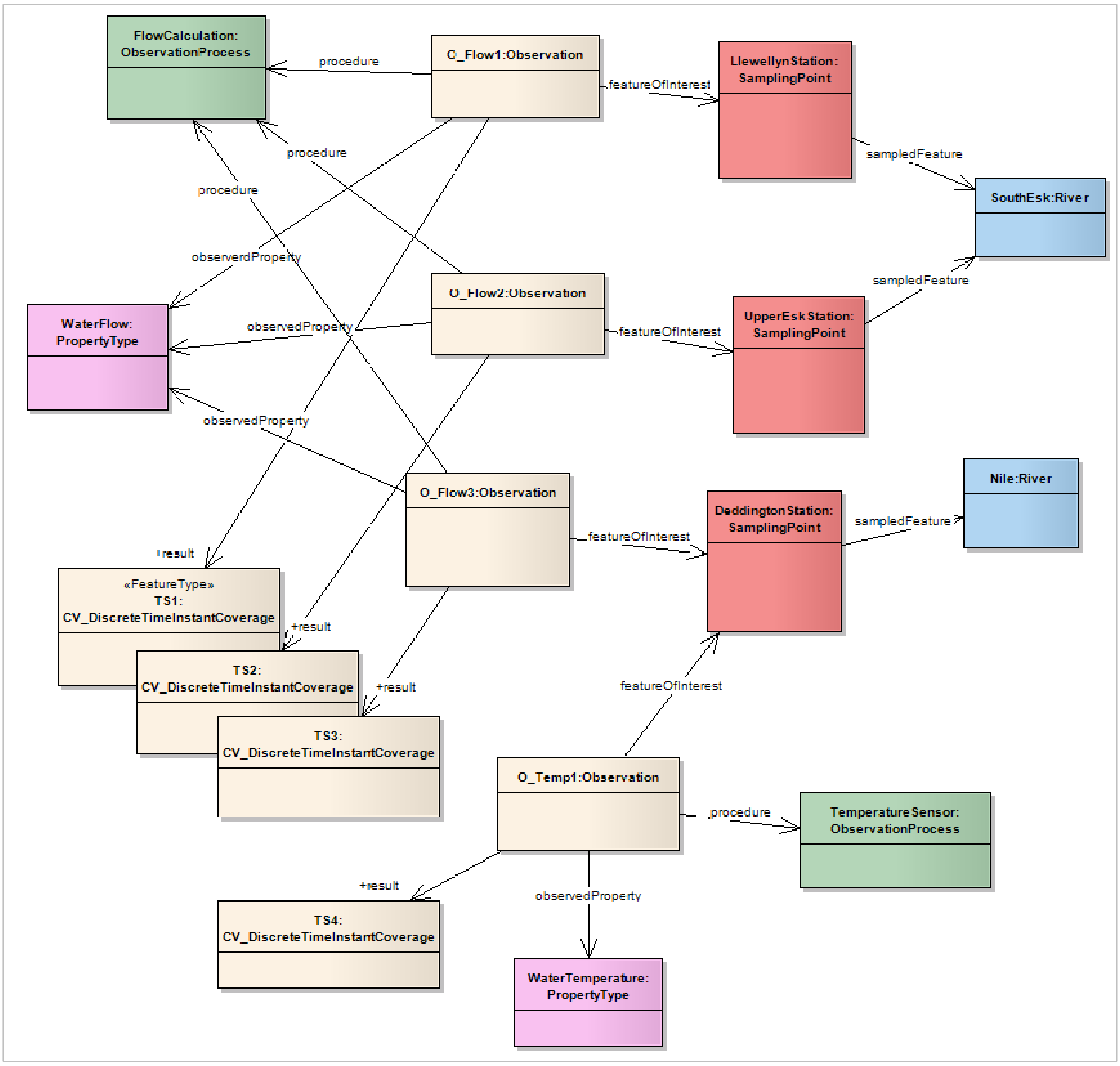


Figure 4 - Example sampling feature relationship

Sampling feature collections are also defined, allowing for logical groupings of sampling points. Such groupings of sampling points are common within hydrology, but are often given different names. Sometimes a 'site' is a collection of 'stations'; another view of a 'site' is a collection of measuring 'locations'. The conceptual grouping is generally similar but naming convention differs widely. These groupings may also have relationships to larger, spatial groupings such as catchments, hydrologic units etc. Defining approaches to correctly capturing such relationships within hydrological data sets is an area requiring further work.

The O&M concepts can easily be adapted to describe hydrological observations. For example, observing salinity in a lake will produce salinity in mg/L (result) which is an estimate of the salinity (observed property) in Eagle Lake (feature of interest) using salinity meter 00435 (procedure) [WAL2009].

O&M can be compared to the model at the core of the Observation Data Model (ODM) [TAR2008]. ODM similarly defines a separation of the core elements involved in the observations, with a specific focus on hydrological observations, shown in Figure 5.

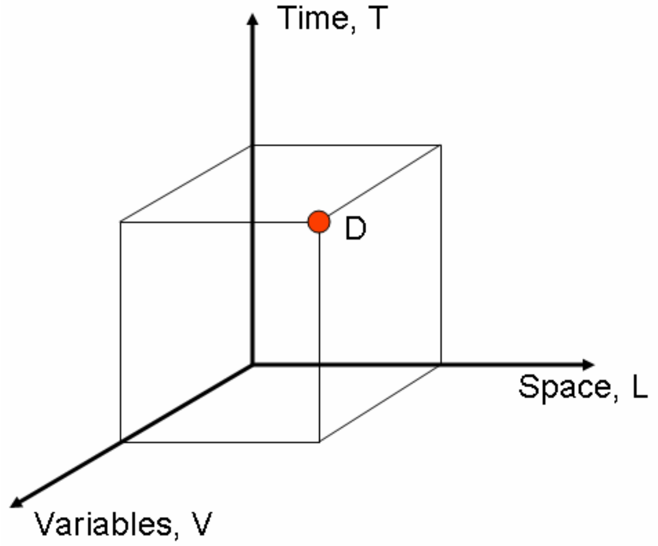


Figure 5 - The ODM model for hydrological observations

The ODM takes a more data centric viewpoint when compared to the broader observation view that O&M takes. Elements such as the description of the observation procedure are built upon the ODM model whereby O&M contains it as a core feature of its model. The differences have implications for how data is interpreted by end users and systems.

7.3 Standards for hydrological information

Existing standards for hydrology data all have a different focus, driven by a particular need for standards in a particular context. This paper examines significant standards of relevance with the aim of capturing core requirements for hydrological observational data. A broader survey of existing standards has been performed in [LEF2008].

Below is a brief summary of existing standards of interest to this paper and how they may be relevant to the harmonisation process.

7.3.1 ArcHydro

ArcHydro [MAI2002] is a data model (Figure 6) for Water Resources and has focused on surface water with input from key state, national, and international contributors. It is implemented as a geodatabase schema. It is widely utilized within the hydrologic community. It is simple and designed to be extended by the users of the data model. A toolset based on the ArcHydro data model is available for ArcGIS desktop applications. The data model presented for time series information only covers the basic information that is needed for analysis.

The ArcHydro data model is undergoing revision to better incorporate “series” into the model. Four additional conceptual information sets are being added: time series (in the original), feature series, attribute series, and raster series.

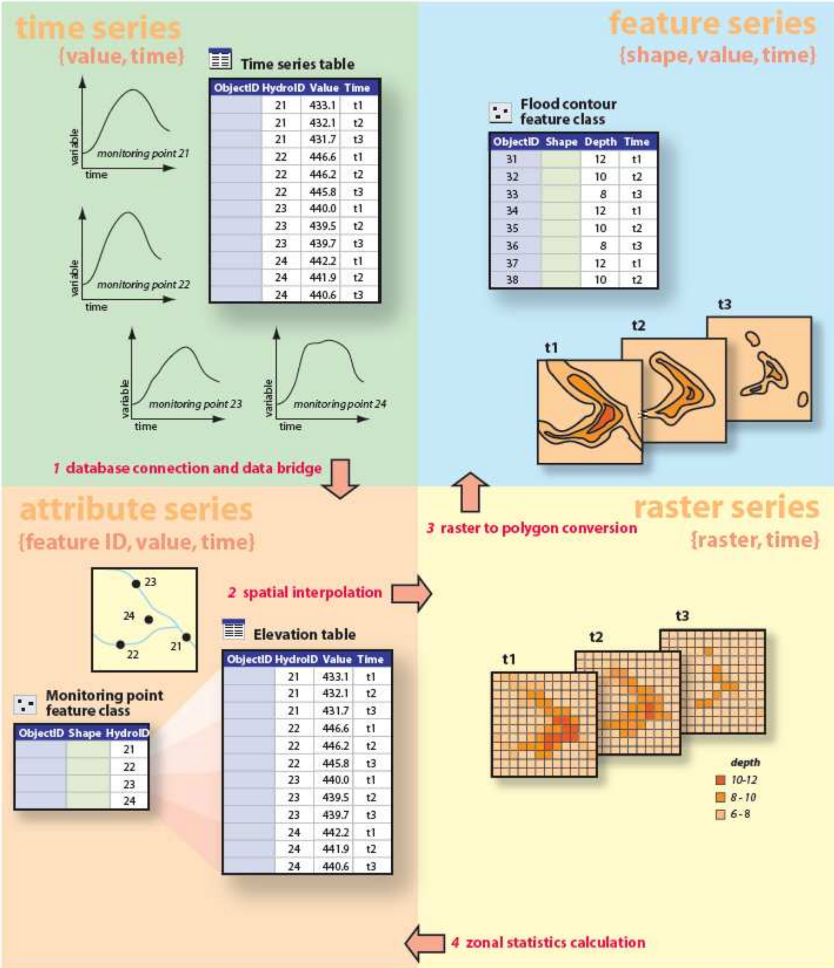


Figure 6. ArcHydro 2 conceptual model

7.3.2 WaterML1.0

The Consortium for the Advancement of Hydrological Sciences Inc (CUAHSI) has developed the WaterML standard, now in version 1.1, which allows for the encoding of hydrological observations via their WaterOneFlow web services. The initial driver for the development of WaterML1.0 was "… to encode the semantics of hydrologic observation discovery and retrieval and implement water data services in a way that is both generic and unambiguous across different data providers, thus creating the least barriers for adoption by the hydrologic research community."

WaterML1.0 is implemented as an XML schema and does not currently make use of OGC or other existing standards. The semantics utilized are from the CUAHSI Observations Data Model [TAR2008]. One of the future goals of developing a harmonised observation model is to allow WaterML1.0 to converge with existing standards.

7.3.3 Australian Water Data Transfer Format

The Water Data Transfer Format (WDTF) is currently being developed by the Bureau of Meteorology and CSIRO as part of the water information research and development alliance. It forms part of the Bureau of Meteorology's AWRIS software. The scope of the format is to allow for the encoding of information that must be supplied to the Bureau from state water agencies or organisations that take hydrological measurements. The standard not only addressed observational data, but also descriptions of features (storages, water courses), transactional information (for synchronising with a data warehouse), conversions (e.g. a rating table conversion) and water quality samples. Version 1.0 is planned to include groundwater observations.

This format makes use of the O&M specification, through a simple features GML profile [ISO19125-1] that restricts certain aspects such as the available geometries and complexity of types. It also uses GML for spatial types.

7.3.4 United States Environmental Protection Agency (EPA) – Water Quality Exchange (WQX)

WQX is focused on the exchange of water quality information. It is based on the Environmental Sampling, Analysis, and Results (ESAR) data standard [ESA2006] which was developed to facilitate the sharing of laboratory result data. EPA also provides a validation service that allows for documents to be validated against the schema definitions.

The WQX standards are developed by the Environmental Data Standards Council (The Council) which is comprised of ten members from Tribes, States and US EPA. The Council's primary function is to develop and adopt Data Standards - documented agreements on terms, definitions, and formats - when there is an environmental business reason. Version 2.0 of the standard is used by the US EPA, and the USGS to deliver water quality information over web services and REST interfaces.

7.3.5 XHydro

The goal of developing XHydro was to allow for the transmission of water level data through web services for the German Federal Waterways and Shipping Administration. It goes slightly further in that it specifies an XML schema for the encoding of generic time series, with an extension that is tailored specifically for water level and discharge data. The time series model is the key point of interest.

The documentation of XHydro also refers to the creation of a generic conceptual model from which other schemas can be created to address particular needs; the core premise of the proposed methodology. The modularity of the model also assists when assessing the standard from a harmonisation point of view.

7.3.6 UK Environmental Agency time series data exchange

The UK Environmental Agency developed the EA Time Series Data Exchange Format (UK-EA-TS) to address the need "…to exchange a variety of sets of time-series data with both internal and external stakeholders". The primary type of time series were hydrological data types such as lake and reservoir levels, river levels and flows, and rainfall.

The standard addresses:

- Rainfall amounts

- River levels and flows
- Tide levels
- Lake and reservoir levels
- Groundwater levels
- Areal modelled evaporation, soil moisture deficits, etc.
- Continuously monitored water quality parameters: e.g. dissolved oxygen and ammonia quantities
- Climate station data: e.g. temperatures, wind speed and radiation.

7.3.7 The French Data Reference Centre for Water (SANDRE)

The SANDRE system provides national infrastructure for sharing water information within France. Its architecture is based on the use of a common language for water information that has defined standards for a number of areas of both spatial and observational hydrological information. It has made use of ISO and OGC standards, using ISO19115 for its metadata definitions and a number of OGC service interfaces for exposing data assets. The information models developed within this project are well developed and in active use and are thus of particular interest to the harmonisation process.

7.4 Other standards of relevance

There are a number of existing standards that are of interest for their approach on either using other standards or solving similar harmonisation issues. The aspect of each model that is relevant is outlined.

7.4.1 SWE Common

Sensor Web Enablement (SWE) is a group within the OGC that develops standards associated with encoding and transmitting sensor data as well as other functions such as providing sensor descriptions, control, alerting and processing. There is a common specification within SWE, known as SWE Common, which defines re-usable data structures and types such as data records, arrays and techniques for defining phenomena and more. O&M makes use of parts of this specification as to other standards within the SWE group.

7.4.2 Climate Science Modelling Language (CSML)

CSML makes use of the ISO coverage model extensively for its modelling of the result sets for climate science. These are often gridded data sets but also cover time series data. CSML also leverages existing OGC standards such as GML and SWE. They also employ the MDA approach to developing information models.

7.4.3 Integrated Ocean Observing System (IOOS)

IOOS is “a federal, regional, and private-sector partnership working to enhance our ability to collect, deliver, and use ocean information.” [IOS2009]. The Data Integration Framework initiative within IOOS is focussed on improving management and delivery of ocean observation data. The project is using a number of open standards for information structure and web service delivery. They have made use of the O&M, GML and SWE Common to develop an information model suitable for ocean observing systems.

7.4.4 Ground Water Mark-up Language (GWML)

GWML makes use of GeoSciML to define a model for capturing information on groundwater, with a focus on the definitions of features. It has used similar model-driven approaches to developing the model and as such can be used as a reference for methodology.

7.4.5 The Global Runoff Data Centre (GRDC) Hydrologic Datasets - metadata profile

The GRDC operates under the auspices of the World Meteorological Organisation (WMO) and have recently developed a metadata profile for hydrological data sets. They have used a model-driven approach to the development of this standard, with close alignment to the ISO19115 metadata specification.

The profile also makes use of the O&M model to define observations and their associated properties. This specification is of interest for its use of standards as well as its definitions of hydrographical features.

7.4.6 Marine Metadata Interoperability

The mission of the Marine Metadata Interoperability (MMI) project is "*To promote the exchange, integration and use of marine data through enhanced data publishing, discovery, documentation and accessibility.*" The project is funded by the National Science Foundation (NSF) but has been supported by a number of other international organisations. The project has published a number of relevant outputs on the description and handling of metadata in distributed environments. These 'guides' provide guidance on the use of metadata standards, URI schemes, controlled vocabularies and semantic techniques for data mark-up.

8 Harmonising core concepts

8.1 Defining existing concepts

In order to harmonise on a model for hydrological observations, this paper will analyse the components of existing standards and define a core set of properties that must be represented in a common model.

The focus of this analysis is on the first category (in-situ, fixed observation style) of data identified in section 6 but other areas such as descriptions of features, processes and other areas will be touched on. Hydrological data sets contained within the other identified categories of data, such as rating curve descriptions, gauging measurements, and water quality, will be addressed in future work.

The following table rates existing standards along axis of interest for harmonisation. The scores are not just based on ability but on *relevance* within each area to the goal of a conceptual model for hydrological observational data. This is not a general qualitative rating for each standard.

The + ratings give a relevance score for each standard against each aspect of the O&M conceptual model. This allows us to focus on particular aspects of each standard to identify concepts for harmonisation. For example, GWML is a standards-driven model with some areas of interest in its definition of features and procedures, but does not have relevance in terms of encoding time series.

Each aspect of the table is described as follows:

Results: the generated values of an observation (e.g. time series) and the metadata describing the result structure. (ODM: DataValues)

Features: descriptions of the real world objects involved in the observation (e.g. gauging stations, rivers, lakes etc.)

Procedures: the process involved in making an observation (e.g., turbidity sensor, laboratory procedure etc.)

Properties: the phenomena that are the subject of observation (e.g. water level, rainfall etc.)

	Use of standards		**Features**	**Procedures**	**Properties**	**Results**
Australian Water Data Transfer Format	++		+	+	+	+++
WaterML1.0	-		+	+	+	+++
XHydro	+		-	-	+	++
UK Environmental Agency time series data exchange	-		-	-	-	++
Climate Science Modelling Language	++		+	-	+	++
Ground Water Mark-up Language (GWML)	++		+	+	-	-
INSPIRE Hydrography model	+++		+++	-	-	-

GRDC Hydrologic Datasets - metadata	++		+++	++	+	+
Integrated Ocean Observing System (IOOS)	++		+	++	+	+
Marine Metadata Interoperability	++		-	-	+++	-

Table 1 - Relevance of standards to the harmonisation process

- = Standard does not contain a relevant approach to the concept

+ = Standard at least contains a reference to the concept

++ = Standard defines the concept partially

+++ = Standard provides mechanisms for full description of the concept

Table 1 indicates that most identified standards can contribute to various aspects of a common model. In this document the focus will be on WaterML and WDTF for results, with EK-EA-TS and XHydro providing extra inputs. Further revisions of this document may incorporate contributions from other standards. It is useful to have a number of standards when looking at the result model to see the variety of interpretations of time series that exist.

For feature definitions, there are broader standards available such as the INSPIRE and GRDC models with their definitions of hydrographical features. Most models make reference to external procedure definitions through identifiers, but a few provide partial descriptions of the underlying process types; such as the GRDC profile and the IOOS Data Integration Framework.

The next sections will address each of the areas of concern in turn, taking into account the appropriate standards for analysis of concepts and approach. The initial focus of harmonisation has been the definition of results.

9 Results

As defined in section 6, the in-situ style of hydrological observations primarily produce a time series of values that represent an estimated value for a given phenomenon at a particular time (or across a time period). The focus of this section is to investigate a consistent model for representing hydrological time series.

9.1.1 Time series

Time series values can have a different relationship to the temporal spacing in which they occur. The differences come about either through different measuring processes (e.g. a sensor) or the result of post-processing a time series (e.g. result of an aggregation calculation). Capturing the relationship between points is important when interpreting the values for analysis or further processing.

These time series types are particularly important when one is to perform interpolation between time series points in order to estimate the value of a phenomenon where no measurement occurs. A re-usable information model within hydrology must make explicit the type of data that is being represented; it must also be sufficiently precise to allow other models to map their structure onto this model.

9.1.2 Existing time series models

The concepts that are captured in the existing standards provide a baseline set of requirements that need to be investigated for a time series model.

The following examples show the time series encoding section of existing standards. The examples shown are not encoding the same data; they are provided from the specifications and only contain time series descriptions. Full examples are listed in Annex A.

9.1.2.1 WDTF

```xml
<wdtf:result>
    <wdtf:TimeSeries>
      <wdtf:defaultInterpolationType>InstVal</wdtf:defaultInterpolationType>
      <wdtf:defaultUnitsOfMeasure>m</wdtf:defaultUnitsOfMeasure>
      <wdtf:defaultQuality>quality-A</wdtf:defaultQuality>
      <wdtf:timeValuePair
        time="2001-07-31T20:12:01+10:00">1.25</wdtf:timeValuePair>
      <!-- This time point is missing -->
      <wdtf:timeValuePair
        time="2001-08-01T20:15:01+10:00"
        comment="text"
        interpolationType="InstVal"
        xsi:nil="true"/>
      <wdtf:timeValuePair
        time="2001-08-02T20:10:01+10:00"
        comment="Sample comment"
        quality="quality-B">1.28</wdtf:timeValuePair>
    </wdtf:TimeSeries>
</wdtf:result>
```

9.1.2.2 WaterML1.1

```xml
<values>
 <value censorCode="nc" dateTime="2007-11-07T13:00:00" timeOffset="-07:00"
dateTimeUTC="2007-11-07T20:00:00" methodCode="25" sourceCode="3" labSampleCode="9188"
qualityControlLevelCode="2">10.5</value>
```

```xml
    <value censorCode="nc" dateTime="2007-11-13T12:30:00" timeOffset="-07:00"
dateTimeUTC="2007-11-13T19:30:00" methodCode="25" sourceCode="3" labSampleCode="9398"
qualityControlLevelCode="2">2.5</value>
    <value censorCode="nc" dateTime="2007-11-21T14:00:00" timeOffset="-07:00"
dateTimeUTC="2007-11-21T21:00:00" methodCode="25" sourceCode="3" labSampleCode="9509"
qualityControlLevelCode="2">7.2</value>
    <value censorCode="nc" dateTime="2007-12-05T11:00:00" timeOffset="-07:00"
dateTimeUTC="2007-12-05T18:00:00" methodCode="25" sourceCode="3"
labSampleCode="G120507-WELL-TSS" qualityControlLevelCode="2">2.5</value>
    <value censorCode="nc" dateTime="2007-12-20T14:05:00" timeOffset="-07:00"
dateTimeUTC="2007-12-20T21:05:00" methodCode="25" sourceCode="3"
labSampleCode="G122007-WELL-TSS" qualityControlLevelCode="2">2.5</value>
    <qualityControlLevel qualityControlLevelID="2">
     <qualityControlLevelCode>2</qualityControlLevelCode>
     <definition>Derived products</definition>
     <explanation>Derived products that require scientific and technical interpretation and may
include multiple-sensor data. An example is basin average precipitation derived from rain gages using
an interpolation procedure.</explanation>
    </qualityControlLevel>
    <method methodID="25">
     <methodCode>25</methodCode>
     <methodDescription>Water chemistry grab sample collected by technicians in the
field.</methodDescription>
    </method>
    <source sourceID="3">
     <sourceCode>3</sourceCode>
     <organization>Utah State University Utah Water Research Laboratory</organization>
     <sourceDescription>Water chemistry monitoring data collected by Utah State University as part
of a National Science Foundation funded test bed project.</sourceDescription>
     <contactInformation>
      <contactName>Amber Spackman</contactName>
      <typeOfContact>main</typeOfContact>
      <email>amber.s@aggiemail.usu.edu</email>
      <phone>1-435-797-0045</phone>
      <address xsi:type="xsd:string">8200 Old Main Hill
,Logan, Utah 84322-8200</address>
     </contactInformation>
     <sourceLink>http://water.usu.edu/littlebearriver</sourceLink>
     <citation>Water chemistry monitoring data collected by Jeff Horsburgh, David Stevens, David
Tarboton, Nancy Mesner, Amber Spackman, and Sandra Gurrero at Utah State University as part of a
National Science Foundation funded WATERS Network Test Bed project.</citation>
    </source>
    <sample sampleID="26">
     <labSampleCode>9188</labSampleCode>
     <sampleType>Grab</sampleType>
     <labMethod>
      <labCode>9188</labCode>
      <labName>USU Analytical Laboratory</labName>
      <labOrganization>Utah State University</labOrganization>
      <labMethodName>EPA 340.2</labMethodName>
     </labMethod>
    </sample>
    <sample sampleID="32">
     <labSampleCode>9398</labSampleCode>
     <sampleType>Grab</sampleType>
```

```
    <labMethod>
     <labCode>9398</labCode>
     <labName>USU Analytical Laboratory</labName>
     <labOrganization>Utah State University</labOrganization>
     <labMethodName>EPA 340.2</labMethodName>
    </labMethod>
   </sample>
   <sample sampleID="38">
    <labSampleCode>9509</labSampleCode>
    <sampleType>Grab</sampleType>
    <labMethod>
     <labCode>9509</labCode>
     <labName>USU Analytical Laboratory</labName>
     <labOrganization>Utah State University</labOrganization>
     <labMethodName>EPA 340.2</labMethodName>
    </labMethod>
   </sample>
   <sample sampleID="83">
    <labSampleCode>G120507-WELL-TSS</labSampleCode>
    <sampleType>Grab</sampleType>
    <labMethod>
     <labCode>G120507-WELL-TSS</labCode>
     <labName>USU Aquatic Biogeochemistry Laboratory</labName>
     <labOrganization>Utah State University</labOrganization>
     <labMethodName>Total Phosphorus</labMethodName>
    </labMethod>
   </sample>
   <sample sampleID="171">
    <labSampleCode>G122007-WELL-TSS</labSampleCode>
    <sampleType>Grab</sampleType>
    <labMethod>
     <labCode>G122007-WELL-TSS</labCode>
     <labName>USU Analytical Laboratory</labName>
     <labOrganization>Utah State University</labOrganization>
     <labMethodName>EPA 340.2</labMethodName>
    </labMethod>
   </sample>
   <censorCode>
    <censorCode>nc</censorCode>
    <censorCodeDescription>not censored</censorCodeDescription>
   </censorCode>
  </values>
```

9.1.2.3 UK-EA-TS

```
<!-- Four days of daily mean flows -->
<SetofValues parameter="Flow" dataType="Mean" period="Day" characteristic="Derived"
  units="m3/s" startDate="2003-04-20" endDate="2003-04-23" dayOrigin="09:00:00">
  <Value date="2003-04-20" flag1="1" flag2="1" percentFlag2="100">15.63</Value>
  <Value date="2003-04-21" flag1="2" flag2="1" percentFlag2="92.5">16.21</Value>
  <Value date="2003-04-22" flag1="1" flag2="1" percentFlag2="87" flag3="2"
    percentFlag3="5.5">16</Value>
  <Value date="2003-04-23" flag1="2" flag2="1" percentFlag2="85.2" flag3="2"
```

```
      percentFlag3="14.8">17.36</Value>
    <Comment startDate="2003-04-22">This daily mean flow was derived from an incomplete set
      of good and suspect data but has been validated and found to be of good overall
      quality</Comment>
    <Comment startDate="2003-04-21" endDate="2003-04-23">This demonstrates that you can have
      nested comments</Comment>
  </SetofValues>
```

9.1.2.4 XHydro

* The element names used in the schema are abbreviations for the actual concept names in the XHydro model.

```
<?xml version="1.0" encoding="UTF-8"?>
<tsel xmlns="http://xhydro.org/minimal/2007/06"
  xmlns:d="http://www.disy.net/device"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://xhydro.org/minimal/2007/06
http://www.xhydro.org/download/schemas/v200706/schemas/XHydro.xsd">
  <ext />
  <tse>
    <!-- snip -->
    <tsmd>
      <tsd>
        <dn>ddts</dn>
        <dd>A dummy dis device to measure time.</dd>
      </tsd>
      <tsq>
        <tsmi>1.5E-6</tsmi>
        <tsqr codeList="disy1" codeListAgency="disy"
          codeListVersion="1.0">
          ownCode
        </tsqr>
      </tsq>
    </tsmd>
    <d>
      <tde>
        <!-- No timestamp is given because isochron -->
        <vls>
          <v>
            <vq>
              <vmi>6E-4</vmi>
              <xvqr>affected</xvqr>
            </vq>
            <vl>
              <pt>
                <xrs>32632</xrs>
                <px>5.0</px>
                <py>6.0</py>
              </pt>
            </vl>
            <vf>4.5</vf>
          </v>
```

```
          <v>
            <vf>4.6</vf>
          </v>
          <v>
            <vq>
              <xvqr>missing</xvqr>
            </vq>
            <va xsi:nil="true" />
          </v>
        </vls>
      </tde>
    </d>
  </tse>
</tsel>
```

9.1.3 Time series metadata comparison

The following table provides a summary of the concepts that each format is capturing with a description of where the property has been modelled. The concepts within the table were identified as being present across a number of the existing standards. The concepts are all related to the ‘result’ structure within the Observations & Measurements model, which we are defining as being a time series for the purpose of the initial phase of definition.

Table 2 - Comparison of time series meta data elements

Name: element name used.

Defined: describes where in schema the property is defined.

Type: typing mechanism used.

Concept	WDTF	WaterML1.0	XHydro	EA UK TS
Qualifiers	Name: valueQualifier. Defined: Per measurement (not time series value). Type: Unconstrained string.	Name: qualifiers Defined: Per point, with ability to define qualifiers as series metadata. Type: xsi:Token	Name: dataValueQualityRemark (grouped quality and qualifiers) Defined: Per point. Type: Locally defined code list element OR externally defined code list element OR free text.	Name: qualifier Defined: Per parameter (series, set of values). Type: Locally defined code lists element only.
Quality	Name: quality Defined: Per point. Type: Locally defined code list.	Name: quality Defined: Per point. Type: Locally defined code list.	Name: dataValueQuality Defined: Per point. Type: Locally defined code list element OR externally defined code list element OR free text.	Name: flag Defined: Per point. Multiple quality levels specified. Type: Locally defined code lists. Provides 10 levels of quality flags.
Comments	Name: comment Defined: Per point. Type: Unconstrained string.	N/A	N/A	Name: comment Defined: Series level but allows a period of record to be specified, allowing multiple comments to be encoded. Type: Unconstrained string.
Interpolation	Name: interpolationType.	Name: dataType.	Name: dataType.	Name: dataType

type	Defined: Default per series or per point. Type: Locally defined code list.	Defined: Per variable (set across a time series). Type: Locally defined code list.	Defined: Per time series. Type: Locally defined code list.	Defined: Per time series. Type: Locally defined code list.
Processing	Name: processingType Defined: Default per series or per point. Type: Locally defined code list.	Name: valueType Defined: Per variable. Type: Locally defined code list.	N/A	Name: characteristic Defined: Per time series. Type: Locally defined code list.
Accuracy	N/A	Name: accuracyStdDev Definition: Per value. Type: double	Name:dataValueMeasurementInaccuracy Defined: Per value. Type: float	N/A
Units	Name: uom Defined: Per point. Type: Locally defined code list.	Name: unit Defined: Per variable. Type: Complex type containing unit code (from code list), description, abbreviation and type (mass, length, velocity etc.).	Name: unit Defined: Per parameter. Type: Locally defined code list OR externally defined code list.	Name: units Defined: Per time series. Type: Locally defined code list.
Offsets	N/A	Name: offset Defined: per value Type: Complex type allowing offset value, type, description and units to be defined.	N/A	N/A
Null Values	Uses xsi:nillable.	Defines a NoDataValue per document to describel a value to	Uses xsi:nillable.	NaN, INF and –INF through the use of the W3C float type.

		indicate null values.		

Locally defined code list: The technique used by the schema is an XML Schema enumeration simple type. This is can be checked against using schema validation.

Unconstrained string: Free text. No validation is implied.

9.1.4 Per value vs. per time series properties

One of the common structural differences of the compared properties for time series is if they are defined per value, or if the property holds across a full time series. WDTF has used a default property pattern where one is able to specify for the whole series the default value, or define individually for each value. The need to define properties per time series point arises from characteristic changes within the series. WDTF employs this for interpolation type, processing code, quality and unit of measure. This model is a useful trade off between flexibility and verbosity of the encoding.

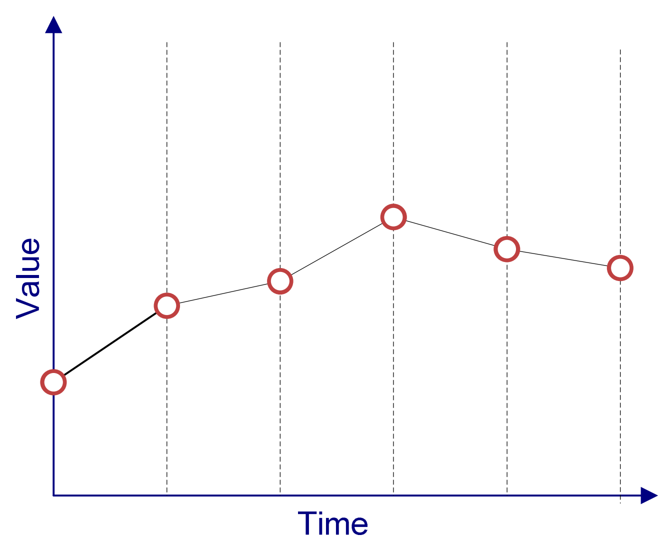
The ways in which per value vs. per series properties are structured have implications for the discovery process. These will need to be investigated further.

9.1.5 Interpolation Types

One of the core aspects of time series is the relationship between the value and its associated time instant (or period). This relationship is determined by the procedure that was used to make the estimate that the value represents. In most data models this is referred to as the data type or interpolation type.

Whilst it is possible to provide a placeholder to allow users to specify the interpolation or data type of a time series within their given context (through code lists), it is important to understand the way existing standards deal with this concept, as it is pivotal for correct interpretation of a time series, consistent post-processing and summary statistics .

Table 2 – Comparison of interpolation/data types



9.1.6 Continuous/Instantaneous

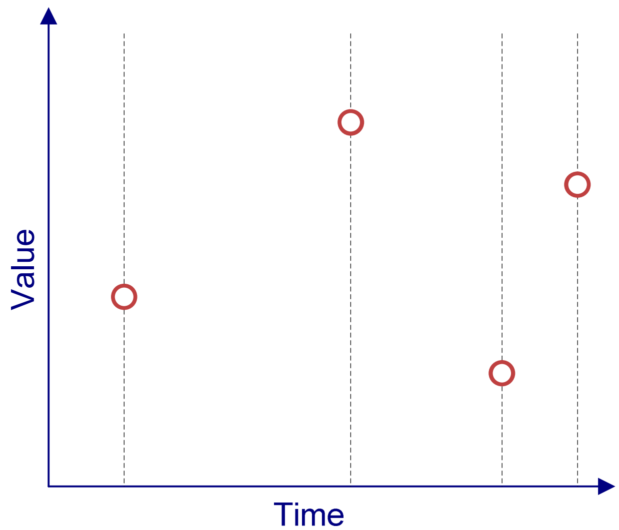
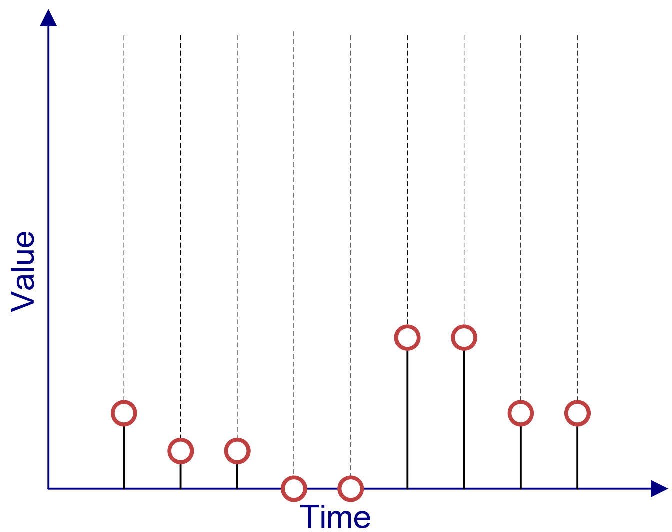
WDTF: InstVal

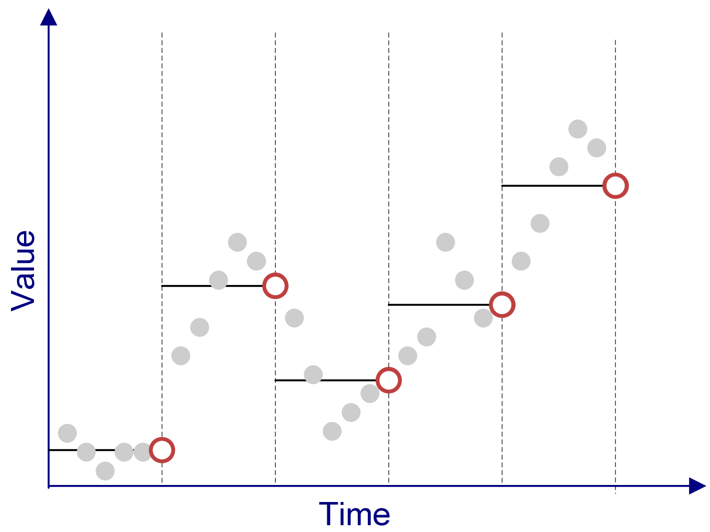
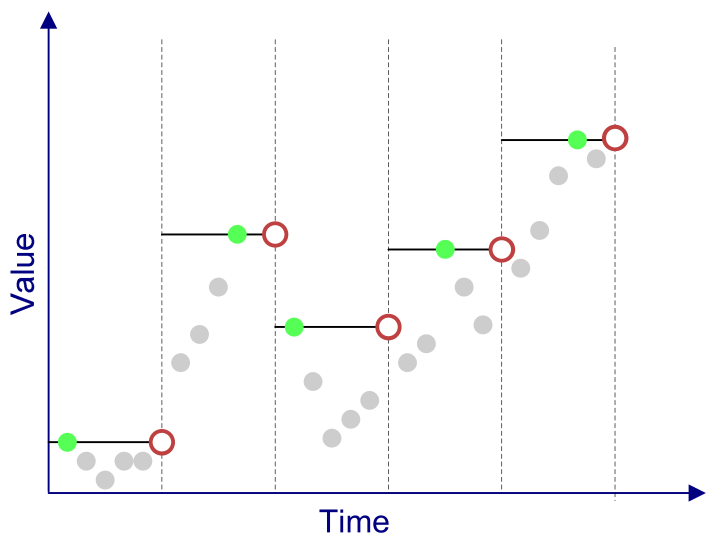
WaterML1.0: Continuous

XHydro: contData

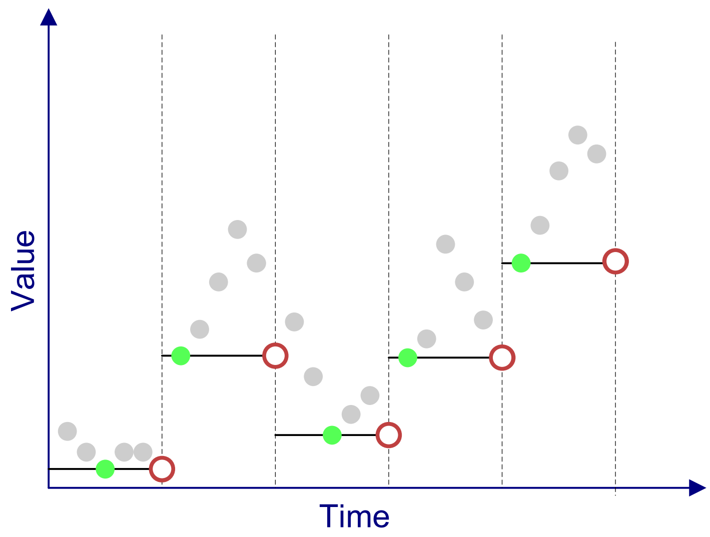
UK-EA-TS: Instantaneous

A continuous time series indicates the observation result is the value of a property at the indicated instant in time. The points are essentially connected and interpolation may occur between points in order to estimate the value of the property between points. The appropriate time spacing between successive points to mimimise interpolation errors is related to rate of change

	(wrt time) of the property.
	**9.1.7 Discontinuous** WDTF: NoJoin WaterML1.0: Sporadic XHydro: N/A UK-EA-TS: Instantaneous The sampling of the property occurs such that it is not possible to regard the series as continuous. The time between samples is too large to classify the measurements as continuous. Example: Infrequent water sample measuring pH.
	**9.1.8 Instantaneous total** WDTF: InstTot WaterML1.0: Incremental XHydro: contTotal UK-EA-TS: Event Value represents a total attributed to a specific time instant. This is normally generated from an event based measuring device such as a tipping bucket rain gauge. Example: An individual tip of a tipping bucket rain gauge.

	**9.1.9 Average in preceding interval** WDTF: PrecVal WaterML1.0: Average XHydro: aggMean UK-EA-TS: Mean[1] Value represents the average value over the preceding interval. Example: Daily mean discharge.
	**9.1.10 Maximum in preceding interval** WDTF: PrecMax WaterML1.0: N/A (diff interval) XHydro: aggMax UK-EA-TS: Maximum Value represents the maximum value that was measured during the preceding time interval. Example: Monthly maximum discharge

[1] Assumption has been made the interval being described is the preceding interval



9.1.11 Minimum in preceding interval

WDTF: PrecMin

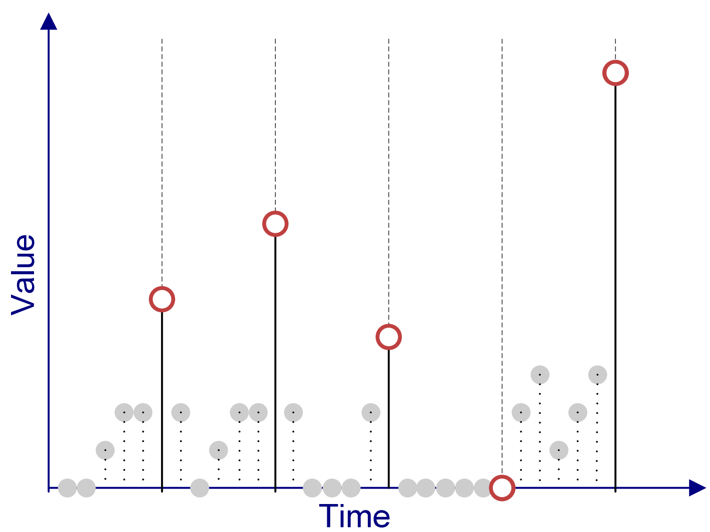
WaterML1.0: N/A (diff interval)

XHydro: aggMin

UK-EA-TS: Minimum

Value represents the minimum value that was measured during the preceding time interval.

Example: Daily minimum temperature.



9.1.12 Preceding total

WDTF: PrecTot

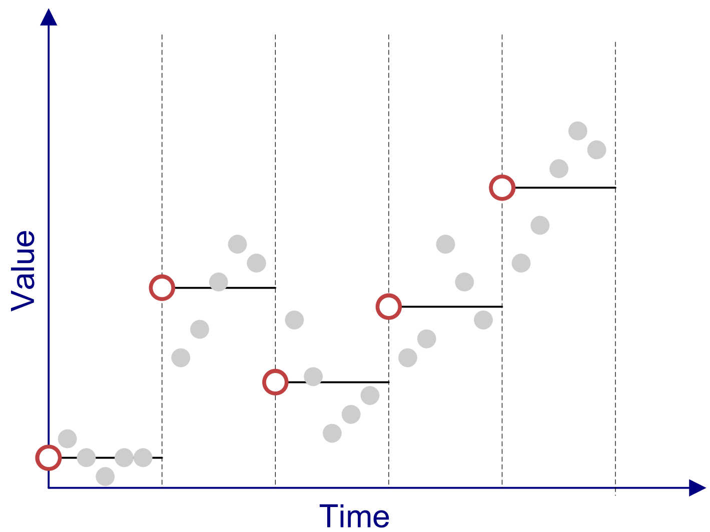
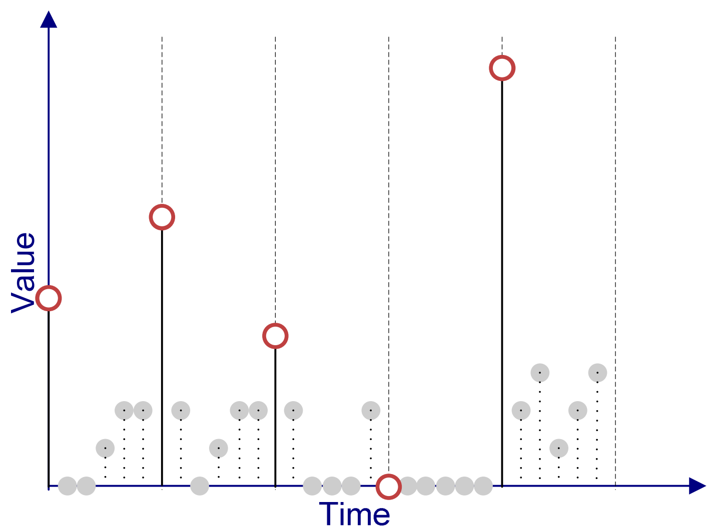
WaterML1.0: N/A (diff interval)

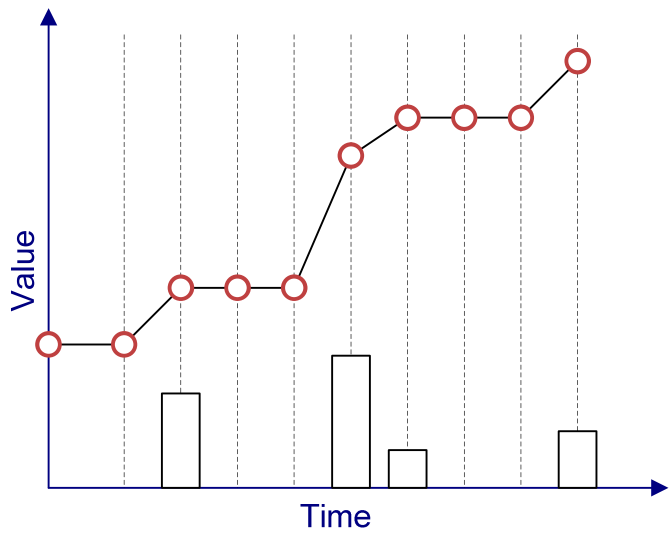
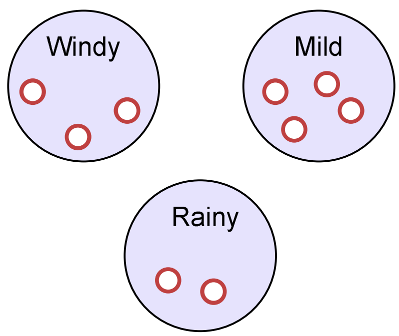
XHydro: aggTotal

UK-EA-TS: Total

Value represents the total of measurements taken within the previous time interval.

Example: Daily pan evaporation

	**9.1.13 Average in succeeding interval** WDTF: SucVal WaterML1.0: N/A XHydro: aggMean UK-EA-TS: N/A Value represents the average value over the following interval. Example: Daily mean discharge encoded as value representing beginning of interval (ODM style).
	**9.1.14 Succeeding total** WDTF: SucTot WaterML1.0: N/A XHydro: aggTotal UK-EA-TS: N/A Value represents the total of measurements taken within the following time interval. Example: Total daily rainfall from 9am to 9am.

	**9.1.15 Cumulative** WDTF: accumulated (metadata) WaterML1.0: Cumulative XHydro: aggTotal UK-EA-TS: Cumulative Total Value represents an accumulated total since a reset time. Example: Total rainfall across a period, total river discharge etc.
	**9.1.16 Categorical** WDTF: Different result type WaterML1.0: Categorical XHydro: Different result type UK-EA-TS: N/A A categorical measurement represents named 'bins' to which values can be assigned. Example: human weather observations: 'mild', 'windy', 'rainy' etc.

9.1.17 Handling cumulative data

Data of type instantaneous total is often accumulated across a period to show the running total since accumulation commenced. This is often the way in which rainfall data is reported (e.g. total rainfall from 9am to 9am – 24 hours of accumulated instantaneous total data). Existing models handle this concept slightly differently; Table 4 and Table 5 give a summary of some of the mechanisms employed to handle these concepts.

WaterML1.0 and UK-EA-TS captures accumulation as a separate data type where as WDTF designates at a time series level whether the series is accumulated. UK-EA-TS and WDTF store the accumulation begin and end points to allow for correct de-accumulation to performed.

Table 3 - Mapping interpolation/data types between models

WDTF	XHydro	UK-EA-TS	WaterML1.0	Comments
InstVal	contData	Instantaneous	Continuous	
InstTot	contTotal	Event	Constant Over Interval	
PrecVal		Mean	N/A	See note on reporting intervals
PrecMax	aggMax	Maximum	N/A	See note on reporting intervals
PrecMin	aggMin	Minimum	N/A	See note on reporting intervals
PrecTot	aggTotal	Total	Incremental*	
PrecDir	N/A	N/A	N/A	
SuccVal	N/A	N/A	Average	
SuccTot	N/A	N/A	Incremental	
NoJoin	N/A	N/A	Sporadic	
N/A	N/A	N/A	Mode	
N/A	N/A	N/A	Categorical	
N/A	aggStdDev	N/A	StandardDeviation	
N/A	N/A	N/A	Unknown	
N/A	N/A	N/A	Minimum	See note on reporting intervals
N/A	N/A	N/A	Maximum	See note on reporting intervals
N/A	N/A	N/A	Best Easy Systematic Estimator	
Captured in the time series meta	N/A	Cumulative Total	Cumulative	See table below

data as Boolean.				
N/A	aggMedian	N/A	Median	
N/A	N/A	N/A	Variance	
N/A	aggMovingMean	N/A	N/A	

Table 4 - Comparision of cumulative data descriptions

Name: element name used.

Defined: describes where in schema the property is defined.

Type: typing mechanism used.

Concept	WDTF	WaterML1.0	XHydro	EA UK TS
accumulationPeriodicAnchorTime	Name: accumulationPeriodicAnchorTime Defined: Per series Type: xsi:time Indicates the base time for the interpolation interval. For example 9am for cumulative rainfall since 9am.	N/A	N/A	Name: dayOrigin Defined: Per Series Type: xsi:time. The time at which a day value begins (eg. 09:00:00 for a water day or a rain day)
accumulationPeriodicIntervalLength	Name: accumulationPeriodicIntervalLength Defined: Per series Type: xsi:duration Indicates the recurring interval from the accumulationPeriodicAnchorTime that the accumulator resets. i.e. readings will be cumulative within the period. For example 1 day for cumulative rainfall since 9am.			

9.1.18 Note on reporting intervals

When values are being attributed to a certain time interval, it must be made explicit which part of the interval the value holds over. It may be the value represents the beginning of the observed value for the interval or the end of the interval.

WDTF makes this explicit in the data type by specifying whether the value holds over the preceding interval or the succeeding interval.

CUAHSI ODM defines the value at a particular time represents the beginning of the interval. This is represented as succeeding interval in WDTF. The justification from ODM is as follows [TAR2008]:

"*Data types 4 to 8 above apply to data values that occur over an interval of time. The date and time reported and entered in to the ODM database associated with each interval data value is the beginning time of the observation interval. This convention was adopted to be consistent with the way dates and times are represented in most common database management systems. It should be noted that using the beginning of the interval is not consistent with the time a data logger would log an observation value. Care should be exercised in adding data to the ODM to ensure that the beginning of interval convention is followed."*

XHydro separates this definition into a time stamp qualifier that specifies whether the time stamp represents the start, middle or end of the interval.

As stated in the ODM specification, loggers generally report the value at the end of the period of measurement but it should be possible to accommodate each type.

Table 5 - Other metadata defining the interpolation interval

Name: element name used.

Defined: describes where in schema the property is defined.

Type: typing mechanism used.

Concept	WDTF	WaterML1.0	XHydro	EA UK TS
Duration	Name: duration Defined: Per point Type: xsi:duration The period over which the measurement applies.	Name: timeSpacing Defined: per series Type: float for regular series	Name: distance Defined: per series Type: xsi:duration For Isochronous series.	Name: period or interval Defined: Per Series Type: Local code list. Expected interval of data particularly applying to rolling accumulations where it is not the same as the data period (eg. 15 min, 1 h, Daily, etc.) - ie. Daily Means may be recorded on an hourly basis.
AnchorPoint	Name: anchorPoint Defined: Per series Type: xsi:dateTime A point indicating the first point in the series so Prec* interpolation types have an earlier bound.	Start time	Start time	Start time

9.1.19 Time

The four major standards considered have similar time stamp fields as show below. These standards all use the Gregorian calendar. Other systems however, such as OpenMI and GPS based systems, do use other calendars such as the Julian calendar.

Table 6 - Comparison of time meta data elements

Name: element name used.

Defined: describes where in schema the property is defined.

Type: typing mechanism used.

Concept	WDTF	WaterML1.0	XHydro	EA UK TS
TimeStamp	Name: time Defined: Per point Type: xsi:dateTime with mandatory time zone. Date only values are required to provide a 00:00:00 time.	Name: dateTime and dateTimeUTC Defined: Per point Type: xsi:dateTime. The dateTime field is mandatory but the UTC one is optional.	Name: timeStampValue Defined: Per point. Type: xsi:dateTime however the value itself is optional. If the isochron element is used the series has a fixed time step (isochronal) and the time stamp is calculated using the isochron "distance".	Name: date and time Defined: Per Point. Type: xsi:date and xsi:time. These are two separate attributes of the value with date being mandatory and time being optional. The time is to the nearest second.
TimeZone	Built into time stamp.	Name: timeOffset Defined: Per point. Type: Locally defined normalised string. There is also a timeZoneInfo block to define the default time zone for the series.	Built into timeStampValue.	N/A
TemporalInaccuracy	N/A	N/A	Name: measurementInaccuracy Defined: Per point with and	N/A

			optional default value. Type: float	
TimeStampQuality	N/A	N/A	XHydro provides extensive metadata on timer quality. These values can also be specified as defaults.	N/A

9.1.20 Timing metadata

Existing standards define a number of timing metadata elements that relate to various temporal properties of time series. It is important to clarify some different concepts in terms of hydrological time series. The names given here are for descriptive purpose only.

- *Reporting frequency*: The time sampling regime of the observation. For a sensor, this would be the frequency of sensor measurements (e.g. 15 minutely); for manual observations this would be how often the observation is recorded, for example daily temperature checks. Termed spacing in the ODM specification [TAR2008].
- *Download/update frequency*: Describes how often a data set is unloaded from a logger or other recording device. This may not be captured and is often not relevant to exchange formats, but is important for understanding update cycles.
- *Regularity*: This describes whether the time distance between points is equidistant. It is often used for performance and compression techniques in systems storing or transmitting data. This can be determined by examining the data set.
- *Spacing (waterml1.0)*: The interval that the sample is measured. For instantaneous, the value is zero. For daily observations, the sampling interval would be 1 day. This is needed because there may be daily datasets which are averaged over a 5 or 7 day period.

In the CUAHSI ODM specification, the idea of support scale is introduced, which contains three components as identified from Blöschl [BLO1995] [BLO1996]. These are extent, spacing and support.

1. Extent is the temporal extent over which the values occur (i.e. start and finish times of the time series).
2. Spacing relates the distance in time between each point. This is essentially a descriptive component as it may be derived by looking at the values. However this information may be useful when discovering data sets to get an idea of the regularity of the measurements. Note that this may actually be different than the sampling frequency that is described in a description of the sensor or procedure making the measurement.
3. Support relates to the time distance between points (as described above).

A harmonised model of the above concepts would allow for clarification of the concepts across standards.

9.1.21 Null Points

It is often the case in hydrological time series that a point will exist but has been marked as a null point. There are a number of reasons that a data point may be flagged as null or missing and sometimes the semantics of this indication is only known to a particular system (e.g. a point may be indicating that between this point and the next null point no interpolation should occur).

Often systems include point comments to indicate the nature of the null point. The system may flag the values as null due to some particular processing that has occurred (e.g. a phenomenon has exceeded allowable limit). Null points may also carry particular semantics within a system, such as being an ‘anchor’ point for the calculation of total values. This point would indicate that it is the first point where the accumulation period has begun.

WDTF allows null values to be expressed using the xsi:Nil attribute (through defining the element as nillable). XHydro also takes this approach.

WaterML1.0 uses the concept named NoDataValue in order to identify a value that will represent null points in the context of a document. This is driven by an often used technique within agencies to use special values to represent null values. Using this technique allows for the particular usage information to be conveyed, but if no technique exists then an arbitrary number must be assigned as a stand-in (typically at either extreme of the supported numeric scale).

In future definitions of null values it would be useful to include a property outlining the reason for a null value. This is something GML captures through its definition of a Nil Reason Type which elaborates on why the value has been omitted.

9.1.22 Values

For hydrological observation results the values for each time series point are generally a measurement that indicates an estimate of the observed phenomenon. This is largely the case across the four core hydrological standards under review with the addition of categorical values that are supported in WDTF, WaterML1.1 and XHydro. XHydro adds the ability to embed binary content such as images through the use of MIME types.

9.1.22.1 Accuracy

WaterML1.0 allows for the specification of accuracy per time series value. This is captured as a double value indicating the standard deviation of the measurement.

XHydro also allows for the specification of accuracy through its data quality definition. It is represented as a decimal value.

SWE Common provides a mechanism for specifying the qualitative values for each measurement. One of these may be a specification of the accuracy of a value.
This gives us the ability to encode “values of precision, accuracy, tolerance, and confidence level.” [BOT2007]

The example below shows two ways in which this approach may be used.


```
<swe:quality>
  <swe:QuantityRange definition=" urn:ogc:def:property:OGC:tolerance2std">
    <swe:value> -0.02 0.02 </value>
  </swe:QuantityRange>
</swe:quality>

<swe:quality>
  <swe:QuantityRange definition= "urn:ogc:def:property:OGC:absoluteAccuracy">
    <swe:uom xlink:href="urn:ogc:unit:percent"/>
    <swe:value>-0.5 0.5</swe:value>
  </swe:QuantityRange>
</swe:quality>
```

This approach is flexible in that it allows users to specify the particular qualitative properties that may be of interest, but this implies there should be a set of well know definitions of accuracy measures that may be used in order for people to interpret the meaning correctly.

9.1.23 Data Quality

Data quality is obviously a commonly represented property. The use of code lists here to constrain possible types is common across the standards, with some providing extra metadata alongside a code. Harmonising on quality code definitions is something that will not be attempted in defining a core model; arguments over common models for data quality have raged long in many data communities. The concept will exist in the model, but will be left for organisations to specify for their context. A comparison of the available types is supplied in Appendix A.

9.1.24 Comments

Per point comments are common occurrences within hydrological time series, describing notes from the field or particular information regarding corrections, shifts or editing that may have occurred to the data. As shown in Table 2, WDTF and UK-EA-TS have the ability to capture time series point comments.

10 Features and sampling features

As described in section 2.1, O&M breaks features into two categories:

- Sampled features: natural, real world features (rivers, storages, dams etc.)
- Sampling features: features involved in the sampling process (gauging stations, bottles, specimens etc.)

Most hydrological data is linked directly to the station where observations are made. The site name will sometimes contain the name of the river being measured, but the connection between the station and the actual river is sometimes implicit in the data (through its coordinates, identifier system etc.) or in an internal system the data is stored in.

For observational data to be incorporated into broader SDIs, it is important to improve definitions of links to actual features that are being measured. This is also important for hydrological modelling, where river networks and flow processes are key in model development.

WDTF has adapted O&M to fit the structure in use by the Bureau where by a site may contain many sensors, all measuring at different locations. This hierarchy is achieved through the use of the SamplingPoint and SamplingGroup concepts. Each sampling group is defined with its properties (name, location, time zone etc.) and each sampling point is associated with this group. This allows for flexible spatial groupings to occur.

WaterML1.1 links observation data to individual sites which represent the location of the measurement being captured. Groupings of sites are achieved through the concept of a network which is defined as “a collection of sites where a particular set of variables is measured” [VAL2009b]. This also allows for similar flexible spatial groupings to occur.

Further work is required on defining features specific to the hydrology domain. There are a number of existing standards that may be utilised here, currently the most relevant are:

- INSPIRE data specification for hydrography [INS2008]
- GRDC metadata profile for hydrologic datasets [DOR2009]
- Australian Hydrologic Geofabric [BOM2008]
- ArcHydro data model [MAI2002]

By making use of similar model-driven and standards-based approaches of these projects, the ability to create connections between observational data sets and the spatial domain which they reference becomes easier.

The GRDC metadata profile describes hydrographic features and their relationships. For example, Figure 7 shows the relationship between basins, catchments and rivers, lagoons, reservoirs etc.

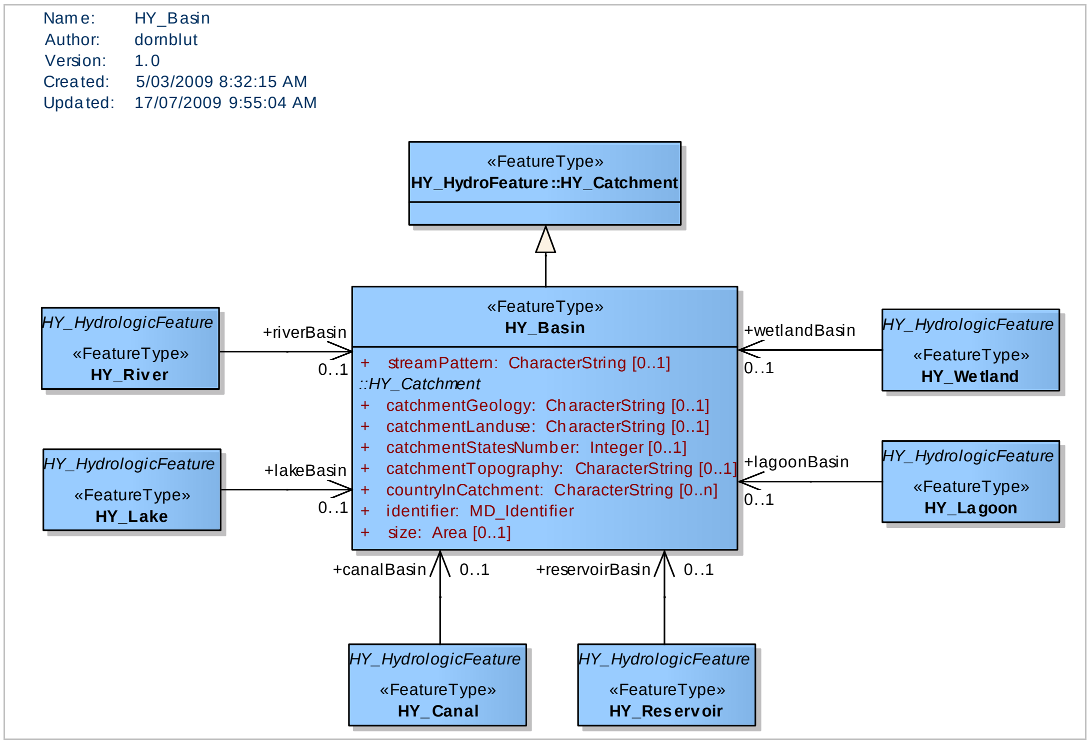


Figure 7 – Basins UML in the GRDC profile

By re-using the GRDC model, O&M descriptions of observations can be linked to the relevant spatial features. Figure 8 shows an example of how this may be done using an example of a flow observation at a site on a river. The elements in green represent components in the observation process from O&M; those in blue are instances of classes from the GRDC metadata profile.

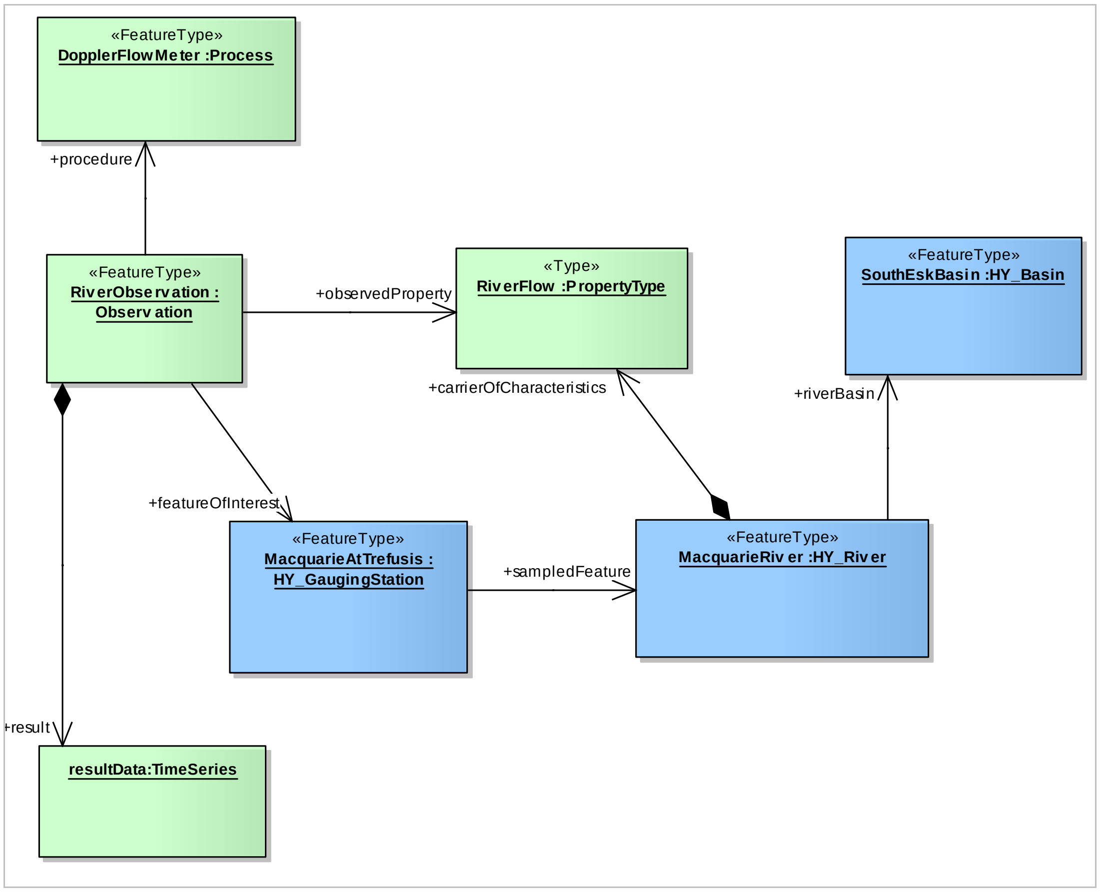


Figure 8 - Example river flow observation in UML

A literal reading of the example gives, "A doppler flow meter was used to make a flow measurement of the Macquarie river at the Macquarie at Trefusis station. The result of this observation was a time series. The Macquarie River is in the South Esk Basin."

The example does not show the properties of all the classes involved (such as observation times, results etc.) but these exist within each of the class definitions and would be described explicitly when encoding occurs.

The process is shown as a generic O&M process instance (being a doppler meter). The GRDC profile does define some specialisation of process types such as instrument, so these classes may also be used. The following section addresses procedure descriptions further.

11 Procedures

The description of the process used to generate an observation result can vary in complexity. Raw observational results generated from sensors undergo internal processing to convert signals into parameter estimations; time series are generated from chains of conversion processes to convert units or estimate related phenomenon; models link complex process chains to provide past and future estimates for phenomenon. This information is useful when interpreting data sets to gain further insight into the nature of the estimation.

The existing standards analysed provide some categorical grouping of the types of processing that occurred on a particular data set, but they generally don't provide a high level of metadata that could be used for further interpretation (i.e. to provide estimates of uncertainty in data).

Within the OGC Sensor Web Enablement suite of specifications the SensorML specification is a schema for describing in detail the processes that occur when sensors and instruments take in creating estimates of phenomenon. It is a very flexible schema, and there is need to investigate its use, a subset of it, or alternate specifications, for describing fully the procedure used to generate hydrologic observation results.

There is an activity within the W3C call the Semantic Sensor Network Incubator Group which is currently developing ontologies to define the capabilities of sensor and sensor networks [W3Cb]. This group is also looking at approaches for providing annotations within existing standards to link to well specified description of sensor capabilities.

The GRDC metadata profile provides some initial work in this area, with its specialisation of the O&M procedure and separating instruments and simulations (models etc.). Further testing of this model will be required to determine whether it is capable of matching current and future requirements.

11.1 Derived time series

The CUAHSI ODM allows for grouping and derived from associations to be created. This concept can be used to provide metadata describing how a particular time series may have been generated. A group of values may be associated with another group from which the series was derived. This association can then be linked to a method identifier which describes the process that was undertaken to generate the derived values. Essentially this is modelling two concepts: process chains and describing dependencies between data sets.

This grouping and dependency tracking is a common requirement in hydrological data, where there are many derived data products that have dependencies both to other data sets and to algorithms (defined as category 3 data in section 6). Maintaining the relationship between time series for when dealing with sharing of data requires a common (or at least transparent) identifier system; this allows series to be cross referenced and resolved appropriately. Further investigation should be done in capturing a common model for such requirements.

There is a difficulty managing the identity of data values when externally processing information in a program or model that does not fully support a data model. Operating the processing chain over a grouping, or series, and maintaining relationships between groups requires less information and allows for external processing operations.

12 Observed properties

Observed properties relates to the definition of the phenomenon that is being observed. Hydrological observed properties are generally phenomenon such as water level, river flow (or discharge), turbidity etc. Exchange formats may want to refer to well defined lists of possible observed properties or describe the nature of the observed property inline in an instance document. Both models should be supported.

There are differences in the definitions of what constitutes an observed property across existing hydrological standards. This appears to be mainly due to the way the underlying model has captured the various components of observations. This is an area that requires further harmonisation and will be within scope of defining a harmonized model.

WaterML1.1 uses the ODM concept of a variable, which links the observed property, units of measure and the temporal sampling regime being used (e.g. regular hourly intervals). It is also possible to specify the medium being sampled (e.g. surface water, air, ground water etc.). Additionally, it includes series information, like interpolation and data type, and meta-information about the variable such as general category and no data value. In the HIS discovery system [HIS2009], an ontology is used to link variables to concepts.

WDTF mainly uses the referencing of controlled terms approach to defining its property types. It does have the ability to provide an inline description that sets a unit against a particular observed phenomenon. WDTF also provides a mechanism for creating compound properties that links to phenomenon through a mapping. This has been used to allow for the definition of rating (conversion) tables (compound property defines the level to flow mapping). It is not currently used to create time series of compound phenomena where one may want to define a single time series structure that applies across a number of observed properties.

12.1 SWE Common

SWE Common provides mechanisms for defining property types (phenomenon) along with the facility to create compound and constrained properties. This model is shown in Figure 9.

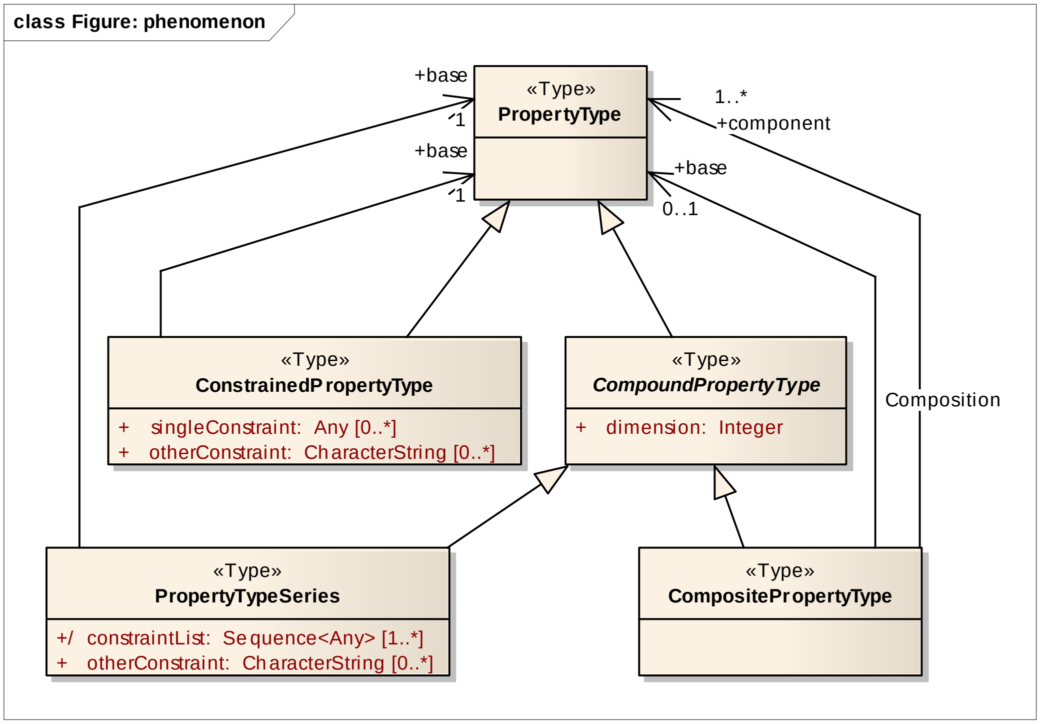


Figure 9 - SWE common phenomenon definition UML

An example use of a constrained phenomenon to limit temperature is given in the SensorML specification [BOT2007]:

```
<swe:ConstrainedPhenomenon gml:id="SurfaceWaterTemperature">
  <gml:name codeSpace="urn:ogc:tc:arch:doc-rp(05-010)">
    urn:ogc:def:property:OGC:SurfaceWaterTemperature
  </gml:name>
  <gml:name>Surface Water Temperature</gml:name>
  <swe:base xlink:href="#WaterTemperature"/>
  <swe:singleConstraint>
    <swe:TypedValue>
      <swe:property codeSpace="./">Depth</swe:property>
      <swe:value>
        <swe:Interval>
          <swe:lowerBound xsi:type="gml:MeasureType" uom="./units.xml#m">
            0.0
          </swe:lowerBound>
          <swe:upperBound xsi:type="gml:MeasureType" uom="./units.xml#m">
            1.5
          </swe:upperBound>
        </swe:Interval>
      </swe:value>
    </swe:TypedValue>
  </swe:singleConstraint>
</swe:ConstrainedPhenomenon>
```


12.2 IOOS approach

IOOS leverages the SWE Common property definitions along with GML dictionaries to allow phenomenon definitions within documents.

```
<om:observedProperty xlink:href="http://www.csc.noaa.gov/ioos/schema/IOOS-DIF/IOOS/0.6.1/dictionaries/phenomenaDictionary.xml#WaterLevel"/>
```

Where the definition of the type is as follows:

```
<gml:definitionMember>
    <swe:Phenomenon gml:id="WaterLevel">
      <gml:description>Level of the water.</gml:description>
      <gml:identifier codeSpace="urn:x-noaa:ioos:def:phenomenonNames">WaterLevel</gml:identifier>
    </swe:Phenomenon>
  </gml:definitionMember>
```

12.3 Potential future approach

When defining local property definitions the SWE Common approach is quite flexible. It allows property definitions to be used that are appropriate for the context of use. For example, a US agency serving information for the National Water Information System (NWIS) may encode their property definition as follows:

```
 <om:observedProperty>
    <swe:Phenomenon gml:id="NWIS_00060">
      <gml:description>Discharge, cubic feet per second</gml:description>
      <gml:name>00060</gml:name>
    </swe:Phenomenon>
  </om:observedProperty>
```

Or alternatively, the definition could be referenced via xlink using an identifier:

```
<om:observedProperty xlink:href="http://nwis.waterdata.usgs.gov/usa/nwis/pmcodes/00060"/>
```

It would be optimal if this identifier could be resolved to an equivalent definition of the property. This will depend somewhat on the governing body of the definitions, but having such definitions available allows for a more consistent use by data providers.

13 Linking to code lists and ontologies

The concept of linking out to definitions is to allow a clear reference to be made to a well governed definition of a concept. Often the concept that is being represented contains information that is important in correctly interpreting a data set, but it is not feasible to define the semantics of this along side the data. This case is commonly seen in the use of controlled vocabularies or code lists where an identifier is used for a well-defined concept within an organisation or its operating context. For example, an organisation may define a set of common data quality identifiers that categorise some qualitative information, such as the USGS qualification for instantaneous values:

Table 7 - USGS codes for instantaneous values

Code	Description
e	The value has been edited or estimated by USGS personnel
A	The value is affected by ice at the measurement site.
B	The value is affected by backwater at the measurement site.
R	The rating is undefined for this value
&	This value is affected by unspecified reasons.
K	The value is affected by instrument calibration drift.
X	The value is erroneous. It will not be used.
<	The value is known to be less than reported value
>	The value is known to be greater than reported value

The link from the identifier to the underlying concept is done through the code identifier (e.g. 'R'). On its own the identifier is fairly meaningless, but it may be contextualised by a link that provides the definition of the concept.

An organisation adopting an existing standard for data exchange will want to continue to use their own set of codes for particular concepts. Re-usable schemas must therefore be able to use a different set of definitions without version changes. A recommended approach to adapting a schema for use should be provided by the standard definition or at least by best practice documents.

True interoperability between information systems can occur only when it is possible to either translate exactly between codes from different organisations or a common set of codes are adopted. While this is the goal, it is realised that significant advances in interoperability can occur even when there is only a common format capable of housing those codes.

The Semantic Web [W3Cb] community is interested in making connections between how data relates to real world objects. An ontology [WIK2009] is one technique available for defining the nature of real world objects and their relationships. A number of information modelling groups are looking towards using ontologies, and linking to them, to allow data to be connected with its conceptual meaning. There are some existing approaches on how best to 'mark-up' data with appropriate connections to such ontological definitions.

The Marine Metadata Interoperability (MMI) project recommends usage [ALE2009] of URLs for defining links to terms within ontologies. If this target URL is resolvable then one could retrieve the definition of the term within a hierarchy of related definitions.

The suggested structure is as follows:
http://{hostdomain}/{ontologiesRoot}/{authority}/{version}/{resourceType}/{shortName}

Following such a definition, an example could be encoded as follows (referencing a CUAHSI ontology for surface hydrology to define stream discharge):

```
<om:observedProperty xlink:href="
https://svn.sdsc.edu/repo/WATER/CUAHSI/OntologyOwl/StarTree_Current/ontology/surfhydrosyn/
dischargeStream"/>
```

Current work within the W3C Semantic Sensor Networks incubator group [W3Cc] is investigating techniques for marking up data with links to its semantic meaning. They provide an example on a weather compound phenomenon [COM2009] linked to an ontology describing the components making up weather observables:

```
<om:observedProperty>
 <swe:CompositePhenomenon dimension="5" gml:id="WEATHER_OBSERVABLES">
 <gml:name>Weather Measurements</gml:name>
  <swe:base xlink:href="http://www.w3.org/2009/Incubator/ssn/ontologies/SensorOntology.owl"/>
   <swe:component xlink:href="http://www.w3.org/2009/Incubator/ssn/ontologies/SensorOntology.owl#AirTemperature"/>
   <swe:component xlink:href="http://www.w3.org/2009/Incubator/ssn/ontologies/SensorOntology.owl#DewPoint"/>
   <swe:component xlink:href="http://www.w3.org/2009/Incubator/ssn/ontologies/SensorOntology.owl#RelativeHumidity"/>
   <swe:component xlink:href="http://www.w3.org/2009/Incubator/ssn/ontologies/SensorOntology.owl#WindSpeed"/>
   <swe:component xlink:href="http://www.w3.org/2009/Incubator/ssn/ontologies/SensorOntology.owl#WindDirection"/>
</om:observedProperty>
```

Figure 10 - Example link to weather phenomena within an ontology

SWE common provides sufficient linking mechanisms (through use of xlink) to define links to ontologies using this approach. Dictionaries that are local to the schema can be defined, and these may be generated from catalogs that are used as the definitive source of the definitions. The actual approach used will depend upon the end user of a schema; descriptions of recommended approaches would be useful.

14 Grouping observations

The O&M model describes individual observations that capture the relationship between the observed property, the feature of interest, the procedure and the ultimate result. Grouping of observations is important when handling transmission of observation series, for example as responses to web services or other query interfaces. Often such groupings are called series (WaterML1.0), datasets (GRDC metadata profile) or collections.

O&M version 1.0 contains an observation collection definition that allows for grouping of multiple observation descriptions. However, this definition may not be included in subsequent versions. The basis for this decision is that the description of collections is not a part of the observation description, merely a convenience for transmission of observational sets. These types of structure are of significance when defining a service interface to the model.

CSML defines a dataset class, allowing for:

- Spatial extent summary of the grouping
- Local dictionary definitions of coordinate reference systems, phenomenon, units of measure
- Description of multiple features (and coverages).

Grouping in WDTF allows for description of

- Features
- Document metadata (versioning, data owner, document generation etc.)
- Local phenomena definitions
- Transactional information
- Specimens
- Conversions
- Observations.

Within WDTF, a grouping does not necessarily imply there is a relationship specific to the type of observational data contained. As it a transfer format, the observations that are contained within the group are merely determined by the approach a user has taken when exporting their data set. The observation data contained within a document explicitly define relationships through the use of identifiers. For example, these identifiers may relate sets of observations through common features of interest (i.e. spatially exist at the same 'site').

Grouping in WaterML1.0 is based around the interface used for discovering data. Observation groupings are called series (a unique combination of site, variable and time intervals). A location has group of series called a series catalog. Since locations can be shared, it is possible to have more than one series catalog, although this has been deprecated. Within the CUAHSI HIS, series are essential to the discovery process. A central metadata catalogue of series allows for the discovery of information across data providers. A series catalog is a group of series that is defined for a site. This is used for discovery purposes through the GetSiteInfo service call in WaterOneFlow. Time series are grouped in time series response documents which results from a GetValues service call.

Grouping structures of observations are often developed from the viewpoint of how people discover observational data sets. Within OGC the service model built around accessing O&M data is the Sensor Observation Service (SOS). This service has a concept of "offering" that allows for grouping of data sets along with interface calls for discovery and filtering of concepts.

Groups are important to the management of the information, and they often reflect how the information is collected and managed. For streaming data, United States data providers manage data streams as "site-variable-begin date-end date" which it calls "period of record". This differs from a water quality collection model which uses a project-site-methods-results grouping. This water quality model is reflected in the WQX standards used both by the EPA, and the USGS. Functionally equivalent to the series/period of record concept are 'availability records' for a particular location-variable. These are used by agencies, such as the National Climate Data Center, and the National Resources Conservation Service. SOS provides a mechanism that is called "offering" but since grouping outlined above can have millions of records it is not possible to use in SOS 1.0 to expose large numbers of "offerings".

A conceptual model for hydrological observations needs to provide a standardized mechanism for discovering time series groupings. Periods of records/offerings/series are presently basic practice when exposing hydrologic information. The CUAHSI HIS system has demonstrated that such groupings allow for discovery of information.

15 Summary of requirements for a core water observation model

This section provides a summary of requirements for a core model and provides guidance for future work on harmonising, and developing, future standards for exchange of water information. It is not a comprehensive list and further engagement by a wider community will help to identify areas that require extra work.

Results: The core properties that were identified in this report include: interpolation types, quality, value qualifiers, accuracy, processing metadata (linked to procedure definitions), textual comments, units, null values, temporal metadata. Adapting these into a specialised O&M result model and testing against a number of exchange requirements is underway. The focus of this report is on time series; other observation types do occur in hydrological observations, such as geometric observations of river cross sections – these will need to be addressed further.

Feature descriptions: mechanisms for linking to feature descriptions should be provided. Full descriptions may be supported by linking to existing information models. Certain usages patterns require different levels of metadata, unpacking these will be important future work. For example, an information model to support a flow forecast model has higher requirements on network descriptions than a rainfall reporting service.

Procedures: Current metadata is limited in existing standards. Linking to external descriptions should at least be possible. Further investigation of a common approach for specific hydrological processes is needed, addressing processes such as hydrologic models, complex conversions (rating curves, volume calculations etc.) and sensor descriptions. The procedural information is often closely related to the result types and associated metadata; understanding different viewpoints and needs of observation data will assist here. Procedure descriptions assist greatly in downstream interpretation of data sets; this is an area that is not currently well supported.

Groupings: Grouping observation sets is important for discovery purposes. Existing standards are built around end user needs of discovery – these should be analysed further.

Flexible code lists: ability to link to existing code lists for particular agencies is a common requirement (see section 13).

Observed properties/phenomenon: SWE common provides an initial approach to capturing the definitions of observed properties. There is a need to harmonise on the way relationships are drawn between observed phenomenon, results and procedures (which O&M provides guidance on) – existing standards group concepts together slightly differently and the core components need to be separated in a consistent manner.

15.1 Encoding types

The existing formats investigated are all based on XML encodings for data exchange. Current practice for data exchange of hydrological observations hinges largely on the exchange of CSV and Excel spreadsheets. These formats often have minimum metadata and often rely on apriori knowledge of opaqe1T.479431(a)3014794215(d0.956417(fd [(t)-2.53658(s)956417(a)

16.1 Identifying core requirements

Existing standards will already have a certain level of semantic alignment of concepts between them, as shown in Figure 11, due to the nature of the domain. Close alignment of concepts occurs in centre of the diagram; areas outside the core *may* require harmonisation – concepts should only be considered for harmonisation if they can be matched to a common requirement. If they can not, they may exist in a specific application of the core model. Familiar concepts such as units of measure and spatial locations could be expected to exist across the standards.

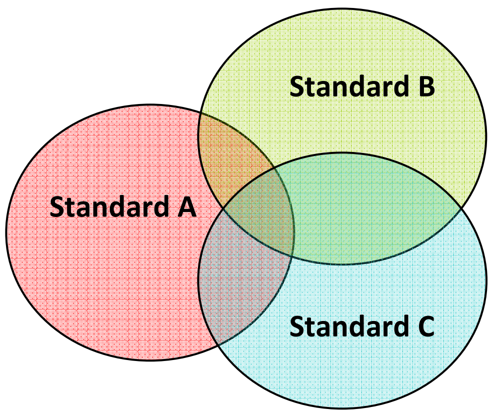


Figure 11 - Alignment of concepts

Analysing the existing sets of standards gives us a proxy for the requirements of exchange formats (as they have been driven by particular needs), but an analysis of further requirements is important.

Building exchange formats around such a core model provides a number of benefits:

- Clear definition of the semantics of each concept within the model and its relationships. This allows agreement to be set at a high level between people, organisations or systems by referring to the model rather than having to resolve semantic differences for each concept. An example from the investigated standards would be reconciling the differences between variable (WaterML), parameter (UK-EA-TS) and phenomenon (WDTF). This benefit is paramount when addressing interoperability of systems.
- Schemas, databases, documentation and code may be generated directly from the model. This allows easier management of versioning as a number of artefacts can be generated when changes occur.
- Sharing of tools that have been developed to support the model. Code need not be re-written for all the standard requirements when using exchange formats such as ingestion, encoding, validation etc.

- Tools such as web services may also be re-used to provide query and access interfaces to both people and other applications. The use of a core model here simplifies the adoption of such services for use in an organisation.
- The parts of an exchange format that are specific to an organisation can be governed and introduced separately to the model. This helps with managing complexity of a format and enhances extensibility.

Having such a core model is especially beneficial when dealing with large scale distribution of stakeholders and systems. This is the case with initiatives mentioned such as INSPIRE, AWRIS and CUAHSI. The issue with data integration in such distributed systems is that if each system is communicating with one another, the number of format translations required increases exponentially with the inclusion of new systems.

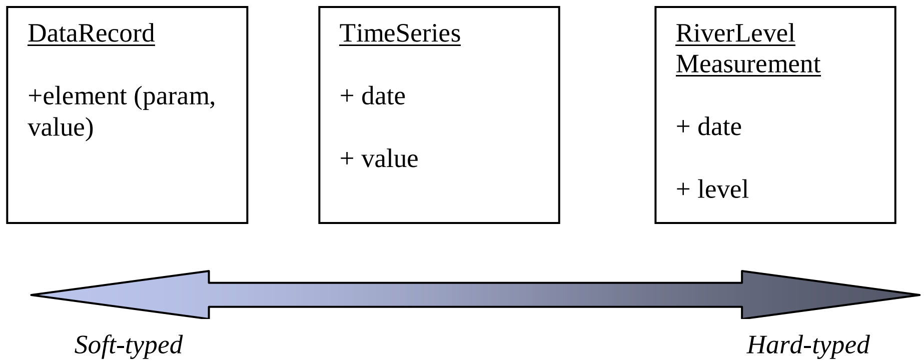
The Canonical Data Model pattern [HOH2003] describes the use of a common information model that all members of the system must subscribe to in order to communicate with other systems. This allows translation to occur only at one point (where the system is introduced), rather than for each communication channel between distributed parties. This seems like an obvious concept when designing systems, but the current state of data exchange is actually multi-channel when considering channels such as FTP, phone, email – all often used to exchange data. For example, a common occurrence is for a data holder to email a CSV data file to an interested party, who subsequently rings up and asks about the metadata (what coordinate system is used, what is phenomenon ID6854 etc.) in order to correctly interpret the file.

Developing a common model assumes that it is actually possible to correctly harmonise on the concepts within the domain. A separation of concerns can help here in packing problems into manageable parcels which capture a set of agreed upon concepts. This allows commitment to these definitions without pulling in a full descriptive model which they may not align to.

16.2 Soft-typing vs. hard-typing

The concept of soft-typing refers to the case where a schema does not make explicit structural definitions for what types should be allowed for particular classes or concepts. The types can essentially be defined at ‘run-time’. The opposite of this is hard typing which defines up front the types and their structure that should be allowed within a schema.

Adapting the explanation provided in the CSML definition [WOO2007] to a description of time series,



An example instance:

```
<DataRecord>
   <element param="time">01/01/2009T09:00:00</element>
   <element param="value">0.3</element>
   <element param="observedProperty">RiverLevel</element>
</DataRecord>

<TimeSeries>
   <date>01/01/2009T09:00:00</date>
   <value>0.3</value>
   <property>RiverLevel</property>
</TimeSeries>

<RiverLevelMeasurement>
   <date>01/01/2009T09:00:00</date>
   <level>0.3</level>
</RiverLevelMeasurement
```

Balancing between hard-typing and soft-typing in descriptions of concepts and types in a conceptual model is important. Soft typing allows flexibility but reduces the specificity of the model, which creates ambiguity, reduces interoperability and affects the validation process of encoded documents; hard-typing tightly defines concepts making semantics clear and validation using existing tools easier, but reduces the ability to extend definitions without revising the schema.
The general approach is that if a concept is core to the domain and can be harmonised to provide a common definition, then it is a candidate to be hard-typed. Concepts that are more specific to particular organisations or contexts should be made available through the use of soft-typed definitions.

16.3 Proposed methodology

The proposed methodology is to address the development of various aspects of a harmonised model in phases that relate to particular types of observations styles define in section 6. The initial phase will address in-situ style observational data sets (category 1 from section 6) in the following areas:

1. Time series structures (results);
2. General metadata for the procedure used in measurement;
3. Minimal metadata data for spatial features (descriptions of stations) and guidance on linking to external descriptions;
4. Techniques for linking to definitions of observed phenomenon.

The target for a proposed harmonised schema is the OGC. This implies a number of operating restrictions in developing a schema. Standards already defined within the OGC should be re-used where possible; relevant standards that should be considered in development include:

- Geography Markup Language (GML)
- Sensor Web Enablement (SWE):
    - i. Observations & Measurements (O&M);
    - ii. Sensor Markup Language (SensorML);
    - iii. SWE Common;
    - iv. Sensor Observation Service (SOS).

The initial phase of development will refine the general approach of using UML to generate XML Schema . This will allow candidate schemas to be quickly developed and deployed for testing purposes. The development cycle will be aimed at being as agile as possible in its ability to respond to comment and inputs from interested parties.

While the initial focus is on structural aspects leading to XML Schemas, it is recognised that including some code lists (such as interpolation types) in the standard may increase the usefulness of common tools. As the XML Schemas are standardised the candidates for common code lists will be examined as well as tools to support the use of local code lists.

The OGC has an interoperability program [OGC2009] that is a light-weight program to test and demonstrate the use of OGC schemas or candidate schemas. These programs involve deployments of services to solve real world problems, often grounded in an organisations particular need for data exchange. They provide a testing ground for schemas and feedback from such projects are extremely beneficial for the development of such standards.

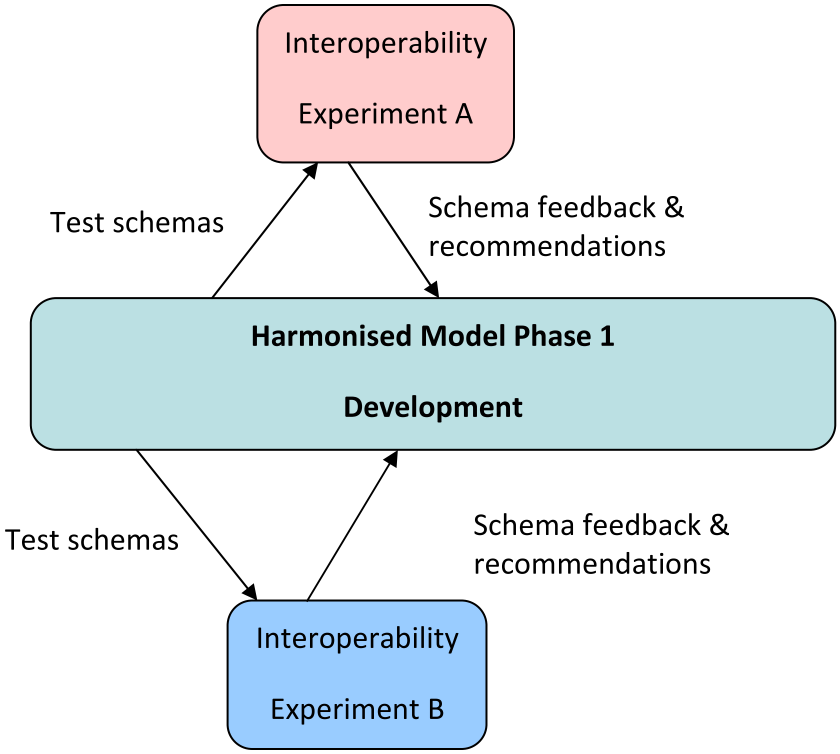


Figure 12 - Interoperability experiment interactions

17 Data Exchange vs. Archival

There are two diverse aspects to sharing hydrological data: sending out minimal information, enough to allow use; or exchanging enough for archival storage.

There are formats that are more archival in nature, e.g. GRDC Hydrologic Datasets metadata profile is based on the ISO metadata standard. Presently, many datasets are available in text formats which enable use. They require the user to infer the details of the information which is often documented outside of the downloaded file.

CUAHSI noted that presently, scientists tend to store retrieved information on disk after retrieval from a data source, even when the source was a method from a web service. Because of this observation, it was determined that a set of results must be returned with data for disconnected use (site information, variable information, and data value attributes).

Future work will need to determine at which point along the spectrum a schema will exist or be developed. While outside the scope of the conceptual model, it will need to determine how ancillary information, such as rating curves and the details of a laboratory analysis, will be made available.

18 Adopting a common model

In order for users to adopt a core conceptual model there would be a process of ‘contextualising’ the model to satisfy an end–user’s need. The process will be determined somewhat by their requirements for data exchange (file based exchange format, web service responses, database schemas etc.) but there are common requirements that could be foreseen. It would be of benefit for best practices or specifications to be developed for various hydrological data exchange patterns. Such work could be envisaged for the Hydrology Domain Working Group (HDWG).

The process may include:

1. **Generating a schema from the core model that suits a deployment platform**. Current tools allows for the generation of full GML schemas from UML models. This will be the initial approach for developing first round schemas from a model. In the future there may be need to investigate methods for generating other encodings, such as a simple features version of GML.
2. **Defining or importing other specific schema requirements**. A user may have requirements that are not a part of the core model such as transactional information, describing ground water structures or other more explicit feature descriptions. It may be possible to import these from other existing standards or they may need to be developed as needed.
3. **Linking a schema to vocabulary definitions**. Some core vocabularies (such as observed properties, units of measure etc.) for the hydrology domain may be established through community groups such as the HDWG. These could be used directly if they have sufficient coverage. Others that are more context-dependent

Annex A

```
      <unitName>milligrams per liter</unitName>
      <unitType>Concentration</unitType>
      <unitAbbreviation>mg/L</unitAbbreviation>
      <unitCode>199</unitCode>
    </unit>
    <noDataValue>-9999</noDataValue>
    <timeScale>
      <unit>
        <unitName>second</unitName>
        <unitType>Time</unitType>
        <unitAbbreviation>s</unitAbbreviation>
        <unitCode>100</unitCode>
      </unit>
      <timeSupport>0</timeSupport>
    </timeScale>
    <speciation>Not Applicable</speciation>
  </variable>
  <values>
    <value censorCode="nc" dateTime="2007-11-07T13:00:00" timeOffset="-07:00" dateTimeUTC="2007-11-07T20:00:00" methodCode="25" sourceCode="3" labSampleCode="9188" qualityControlLevelCode="2">10.5</value>
    <value censorCode="nc" dateTime="2007-11-13T12:30:00" timeOffset="-07:00" dateTimeUTC="2007-11-13T19:30:00" methodCode="25" sourceCode="3" labSampleCode="9398" qualityControlLevelCode="2">2.5</value>
    <value censorCode="nc" dateTime="2007-11-21T14:00:00" timeOffset="-07:00" dateTimeUTC="2007-11-21T21:00:00" methodCode="25" sourceCode="3" labSampleCode="9509" qualityControlLevelCode="2">7.2</value>
    <value censorCode="nc" dateTime="2007-12-05T11:00:00" timeOffset="-07:00" dateTimeUTC="2007-12-05T18:00:00" methodCode="25" sourceCode="3" labSampleCode="G120507-WELL-TSS" qualityControlLevelCode="2">2.5</value>
    <value censorCode="nc" dateTime="2007-12-20T14:05:00" timeOffset="-07:00" dateTimeUTC="2007-12-20T21:05:00" methodCode="25" sourceCode="3" labSampleCode="G122007-WELL-TSS" qualityControlLevelCode="2">2.5</value>
    <qualityControlLevel qualityControlLevelID="2">
      <qualityControlLevelCode>2</qualityControlLevelCode>
      <definition>Derived products</definition>
      <explanation>Derived products that require scientific and technical interpretation and may include multiple-sensor data. An example is basin average precipitation derived from rain gages using an interpolation procedure.</explanation>
    </qualityControlLevel>
    <method methodID="25">
      <methodCode>25</methodCode>
      <methodDescription>Water chemistry grab sample collected by technicians in the field.</methodDescription>
    </method>
    <source sourceID="3">
      <sourceCode>3</sourceCode>
      <organization>Utah State University Utah Water Research Laboratory</organization>
      <sourceDescription>Water chemistry monitoring data collected by Utah State University as part of a National Science Foundation funded test bed project.</sourceDescription>
      <contactInformation>
        <contactName>Amber Spackman</contactName>
```

```xml
        <typeOfContact>main</typeOfContact>
        <email>amber.s@aggiemail.usu.edu</email>
        <phone>1-435-797-0045</phone>
        <address xsi:type="xsd:string">8200 Old Main Hill
,Logan, Utah 84322-8200</address>
      </contactInformation>
      <sourceLink>http://water.usu.edu/littlebearriver</sourceLink>
      <citation>Water chemistry monitoring data collected by Jeff Horsburgh, David Stevens, David
Tarboton, Nancy Mesner, Amber Spackman, and Sandra Gurrero at Utah State University as part of a
National Science Foundation funded WATERS Network Test Bed project.</citation>
    </source>
    <sample sampleID="26">
      <labSampleCode>9188</labSampleCode>
      <sampleType>Grab</sampleType>
      <labMethod>
        <labCode>9188</labCode>
        <labName>USU Analytical Laboratory</labName>
        <labOrganization>Utah State University</labOrganization>
        <labMethodName>EPA 340.2</labMethodName>
      </labMethod>
    </sample>
    <sample sampleID="32">
      <labSampleCode>9398</labSampleCode>
      <sampleType>Grab</sampleType>
      <labMethod>
        <labCode>9398</labCode>
        <labName>USU Analytical Laboratory</labName>
        <labOrganization>Utah State University</labOrganization>
        <labMethodName>EPA 340.2</labMethodName>
      </labMethod>
    </sample>
    <sample sampleID="38">
      <labSampleCode>9509</labSampleCode>
      <sampleType>Grab</sampleType>
      <labMethod>
        <labCode>9509</labCode>
        <labName>USU Analytical Laboratory</labName>
```



```
      <labSampleCode>G122007-WELL-TSS</labSampleCode>
      <sampleType>Grab</sampleType>
      <labMethod>
       <labCode>G122007-WELL-TSS</labCode>
       <labName>USU Analytical Laboratory</labName>
       <labOrganization>Utah State University</labOrganization>
       <labMethodName>EPA 340.2</labMethodName>
      </labMethod>
     </sample>
     <censorCode>
      <censorCode>nc</censorCode>
      <censorCodeDescription>not censored</censorCodeDescription>
     </censorCode>
    </values>
  </timeSeries></timeSeriesResponse>
```

19.2 Water Data Transfer Format

```
<?xml version="1.0"?>
<wdtf:HydroCollection
 xmlns:sa="http://www.opengis.net/sampling/1.0/sf1"
 xmlns:om="http://www.opengis.net/om/1.0/sf1"
 xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
 xmlns:xlink="http://www.w3.org/1999/xlink"
 xmlns:gml="http://www.opengis.net/gml"
 xmlns:wdtf="http://www.bom.gov.au/std/water/xml/wdtf/1.0"
 xmlns:ahgf="http://www.bom.gov.au/std/water/xml/ahgf/0.2"
 xsi:schemaLocation="http://www.opengis.net/sampling/1.0/sf1 ../sampling/sampling.xsd
 http://www.bom.gov.au/std/water/xml/wdtf/1.0 ../wdtf/water.xsd
 http://www.bom.gov.au/std/water/xml/ahgf/0.2 ../ahgf/waterFeatures.xsd"
 gml:id="HC-t1">
 <!--  Change log     -->
 <!--  snip -->
 <gml:description> This document encodes timeseries. Documentation snipped.
 </gml:description>

 <gml:name
codeSpace="http://www.bom.gov.au/std/water/xml/wio0.2/feature/HydroCollection/w00001/">HC-
t1</gml:name>

 <wdtf:metadata>
  <wdtf:DocumentInfo>
   <!-- specify the version of the data package -->
   <wdtf:version>wdtf-package-v1.0</wdtf:version>
   <wdtf:dataOwner
codeSpace="http://www.bom.gov.au/std/water/xml/wio0.2/party/person/bom/">w00001</wdtf:dataO
wner>
   <wdtf:dataProvider
codeSpace="http://www.bom.gov.au/std/water/xml/wio0.2/party/person/bom/">w00001</wdtf:dataPr
ovider>
   <!-- All dates and time should include a time zone or terminate in Z for UTC -->
   <wdtf:generationDate>2008-07-11T00:00:00+10:00</wdtf:generationDate>
```

```xml
    <wdtf:generationSystem>AWRIPS</wdtf:generationSystem>
   </wdtf:DocumentInfo>
  </wdtf:metadata>

  <!-- transaction Members indicate the transactions associated with the document -->
  <wdtf:transactionMember>
   <!-- A synchronizationTransaction is used to replace a block in the time period -->
   <wdtf:SynchronizationTransaction
    gml:id="synch1">
    <!-- a nil period would indicate that the new data replaces the entire existing data set -->
    <wdtf:period>
     <om:TimePeriod>
      <om:begin>2001-07-31T20:12:01</om:begin>
      <om:end>2001-08-02T20:10:01</om:end>
     </om:TimePeriod>
    </wdtf:period>
   </wdtf:SynchronizationTransaction>
  </wdtf:transactionMember>

  <!-- defining data time series -->
  <!-- one Time series observation per observation  member but any number of observation members -->
  <wdtf:observationMember>
   <wdtf:TimeSeriesObservation
    gml:id="TS1">
    <!--  comment snip -->
    <gml:description>Telemetry data, and other unstructured information</gml:description>
    <gml:name
codeSpace="http://www.bom.gov.au/std/water/xml/wio0.2/feature/TimeSeriesObservation/w00001/41
0729/1/level/validated/">1</gml:name>

    <!-- resultTime indicates the time at which this time series was last revised (validation, annotation, etc)
as opposed
     to when the data was collected -->
    <om:resultTime>2008-07-10T10:30:00</om:resultTime>

    <!-- here a procedure unique across all sites for w0001 or a generic procedure type is used so it is not
     qualified by either the sampling point or group ids -->
    <om:procedure
     xlink:href="http://www.bom.gov.au/std/water/xml/wio0.2/procedure/Sensor/w00001/gaugeABC"/>
    <om:observedProperty
     xlink:href="http://www.bom.gov.au/std/water/xml/wio0.2/property//bom/WaterCourseLevel_m"/>
    <!-- mandatory link back to the Sampling point or location -->
    <om:featureOfInterest

xlink:href="http://www.bom.gov.au/std/water/xml/wio0.2/feature/SamplingPoint/w00001/410729/1"/>
    <!-- overall quality  -->
    <om:resultQuality>quality-A</om:resultQuality>
    <!-- optional link back to the sampling Group or site -->
    <wdtf:relatedSamplingFeature

xlink:href="http://www.bom.gov.au/std/water/xml/wio0.2/feature/SamplingGroup/w00001/410729"/>
    <wdtf:metadata>
```



```
      <wdtf:TimeSeriesObservationMetadata>
       <wdtf:relatedTransaction xlink:href="#synch1"/>
       <!-- the regulation property  these measurements relate to -->
       <wdtf:regulationProperty>Reg200806.s3.1a</wdtf:regulationProperty>
       <wdtf:securityConstraints>CommerciallySensitive</wdtf:securityConstraints>
       <!-- the next report is expected in a day -->
       <wdtf:reportingFrequency>P1D</wdtf:reportingFrequency>
       <wdtf:status>validated</wdtf:status>
      <wdtf:dataum>urn:ogc:def:datum:bom::GaugeDatum</wdtf:datum>
      </wdtf:TimeSeriesObservationMetadata>
     </wdtf:metadata>

     <wdtf:result>
      <wdtf:TimeSeries>
       <wdtf:defaultInterpolationType>InstVal</wdtf:defaultInterpolationType>
       <wdtf:defaultUnitsOfMeasure>m</wdtf:defaultUnitsOfMeasure>
       <wdtf:defaultQuality>quality-A</wdtf:defaultQuality>
       <wdtf:timeValuePair
        time="2001-07-31T20:12:01+10:00">1.25</wdtf:timeValuePair>
       <!-- This time point is missing -->
       <wdtf:timeValuePair
        time="2001-08-01T20:15:01+10:00"
        comment="text"
        interpolationType="InstVal"
        xsi:nil="true"/>
       <wdtf:timeValuePair
        time="2001-08-02T20:10:01+10:00"
        comment="example comment"
        quality="quality-B">1.28</wdtf:timeValuePair>
      </wdtf:TimeSeries>
     </wdtf:result>
    </wdtf:TimeSeriesObservation>
  </wdtf:observationMember>
</wdtf:HydroCollection>
```

19.3 UK Environmental Agency Time Series Exchange format

```
<?xml version="1.0" encoding="UTF-8"?>
<EATimeSeriesDataExchangeFormat xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
   xsi:schemaLocation="http://www.environment-
agency.gov.uk/XMLSchemas/EATimeSeriesDataExchangeFormat
EATimeSeriesDataExchangeFormat.1.2.xsd"
   xmlns="http://www.environment-agency.gov.uk/XMLSchemas/EATimeSeriesDataExchangeFormat"
xmlns:md="http://www.environment-agency.gov.uk/XMLSchemas/EAMetadataFormat">
   <md:Publisher>Environment Agency</md:Publisher>
   <md:Source>Plain English Document</md:Source>
   <md:Description>Mixed data file</md:Description>
   <md:Date>2003-06-20</md:Date>
   <md:Time>15:30:15</md:Time>
```

```xml
  <Station region="Thames" stationReference="2200" stationName="RIVER THAMES AT READING"
    ngr="SU71807406">
    <!-- Four days of daily mean flows -->
    <SetofValues parameter="Flow" dataType="Mean" period="Day" characteristic="Derived"
      units="m3/s" startDate="2003-04-20" endDate="2003-04-23" dayOrigin="09:00:00">
      <Value date="2003-04-20" flag1="1" flag2="1" percentFlag2="100">15.63</Value>
      <Value date="2003-04-21" flag1="2" flag2="1" percentFlag2="92.5">16.21</Value>
      <Value date="2003-04-22" flag1="1" flag2="1" percentFlag2="87" flag3="2"
        percentFlag3="5.5">16</Value>
      <Value date="2003-04-23" flag1="2" flag2="1" percentFlag2="85.2" flag3="2"
        percentFlag3="14.8">17.36</Value>
      <Comment startDate="2003-04-22">This daily mean flow was derived from an incomplete set
        of good and suspect data but has been validated and found to be of good overall
        quality</Comment>
      <Comment startDate="2003-04-21" endDate="2003-04-23">This demonstrates that you can have
        nested comments</Comment>
    </SetofValues>
    <!-- 1 and a half hours of recorded levels (e.g. from telemetry) -->
    <SetofValues parameter="Water Level" qualifier="Stage" dataType="Instantaneous"
      period="15 min" characteristic="Measured" productRef="H12" units="m"
      startDate="2003-04-20" startTime="12:00:00" endDate="2003-04-20" endTime="13:30:00"
      dayOrigin="09:00:00" valuesPerDay="96">
      <Value date="2003-04-20" time="12:00:00">3.125</Value>
      <Value date="2003-04-20" time="12:15:00">3.126</Value>
      <Value date="2003-04-20" time="12:30:00">3.125</Value>
      <Value date="2003-04-20" time="12:45:00">3.127</Value>
      <Value date="2003-04-20" time="13:00:00" flag1="25">8.568</Value>
      <Value date="2003-04-20" time="13:15:00">3.127</Value>
      <Value date="2003-04-20" time="13:30:00">3.126</Value>
    </SetofValues>
  </Station>
  <Station region="Thames" stationReference="265922" stationName="CAVERSHAM LOCK"
ngr="SU72067403">
    <!-- 1 monthly rainfall total -->
    <SetofValues parameter="Rainfall" qualifier="Storage Raingauge" dataType="Total"
      period="Month" characteristic="Measured" units="mm" startDate="2003-04-01"
      endDate="2003-04-01" dayOrigin="09:00:00">
      <Value date="2003-04-01" flag1="4">36.5</Value>
    </SetofValues>
  </Station>
</EATimeSeriesDataExchangeFormat>
```


19.4 XHydro


```xml
<?xml version="1.0" encoding="UTF-8"?>
<tsel xmlns="http://xhydro.org/minimal/2007/06"
 xmlns:d="http://www.disy.net/device"
 xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
 xsi:schemaLocation="http://xhydro.org/minimal/2007/06
```

```
http://www.xhydro.org/download/schemas/v200706/schemas/XHydro.xsd">
 <ext />
 <tse>
  <xids>
   <xid>
    <ext />
    <xk>disy</xk>
    <xv>test</xv>
   </xid>
  </xids>
  <org>
   <on>disy</on>
   <od>A company.</od>
  </org>
  <iso>
   <dst>P1D</dst>
   <sts>
    <tsq>
     <tsmi>1.5E-6</tsmi>
    </tsq>
    <tsv>2001-12-31T12:00:00</tsv>
   </sts>
  </iso>
  <pmdl>
   <pmd>
    <xp>W</xp>
    <c>This is a non-standard category code/remark.</c>
    <xu>m</xu>
    <tl>
     <ldn>
      <ln>Europe/Germany/Karlsruhe</ln>
      <ld>Karlsruhe, a german city.</ld>
     </ldn>
    </tl>
    <dt>
     <xdtc>aggMean</xdtc>
     <ag>
      <it>P1D</it>
      <ot>P15M</ot>
      <f>P1M</f>
      <xtsp>begin</xtsp>
      <l>1.5</l>
     </ag>
    </dt>
    <pd>
     <ext>
      <d:serial>ABCDEFG</d:serial>
     </ext>
     <dn>dd</dn>
     <dd>A dummy disy device.</dd>
     <dl>
      <ldn>
```



```xml
    <ln>Europe/Germany/Karlsruhe</ln>
    <ld>Karlsruhe, a german city.</ld>
   </ldn>
  </dl>
 </pd>
 <vq>
  <vmi>5E-3</vmi>
  <vqr>
   It is a quite imprecise device, isn't it? This quality remark
   demonstrates that free-text remarks are possible, too.
  </vqr>
 </vq>
</pmd>
</pmdl>
<tsmd>
 <tsd>
  <dn>ddts</dn>
  <dd>A dummy dis device to measure time.</dd>
 </tsd>
 <tsq>
  <tsmi>1.5E-6</tsmi>
  <tsqr codeList="disy1" codeListAgency="disy"
   codeListVersion="1.0">
   ownCode
  </tsqr>
 </tsq>
</tsmd>
<d>
 <tde>
  <!-- No timestamp is given because isochron -->
  <vls>
   <v>
    <vq>
     <vmi>6E-4</vmi>
     <xvqr>affected</xvqr>
    </vq>
    <vl>
     <pt>
      <xrs>32632</xrs>
      <px>5.0</px>
      <py>6.0</py>
     </pt>
    </vl>
    <vf>4.5</vf>
   </v>
   <v>
    <vf>4.6</vf>
   </v>
   <v>
    <vq>
     <xvqr>missing</xvqr>
    </vq>
```

```
      <va xsi:nil="true" />
     </v>
    </vls>
   </tde>
  </d>
 </tse>
</tsel>
```

Bibliography

[AND2009] AndroMDA, Open Source MDA toolkit - http://andromda.org.

[ALE2009] Alexander, P. 2009. "Citing Terms." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org/guides/vocabs/ont/existing/citing. Accessed: 08/18/2009

[AUSWA2007] Australian Government, Water Act 2007, accessible from http://www.comlaw.gov.au/comlaw/management.nsf/lookupindexpagesbyid/IP200733064?OpenDocument

[BAC2007] Bacharach, S (2007) New Implementations of OGC Sensor Web Enablement Standards Sensors Magazine Dec. 2007 http://www.sensorsmag.com/sensors/article/articleDetail.jsp?id=480557

[BLO1996] Blöschl, G., (1996), Scale and Scaling in Hydrology, Habilitationsschrift, Weiner Mitteilungen Wasser Abwasser Gewasser, Wien, 346 p.

[BLO1995] Blöschl, G. and M. Sivapalan, (1995), "Scale Issues in Hydrological Modelling: A Review," Hydrological Processes, 9(1995): 251-290.

[BOM2009a] Australian Water Data Transfer Format, http://www.bom.gov.au/water/WDTF_technical_information.shtml

[BOM2008b] Australian Water Resources Information System, Geofabric concept http://www.bom.gov.au/water/publications/awris.pdf

[BOM2008c] Australian Water Regulations, http://www.bom.gov.au/water/regulations/regulations.php

[BOT2006] Botts, M, Robin, A, Davidson, J & Simonis, I (2006), Sensor Web Enablement Architecture Document, Version 1.0, OpenGIS® Discussion Paper 06-021r1, 4 Mar 2006, Available from http://www.opengeospatial.org/standards/dp

[BOT2007] OpenGIS® Sensor Model Language (SensorML) Implementation Specification, OGC document 07-000

[COM2009] Compton, M et. al. A Survey of the Semantic Specification of Sensors, http://lists.w3.org/Archives/Public/public-xg-ssn/2009Aug/att-0037/SSN-XG_StateOfArt.pdf

[COX2008] Cox S, 2008, GML Encoding of Discrete Coverages (interleaved pattern), OGC document 06-188r2.

[COX2007a] Cox S & Brodaric, B 2007 Water Resources Information Model Workshop Canberra, 25-27 September, 2007 http://wron.net.au/documents/WaterML_workshop_report.pdf

[COX2007b] Cox, S (ed.) 2007a Observations and Measurements – Part 1 - Observation schema Version 1.0 OGC document 07-022r1 8 Dec. 2007 Accessible from http://www.opengeospatial.org/standards/o%2526m

[COX2007c] Cox, S (ed.) 2007b Observations and Measurements – Part 2 - Sampling Features Version 1.0 OGC document 06-188r1 8 Dec. 2007 Accessible from http://www.opengeospatial.org/standards/o%2526m

[DOR2009] Dornblut, I. Report 39 - Hydrologic Information – Metadata). Version 1, June 2009. Accessible from http://www.bafg.de/nn_317460/GRDC/EN/02__Services/04__Report__Series/39__metadata.html?__nnn=true

[ENE2009] State Energy Conservation Office,

http://www.energyeducation.tx.gov/renewables

[ESA2006] Environmental Sampling, Analysisand Results Data Standards. http://www.exchangenetwork.net/standards/ESAR_Overview_01_06_2006_Final.pdf

[EU2007] Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE).

[FUL2009] FullMoon tool for standards-based, model-driven development of information models, http://projects.arcs.org.au/trac/fullmoon/

[GAL2009] Geo-interface for Atmospheric, Land, Earth and Ocean netCDF project (GALEON). http://www.opengeospatial.org/projects/initiatives/galeonie

[GCO2000] Cihlar, J et al. WMO/TD – 1047, Establishment of a Global Hydrological Observation Network for Climate, 2000

[GEO2005] Global Earth Observation System of Systems (GEOSS) 10-year implementation plan, available at http://earthobservations.org/docs/10-Year%20Implementation%20Plan.pdf.

[HIS2009] HydroSeek data discovery tool, http://his.cuahsi.org/hydroseek.html.

[HOH2003] Hohpe, G, Woolf B, Enterprise Integration Patterns: Designing, Building, and Deploying Messaging Solutions. Addison-Wesley Professional (October 20, 2003).

[HOL2009] HollowWorld UML tool for assisting development of GML application schemas, https://www.seegrid.csiro.au/twiki/bin/view/AppSchemas/HollowWorld

[INS2008] INSPIRE Thematic Working Group Hydrography, D2.8.I.8 INSPIRE Data Specification on *Hydrography* – Draft Guidelines, 19th December, 2008.

[INS2007] INSPIRE Drafting Team "Data Specifications", deliverable D2.6: Methodology for the development of data specifications, 23rd August 2007.

[IOS2009] Integrated Ocean Observing System, National Oceanic and Atmospheric Administration (NOAA), http://ioos.gov/

[ISO19125-1] ISO/TC-211 2004, 19125-1 Simple feature access -- Part 1: Common architecture, International Organization for Standardization.

[KEN2009]

http://www.water.ky.gov/dw/consumer/dwed/lessons.htm

[LEF2008] Lefort, L 2008, Towards defining a water data exchange format: State of the art review on water data standards, CSIRO ICT Centre Technical Report 08/111.

[LEM2008] Hydrology Domain Working Group Charter, OGC document 08-095r5, 26th April 2008 updated 2 September 2009.

[MAI2002] Maidment, D, 2002. Arc Hydro: GIS for Water Resources. ESRI Press.

[MSEF] Microsoft Entity Framework, http://msdn.microsoft.com/en-us/library/aa697427(VS.80).aspx

[OGC2009] Open Geospatial Consortium Interoperability Program, http://www.opengeospatial.org/ogc/programs/ip

[OMGa] Model Driven Architecture, Object Management Group, http://www.omg.org/mda

[OMGb] Unified Modeling Language, Object Management Group, http://www.uml.org/

[OOM2009] Hibernate OOMEGA, Model-driven software engineering - http://www.oomega.net/

[TAR2008] Tarboton DG, Horsburgh JS, Maidment DR. CUAHSI Community Observations Data Model (ODM) Design Specifications. Version 1.0 May 2008. Accessible from http://his.cuahsi.org/documents/ODM1.1DesignSpecifications.pdf

[UNI2009] network Common Data Form (netCDF), Unidata: http://www.unidata.ucar.edu/software/netcdf/

[W3Ca] World Wide Web Consortium (W3C) XML Schema, http://www.w3.org/XML/Schema

[W3Cb] World Wide Web Consortium (W3C), Semantic Web concept http://www.w3.org/2001/sw/

[W3Cc] World Wide Web Consortium (W3C), Semantic Sensor Networks incubator group, http://www.w3.org/2005/Incubator/ssn/charter

[WCS2007] OpenGIS Web Coverage Service (WCS) Implementation Standard, OGC document 07-067r5.

[WAL2009] Walker G, Taylor P A, Cox S, Sheahan P. Water Data Transfer Format (IWDTF): Guiding principles, technical challenges and the future. MODSIM 2009.

[WIK2009] Ontology definition (information science), http://en.wikipedia.org/wiki/Ontology_(information_science)

[WOO2007] Woolf, A & Lowe, D 2007, CSML User's Manual, v2.doc Accessible from http://ndg.nerc.ac.uk/csml/

[WMO1994] WMO, WMO-168: Guide to Hydrological Practices, Fifth edition, 1994.

[WMO2006] WMO, WMO-49: Technical Regulations, Hydrology volume III, 2006.

[VAL2009b] Valentine, D & Zaslavsky, I WaterML1.1 Draft Specification, Part 2. 5 June 2009. Accessible from http://his.cuahsi.org/documents/WaterML_1.1_part2_v1.docx [2] IEC 60027 (all parts), *Letter symbols to be used in electrical technology*.

