

WaterML2.0: development of an open standard for hydrological time-series data exchange

P. Taylor, S. Cox, G. Walker, D. Valentine and P. Sheahan

ABSTRACT

The increasing global demand on freshwater is resulting in nations improving their terrestrial water monitoring and reporting systems to better understand the availability, and quality, of this valuable resource. A barrier to this is the inability for stakeholders to share information relating to water observations data: traditional hydrological information systems have relied on internal custom data formats to exchange data, leading to issues in data integration and exchange. Organisations are looking to information standards to assist in data exchange, integration and interpretation to lower costs in use, and re-use, of monitoring data. The WaterML2.0 Standards Working Group (SWG), working within the Open Geospatial Consortium (OGC) and in cooperation with the joint OGC-World Meteorological Organisation (WMO) Hydrology Domain Working Group (HDWG), has developed an open standard for the exchange of water observation data. The focus of the standard is time-series data, commonly used for hydrological applications such as flood forecasting, environmental reporting and hydrological infrastructure, where a lack of standards inhibits efficient re-use and automation. This paper describes the development methodology and principles of WaterML2.0, key parts of its information model, implementation scenarios, evaluation and future work. WaterML2.0 was adopted by the OGC as an official standard in September 2012.

Key words | data standards, information modelling, open geospatial consortium, time-series data

P. Taylor (corresponding author)
CSIRO, Hobart, TAS,
Australia
E-mail: peter.taylor@csiro.au

S. Cox
CSIRO, Highett, VIC,
Australia

G. Walker
CSIRO, Canberra, ACT,
Australia

D. Valentine
San Diego Supercomputer Center,
University of California,
San Diego,
USA

P. Sheahan
Australian Bureau of Meteorology,
Canberra, ACT,
Australia

INTRODUCTION

Increasing pressure on the natural environment is resulting in nations working on improving environmental monitoring and management, with a view to long-term sustainability. Monitoring of fresh water resources traditionally consists of regional monitoring programmes, designed for specific purposes: flood warnings, providing water management guidelines (applying restrictions, setting allocations, etc.), drought management, etc. Nations are now moving towards more holistic monitoring systems in order to provide multi-level reporting on water availability.

In addition to national reporting, there is an increasing need to exchange data across domains and into global-level reporting frameworks, such as Global Earth Observation System of Systems (GEOSS), the Global Climate Observing System (GCOS) and the World Meteorological Organization's (WMO's) Water Information System.

Providing data to national and global information systems is hindered by the use of different monitoring techniques and underlying information systems. Providing consistent information from multiple reporting parties involves understanding the way in which data are collected, stored and managed; no two systems will be the same. This can be alleviated through the use of standards, both in the monitoring technique, and how data are represented and exchanged.

There is, however, a lack of standards for exchanging hydrological data at national and global levels. A report (GCOS 2005) from the GCOS highlighted that there are no international standards for hydrology data, and '*Different institutions responsible for sectors like water, energy or agriculture, often keep their own data records without exchanging them at national levels. Common metadata*

standards are, even on national scales, the exception rather than the rule.' An IBM report (IBM 2009) on 'Water Management Pains' in the USA recommended to '*...implement open standards to improve interoperability across internal systems and external systems*'.

As an initial step towards development of an international standard for hydrological observational data, representatives from the Open Geospatial Consortium's (OGC) Hydrology Domain Working Group (HDWG) (Lemon 2011) identified, compared and contrasted a number of existing national exchange formats (Taylor *et al.* 2011). The report also proposed a methodology for developing a harmonised standard to address the common requirements identified within existing formats. Subsequently, a standards working group (SWG) to develop WaterML2.0 was formed in the OGC, with representatives from organisations around the world. Together the SWG and HDWG are aiming at building an international community of practice to foster data sharing principles and associated tools for sharing hydrological data. This paper gives a detailed account of the sustained research and development effort (Taylor *et al.* 2010b; Walker & Taylor 2011; Taylor *et al.* 2012; Valentine *et al.* 2012) that led to adoption of WaterML2.0 by the OGC in September 2012.

This paper describes part 1 of the WaterML2.0 information model, which focuses on time-series. The paper covers a number of topics in order: key use cases and requirements; an overview of related work; development methodology, principles and evaluation criteria; overview of key concepts within the information model; discussion of example implementations; evaluation against the defined criteria; challenges faced in development; and conclusions and areas for future work.

USE CASES AND REQUIREMENTS

Water observations data are used in many ways, from day to day management of rivers and storages, to provision of services such as water supply, irrigation, flood forecasting and hydropower. This paper outlines three key use cases/scenarios that were used in development of WaterML2.0 and assisted in determination of its scope.

Exchange of hydrological observations to support infrastructure and emergency services

A river-forecasting agency requires near real-time information relating to river conditions from operators of infrastructure, such as reservoirs, dams, etc. The forecasting agency requires operators to provide a machine-readable point of data access to time-series of conditions relating to reservoir levels and current discharge at outlets and/or inflows. The time-series must contain estimations of the accuracy of measurements and any quality assurance that has been performed by the providing agency. Influencing environmental conditions that can affect measurement values must be provided, where possible, to enable models to determine whether values are appropriate for use.

This scenario focuses on the base level representation of time-series data for the purposes of continuous monitoring and forecasting. The scenario presents requirements for information content of time-series as well as querying and access requirements.

Exchange of hydrological observations data to national repository

A regional hydrological monitoring agency maintains a network of river gauging stations that continuously report on river conditions, with a range of phenomena being measured per monitoring location. The agency is required to exchange this information with a national data repository that is used to produce national reports on hydrological conditions. The agency is required to provide a daily export of all new data to the national body for import into their system. The agency must supply local site identifiers and those provided by the national body in order to unambiguously identify monitoring locations. The data must be exchanged along with any quality assurance, processing and/or qualifying information that will assist in determination of the inherent quality and uncertainty in the data.

This scenario introduces a larger scope and number of involved parties, bringing with it requirements relating to multiple identifiers, locations of sites and information relating to the processing performed on data.

Exchange of groundwater levels across international borders

A groundwater resource monitoring body actively monitors an aquifer that crosses an international border and thus has limited understanding of water levels within the aquifer outside their responsible area. In order to enhance understanding of the state of the aquifer, a data exchange network is setup with the responsible agency across the border. The agencies must provide information relating to observations of the water levels in wells, details of the wells themselves and the reference datum used for measuring levels. Information on the measurement process (e.g. automated sensors, manual readings, etc.) should be provided, if available.

This scenario adds complexity to the spatial features being monitored (groundwater) as well as additional information relating to the measurement process. The cross-border nature of the exchange adds requirements relating to different vocabularies, identifier and naming schemes, units of measure, etc.

Requirements

Requirements were detailed as part of the development of WaterML2.0 by analysing the use cases and using existing formats as a proxy for requirements. The organisations involved all had domain experience relating to these use cases, and were able to fully describe and expand requirements resulting from these scenarios. The requirements are not replicated fully here, but are referred to within the evaluation and results section; generally, they apply to two areas: the content of the information model and the service used in implementing the transmission and querying of XML documents. The focus of the evaluation is on the content of the model, whereas the service aspects relate to the binding of WaterML2.0 with a particular web service such as OGC's Sensor Observation Service (SOS) (OGC 2012a).

RELATED WORK

The increasing need for data exchange within the environmental sciences has led to an increase in the development of open data formats and exchange protocols. These are

generally based on domain specific formats, where the underlying information model relates closely to the way data are understood and used within a domain. As science is becoming more cross-disciplinary, there are standards emerging that define concepts at higher levels of abstraction. For example, the OGC-ISO Observations and Measurements (O&M) (OGC 2010; ISO 2011) standard defines a cross-domain vocabulary for observational data and meta-data (extended discussion below). These standards allow more specific information models to relate to a common model and thus make it easier to share data across domains.

Related topics in hydrology

Recent work within Spatial Data Infrastructures (SDIs) has seen the development of some relevant water resources and hydrology information models. Specific hydrological time-series formats are covered in the description of the harmonisation process, as these were directly analysed as part of the development.

The Groundwater Markup Language (GWML) (Boisvert & Brodaric 2011) focuses on description of **hydrogeologic features**, such as aquifers, and water wells. GWML employs a similar information model design strategy by leveraging relevant standards from OGC. WaterML2.0 was shown to be compatible and usable alongside GWML for capturing observations relating to groundwater – this is detailed further in the results section.

A conceptual model for the definition of **hydrological features** (Atkinson *et al.* 2012) provides future means to relate water observations to common representations of features such as catchments, river networks and their constituent parts. Future testing will involve coupling WaterML2.0 with such a model, promising an advanced reporting architecture for hydrological systems.

Related domains

The Network Common Data Form (NetCDF) is a set of libraries and, primarily, binary file formats for the description of array-oriented scientific data. The Climate and Forecast (CF) conventions define how NetCDF formats should be used for common description of variables, spatial and temporal properties and grids. Together they

form one of the most common exchange formats for gridded data within the earth sciences. The format supports many forms of data, and the recent definition of discrete sampling geometries within the CF conventions directly addresses the various types of time-series generated from observational data. Hydrological use is not as extensive but there is work underway on mapping the WaterML2.0 conceptual model onto NetCDF (Palmer 2012). This would result in ability to translate between two models and may provide an option for a binary (well compressed) encoding of the WaterML2.0 information model.

The Climate Science Modelling Language (CSML) (Woolf *et al.* 2005) defines an abstract data model for addressing the semantics of climate data, which covers a number of types of spatio-temporal coverages, including time-series. Version 3.0 includes improved harmonisation with a number of existing standards including NetCDF climate forecasts discrete sampling geometries, coverages and O&M. Bringing these data types together into a harmonised model for cross-domain sharing of coverage and observational data is an area of future research.

The Weather Information Exchange Model (WXXM) is a proposed standard for the exchange of aeronautical

weather information for air transport systems. It makes use of O&M. Related to WXXM is the Aeronautical Information Exchange Model (AIXM) that focuses on definitions of features such as airspaces, aerodromes, restricted fly areas, etc.

The GeoScience Markup Language (GeoSciML) (Simons *et al.* 2006) addresses exchange of geological map data. It also developed a number of information modelling methods that were used in the development of WaterML2.0.

None of the above work addresses definition of a conceptual model for time-series that is flexible enough to support the identified requirements. While some provide the means to encode time-series in a particular way, the model is generally coupled to a specific domain (such as climate) or misses crucial aspects of time-series, such as support for metadata at varying levels of granularity.

METHODOLOGY AND PRINCIPLES

Harmonisation

A high level overview of the WaterML2.0 development process is shown in Figure 1. The development began with

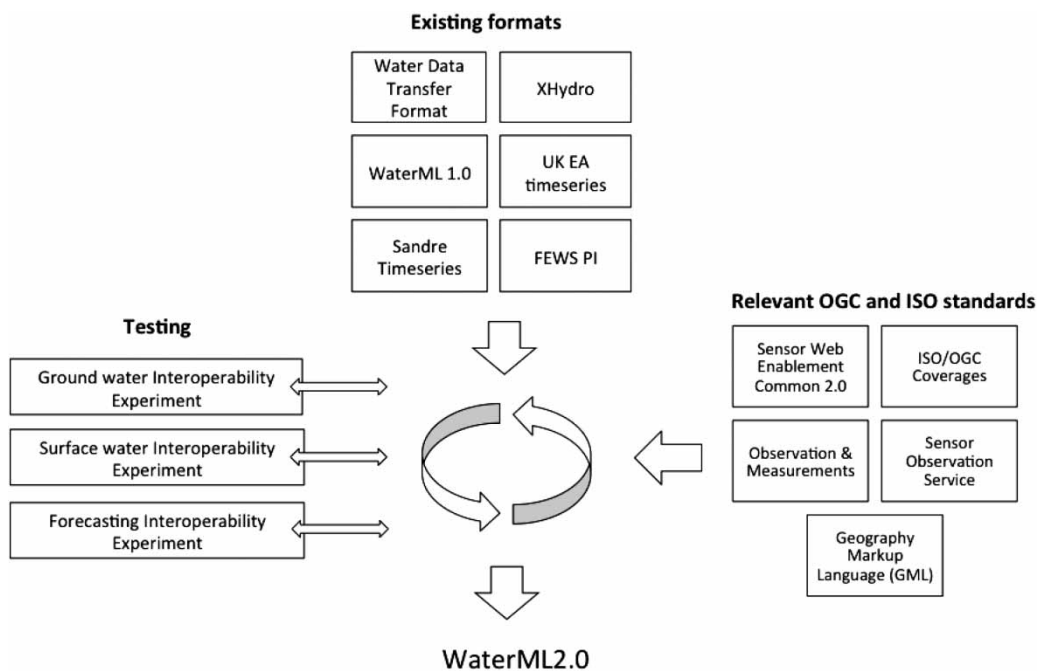


Figure 1 | Development methodology.

analysis of other existing formats identified internationally; there are no official international standards for this type of data, and national formats were generally not at the level of official standards. The formats that were analysed were: WaterML1.0 (Zaslavsky, I., Valentine, D., & Whiteaker, T. 2007 CUAHSI WaterML. Submitted to Open Geospatial Consortium as discussion document. Unpublished work), the Water Data Transfer Format (Walker *et al.* 2009), XHydro, UK Environmental Agency (EA) time-series format, Sandre timeseries format and the Flood Early Warning System's (FEWS) published interface from the USA, Australia, Germany, UK, France and the Netherlands respectively.

The analysis looked at the scope and conceptual overlap of the content of the formats, concluding that the formats contained significant overlap of core concepts and were addressing similar data exchange requirements. The formats had varying levels of scope: while most directly address exchange of time-series, others branch in to related areas, such as gridded data, cross sections, ratings and conversions, etc. A staged approach to develop WaterML2.0 was suggested to manage scope, with each part addressing particular aspects of hydrological data, with the time-series component making up part 1. As OGC was the target standards body, it was desirable that the development considered re-use, or at least alignment, with existing standards in the OGC suite. This was one of the key principles that guided development and is expanded below.

The development required interaction with two technical groups in OGC: the HDWG, which provides coordination and communication on hydrology topics in OGC; and the WaterML2.0 Standards Working Group (WaterML2.0 SWG), which focused on the preparation of the WaterML2.0 standard.

Standards basis

OGC architecture

The OGC architecture is specified in the OpenGIS Service Architecture (republished from ISO 19119). The underlying architecture principles for the OGC stack are defined within this specification; its basis is in the specification of Service Oriented Architectures for sharing geospatial information. Thus the majority of OGC

standards address either the information content (what is being exchanged) or services (how to access and query information). WaterML2.0 focuses on the information content, with the intention that the delivery and query mechanism makes use of existing services, either from OGC or those defined elsewhere.

Application schemas

OGC standards are based primarily on abstract models for geographic information described in the ISO 19100 suite of standards. The reference model, ISO 19101, defines principles for the definition and operation of the standards, with the key goal of interoperability within distributed computing environments.

The core principles for defining a domain information model and encoding, such as WaterML2.0, are provided in ISO 19109 'Rules for application schemas', which defines two base concepts for the representation and exchange of digital geographic data: The General Feature Model (GFM), and application schema.

An application schema defines the '*...content and structure of data and specifications for operations for manipulating and processing data by an application*'. The GFM is a meta-model for geographic 'features', which are the fundamental unit of geographical information. There are three key concepts: feature instances (objects with identity), feature types (classes of objects with common characteristics), and feature type properties (attributes, associations and operations of features). Together, the GFM and application schema provide a platform under which information can be consistently described and exchanged between parties.

Geography Markup Language

The Geography Markup Language 3.2 (GML) (ISO 2007) provides an XML encoding of the abstract feature concept, and other key classes from the standards, in particular:

- ISO 19107 – Spatial schema (spatial geometry and topology);
- ISO 19108 – Temporal schema (temporal geometry and topology, temporal reference systems);
- ISO 19111 – Spatial referencing by coordinates (coordinate reference systems);

- ISO 19123 – Schema for coverage geometry and functions (coverages, grids).

The XML syntax in GML reflects the nature of the GFM: all objects are expressed as features with properties that have values, which may also be features. This is derived from object-oriented modelling principles, and the encoding was inspired by the subject-predicate-object triples of the Resource Description Framework (RDF) formalisation of knowledge representation. As in RDF, GML features and properties of features may appear inline, or as references to values provided remotely. This adds significant flexibility, but at the cost of complexity in both producing and consuming applications.

Observations and measurements

ISO 19156 – O&M specifies ‘...a conceptual schema for observations, and for features involved in sampling when

making observations’. This is a high-level model that standardises certain terms for observation metadata, in particular: feature-of-interest, observed-property, procedure, temporal metadata and result. This allows data from disparate sources to be mutually understood and reused, and is applicable to many exchange scenarios. The core OM_Observation type is shown in Figure 2 using Unified Modelling Language (UML). O&M was developed as part of OGC’s Sensor Web Enablement (SWE) initiative (Bröring et al. 2011) and progressed into an ISO standard.

O&M also standardises the description of sampling artefacts such as sampling-points (e.g. station), sampling-curves (e.g. transect) and specimens (i.e. sample for *ex-situ* analysis), and the relationships between these (e.g. specimens retrieved from sampling-points).

Alignment of WaterML2.0 with O&M was identified as a key principle of development given its description of not only relevant observational-related data but also the

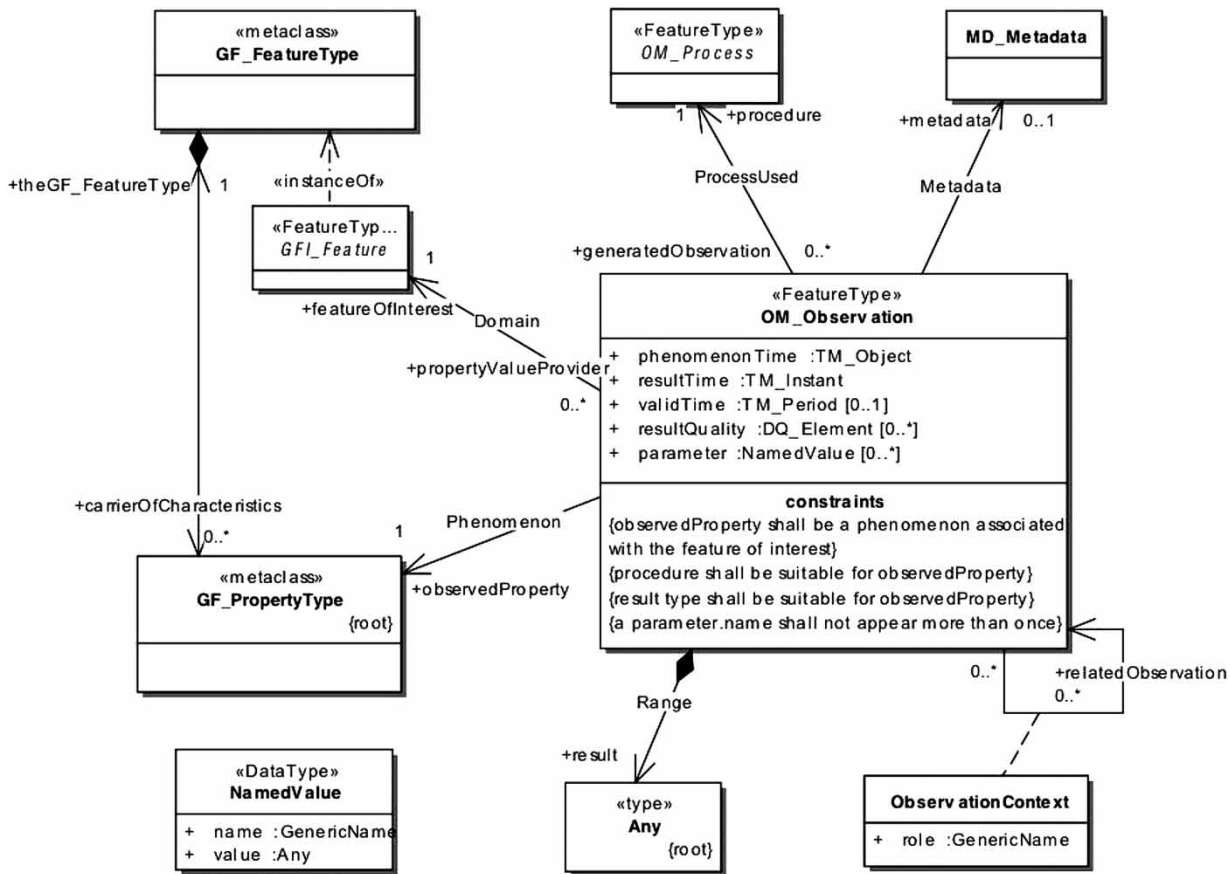


Figure 2 | OM_Observation type.

sampling process. It is also emerging as the preferred basis for data exchange applications from multiple domains and has been recommended for use within the European Infrastructure for Spatial Information in the European Community initiative (Schleidt *et al.* 2012).

Sensor Web Enablement (SWE) Common 2.0

The SWE Common Data Model 2.0 (OGC 2011) provides base types that are directly relevant to time-series, while also providing some custom XML encoding techniques specific for sensor-related data. The available types include concepts such as quantities, text expressions, categorical and Boolean values and data structures (arrays, matrices, etc.); these are re-used by WaterML2.0.

Model-driven methodology

The development of WaterML2.0 made use of a *model-driven* approach that is defined within ISO 19101, the geographic information reference model, which describes methods for developing geospatial information exchange systems. A key aspect of the method is in the expression of *conceptual models* that are abstract descriptions of a set of related concepts. These conceptual models only exist within the minds of people who communicate them verbally or textually, and are subject to misinterpretation; a *conceptual schema language* is required to express a conceptual model in a consistent way. The conceptual schema language is based on a *conceptual formalism*, for which the ISO 19000 series of standards makes use of version 1.0 of the Object Modelling Group's (OMG) UML (OMG 2009). UML is a standardised modelling language that provides graphical notation to express models and is used in this paper to describe aspects of the WaterML2.0 model. A conceptual model expressed using UML is referred to as a *conceptual schema*, and for the purposes of this paper is synonymous with an information model.

OGC reflects this practice, and many of the OGC standards use UML to express conceptual schemas. From the UML conceptual schema it becomes possible to generate GML conformant XML schema using automated tools (Golodoniuc & Cox 2010; ShapeChange 2012) providing

they follow the encoding rules defined within the GML Annex E.

Evaluation criteria

To evaluate the effectiveness of WaterML2.0 as a basis for hydrological time-series exchange, the following assessment criteria are used.

Compliance with OGC and ISO standards

Given the group was operating within a joint WMO/OGC domain working group, this was deemed a crucial criteria for acceptance as standards within these bodies. Failure to re-use existing standards would likely result in rejection of a proposed standard or reduced adoption through non-compliance with existing toolsets, such as those providing GML compliant software.

Satisfaction of use cases and requirements

The information model must be able to suitably address the domain-level requirements for time-series exchange. Two approaches are used to evaluate fulfilment of the requirements: the standard addresses the requirements derived from the existing time-series exchange formats, and the implemented standard satisfies the specified use cases.

The report generated from the harmonisation process (Taylor *et al.* 2010a) contains a detailed comparison of the domain-level requirements for expressing time-series and related concepts. The report includes a number of comparison matrices that compare features of each exchange format. These are used as requirements for assessment of the standard.

The results of three Interoperability Experiments (IEs) and another implementation are used to verify satisfaction of the use cases.

Usability, verifiability and performance

Putting the standard to use in the real world requires that the standard be implementable in an acceptable timeframe, understandable to developers and users, and verifiable against a set of technical and business rules that reflect the

semantics of this type of data (i.e. the requirements of the standard). The IEs are used as a basis for evaluation against these criteria along with the development of a validation service that supports verifiability against the standard.

INFORMATION MODEL OVERVIEW

Scope

Hydrological observations are performed using a variety of sampling and processing techniques, each requiring distinct data structures to capture results. Table 1 provides some broad categorisations and examples, grouping the styles to reflect common structures. The definitions are only indicative, as they represent a generalisation of operational practices: there are cases where the spatial and/or temporal resolutions vary for specific usages.

Part 1 of WaterML2.0 focuses on single parameter time-series, which fit primarily into categories 1, 2 and 4. Water quality data generally fits into category 3, which will be addressed in future work, as discussed below in the future work section.

Specification components

The WaterML2.0 specification (OGC 2012b) consists of multiple artefacts that together capture the information model, constraints and encodings. OGC rules for technical specifications (OGC 2009) require a modular structure based on requirements and conformance classes, to facilitate management of dependencies both within and between standards

(Coetzee *et al.* 2011). Each requirements-class defines a set of requirements for a particular standardisation target type. For example, a conceptual model defines requirements that must be satisfied by an implementation, and an XML schema defines requirements for an XML document. Each conformance-class packages conformance-tests for a single requirements-class. The WaterML2.0 standard is formalised with this structure. Figure 3 shows how the elements listed above relate to the modularity principles.

WaterML2.0 includes the following:

1. A conceptual information model, defined using UML, and based on O&M.
2. An XML encoding defined using W3C XML schema and derived from the information model using the rules described in the GML standard.
3. A small set of vocabularies, providing standard values for particular fields in WaterML2.0 datasets.
4. Schematron rules that capture additional validity requirements arising from business rules. Schematron is a constraint language for XML that allows expression of complex rules that are not fully covered by XML schema.

Key components

Time-series. Time-series are a key concept in (environmental) monitoring applications, including hydrology allowing continuous reporting and analysis of parameters through space and time. Time-series are transformed, corrected, interpolated and extrapolated to allow accurate and continuous reporting of conditions.

WaterML2.0 is limited to point-based time-series as a coverage that has a fixed spatial but varying temporal domain,

Table 1 | Observation styles

Observation style	Spatial resolution	Temporal resolution	Examples
1. <i>In-situ</i> , fixed	Sparse	Dense	River level, discharge, storage level, rainfall, pH, turbidity.
2. <i>In-situ</i> , manual	Potentially dense	Sparse	Groundwater observations made during pump tests at well sites.
3. <i>Ex-situ</i> , complex processing	Sparse	Sparse	Nutrients (nitrate, phosphorus, etc.), pesticides (atrazine, glyphosate, etc.), biological, etc.
4. Data products. Processed or synthesised data, often fusing multiple observation sources	Varied	Varied	Outputs from models, algorithms, water storage estimates, biological indices, etc.

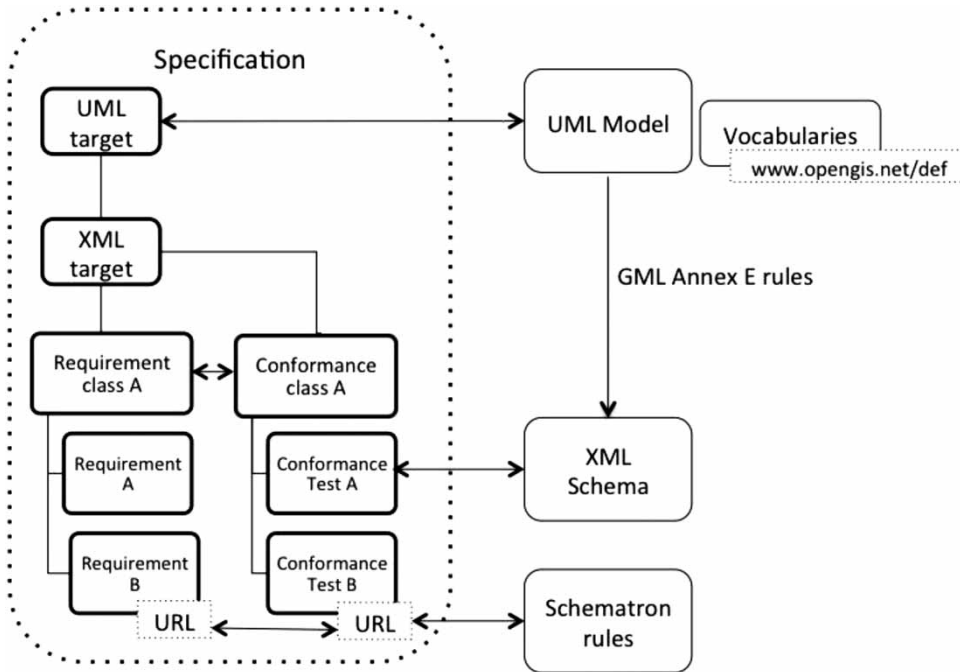


Figure 3 | WaterML2.0 specification artefacts.

where the range defines the varying parameter space (e.g. river level), but the definition is explicitly in terms of the coverage model from geographic information standards. In this terminology, time-series are a type of *coverage*. This a cross-domain concept, defined as ‘*a feature that acts as a function to return values from its range for any direct position within its spatial, temporal or spatiotemporal domain*’ (ISO 2005). A cross-domain conceptualisation for time-series is important for interoperability with related domains, such as meteorology and oceanography, and assists in integration with other coverage representations, such as those in NetCDF or other multi-dimensional structures.

Two structures of coverage representation or access are available in WaterML2.0, and are shown using UML in Figure 4: an interleaved approach in which the position and values are interleaved, and a ‘domain-range’ structure where the domain (set of positions) and range (set of values) are described separately, with a function that describes the relationship between the two parts. This can provide efficiencies when encoding as well as closer integration with other coverage types.

Existing practices within hydrology commonly represent time-series according to the interleaved approach: related

time-value pairs are encoded with associated metadata for individual points. Per point metadata are used within hydrology for making assertions about individual points, such as describing quality, accuracy, interpolation types, comments, event flags, influencing environmental conditions and so on. Traditional data exchanges often simplify the data being exchanged to simple time-value pairs, often resulting in a loss of this useful, and sometimes crucial, information.

The generic time-series structure captures the common nature of time-series irrespective of the phenomenon that is being observed through time. Two common specialisations are defined: measurements (shown in Figure 5) and categorical observations (e.g. ‘cloudy’). These specialisations also scope specific metadata for the value-type (e.g. measurement accuracy). They are single-parameter series where each time instant is associated with a single observational value only; support for encoding multiple parameters with a time instant will be addressed in future work. The single-parameter approach simplifies implementation and covers multiple parameters through collections of single parameter time-series.

Interpolation types. An important contribution was a harmonised set of definitions for interpolation types of individual

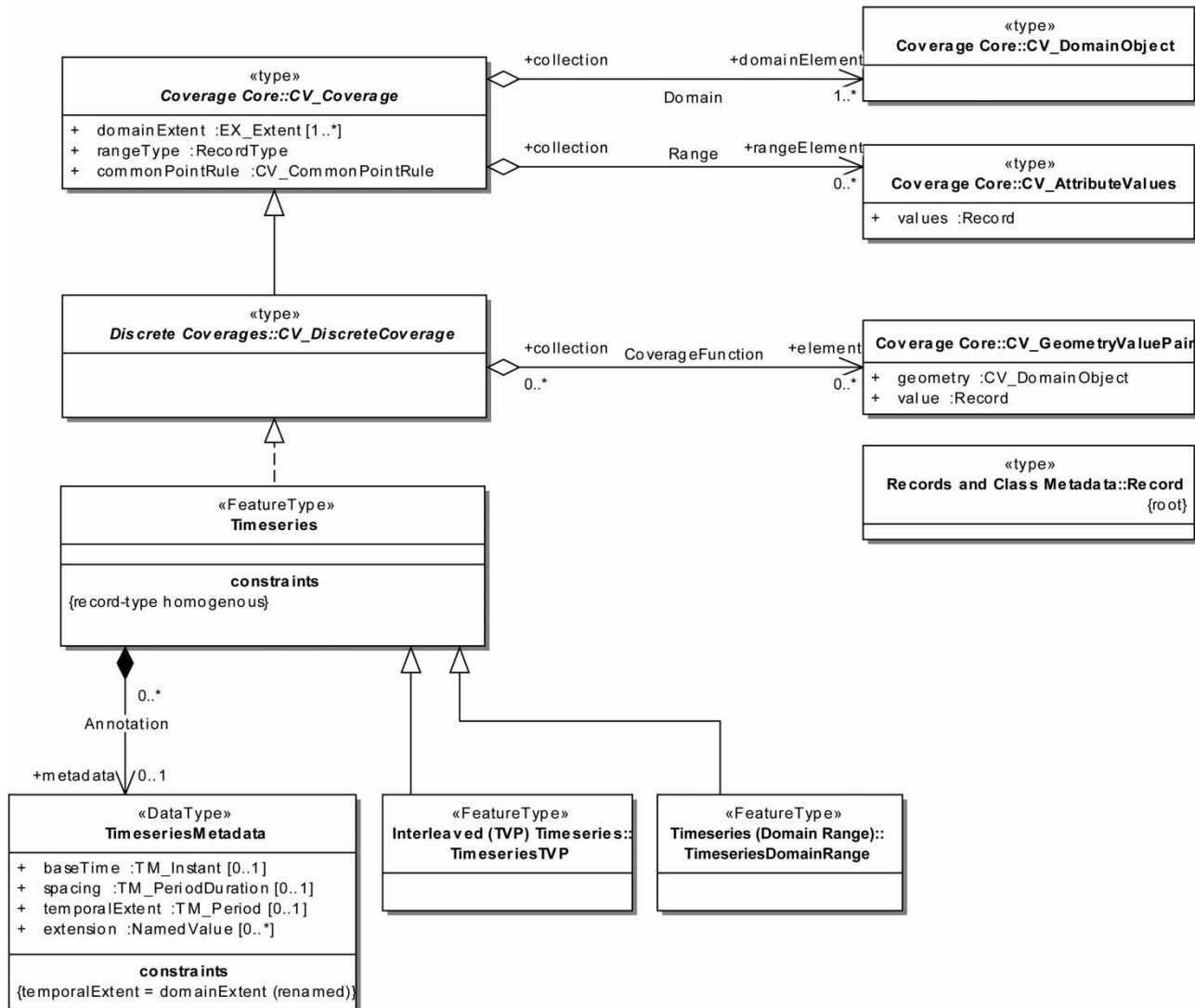


Figure 4 | Times-series relationship to coverages.

values within time-series (examples shown in Figure 6). These allow for correct interpretation of the relationship between time instants and the recorded value, which is determined by the procedure that was used to record the estimated value. WaterML2.0 defines the interpolation types using permanent URLs (e.g. <http://www.opengis.net/def/waterml/2.0/interpolationType/Continuous>), allowing the concepts to be directly referenced by encodings.

Accuracy. There are numerous approaches to quantification of accuracy (and uncertainty) of measurements – WaterML2.0 provides a flexible accuracy property that makes use of the

SWE Common data model to support different approaches. This makes it possible to encode accuracy using ranges, numerical statements or categorical statements. The semantics of the accuracy statement can be associated through a definition field within the SWE Common type. This may link to a controlled vocabulary of appropriate terms, such as those provided by UncertML (Williams *et al.* 2008).

Monitoring points

Within hydrology, terms such as site, station and location identify the spatial context in which observations are

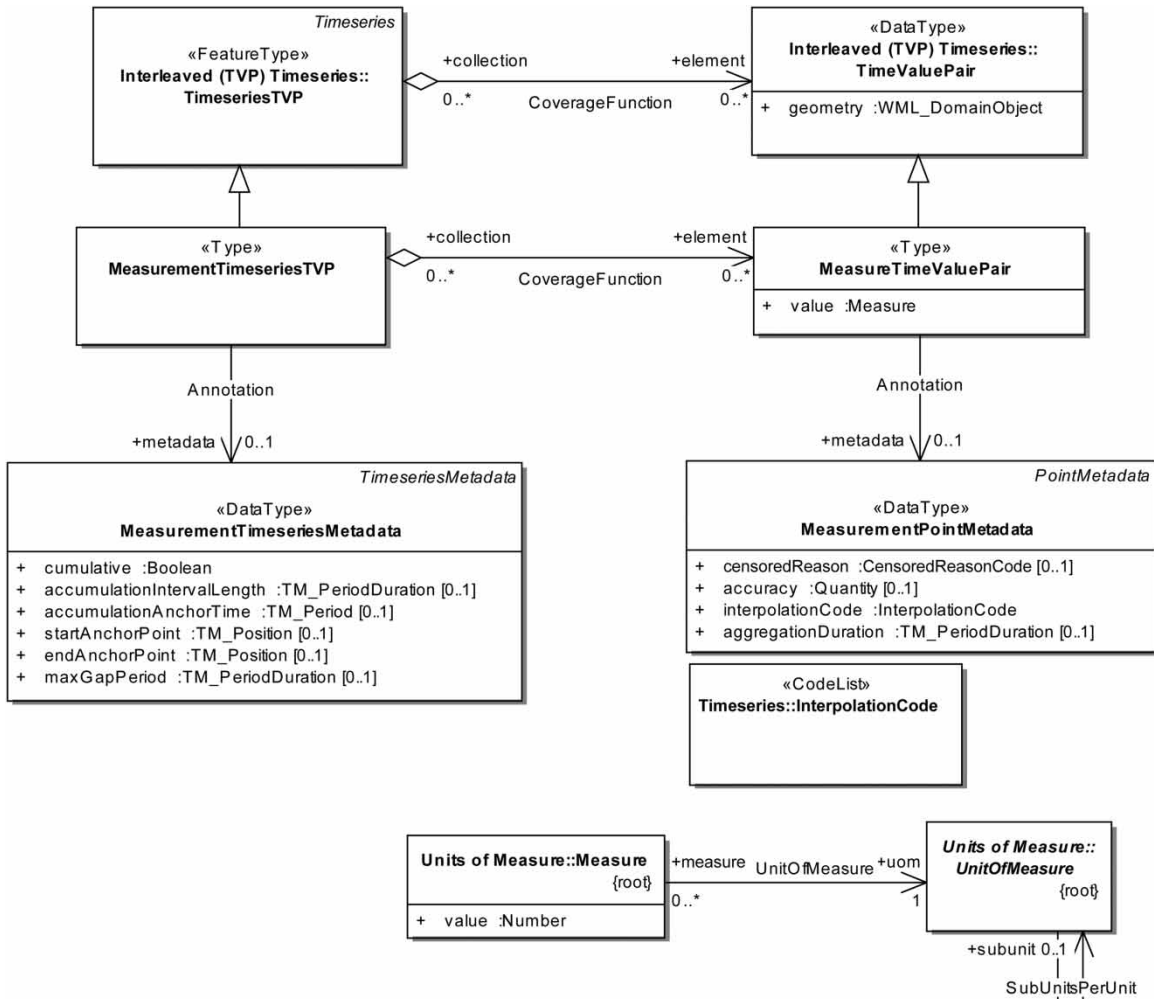


Figure 5 | Measurement time-series specialisation.

made. However, the semantics associated with these terms are not consistent: a site may contain many locations in which observations are made, or vice-versa; a site may be the place in which sensors are deployed, with each site containing just one sensor; a station may have many locations, and so on. The definition and nesting of these concepts is often reflected in the structure of feature identifiers, providing a shortcut in associating databases with the operation and maintenance of infrastructure.

The Sampling Feature model of O&M recognises that features created for the purpose of supporting observations have very specific relationships with the domain-features, with the observations, and with each other, and provides for these relationships to be recorded explicitly. A sampling

feature is a proximate feature that is representative of the larger domain feature. For example, river water quality is being measured at a point on the river reach to give an estimate of the quality of the water of the whole river. In hydrology, domain features would typically be spatial features like river reaches, reservoirs, dams, etc. The definition of hydrological domain features was outside of the scope of WaterML2.0 part 1. Current research on a conceptual model for hydrological features (Atkinson et al. 2012) provides a basis for future domain feature definitions.

A WaterML2.0 MonitoringPoint class was defined as a point-based sampling feature – a SamplingPoint as defined by O&M – that is used to observe the property of domain

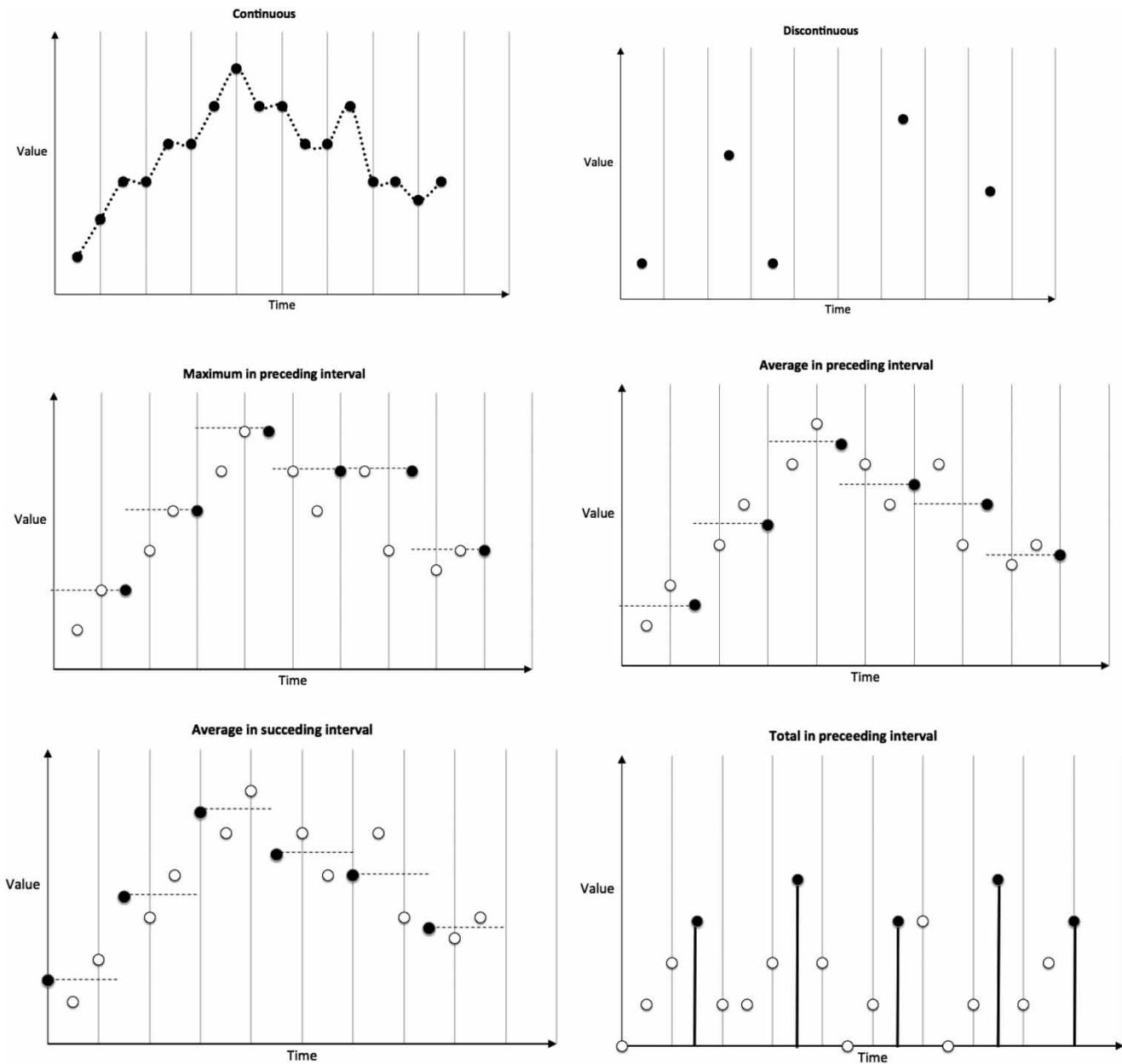


Figure 6 | Example interpolation types.

features. The UML for the WaterML2.0 monitoring point type is shown in Figure 7.

Procedures

Correct interpretation of time-series almost always requires some understanding of the procedure used in creation of

the data. O&M requires that the procedure used in making the observation be identified. The definition of specific types for procedures is delegated to more specific process models, such as OGC's SensorML (Botts & Robin 2007). However, SensorML was considered too general and too expressive for WaterML2.0, so a simple ObservationProcess type is provided (shown in Figure 8) that

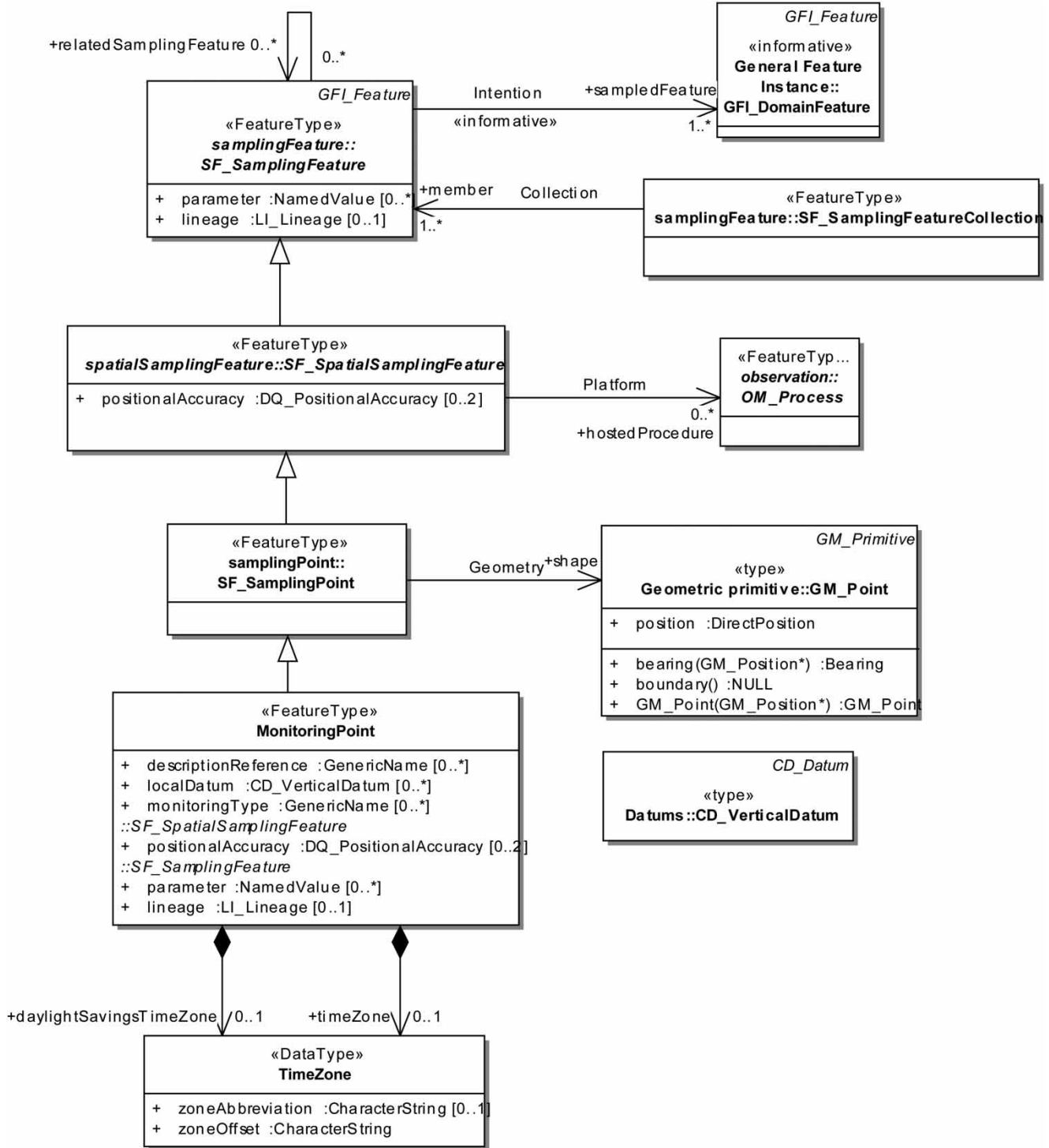


Figure 7 | MonitoringPoint type.

allows capturing of some key metadata relevant for interpreting time-series. While only providing base level information relating to the procedure, this is an improvement in the expressiveness when compared to the existing formats that largely don't track the process used in generating the time-series. This enhances the traceability of data through processing chains.

An associated vocabulary defines several observation process types: Simulation, Manual Method, Sensor, Algorithm and Unknown. The ObservationProcess type could be recast as a profile of SensorML, where the type is expressed as a set of constraints on SensorML to reduce its expressiveness; this may be addressed in future work.

XML encoding

The XML encoding of WaterML2.0 is a GML compliant encoding, thus allowing compatibility with GML compliant software. GML defines specific use of XML schema that affects the style in which XML is encoded, specifically in the way types are defined, a feature-property encoding pattern, use of XML elements over attributes and the use of XLink (W3C 2001) to allow external references from documents.

In addition to the GML encoding there are additional encoding options, such as SWE Common, that use custom techniques for compressing data into XML

elements. It was decided to use standard XML elements to encode the model to allow direct compatibility with off the shelf tools. The WaterML2.0 XML schema makes minor use of the SWE Common Data Model types, such as the Quantity type.

Code lists

Codes, abbreviations and other symbols are often used in standard formats to refer to concepts that hold specific meaning; for example, a qualifier 'p' may be associated to a value to indicate the value is provisional. In order to correctly interpret codes there must be a definition available describing the intended meaning of the code. E.g. For example, 'p indicates that these data are provisional and is subject to revision until formally approved'. When definitions are not available, the interpretation of a code is dependent on having direct contact with data owners or regular users, resulting in difficulties interpreting data from other parties.

Definitions may be provided on the Web with a URL identifier: the identifier is used in the exchange document and data consumers may check the definition by resolving the URI through HTTP. The definition returned may contain links to cited resources or related concepts, resulting in hierarchies of concepts such as thesauri, classification schemes and taxonomies.

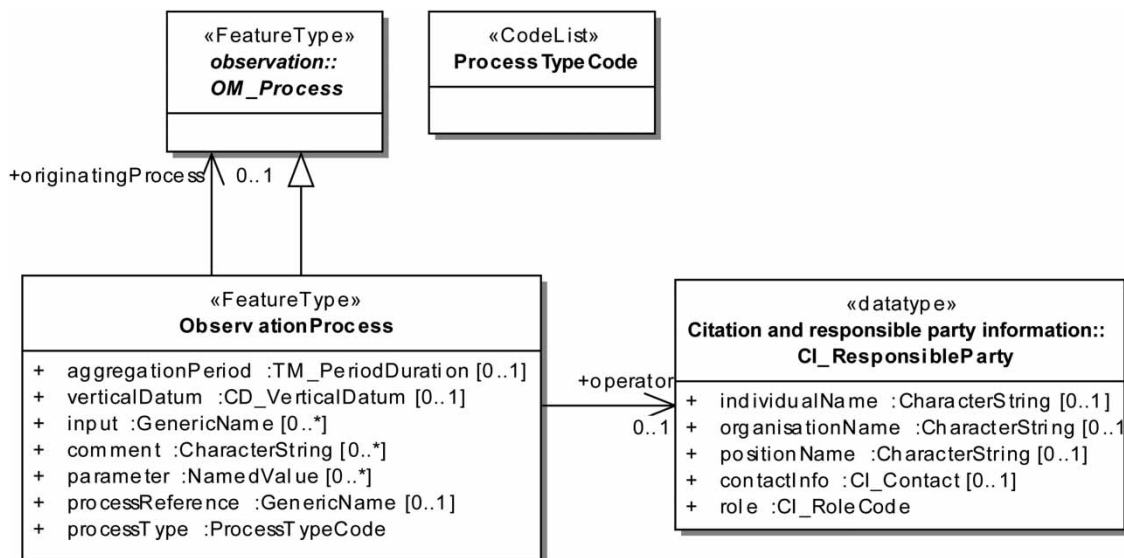


Figure 8 | ObservationProcess type.

The OGC Naming Authority (OGC-NA) subcommittee is responsible for managing identifiers for standards under OGC governance, which include the HTTP identifiers. This committee allows standards development groups to request permanent URLs for specific concepts that require definition. These identifiers are then associated with a persistent, online definition of the term. WaterML2.0 defines a number of such identifiers relating to time-series exchange, with more planned inclusions in future iterations.

IMPLEMENTATION

The most common web service for delivery of WaterML2.0 XML is OGC's SOS. This service provides standard interfaces for querying a broad range of observational data, from static observations to mobile observing platforms. It is a flexible service that may be used to provide access to data for many varied disciplines. Adapting SOS to serve WaterML2.0 encodings is made simpler by its use of OGC standards.

Interoperability experiments

OGC's Interoperability Programme offers a number of initiatives for testing existing or developing standards. IEs are '... are brief, low-overhead, formally structured and approved initiatives led and executed by OGC members to achieve specific technical objectives that further the OGC Technical Baseline.' Through the HDWG, three IEs were run to test WaterML2.0 implementations in various exchange scenarios. The results of the experiments were fed into the WaterML2.0 SWG, resulting in issues being addressed and new features included. The IEs all implemented WaterML2.0 using the SOS as the primary web service interface.

Groundwater interoperability experiment

The groundwater IE focused on the use of WaterML2.0 with other OGC standards and the best practice information model for groundwater information, the GWML (Boisvert & Brodaric 2011). The experiment focused on the first use case described in the use cases section: data sharing of aquifer information across the US-Canada border. Understanding

the aquifers that run across the border is hindered by the use of different information systems from the two countries.

The experiment implemented WaterML2.0 with SOS version 1.0, a Web Feature Service (WFS) for transmission of groundwater features, a Web Map Service (WMS) for displaying maps with features of interest and a number of clients to consume these services. The implementation used a number of parts of WaterML2.0 to satisfy its use cases: the time-series encodings for describing changing water levels of wells, time-series metadata, references to spatial features contained with the WFS to link observations to features, and references to the properties being observed to allow search and discovery. The developed clients demonstrated functional scenarios that allow users to discover, query, display and download both the time-series and groundwater feature data.

The key technical outcome from the groundwater IE was a detailed report (Brodaric *et al.* 2011) that identified 27 profile recommendations for the use of OGC standards, 10 standards-related issues and 7 standards recommendations. It was shown that WaterML2.0 was able to meet all the requirements for capturing time-series data, as well as relating to the groundwater features, and general compatibility with SOS.

Surface water interoperability experiment

The surface water IE focused on three specific use cases: cross-border data exchange for the Rhine river; use of WaterML2.0 for hydrological forecasts; and estimation of global runoff from large rivers discharging to oceans. This covers the other two of the base WaterML2.0 use cases: supporting infrastructure and exchange with a central repository.

The forecasting use case explored the performance requirements for near real-time hydrological forecasting systems. The work executed a number of performance benchmarks and comparisons with existing formats, with a goal of identifying potential issues with using WaterML2.0 for high frequency data. Performance is discussed in the evaluation section below.

The cross-border exchange and estimation of global runoff use cases presented requirements in the representation of time-series and specifically how well WaterML2.0 relates to existing domain software products, as these were

used as integration points. The results were positive and WaterML2.0 was able to capture the majority of required domain-level information (Fitch 2012). The IE proposed a number of additions to WaterML2.0, such as approaches for specifying multiple sensors at monitoring points, varying depths, and additional qualifying information for values. This resulted in additions to the standard once the requirements were analysed by the SWG.

In addition, the IE identified issues with adaption of large hydrological data sets to the SOS, specifically with identification of individual time-series streams within SOS implementations. A discussion paper has suggested a hydrological profile that couples WaterML2.0 with SOS to address this issue (Fuest & Utech 2012).

Forecasting interoperability experiment

The forecasting IE extends on the forecasting use case from the surface water IE and focused on investigation in a number of domain areas: different data types, such as dam and reservoir gate settings (categorical); snow pack information; field visits (stage, discharge metadata); overlapping results from hydrological forecasts and ensembles; per point metadata usage (e.g. grades, flags, notes, approval levels, etc.) and response times from implemented services. The experiment is still in progress, but thus far WaterML2.0 has been able to address all the requirements.

Soil moisture monitoring network

An implementation was developed in addition to the IEs to support data exchange within a cosmic radiation soil moisture network in Australia. A web service was developed (Taylor *et al.* 2011) based on WaterML2.0 and the Representational State Transfer (REST) principles (Fielding 2000), and provides both XML and JSON encodings. The implementation successfully satisfied querying of time-series across multiple phenomena, structuring of collections of series, descriptions of monitoring points and networks, and encoding of the time-series data itself. A dashboard was developed that consumed the service to allow researchers to access, visualise and download data. Informal feedback on the service from both developers and

researchers was positive, indicating it provided a natural fit for access to data.

EVALUATION

The evaluation of WaterML2.0 is based on results from the implementations described, the OGC and public review of WaterML2.0, and satisfaction of requirements expressed in the use cases and requirements section of this paper, and those coming from existing formats.

Compliance with OGC and ISO standards

WaterML2.0 makes use of GML 3.2, O&M 2.0, SWE Common 2.0 and ISO/GML Coverages. It has extensively reused both the conceptual models and XML implementations of these models in all areas of its specification. It deviates in minor areas that have been identified explicitly within the specification and are proposed as extension or revision to the related standards.

The public request period resulted in submission of 107 comments on the specification (OGC 2012c). None of these related to the re-use of existing standards. Similarly, the review from the OGC Architecture Board did not identify any issues relating to use of existing standards. The IEs all tested compliance with GML 3.2, O&M and SWE Common 2.0 through positive validation of schema. These results suggest that the methodology and developed information model was sound when considering its use of existing standards.

Satisfaction of use cases and requirements

Use cases

The WaterML2.0 information model held up well when addressing the described use cases. It rarely required addition of new domain-level features to support any of the requirements. This also suggests the possibility that the use cases are not fully testing the concepts supported within WaterML2.0. Future implementations should address more complex examples of time-series encoding, such as varying datum for level measurements, detailed

information relating to the process of observation, additional value-based metadata such as accuracy statements, quality assurance indicators and thresholds. This would more deeply test the core of WaterML2.0, which is in time-series with values that carry enhanced contextual information allowing for more precise interpretation.

General concepts

WaterML2.0 provides more expressiveness in the definition of the process used to generate time-series than all the other identified hydrological formats. This is essential to building in more transparency in data, especially in derived hydrological data, such as river discharge. This is becoming increasingly important both within hydrology (Beven *et al.* 2012) and more broadly for environmental information (Whitfield 2012).

The ability to reference features, properties and vocabularies through the use of XLink is aligned with the principles of Linked Data (Bizer *et al.* 2009). This ensures WaterML2.0 is consistent with emerging approaches for linked open data, whereby the web is becoming a web of linked data rather than individual web pages. Other hydrological formats tend to be more closed, with references generally tied to a single organisation's data or vocabularies. The ability to reference data elements across the web has many implications (and challenges), including referencing data across organisations and to authoritative vocabularies.

Being based on OGC standards, WaterML2.0 is natively compatible with OGC web services, and other standards based on the OGC stack. Most of the existing formats (WDTF being the exception) are not based on adopted standards, resulting in formats that redefine many base concepts, such as spatial referencing, geometries and base types like measurements, units, etc. The benefits of such compatibility were shown within the groundwater IE, where WaterML2.0 was coupled with GWML for expressing groundwater structures and observations relating to aquifers. As these approaches mature, the community can leverage off well-established implementations of services and information models rather than having many bespoke technologies.

WaterML2.0 was designed as an extensible format; it provides extension points that allow for encoding of

non-standard information. This may be used to carry organisation-specific data or as extension to the general functionality. Other formats were generally locked down to the designed schema and did not provide such extension.

Time-series requirements

The requirements identified within the harmonisation process compared the support for time-series metadata down to per value level for existing formats. Each concept (or feature in the software usage) has been enumerated and rated based on the support within each format with results shown in Table 2. Each format is given a rating based on its support for the concept (0 for no support, 1 for support at a whole of time-series level, 2 for support within temporal bounds and 3 for per value support).

The concepts within the table are defined differently across the existing formats, resulting in a comparison that is subjective to the granularity of definitions. Given WaterML2.0 was developed as a harmonisation of these formats, it is logical that it fully supports these concepts. The comparison shows WaterML2.0 is an information model capable of expressing a number of time-series concepts. However, this does result in a format that is more complex to encode and decode.

There was an ongoing tension in how specific the semantics of each time-series metadata element should be. The existing formats all have slightly varied definitions of what constitutes a particular concept. For example, some formats would combine value quality and qualifiers into one metadata element and loosen the semantics of the term. Loose semantics allows flexibility at the cost of interpretation for software at the encoding/decoding levels. WaterML2.0 attempts to prioritise the key metadata for time-series values, but there still remains some ambiguity that may result in varied usage. The definitions do, however, provide noticeable improvement in clarification over existing formats.

WaterML2.0 also clarified a number of concepts relating to expression of time-series, including the interpolation types (discussed in the information model overview) and temporal metadata, such as anchor points and accumulation types. These existed in varying levels across the existing formats; WaterML2.0 presents a harmonised and expressive

Table 2 | Results of comparison of requirements across formats

Concept	WaterML1.0	XHydro	WDTF	Sandre	EA UK TS	FEWS PI	WaterML2.0
Qualifiers	3	3	3	3	1	3	3
Quality	3	3	3	3	3	3	3
Comments	0	0	3	0	2	3	3
Interpolation type	1	3	3	3	1	0	3
Processing	1	0	3	1	1	0	3
Accuracy	3	3	0	0	0	0	3
Changing vertical	3	0	0	1	0	0	1
Null values	3	3	3	3	3	3	3
Total	17	15	18	14	11	12	22

0 – no support.

1 – defined for whole time-series.

2 – defined for temporal bound.

3 – defined per time-series value.

definition that should allow better common understanding of the underlying concepts. It also provided a number of optimisations for time-series, such as defaulting of value metadata and specification of equidistant series; these were only available in some of the existing formats.

Usability, verifiability and performance

WaterML2.0 was implemented a number of times by members of the HDWG. While no objective measures of implementation effort are available, the implementations proceeded without major investments in development time – IEs are run as in-kind contributions that align with existing work, and thus cannot incur too much cost for participants.

The benefits of being based on open standards will emerge through time; implementations may be shared and built on existing OGC services which should result in cut development times and improved functionality for users.

Validation

A validation engine was implemented (Yu *et al.* 2012) that provides extended validation of WaterML2.0 XML instances. The approach provides validation against the requirements specified within the specification and associates an implementation to achieved conformance classes.

This supports implementations that are more compliant and assists in decreasing development times as developers gain access to instant feedback on conformance. Within the identified existing formats, only WDTF provided this advanced level of validation against complex rules.

Performance

Within the IEs, there were a number of organisations that investigated performance in encoding and decoding WaterML2.0 XML (Gijsbers & de Rooij 2011; Utech & Fuest 2012). The main findings were that WaterML2.0 XML files are more complex and larger than most of the existing formats, and thus slower to produce and parse using standard XML tools. The complexity of the encoding is influenced by two factors: the GML rules for XML application schema (such as not using attributes, and the feature-property pattern discussed in the GML section) and the fully expressive nature of WaterML2.0 (multi metadata elements per value). The decision was made not to prematurely optimise the WaterML2.0 XML encoding in order to save a few XML elements. The IEs demonstrated that use of XML tools such as FastInfoset (a binary encoding of XML) could significantly reduce issues relating to XML generation and parsing. It was also shown that simple compression of XML removed issues relating to file size.

CHALLENGES

Harmonisation

The first stage of harmonisation required analysis of existing standards to identify overlapping concepts and come to agreement on a harmonised definition. Generally the formats captured similar requirements, but with subtle differences, such as how time-series point metadata were handled. There were varying opinions on how this should be handled. For example, a time-series may be described using a separate metadata element that describes metadata blocks applying within particular time bounds. This is a valuable feature, but does put a load on encoders and decoders needing to construct and parse such descriptive blocks. A consensus approach resulted in the single time-series default metadata feature.

The focus of development was on structural definitions for the transfer of time-series, with less work on harmonisation of vocabulary items such as quality codes. The work did harmonise some vocabularies, focusing on those important to the semantics of time-series structures, such as interpolation types. Many vocabularies tend to be specific to local organisation context, so the harmonisation process can take time. Future work within the HDWG will progress development and management of shared vocabularies.

Relating time-series to observations and measurements

The Observation type from O&M provides a common structuring of the core aspects of observations data: spatial, temporal, observed phenomenon, process and other related metadata, to facilitate interoperability. Time-series may be modelled two ways using the O&M Observation type: using a collection of individual Observation objects that have a specific, and most likely homogenous, result type; or by defining a time-series result type where the Observation acts as carrier of all contextual information to identify that particular time-series.

The second approach was taken in WaterML2.0 as this more closely matches how time-series are viewed within the hydrology domain. The time-series structure defined

in the previous section describes the result type of these observations. Following the specialisation mechanism defined by O&M, WaterML2.0 defines additional subclasses of Observation that have time-series as their result type: TimeseriesObservation (abstract), InterleavedTimeseriesObservation and DomainRangeTimeseriesObservation. This results in naming that may confuse some users: the term 'observation' is often interpreted as a single measurement, whereas a TimeseriesObservation has a time-series result that is likely to be many measurements.

Consistent use of standards

Re-using and maintaining consistency with existing standards within the OGC standards stack presented a challenge: there are a number of relevant information models and services within OGC, some with overlapping functionality. For example, the SWE Common standard has defined base types that override some types previously defined in GML; identifying the correct types to use in particular circumstances requires care.

There is a lack of a common definition of time-series within OGC. While in principle 'ISO 19123 – Coverages' deals with spatio-temporal coverages and is in the OGC baseline standards, there is little prior implementation experience with temporal coverage domains. SWE Common provides data structures that may be used to encode time-series, but a formal information model of what constitutes a time-series is not provided.

WaterML2.0, being a domain-focused information model, should not necessarily contain the conceptual definition of time-series themselves, but rather how they are to be exchanged for hydrological applications. Future work should address the definition of an abstract model for time-series that is consistent and reusable across domains. WaterML2.0, NetCDF and CSML provide a good basis for this progression.

Profiling

Being a flexible model for use in hydrological time-series exchange, WaterML2.0 implementers face a number of design decisions when using it in particular exchange scenarios. These decisions relate to the use of code lists

(controlled vocabularies), soft-types (extensions), which conformance classes provide, adapting to web services, mapping of identifier systems and selection of domain features to describe relevant spatial features. Within OGC this is commonly referred to as 'profiling' an information model. Ensuring that WaterML2.0 is profiled consistently will result in enhanced interoperability and compatibility of software tools.

Scope of XML implementation

The XML implementation of WaterML2.0 provides the interleaved encoding in the core schema, and a domain-range style as a proof-of-concept only. This was because the latter had not been tested sufficiently and there was no well-defined approach for associating per value metadata with time and values encoded separately. The implementation used an approach where the time-series could consist of annotation coverages that had the same domain as the domain-range time-series, but the values were the metadata elements. For example, a time-series may have an associated 'quality coverage' that encodes quality as it changes through time. These annotation coverages would often be highly repetitive for irregularly changing metadata – a default metadata approach may be desirable, such as was implemented for the time-value pair encoding. Future work will consolidate the domain-range encoding with existing practices and requirements.

Model and specification development

Keeping the specification artefacts in synch during development was an ongoing challenge. Once decisions on the information model or extra constraints were agreed on, changes to artefacts would result in related changes to dependent parts; for example, a change in a model element within the UML would change the specification document, XML Schema and potentially the Schematron and vocabularies.

This dependency management issue was alleviated through the described model-driven approach to the XML schema development, but this only addresses the UML and XML components, not the specification, XML examples and Schematron rules. Managing the whole specification

package using a model-driven approach is an area of active research.

CONCLUSIONS AND FUTURE WORK

WaterML2.0 represents the first official international standard for hydrological observations data. Its scope has been carefully managed and there remain areas where it may be extended and refined as it evolves, many of which have been outlined in this paper. The results of evaluation show that WaterML2.0 is a robust information model that holds up to many requirements for hydrological data exchange. Its foundation and consistency with other standards provide a basis for building advanced information exchange systems that are able to leverage OGC standards and services.

A major aspect of the WaterML2.0 development was the identification of the key concepts and requirements for information models supporting exchange of hydrological time-series data. The level of interest in implementation and use of WaterML2.0 suggests that it is filling an under addressed area in knowledge representation for the domain. One of the key novelties of the work is the harmonised time-series model, which has the potential for re-use across multiple domains.

The OGC's domain working groups provide an active forum in which domain experts can collaborate to address data exchange and interoperability issues. There is a learning curve associated with new members in both process and technical background, but the access to experts, re-usable standards and the initiatives, such as the IEs, outweigh this. There are other standards bodies, but none – that the authors are aware of – that address exchange of environmental information, such as hydrological data.

A number of areas of work will continue for WaterML2.0. The first area involves extension of the information model to cover specific types of hydrological data. Two core types have been identified: rating and gauging information, and water quality. Rating curves, gaugings and river cross-sections are commonly exchanged in addition to standard time-series data to allow information relating to conversions, such as river level to discharge. Members of the HDWG initiated this work in early 2012. Water quality data are varied in the way it is processed

and in the number of phenomena it measures. It will require specific components of extension to the WaterML2.0 model, making use of the specimen types within O&M and extensive use of controlled vocabularies.

A second area of future work involves different target encodings for the WaterML2.0 conceptual model. XML provides a high fidelity encoding that is optimised for machine interpretation and longevity by being fully explicit in its definition of data items. Other encodings are available that address particular needs, such as efficiency in size of the encoding and more efficient parsing of structure. The community requirements will drive future developments in this area.

Support for multi-parameter time-series is also an area of growing interest. This may be implemented with an extension schema using the base time-series types. Some background work has begun in this area by members of the community.

There remain many challenges to achieve fully expressive and open hydrological data where many parties contribute to an increased global awareness of hydrological systems. These include technical barriers such as interoperability, but there are many cultural issues around data access, timeliness and liability. Organisational methods and procedures for measurement vary widely – standards need to be accommodating for different methods, but not so flexible as to become overly complex. WaterML2.0 attempts to find a middle ground by capturing the key time-series concepts, allowing re-use across many data exchange scenarios.

ACKNOWLEDGEMENTS

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 National Science Foundation grant EAR 06-22374 for the development of Hydrologic Information Systems and the Consortium for the Advancement of Hydrological Sciences Inc. (CUAHSI). The work is operating under the joint World Meteorological Organization (WMO) and Open

Geospatial Consortium (OGC) Hydrology Domain Working Group. The authors would like to thank members of the WaterML2.0 standards working group and Hydrology Domain Working Group for their inputs and enthusiasm in the development of WaterML2.0. They would also like to thank the anonymous reviewers for their constructive reviews.

REFERENCES

- Atkinson, R., Dornblut, I. & Smith, D. 2012 *An international standard conceptual model for sharing references to hydrologic features*. *J. Hydrol.* **424**–**425**, 24–36.
- Beven, K., Buytaert, W. & Smith, L. A. 2012 *On virtual observatories and modelled realities (or why discharge must be treated as a virtual variable)*. *Hydrol. Process.* **26** (12), 1905–1908.
- Bizer, C., Heath, T. & Berners-Lee, T. 2009 *Linked data-the story so far*. *Int. J. Semant. Web Inf. Syst. (IJSWIS)* **5** (3), 1–22.
- Boisvert, E. & Brodaric, B. 2011 *GroundWater Markup Language (GWML) – enabling groundwater data interoperability in spatial data infrastructures*. *J. Hydroinf.* **14** (1), 93–107.
- Botts, M. & Robin, A. (eds.) 2007 *OpenGIS Sensor Model Language (SensorML) Version 1.0 OGC Document 007-000* 17 Jul. 2007, Accessible from <http://www.opengeospatial.org/standards/sensorml>.
- Brodaric, B. & Booth, N. (eds.) 2011 *OGC Groundwater Interoperability Experiment Final Report*. OGC 10-194r3. Available http://external.opengis.org/twiki_public/pub/HydrologyDWG/GroundwaterInteroperabilityExperiment/OGC_GWIE_Final_Report_Mar2011.pdf (accessed 10 July 2013).
- Bröring, A., Echterhoff, J., Jirka, S., Simonis, I., Everding, T., Stasch, C., Liang, S. & Lemmens, R. 2011 *New generation sensor web enablement*. *Sensors* **11** (3), 2652–2699.
- Coetzee, S., Cox, S. J. D. & Herring, J. 2011 *Configuration management of a system of interdependent standards*. In: *7th International Conf. Standardization and innovation in information technology*. <http://dx.doi.org/10.1109/SIIT.2011.6083604>
- Fielding, R. T. 2000 *Architectural Styles and the Design of Network-based Software Architectures*. Doctoral Dissertation, University of California, Irvine.
- Fitch, P. (ed.) 2012 *OGC Surface Water Interoperability Experiment – final report*, Open Geospatial Consortium, OGC 12-018r2.
- Fuest, S. & Utech, M. 2012 *A Hydrology Profile For the SOS*. Presentation to the Hydrology Domain Working Group, Reading, UK.
- Gijsbers, P. & de Rooij, E. 2011 *Comparison of time-series ingest performance of various standardized file formats using*

- DelftFEWS. In: OpenWater symposium, 18–19 April 2011, Delft, The Netherlands, p. 32.
- Global Climate Observing System (GCOS) 2005 Analysis of Data Exchange Problems in Global Atmospheric and Hydrological Networks, GCOS - 96, February.
- Golodoniuc, P. & Cox, S. 2010 *Geospatial information modelling for interoperable data exchange – Application schema modelling: From concept to implementation*. In: *IEEE Sixth International Conference on eScience*, 7–10 December 2010 Brisbane, Australia, pp. 102–105.
- IBM 2009 IBM Water Management Pains Report: Survey Findings and Recommendations. <http://www-935.ibm.com/services/us/gbs/bus/pdf/ibm-water-pains-report-jan09.pdf>
- ISO 19123 2005 *Geographic Information – Coverage Geometry and Functions*. Int. Org. Standardization, Geneva.
- ISO 2007 *Geographic Information – Geography Markup Language v3.2*. Int. Org. Standardization, Geneva. (Also published as OGC 07-036, 2007.)
- ISO 19156 2011 *Geographic Information – Observations and Measurements v2.0*. Int. Org. Standardization, Geneva.
- Lemon, D. 2011 The What, Why and How of the WMO/OGC Hydrology Domain Working Group. In: OpenWater symposium, 18–19 April 2011, Delft, The Netherlands, p. 5.
- OMG 2009 Unified Modelling Language (UML). Accessible from <http://www.uml.org/>
- Open Geospatial Consortium 2009 The Specification Model – A Standard for Modular specifications, OGC 08-131r3.
- Open Geospatial Consortium 2010 Observations and Measurements – XML Implementation. OGC 10-025r1. <http://www.opengeospatial.org/standards/om>.
- Open Geospatial Consortium 2011 Sensor Web Enablement Common Data Model 2.0. 08-094r1. <http://www.opengeospatial.org/standards/swecommon>.
- Open Geospatial Consortium 2012a Sensor Observation Service 2.0. OGC 12-006. <http://www.opengeospatial.org/standards/sos>.
- Open Geospatial Consortium 2012b WaterML2.0 – part 1 timeseries. <http://www.opengeospatial.org/standards/waterml10-126r3>.
- Open Geospatial Consortium 2012c WaterML2.0 RFC Comments and Responses, 12–042.
- Palmer, D. 2012 WaterML2.0 time-series – NetCDF Encoding – Discussion paper, Open Geospatial Consortium 12-031r2.
- Schleidt, K., Cox, S., Grellet, S., Lowe, D., Lutz, M., Portele, C., Sarretta, A. & Ventouras, S. 2012 Observations and Measurements Design Patterns within INSPIRE. In: *EGU General Assembly Conference Abstracts*, vol. 14, p. 14005
- ShapeChange 2012 Available at <http://shapechange.net/>. Accessed September 2012.
- Simons, B., Boisvert, E., Brodaric, B., Cox, S., Duffy, T. R., Johnson, B. R., Laxton, J. L. & Richard, S. 2006 GeoSciML: Enabling the Exchange of Geological Map Data. In: Australian Earth Sciences Convention, 2–6 July 2006, Melbourne, Australia.
- Taylor, P., Cox, S., Walker, G. & Valentine, D. 2012 Adapting Standard Information Models for Water Data. In: *10th International Conference on Hydroinformatics*, 14–18 July, 2012, Hamburg, Germany.
- Taylor, P., Terhorst, A. & Cox, S. 2011 RESTful services for sensor data. *Geophys. Res. Abstr.* **Vol. 13**, EGU2011-13469-4.
- Taylor, P., Valentine, D., Walker, G., Sheahan, P., Pratt, A., Dornblut, I., Grellet, S., Heidmann, C., Wilhelmi, J., Michl, C., Gijsbers, P. & Natschke, M. 2010a Harmonising Standards for Water Observation Data – Discussion Paper, Open Geospatial Consortium 09-124r2.
- Taylor, P., Walker, G., Valentine, D. & Cox, S. 2010b WaterML2.0: Harmonising standards for water observations data. In: EGU General Assembly 2010, 2–7 May, 2010, Vienna, Austria.
- Utech, M. & Fuest, S. 2012 *WaterML2.0 Implementation Performance Results*. Presented to Hydrology Domain Working Group, Reading, UK, June 2012.
- Valentine, D., Taylor, P. & Zaslavsky, I. 2012 WaterML, an Information Standard for the exchange of *in-situ* hydrological observations. *EGU Gen. Assem. Conf. Abstr.* **14**, 13275.
- Walker, G. & Taylor, P. 2011 WaterML2.0: Enabling water information exchange. In: OpenWater symposium, 18–19 April 2011, Delft, The Netherlands, 47.
- Walker, G., Taylor, P., Cox, S. & Sheahan, P. 2009 Water Data Transfer Format (WDTF): Guiding principles, technical challenges and the future. In: 18th World IMACS Congress and MODSIM09 International Congress on Modelling and Simulation, Modelling and Simulation Society of Australia and New Zealand and International Association for Mathematics and Computers, in Simulation, Cairns Australia, pp. 2377–2383.
- Whitfield, P. H. 2012 Why the provenance of data matters: assessing ‘fitness for purpose’ for environmental data. *Can. Water Resour. J.* **37** (1), 23–36.
- Williams, M., Cornford, D., Bastin, L. & Ingram, B. 2008 UncertML: an XML schema for exchanging uncertainty. In: *Proceedings of GISRUK 2008*, pp. 275–279.
- Woolf, A., Lawrence, B., Lowry, R., van Dam, K. K., Cramer, R., Gutierrez, M. & Stephens, A. 2005 Climate Science Modelling Language: standards-based markup for metocean data. In Proceedings of 85th meeting of American Meteorological Society http://ams.confex.com/ams/Annual2005/techprogram/paper_86955.htm (accessed 10 July 2013).
- W3C 2001 XML Linking Language (XLink) version 1.0. Accessible from <http://www.w3.org/TR/xlink>
- Yu, J., Taylor, P., Cox, S. & Walker, G. 2012 Towards validating observation data in WaterML 2.0. In: *10th International Conference on Hydroinformatics*, 14–18 July, 2012, Hamburg, Germany.