Metadata and Data Standards. Sharing Data in Hydrology: Best Practices

Ilya Zaslavsky
San Diego Supercomputer Center

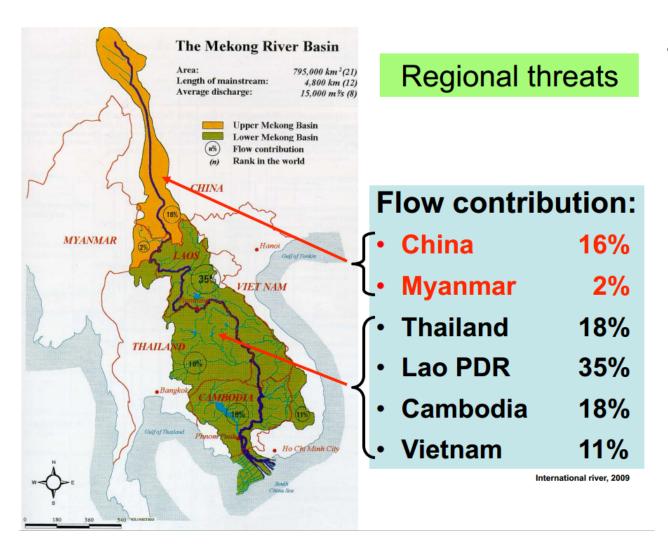
LMI Workshop, Hanoi, August 18-22

/ With several slides from last week's HDWG workshop, presented by HDWG members Irina Dornblut, Paul Sheahan, and others/

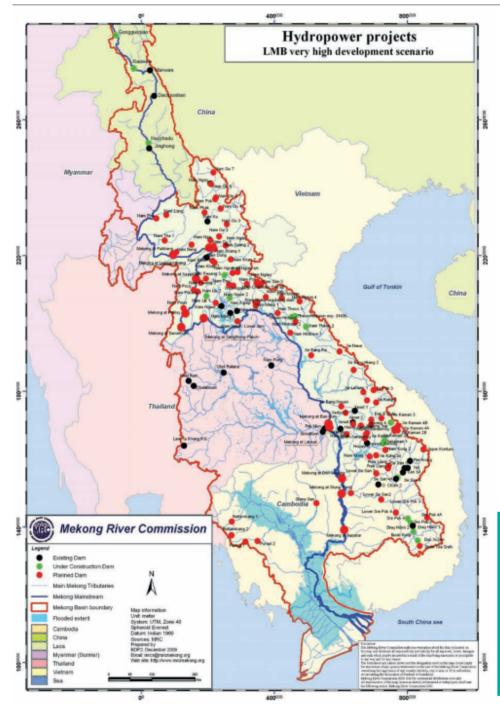
Outline

- Why use standards?
- Open Geospatial Consortium, and spatial data standards
- Standards for water data, and the OGC/WMO Hydrology Domain Working Group
 - history, activities, WMO connection, workshop last week
 - Suite of water data standards
- WaterML 2.0 in detail (optional)
- Assessing compliance, and the CINERGI project (optional)

Why sharing data in LMI?



 Several countries rely on the Mekong but data sharing is complicated



Existing or planned dams on the Mekong river

Challenges:

Habitat alteration Pollution

Extreme weather events
Over-exploitation of resources
Diseases and invasive species
Poverty and social instability

Hydro-power dams on the Mekong river:

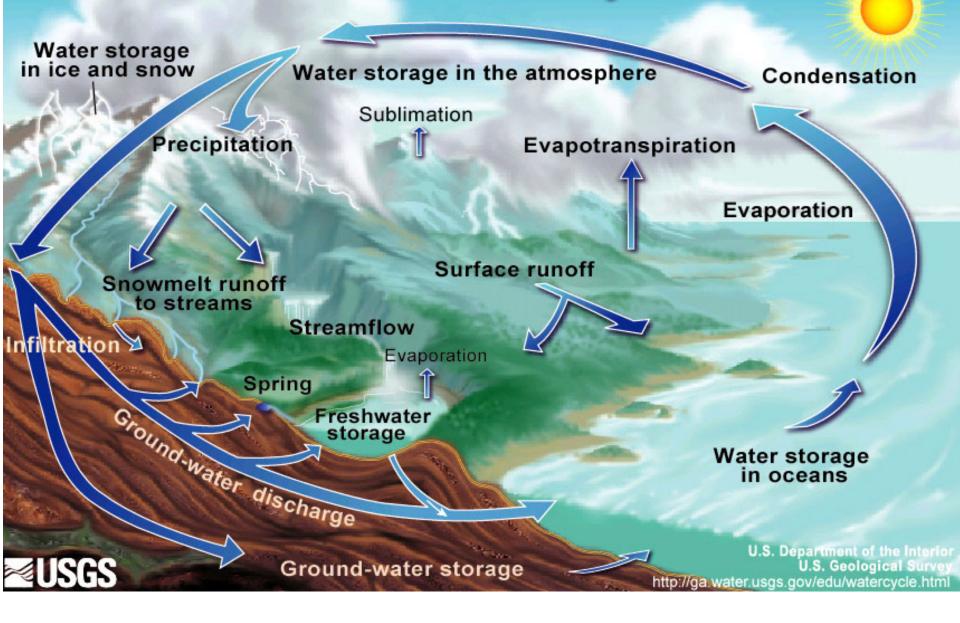
- Existing
- Under construction
- Planned

•

Water - our most valuable asset but ...

- In many places we can't assess
 - How much we have
 - Where it is
 - Who owns it
 - What it is fit for
 - How much we will have
 - Where it will be
- We certainly can't yet share information in a useful timeframe
 - In particular given the complexity of water cycle

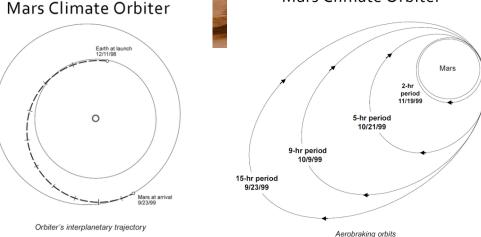
The Water Cycle



Why is it important to coordinate?



Mars Climate Orbiter

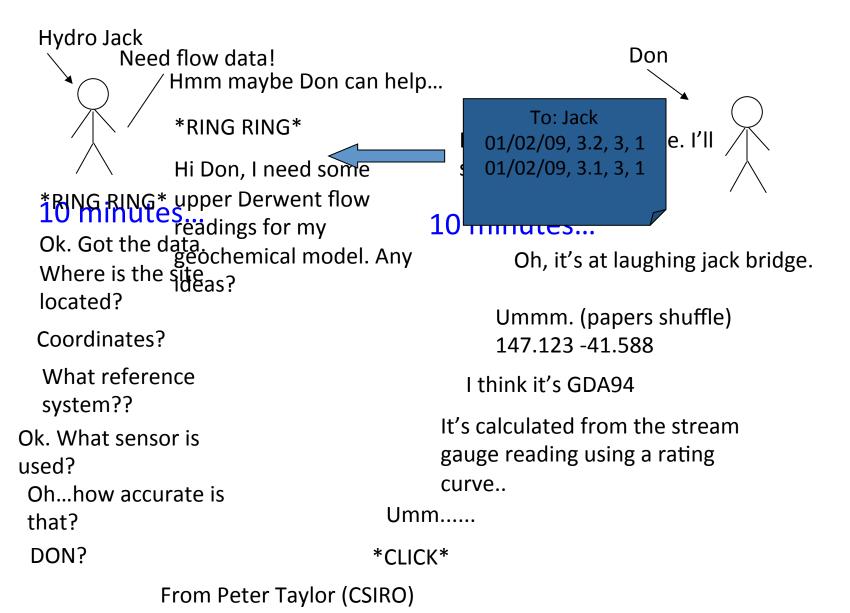


 The orbiter was taken within 57 km of the surface where it likely disintegrated

Why?

 The flight system software used metric units (Newtons); software on the ground used the Imperial system (pound-force, or lbf)

A common situation in hydrology...





WaterML2 Adoption

porate site at www.kisters.d

& Events Contact Us



UNITED STATES

7777 Greenback Lane, Suite 209 Citrus Heights, CA 95610-5800

WaterML 2.0 Adopted as an Official OGC Standard

The results of a public voting process have confirmed





The process of standardizing WaterML 2.0 started in 2009/2010 with a "Harmonization Paper", followed by a "Specification Document" in 2011/2012. As an active long-term supporter of open standards KISTERS has provided both its unique global experience in hydrology, as well as specific expertise in time series data management during the development and testing phases of WaterML 2.0.

Now that WaterML 2.0 has been chosen as a standard other regional formats that have been developed, like the Water Data Transfer Format (WDTF) in Australia, the WaterOneFlow and WaterML 1.0 developed by the Consortium of Universities for the Advancement of Hydrologic Science, Inc. (CUAHSI) in the U.S., the xHydro standard in Germany, and the EA XML standard in the United Kingdom may soon be replaced. "In my opinion, the standardization of WaterML 2.0 is truly a watershed event that stands to define the future of data sharing in the hydrology community", says Fuest.

se tin	0.3.0	CUAHSI WaterML	07-041r1	DP
		Harmonising Standards for Water Observation Data - Discussion Paper	09-124r2	DP
Us Wa		WaterML 2.0 - Timeseries - NetCDF Discussion Paper	12-031r2	DP

End of theme 1, start on HDWG

Interoperability: Definitions

 "the ability of two or more systems or components to exchange information and to use the information that has been exchanged"

IEEE Standard Computer Dictionary. A Compilation of IEEE Standard Computer Glossaries: 610. IEEE, New York (1991)

 "the capability to communicate, execute programs, or transfer data among various functional units in a manner that requires minimal knowledge of the unique characteristics of those units"

ISO/IEC 2382:2001. Information Technology Vocabulary — Fundamental Terms

Interoperability is about agreements

- <u>Technical</u> agreements cover formats, protocols, security systems, so that messages can be exchanged.
- <u>Semantic</u>: Content agreements cover the data and metadata, and include semantic agreements on the interpretation of the information.
- Organizational agreements cover the ground rules for access, preservation of collections and services, payments, authentication, etc.

Also adopted by IDABC: European Interoperability Framework for pan-European eGovernment Services. European Commission, Luxembourg (2004), now introducing political and legal level (here subsumed in organizational).

Benefits of open standards

- Prevents a single group from controlling a standard
- Facilitates competition
- Stimulates innovation
- Customers benefit from not being locked into a particular supplier
- Lower costs and learning curves for equipment, software development, tools, training, and maintenance

But: learning and applying standards initially adds a layer of complexity and coordination

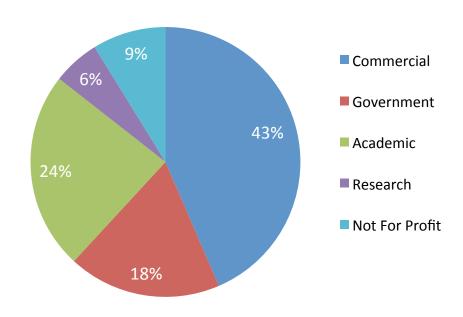
- an important consideration for researchers working on term-limited projects, especially smaller projects

OGC Snapshot



- A Voluntary Consensus Standards Organization, founded in 1994.
- 480 members
- 38 adopted standards
- Hundreds of product implementations
- Broad user community implementation worldwide
- Alliance partnerships with 30+ standards & professional orgs

OGC Membership Distribution



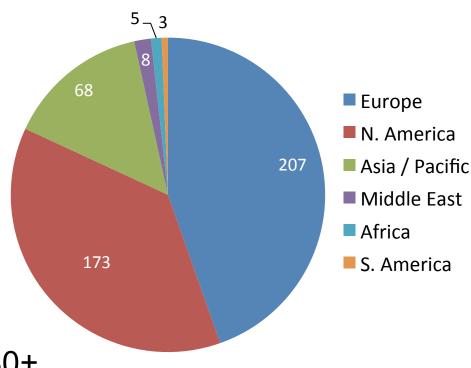


OGC Snapshot



- A Voluntary Consensus Standards Organization, founded in 1994.
- 480 members
- 38 adopted standards
- Hundreds of product implementations
- Broad user community implementation worldwide
- Alliance partnerships with 30+ standards & professional orgs

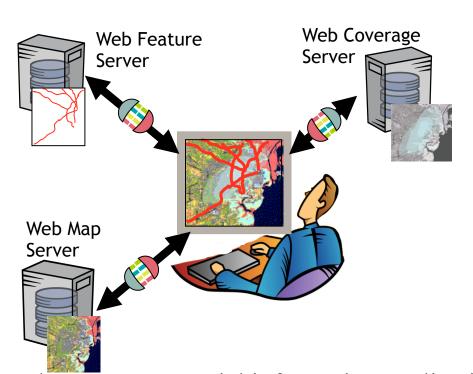
OGC Membership Distribution





OGC Web Services (OWS)

Just as http:// is the dial tone of the World Wide Web, the **geospatial web** is enabled by OGC standards:



Web Map Service (WMS)

Web Map Tile Service (WMTS)

Web Feature Service (WFS)

Web Coverage Service (WCS)

Catalogue (CSW)

Geography Markup Language (GML)

KML

GeoSMS

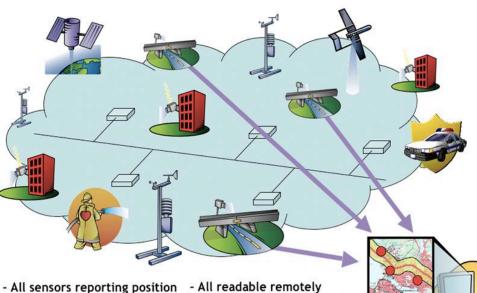
Others...

Relevant to geospatial information applications: Critical Infrastructure, Emergency Management, Weather, Climate, Homeland Security, Defense & Intelligence, Oceans Science, others



OGC Sensor Web Enablement Standards

Discovery and tasking of sensor assets, and the access and application of sensor observations for enhanced situational awareness



Sensor Model Language (SensorML)

Observations & Measurements (O&M)

Sensor Planning Service (SPS)

Sensor Observation Service (SOS)

Catalogue Service

Sensor Alert Service (SAS)

PUCK

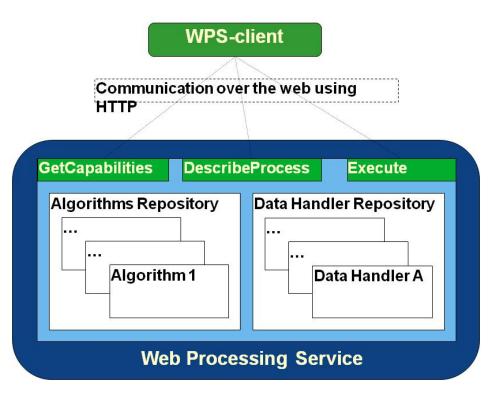
- All connected to the Web
- All with metadata registered
- Some controllable remotely



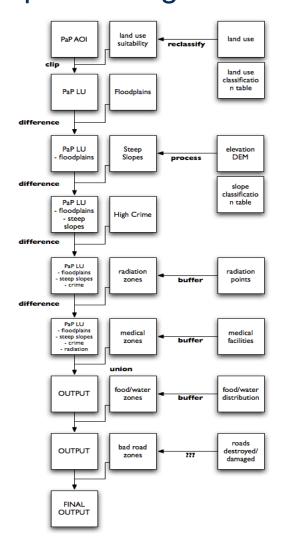
Geospatial Processing, Analysis, Workflow

Web Processing Service – WPS

- OGC Web Service access to algorithms
- Change detection, coordinate transformation, modeling and simulation...



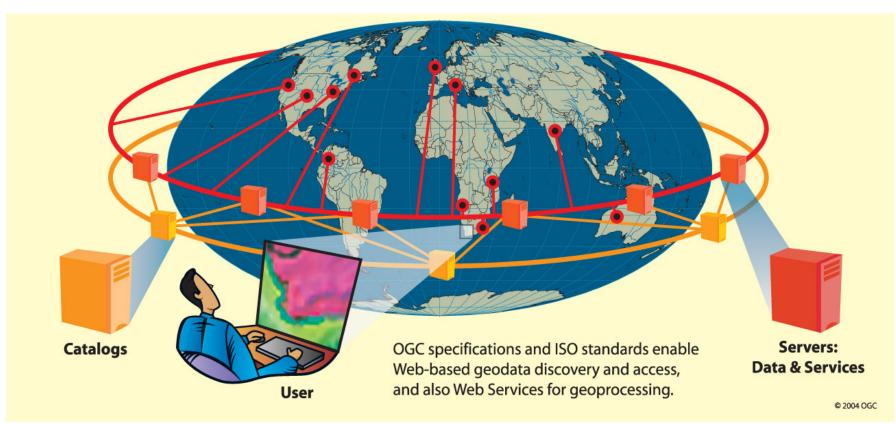
Geoprocessing Workflow





The OGC vision is global





Composed of many collaborating organizations... authoring and publishing open standards for geospatial interoperability



700+ implementing and certified products



http://www.opengeospatial.org/resource/products

1) Select a specification

City Geography Markup Language (CityGML) Encoding Standard v.1.0.0

OpenGIS City Geography Markup Language (CityGML) El Standard 1.0.0

2) Jump to Organization - [-

Bentley Systems Inc.

Product Name OGC Spec

Bentley Map GML 2.1.2, GML 3.1.1, GMLsf 1.0.0,

v8i 1.0.0

ESRI

Product Name OGC Spec

WMS 1.3.0, WMS 1.1.1, WMC 1.0, V

ArcGIS 9.3 WCS 1.1.0, WCS 1.0, SLD 1.0, GML

CAT CS/W 2.0.1, CAT 2.0.2

interactive instruments GmbH

Product Name OGC Spec

XtraServer
3.2

GML 2.1.2, WMS 1.1.1 (compliant 1.0, Filter 1.0, GML 3.1.1, Filter 1.1, 3.2.1, CityGML 1.0.0, UTDS-CityGML

1) Select a specification

Sensor Observation Service v.1.0.0

OpenGIS Sensor Observation Service 1.0.0

2) Jump to Organization - -

1Spatial Group Ltd

Product Name OGC Spec

OSCAR

Sensor

Observation SOS 1.0.0, SensorML Corr 1 1.01, OM 1.0

Service

(SOS) 1.0.0

52 North

Product Name OGC Spec

52N OX- WMS 1.1.1, WMS 1.1, WMS 1.0, WCS 1.1.1 c1, WCS 1.0.0, SPS 1.0.0,

Framework SOS 1.0.0, SAS 0.9

52N Sensor

Observation OM 1.0, OM Sampling 1.0, SensorML 1.0.0, SensorML Corr 1 1.01, SOS



‡

Type

Server

Type

Client

昌 Print

Home > Products > Cloud Services

Search

Cloud Services



DigitalGlobe's Cloud Services unlock the power of the world's largest online library of highresolution Advanced Ortho imagery. With our extensive suite of OGCcompatible web services combined with our powerful hosting infrastructure, we can instantly deliver current, high-resolution imagery and geospatial information to desktops, portals, intranets and mobile devices around the world.

DELIVERING THE WORLD

With our Advanced Ortho Content Programs, our powerful



EASIER INTEGRATION

Satellite imagery and metadata are accessible through a full

Resources

Supporting Documents



Cloud Services

Overview of DigitalGlobe's Cloud Services 3.0

> Download

Web Services Available

Web Feature Service (WFS)

- · Query against metadata
- · Real-time coverage display

Web Map Service (WMS)

- Natively interoperable with most GIS Software
- Images generated upon request per specification

Web Map Tile Service (WMTS)

- Image tiles delivered rapidly
- Renders rapidly for panning and zooming

Web Coverage Service (WCS)

- Easily download GeoTIFFs and other formats
- Perform multispectral analysis and visualization

The World Meteorological Organization (WMO)



'... the UN system's authoritative voice on the state and behaviour of the Earth's atmosphere, its interaction with the oceans, the climate it produces and the resulting **distribution of water resources**'

'... facilitates the free and unrestricted exchange of data and information, products and services in real- or near-real time ...'



International Standardization for Water Data

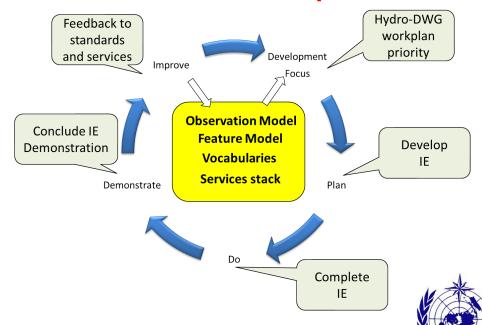


- Hydrology Domain Working Group
 - standards for water data: WaterML 2.0 suite
 - organizing Interoperability Experiments (IEs)
 focused on different sub-domains of water

Chairs:

- Ilya Zaslavsky (USA)
- Tony Boston (Australia)
- Silvano Pecora (Italy)

Iterative Development





Hydrology DWG: History



March 2007

- CUAHSI (Consortium of Universities for the Advancement of Hydrologic Science, Inc.) submits WaterML as OGC Discussion Paper (Zaslavsky, Valentine, Whiteaker)
- September 2007 (Canberra, Australia)
 - Agreed time was right to establish working group to move forward
- 2008
 - Much discussion between CSIRO, CUAHSI, OGC and WMO
- 2009: OGC/WMO Hydrology Domain Working Group is formed
- 2010-2014: regular meetings, annual workshops, IEs
- 2009: Groundwater IE
- 2010: Surface Water IE
- 2011: WaterML 2.0 SWG established
- 2012: OGC® has adopted the OGC WaterML 2.0 Part 1: Time Series Encoding Standard as an official OGC standard



Adoption by WMO



- November 2012
 - Adoption of Resolution 3 at CHy-14
 "Proposed adoption of WaterML 2.0 as a standard"

CHy... decides to commence a process...that could see the potential adoption of the WaterML 2.0 as a WMO standard for information exchange managed by WMO (supported by the WMO/OGC Memorandum of Understanding), and to register this standard as a joint WMO/ISO standard...







HDWG: Selected Past Achievements



- WaterML 2.0 Part 1: Time Series as an official OGC standard
- HY_Features: Common Hydrological Feature Model as an OGC Discussion Document
- WaterML2.0 part 2: Ratings, Gaugings and Sections as an OGC Discussion Document
- WaterML-WQ an O&M and WaterML 2.0 profile for water quality data as an OGC Best Practice Document
- Sensor Observation Service 2.0 Hydrology Profile" as an OGC Best Practice Document
- Numerous IE's (Groundwater, Surface Water, Water Quality...)
- Participation in several cross-domain projects (AIP's, OWS's, CHISP-1 (Climatology-Hydrology Information Sharing Pilot), etc.
- Platform for showcasing numerous activities in the water and related domains
- …list too long to mention everything





HDWG: tasks ahead



- Standard development WaterML2.0 part 2: Ratings, Gaugings and Sections
- Standard development RiverML (Channel and floodplain geometry)
- Standard development GroundWaterML2
- Standard development WaterML2-WaterQuality
- Standard development HY_Features
- Get IE's going (Forecasting, gaugings and ratings...)
- Addressing a Water Suite of standards
- Many other issues waiting for champions to take them further...





Suite of Water Information Standards

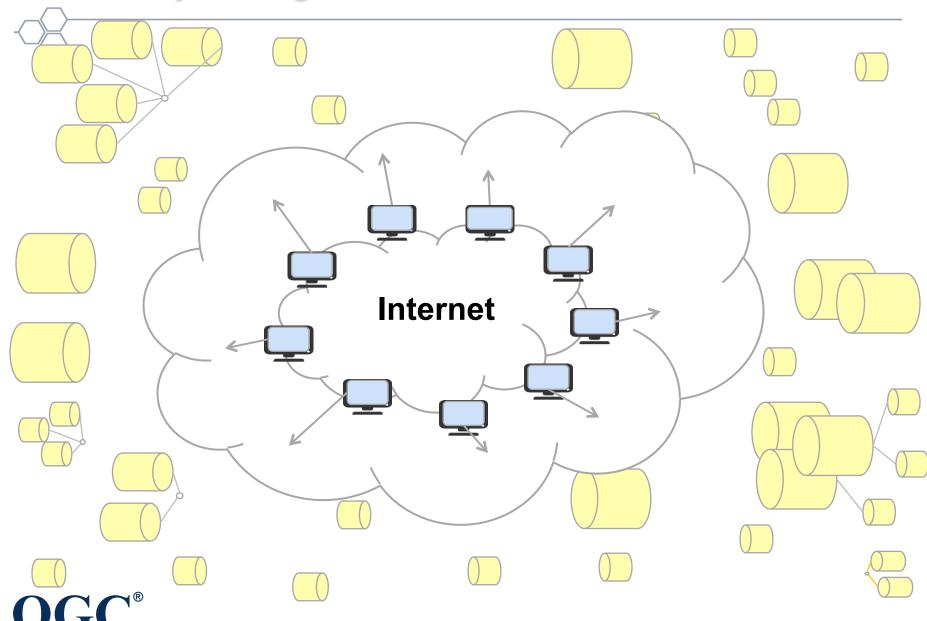
HIC-11 Tutorial: Standardization of Water Data Exchange WMO/OGC Hydrology Domain Working Group Irina Dornblut, GRDC of WMO at BfG New York, CCNY, August 16, 2014



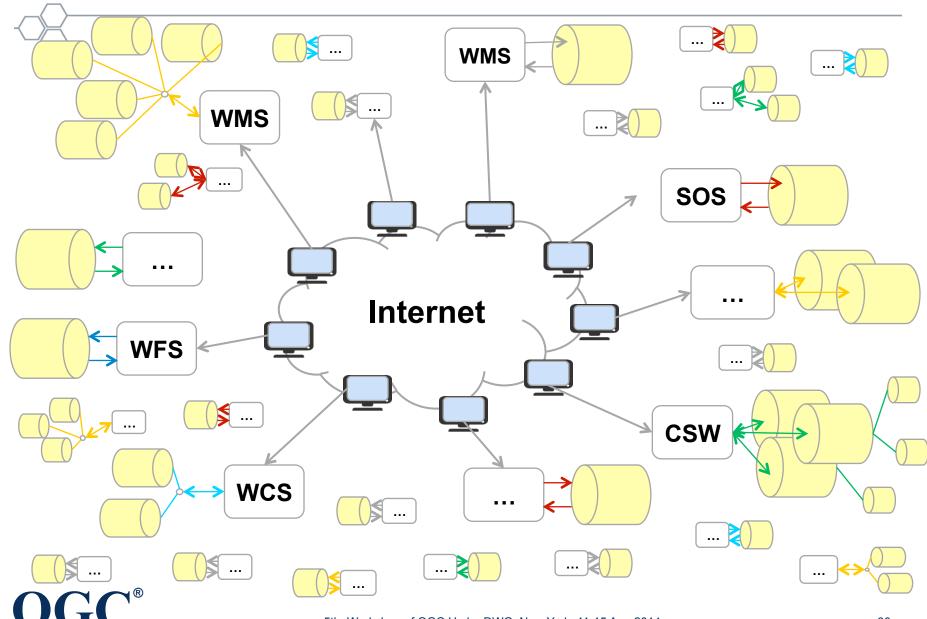




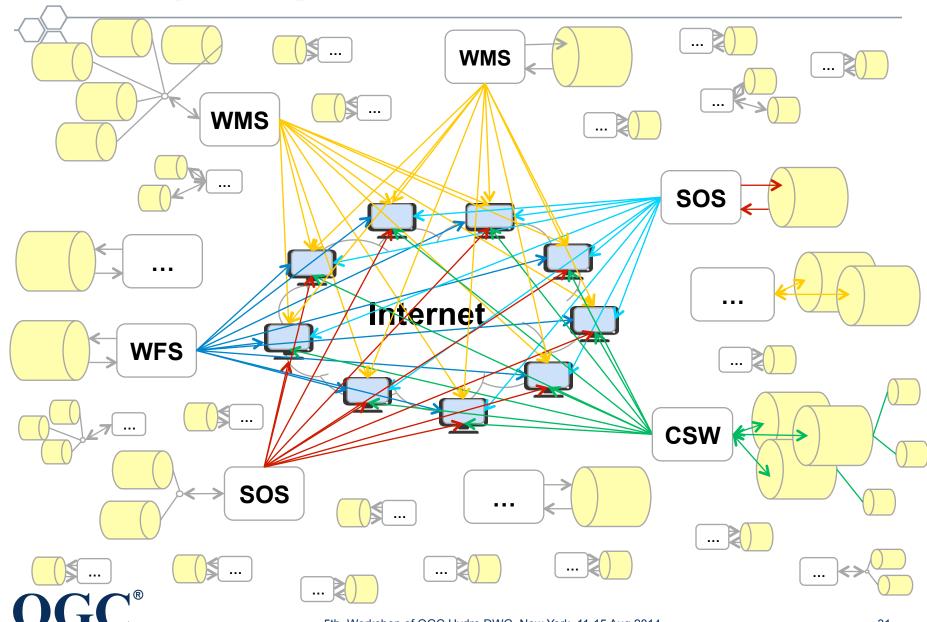
Hydrologic data via Web services



Hydrologic data via Web services



Hydrologic data via Web services



Suite of Water Information standards

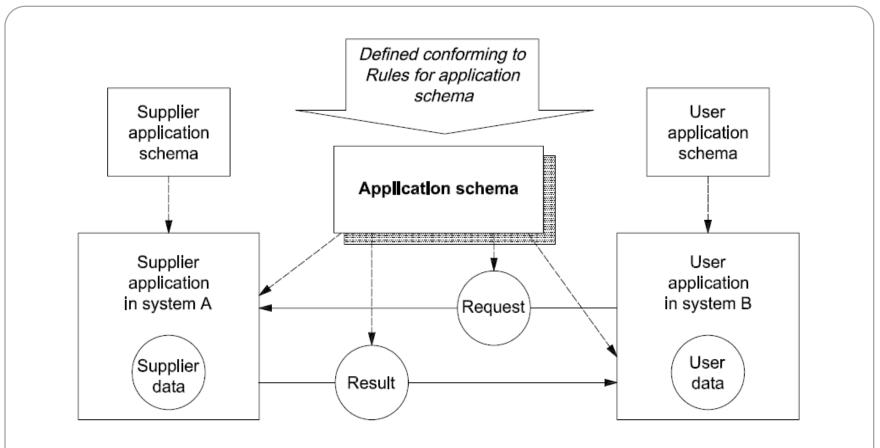


Interoperability ensures communication among Web services.

Rules for application schemas ensure interoperability between technical systems.



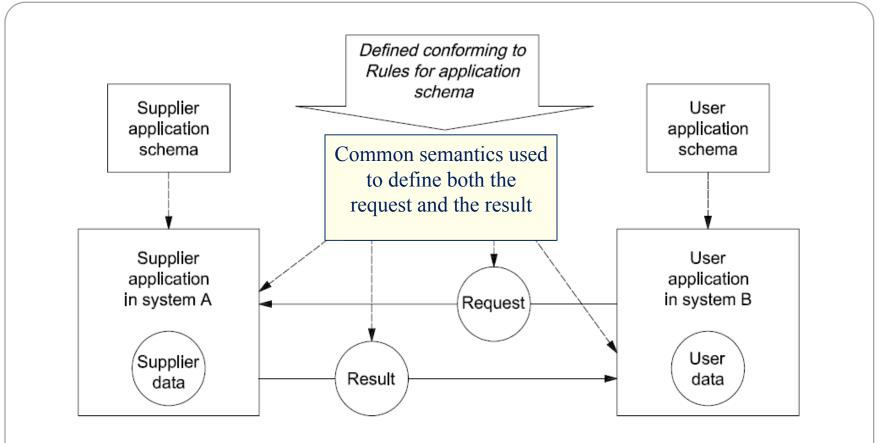
Water data exchange using standards



NOTE The unbroken lines show the flow of data. Broken lines denote the role of the application schema on the data interchange.

[ISO19109:2005, Figure 2]— Data interchange by transactions

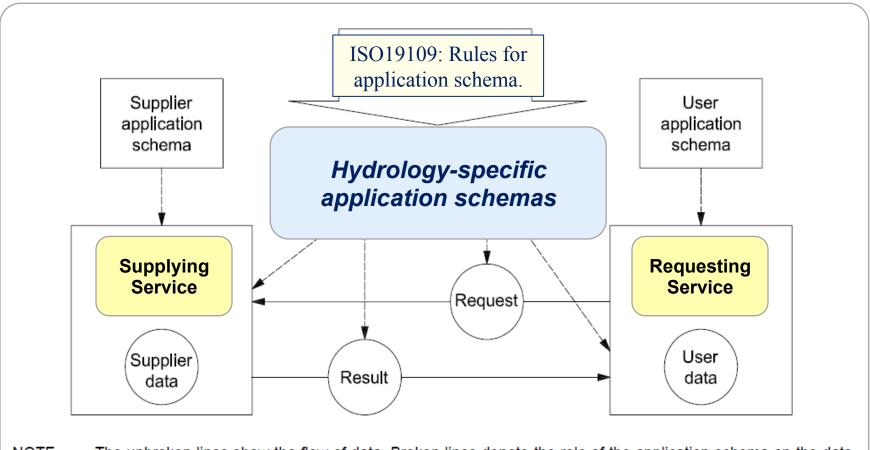
Water data exchange using standards



NOTE The unbroken lines show the flow of data. Broken lines denote the role of the application schema on the data interchange.

[ISO19109:2005, Figure 2]— Data interchange by transactions

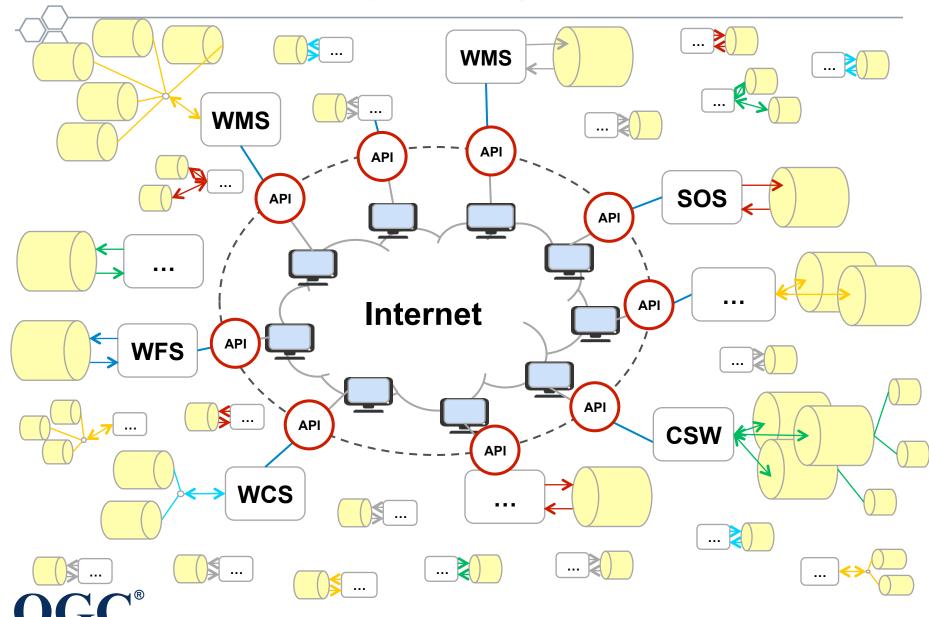
Water data exchange using standards



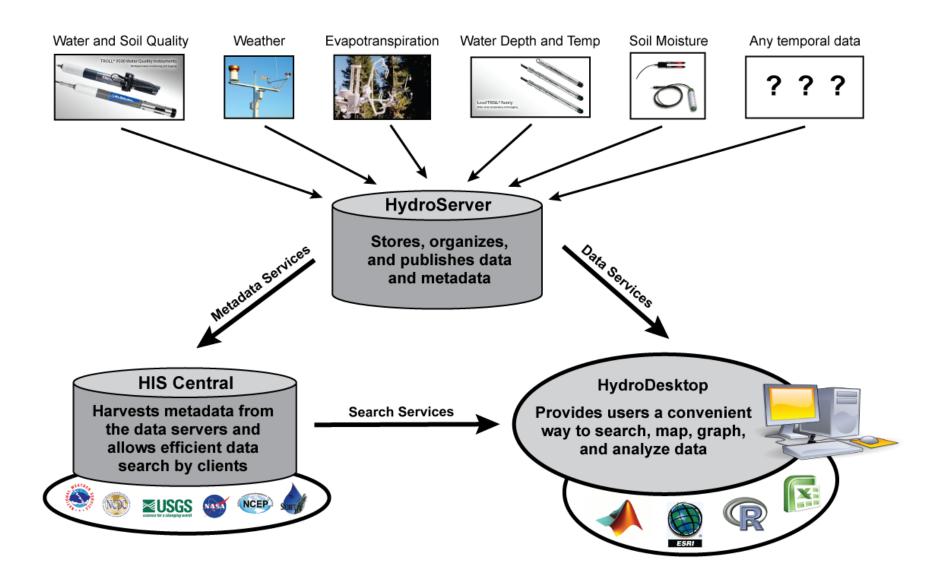
NOTE The unbroken lines show the flow of data. Broken lines denote the role of the application schema on the data interchange.

[ISO19109:2005, Figure 2]— Data interchange by transactions

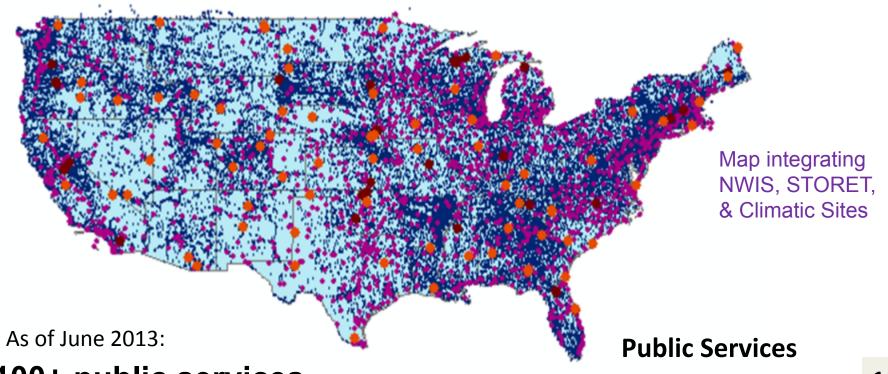
Technical interoperability of Web services



CUAHSI Hydrologic Information System



HIS HydroCatalog Content



100+ public services 32.8 thousand variables

3.25 million sites

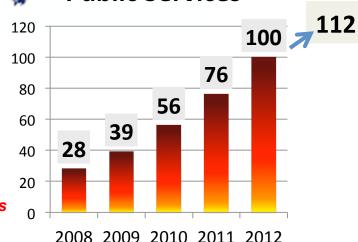
28.9 million series

Referencing 340+ billion data values

Available via GetValues requests

Available via HISCentral

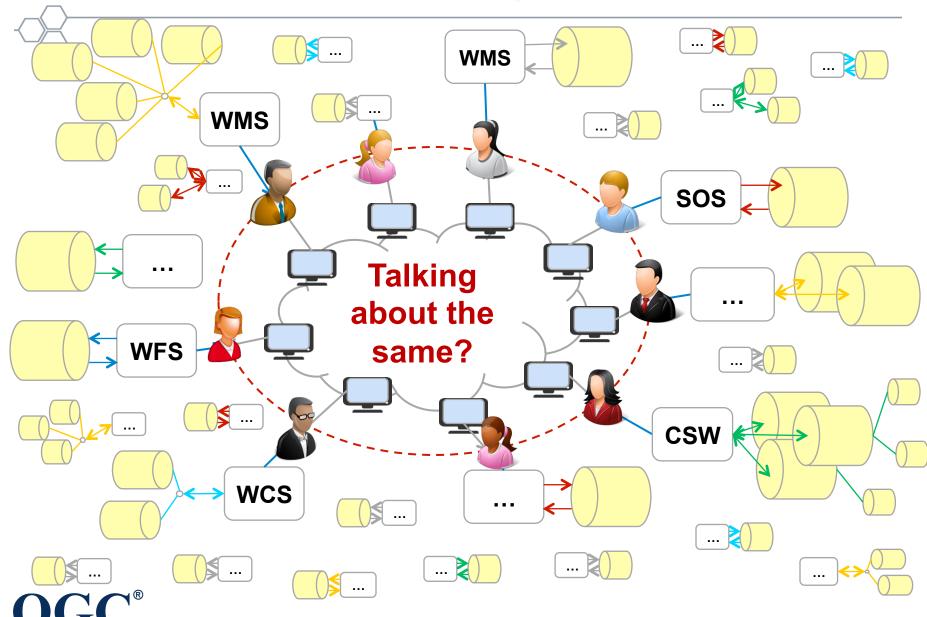
discovery services

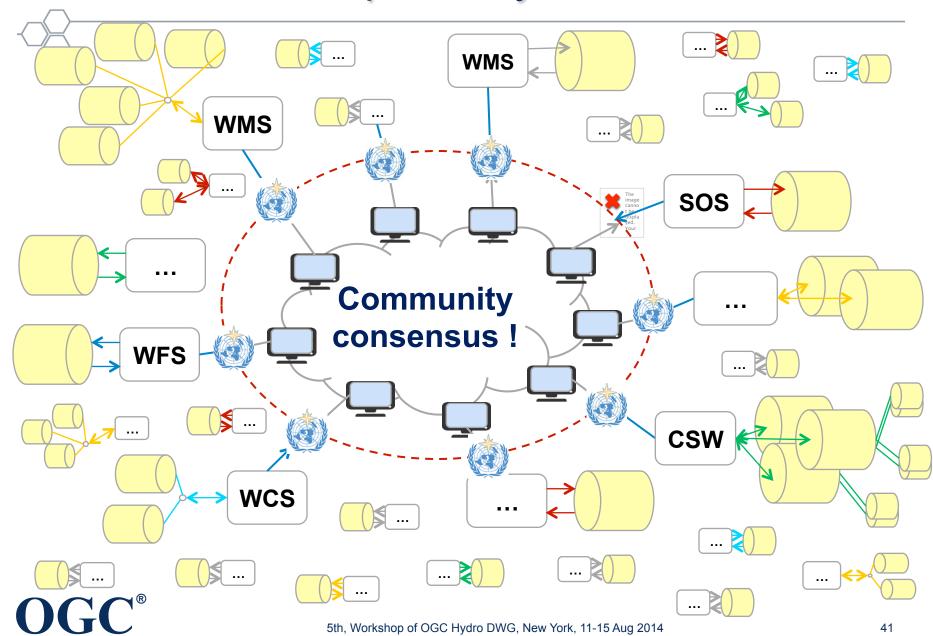


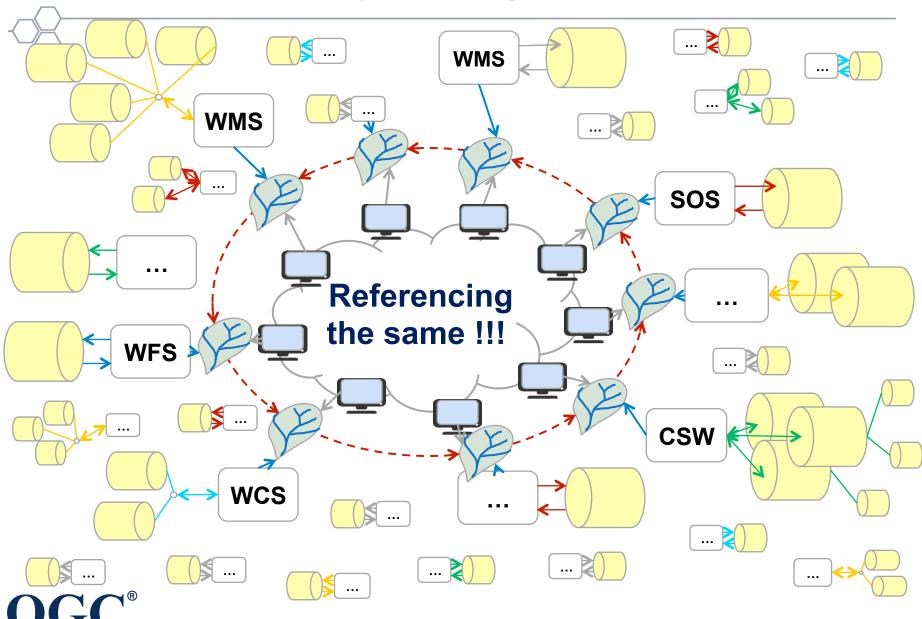
Semantic interoperability between Web services

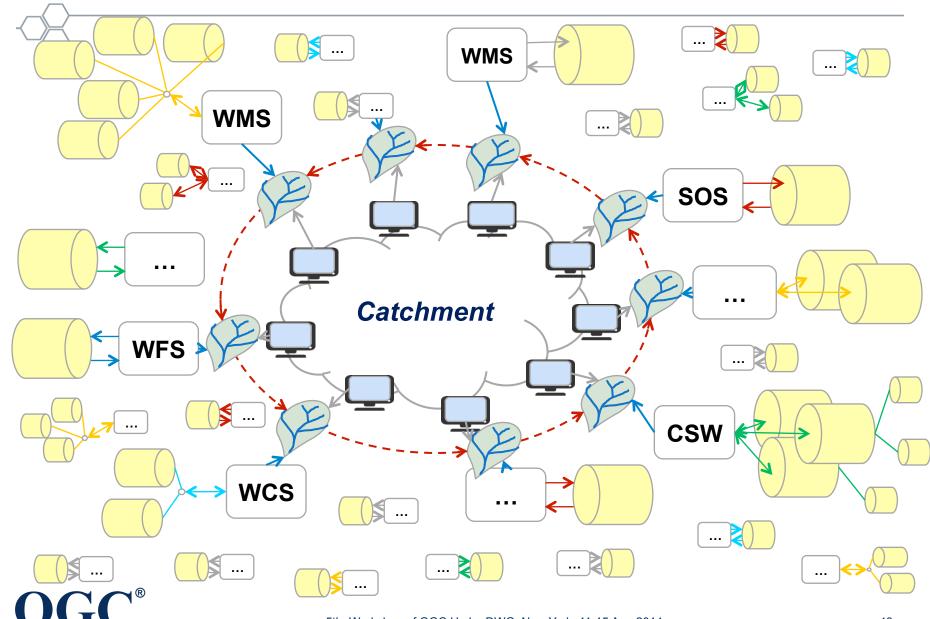
- Concerned with understanding the context and correct meaning of exchanged information
- Requires representation of context (e.g. ontology) and concepts (terms)
- May be achieved by linking references
 - Referencing shared concepts within a community / domain
 - Using an reference model / ontologies as mediator between conceptually overlapping perspectives (mapping semantics)
 - Mapping corresponding ontologies, assisted by a set of formal mapping rules based on the description logic of Semantic Web (semantic mapping)











Suite of Water Information standards



Consensus on context and meaning requires mediation by referencing shared concepts.

Common concepts mediate among overlapping concepts and multiple representations.



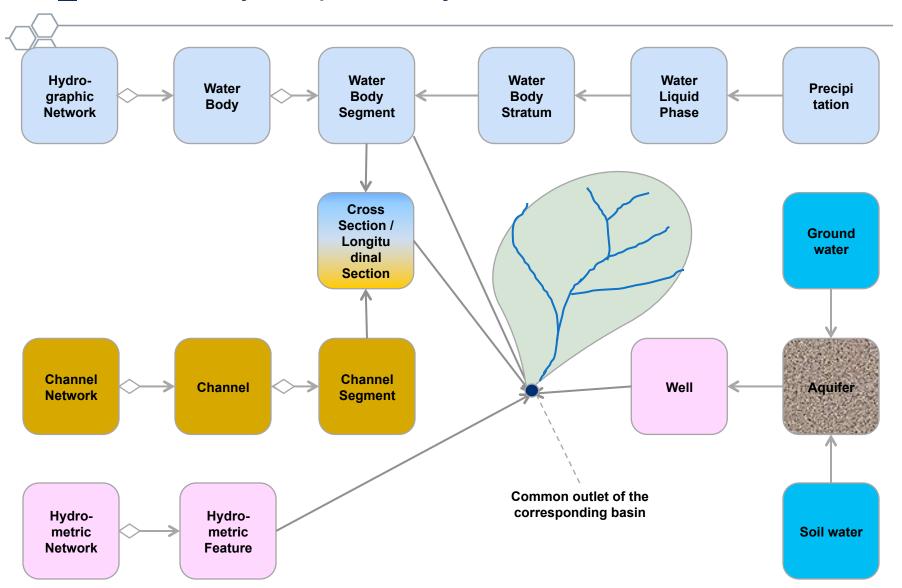
HY_Features, common hydrologic feature model

HY_Features, common hydrologic feature model

- describes major components of the hydrosphere and their fundamental relationships incl. a segmentation of watercourses,
- concepts which reflect hydrologic significance and network connectivity,
- compatible concepts, based on definitions endorsed by WMO-CHy (documented in "WMO International Glossary of Hydrology")

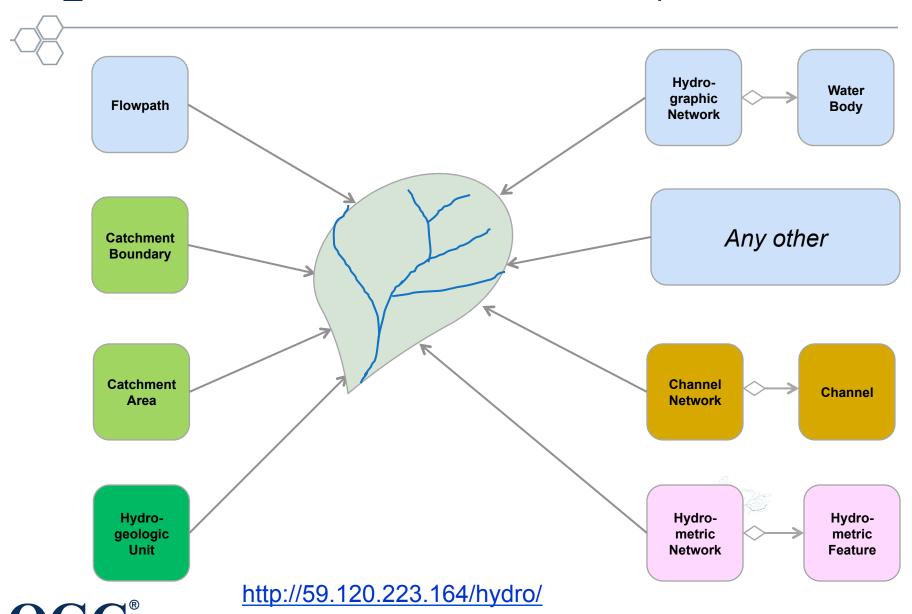


HY_Features, hydrosphere objects related to the basin outlet

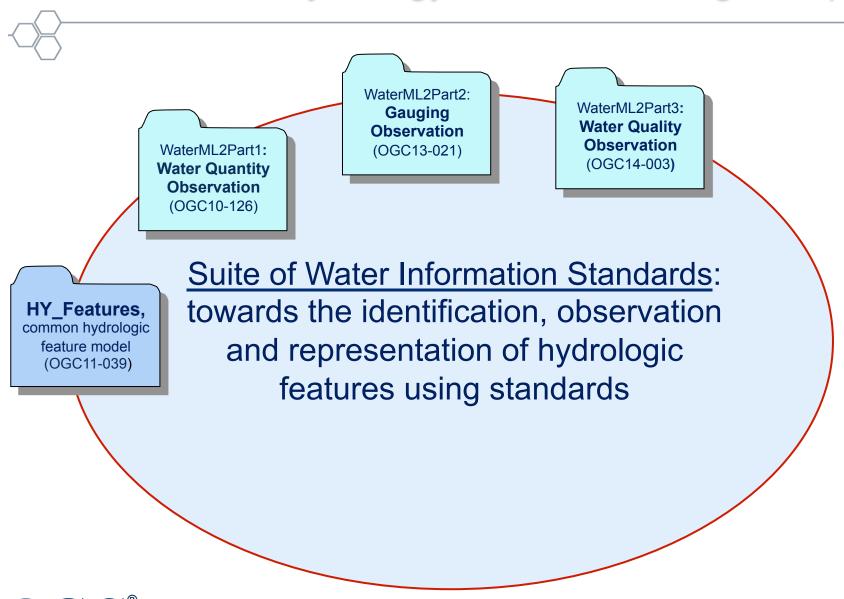




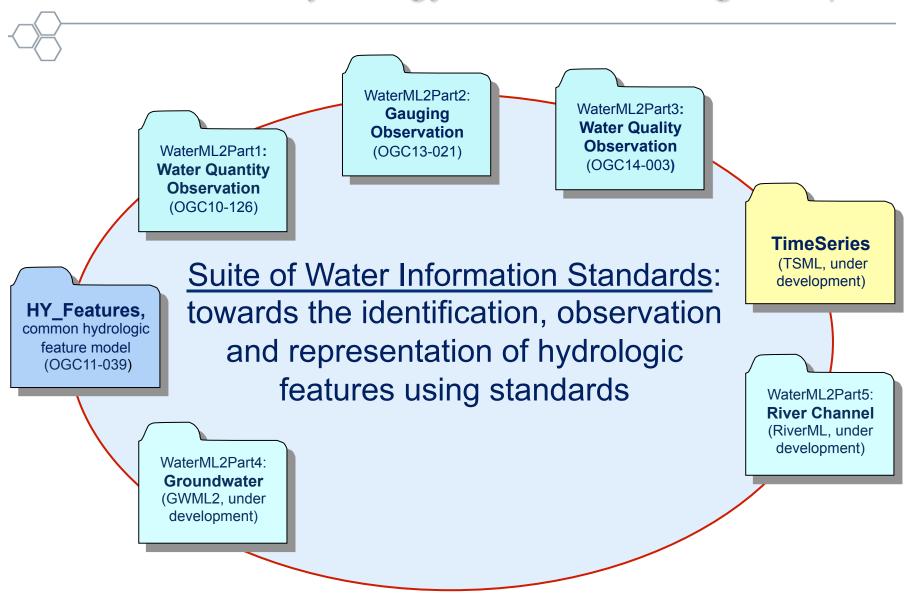
HY_Features associates datasets with the represented basin



WMO/OGC Hydrology Domain Working Group

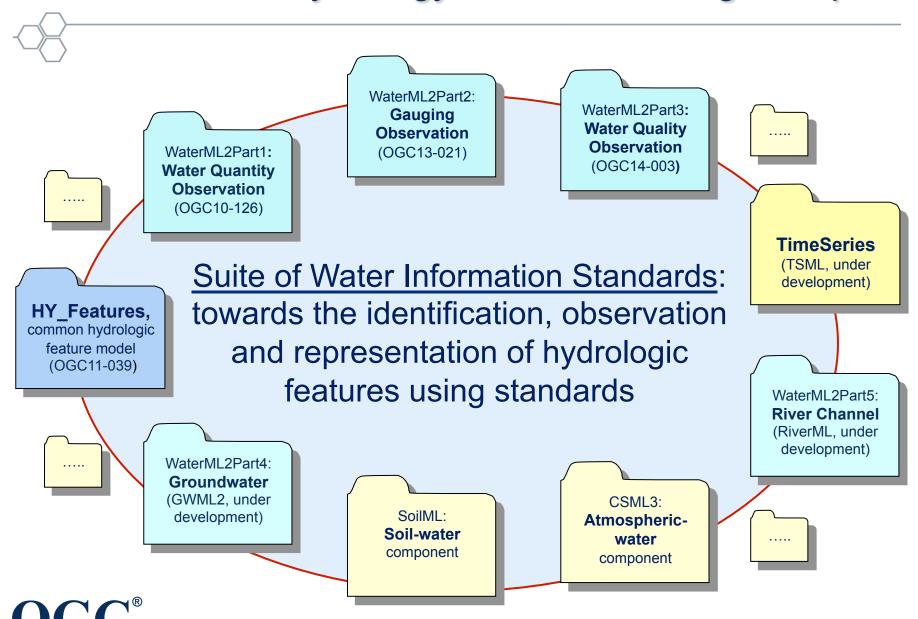


WMO/OGC Hydrology Domain Working Group





WMO/OGC Hydrology Domain Working Group



WaterML2.0 (Optional Section)



WaterML2.0 is an interoperability contract that facilitates data exchange.

- A conceptual model expressed in UML
- An XML schema



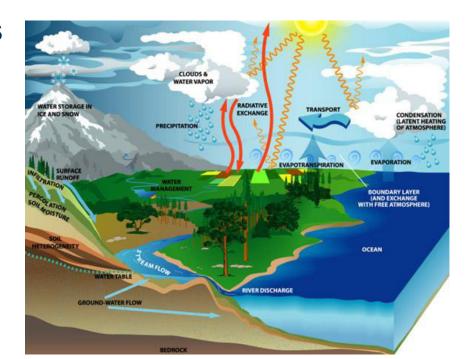
Hydrological observations



Broad categories of water observations

- 1. In-situ, fixed observation style
- 2. In-situ, manual observations
- 3. Ex-situ, complex processing observations
- 4. Remote-sensed observations
- 5. Complex data products

Generally, the more complex the process of making the measurement, the less likely it is to be available as a continuous observation.





Concepts harmonisation



- Discussion Paper
 - Harmonising Standards for Water Observation Data (OGC 09-124r2)
- Considered data exchange formats from:
 - 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
 - Sandre Surface Water Quantity exchange
 - OpenMI
 - FEWS PI



O&M Feature mapping



Hydrological term	ISO19156 – Observations & Measurements
Monitoring station, gauging station, site	SF_SamplingPoint
Borehole, observation well, river profile	SF_SamplingCurve
River cross-section	SF_SamplingSurface



XML schema overview



- Collection
- MonitoringPoint
- ObservationProcess
 - ObservationMetadata
- Timeseries
 - MeasurementTimeseries
 - MeasurementTimeseriesMetadata
 - CategoricalTimeseries
 - TimeseriesMetadata
- TimeValuePair
 - TVPMetadata
 - MeasurementTVP
 - TVPMeasurementMetadata
 - CategoricalTVP



Time series data



Is it all about the dots on a graph?





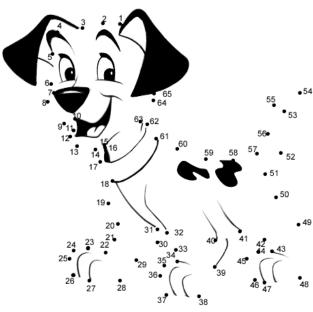
Discrete observations



 The discrete observation is an accurate observation of a feature at a particular point in time.

 The observation is a stand alone item, it is not related to other results.

Typically ex-situ analysis





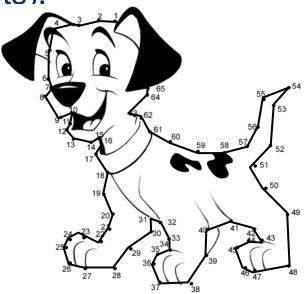
Continuous observations



 Results are collected with a time resolution at which it is appropriate to consider the record to be a continuous representation of the observed phenomena.

 The high data density allows users to fill in the gaps between the observations (dots).

Time series data is not about the dots, its about the lines.





Time series point metadata



- quality
- nilReason
- comment
- qualifier
- processing
- source
- censoredReason
- accuracy
- interpolationCode
- aggregationDuration



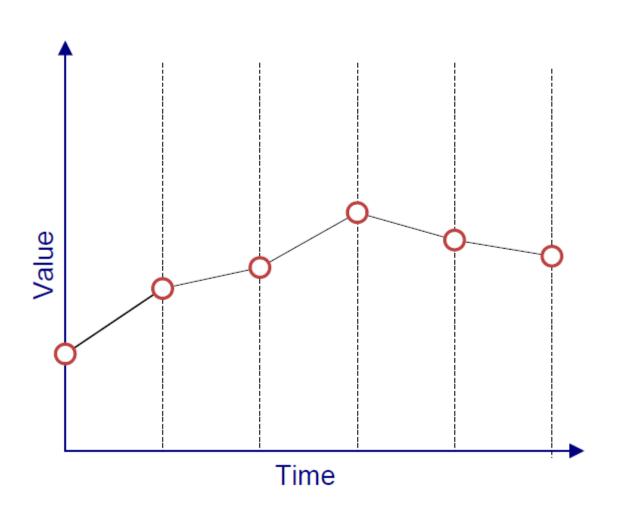
Interpolation code



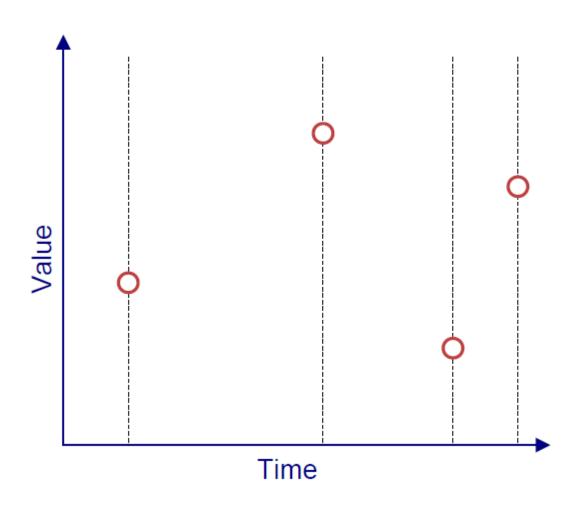
- Its all about the lines....
- Continuous/Instantaneous
- Discontinuous
- Instantaneous total
- Average in preceding interval
- Maximum in preceding interval
- Minimum in preceding interval
- Preceding total
- Average in succeeding interval
- Succeeding total
- Minimum in succeeding interval
- Maximum in succeeding interval
- Constant in preceding interval
- Constant in succeeding interval
- Statistical



Continuous/Instantaneous



Discontinuous





The following may not be suitable for all users.

<It contains="explicit"> material < It/>

```
<wml2:point>
 <wml2:MeasurementTVP>
   <wml2:value xsi:nil="true"></wml2:value>
   <wml2:metadata>
     <wm12:TVPMeasurementMetadata>
       <wml2:qualifier>
         <swe:Quantity definition="http://www.example.com/sensors/lower threshold">
           <swe:description>Lower limit for sensor</swe:description>
           <swe:uom code="m"/>
           <swe:value>1.0</swe:value>
         </swe:Quantity>
       </wml2:qualifier>
       <wml2:censoredReason xlink:href="http://www.opengis.net/def/nil/OGC/0/BelowDetectionRange"</pre>
         xlink:title="Below threshold of sensor"/>
     </wm12:TVPMeasurementMetadata>
   </wml2:metadata>
 </wml2:MeasurementTVP>
</wml2:point>
```

```
<wml2:samplingFeatureMember>
    <wml2:MonitoringPoint gml:id="monitoring-point.1">
      <gml:name codeSpace="http://www.csiro.au/">Deddington
      <sam:sampledFeature xlink:href="http://csiro.au/features/rivers/south esk/deddington"</pre>
xlink:title="Deddington"/>
      <sams:shape>
        <gml:Point gml:id="location deddington">
          <gml:pos srsName="urn:ogc:def:crs:EPSG::4326">-41.814935 147.568517 /gml:pos>
        </gml:Point>
      </sams:shape>
      <wml2:verticalDatum xlink:href="urn:ogc:def:crs:EPSG::5711" xlink:title="Australian height</pre>
datum"/>
      <wml2:timeZone>
        <wml2:TimeZone>
          <wml2:zoneOffset>+11:00</wml2:zoneOffset>
          <wml2:zoneAbbreviation>AEDT</wml2:zoneAbbreviation>
        </wml2:TimeZone>
      </wml2:timeZone>
    </wml2:MonitoringPoint>
  </wm12:samplingFeatureMember>
```

vocabulary



Assessing standards compliance (Optional)



Community nventory of **E**arthCube Resources for Geoscience Interoperability

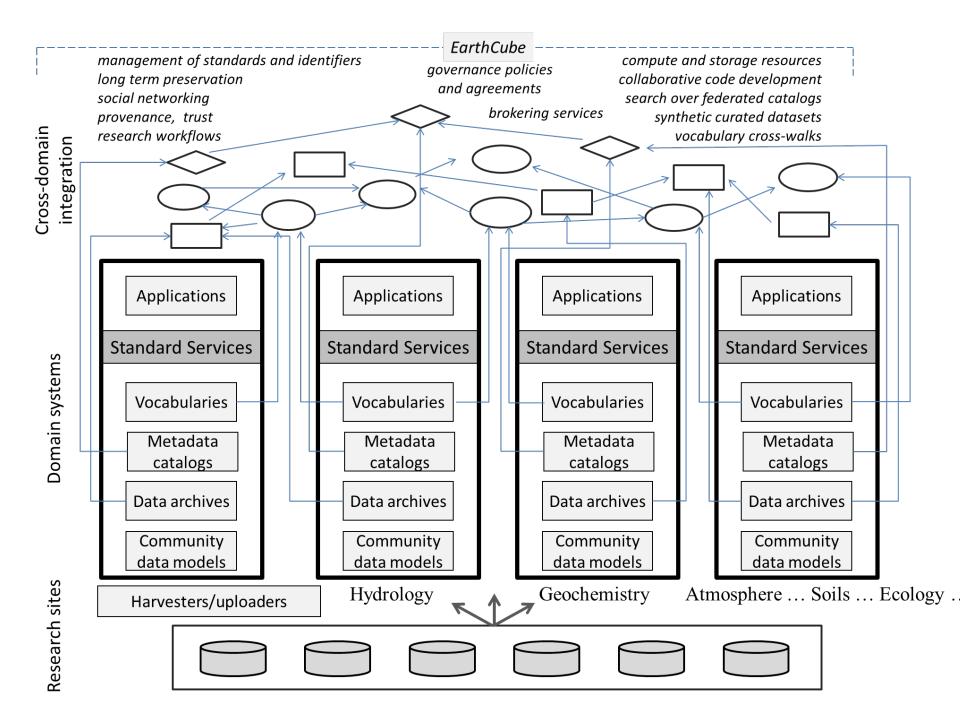
CINERGI



data discovery is the most often cited issue in executive summaries on the EarthCube web site

Goals

- Large inventory of high quality information resources across disciplines, with traceable provenance, usable across EarthCube research scenarios:
 - datasets, catalogs, vocabularies, information models, services, process models, repositories, etc.
- Make it open to the community
- Organize it to enable search and integration across domains and linking between information objects
 - Plus links between resources, people/organizations, publications, models, workflows, software, activities, etc.



Interoperability Readiness

Catalog Metadata			
Has a data listing			
Uses minimal metadata standard, such as Dublin			
Core			
Uses metadata standard, such as FGDC, or INSPIRE			
Catalog Search			
Search Interface			
Search API, not following a standard			
Complies with Opensearch API			
Complies with OGC CSW API		Informati	
Catalog Harvest	CO	Unspecifi	
	Has a data listing Uses minimal metadata standard Core Uses metadata standard, such as Catalog Search Search Interface Search API, not following a standard Complies with Opensearch API Complies with OGC CSW API	Has a data listing Uses minimal metadata standard, such as Core Uses metadata standard, such as FGDC, or Catalog Search Search Interface Search API, not following a standard Complies with Opensearch API Complies With OGC CSW API Catalog Harvest	

	Vocabulary – Control and Access
V1	Uses controlled terminology
V2	Community Managed Terminology
V3	SPARQL
	Vocabulary Representation
	Listing of terminology, such as web
T1	pages
T2	Uses ontology or SKOS

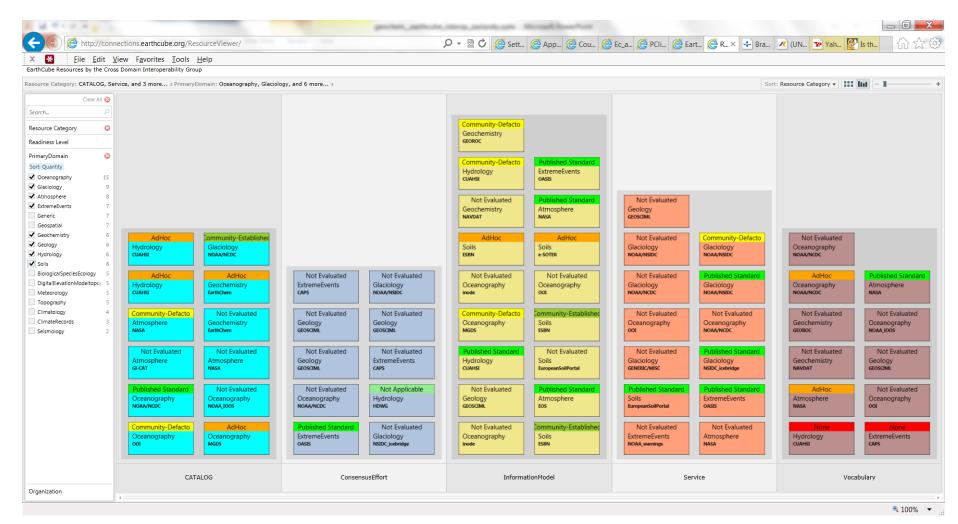
H2		OAI API	
Н3		OGC CSW API	
	Da	ta Access API	
A1	Bulk download		
A2	Sta	Static URL	
А3	Web Service		
Data Query API			
Q1	Sin	nple query subset	
Q2	Co	mplex query	
Q3	Pro	cessing Subset	

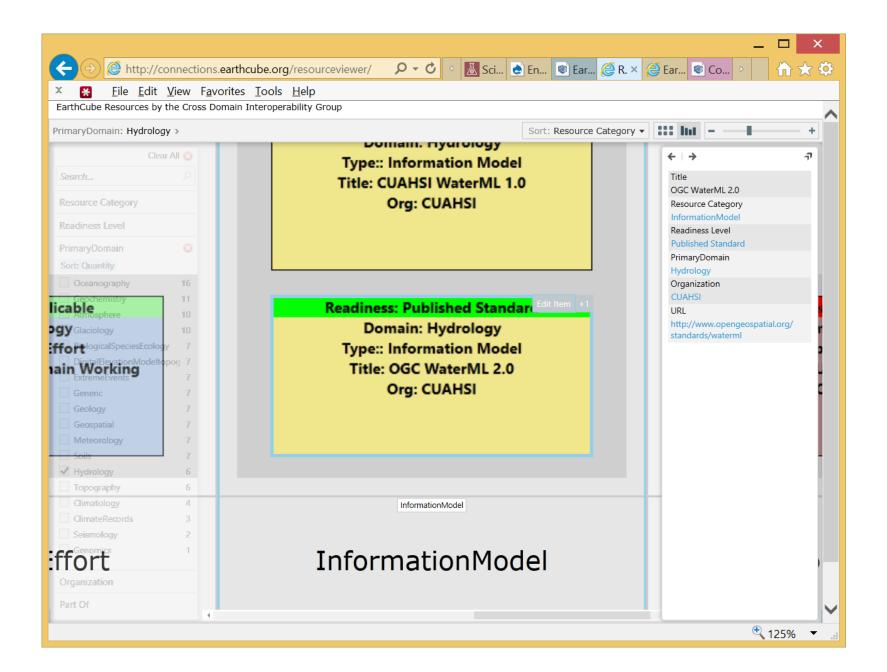
Has a harvest API

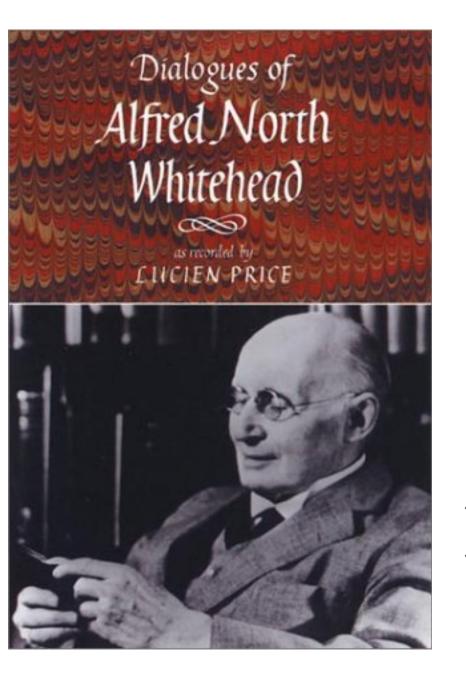
H1

	Information Model Conceptual
CO	Unspecified
C1	Domain/Conceptual Model using UML
C2	Domain/Conceptual Model using UML based on OGC or ISO standards
	Information Model as XML
X1	XML Format. Schema may not be specified
X2	Xml Schema
	Information Model as SQL
S1	Provides an SQL Schema

Community curated inventory and readiness assessment







"Civilization advances by extending the number of important operations which we can perform without thinking of them..."

Alfred North Whitehead.
An Introduction to
Mathematics (1911), ch. 5