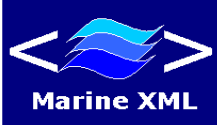




ICES Science Conference 2006



Semantic interoperability: A Goal for Marine Data Management

Roy Lowry

**British Oceanographic Data Centre
SeaDataNet Project and IOC MarineXML**

Luis Bermudez and John Graybeal

**Monterey Bay Aquarium Marine Institute
Marine Metadata Interoperability Project**





Presentation Overview



SeaDataNet



Marine XML



- Introduction
- Development of Syntactic Interoperability
- GF3 and Semantic Interoperability
- How to Achieve Semantic Interoperability (Ideal World)
- Vocabulary Governance
- How to Achieve Semantic Interoperability (Real World)
- Conclusions





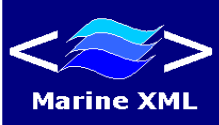
Introduction



- **Interoperability is the ability to share data from multiple sources as a common resource from a single tool**
- **Interoperability has four levels (Bishr, 1998)**
 - **System – protocols, hardware and operating systems**
 - **Syntactic/Structural – loading data into a common tool (reading each others' files)**
 - **Semantic – understanding of terms used in the data by both humans and machines**



Introduction



- **Interoperability has been the elusive Nirvana for oceanographic data management since IODE was established in 1961**
- **Modern Semantic Web knowledge management technologies are bringing this dream within reach**
- **But without a foundation of published well-defined, well-governed and versioned base vocabularies the technology will achieve nothing**



Syntactic Interoperability Development



- There was a time when data were exchanged primarily on 9-track magnetic tape
- These had different byte lengths, word lengths, character encodings and physical file constructions
- Listing a file took an expert 2-3 days
- After which, there was the format to decipher



Syntactic Interoperability Development



SeaDataNet



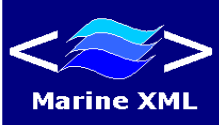
Marine XML



- Now standards have become established making life easier
 - The Excel spreadsheet is a universally accepted currency for biological and chemical oceanographic data exchange
 - NetCDF, particularly standardised variants like CF over protocols such as OpenDAP gives interoperable high volume data exchange
- Scientists without programming skills can visualise data from multiple sources in minutes
- But can they understand what it means?



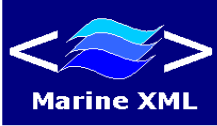
GF3 and Semantic Interoperability



- **GF3 was an oceanographic data exchange format developed by IODE in the 1980s**
- **It was primarily targeted at system, structural and syntactic interoperability issues, particularly 9-track tape harmonisation**
- **But it also addressed semantic interoperability through the introduction of code tables covering countries, platforms, instruments and especially measured phenomena**



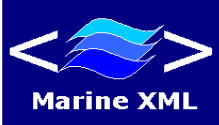
GF3 and Semantic Interoperability



- **Code tables generally not taken up by science, but were adopted by the IODE data centre network**
- **A study of their use, and abuse, by this community can provide valuable lessons for those developing semantic interoperability today**



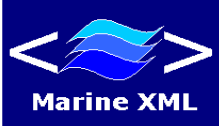
GF3 and Semantic Interoperability



- **Content Governance**
 - Initially the GF3 code tables had strong content governance through IODE GETADE
 - This fell apart when it was realised that effort put into development of vocabularies distributed by paper publication was wasted
 - **Consequently, code tables became dated**
 - * IOC Country codes last revised in 1981
 - * A lot has happened in the world since then
- **Breakdown of content governance leads to a proliferation of local vocabularies**
- **Like finches on Galapagos, these all start the same, but then diversify**



GF3 and Semantic Interoperability

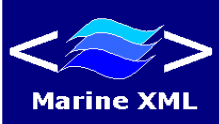


- **GF3 code tables included no definitions of terms or the entity represented**
- **Words have different meanings to different people and even common words like ‘cruise’ gained multiple meanings**
- **Code (key) syntax was deemed important**
 - **Not scalable – there is a limit to how many meaningful combinations can be squeezed from 4 or even 8 bytes**
 - **Diverts attention onto labels and away from the real issue of meaning**



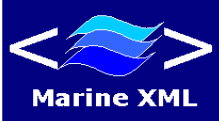
GF3 and Semantic Interoperability

- **Vocabularies were used to ‘fix’ data models**
 - **One-to-one relationships turned into one-to-many by creating vocabulary entries that are themselves lists**
 - * Examples from a ‘data type’ vocabulary
 - Moored profiling CTD + acoustic current meter
 - Turbulence/dissipation data
 - **Adding metadata elements by mixing entities**
 - * Examples from the same ‘data type’ vocabulary introducing ‘platforms’
 - Multiple data types – aircraft
 - Multiple data types – ship
 - **Makes mapping much more difficult or even impossible**

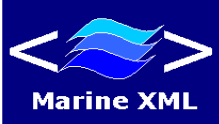




Semantic Interoperability (Ideal World)



- **Comments directed at those building marine data systems without the millstone of legacy data**
 - **Take note of the work of others, learn their lessons and don't reinvent wheels**
 - **Avoid free text in metadata except for abstracts**
 - **Populate fixed fields from existing vocabularies wherever possible**
 - * Spend time looking rather than inventing
 - * Build any new lists in collaboration and establish governance
 - **Always allow for vocabulary development**
 - * Use on-line declarations wherever possible
 - * Don't implement local copies without synchronisation
 - **Ensure your metadata always accompanies your data**



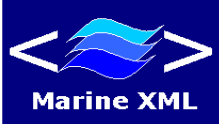
Semantic Interoperability (Ideal World)

- The guidance being given is to adopt vocabulary standards wherever possible
- This has been preached but not practiced in the past because
 - **Interoperability was not top of some agendas**
 - * Data hoarding preferred to data sharing
 - * Interoperability difficulties provided an excuse not to share
 - **Local needs trumped community efforts**
 - * Community development takes time and effort
 - * Quick fixes needed at the coal face
 - **The standards never really delivered**
 - * Vocabulary content lacked adequate governance
 - * Vocabulary updates slow, confusing and unreliable





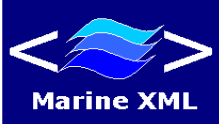
Vocabulary Governance



- For vocabulary standards to be adopted we need content governance (the mechanism for making decisions on vocabulary change) that delivers:
 - * Change with no compromise in vocabulary standards or internal consistency
 - * Change on a timescale that matches the needs of the scientific community
- The CF community have come a long way towards achieving this for their Standard Names
- Work is beginning to establish an infrastructure for the development of vocabularies within the EU SeaDataNet project and the IOC data centre network
- NASA's SWEET project is developing a forum for discussing the management and development of environmental ontologies



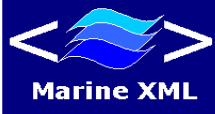
Vocabulary Governance



- For vocabulary standards to be adopted we need technical governance (vocabulary storage, maintenance and serving) that delivers:
 - * Convenient access to up to date vocabularies
 - * Clear, rigorous vocabulary versioning
 - * Version history through audit trails
 - * Maintenance that doesn't break user systems
 - * Semantically neutral keys to represent entries
- **BODC has achieved this with the NERC DataGrid Vocabulary server**
- **ICES is coming along fast with the proposed upgrades to its RECO system**



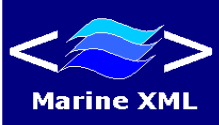
Semantic Interoperability (Real World)



- In the real world we have legacy metadata populated using many similar, but significantly different vocabularies
- The only way to address semantic interoperability is by building maps between the terms of these vocabularies
- Drive and effective tooling to do this is being delivered by the Marine Metadata Interoperability Initiative (MMI)



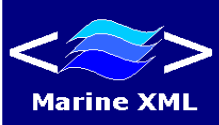
Semantic Interoperability (Real World)



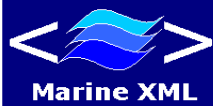
- **MMI (<http://marinemetadata.org>)**
 - **MMI received first funding in 2004 to create advanced, interoperable community-based metadata practices**
 - **MMI has developed sophisticated tools based on Semantic Web technologies to support the vocabulary mapping process**
 - **MMI has just received a 3-year NSF grant to continue to develop semantic interoperability in the marine community**



Semantic Interoperability (Real World)



- **The MMI approach to mapping**
 - **Vocabularies are syntactically harmonised into proto-ontologies using the Voc2OWL tool**
 - **Terms from vocabularies are mapped using the VINE tool resulting in an OWL ontology**
 - **Mapping is based on three relationships**
 - * Broader than
 - * Narrower than
 - * Same as
 - **Gives greater semantic richness than the inferred 'equivalent' relationship used in mappings to date**



VINE - Vocabulary Integration Environment

cf-gcmd.owl

Click on a vocabulary to select it

all none **cf** gcmd

Search for: use REGEX

sea.*(surface|water).*temp find

- sea.*(surface|water).*temp
- water temp -potential -freezing
- water temp -potential
- water temp

integral_of_sea_water_temperature_
 ocean_integral_of_sea_water_tempe
 product_of_eastward_sea_water_vel
 product_of_northward_sea_water_ve

▼ sea_surface_temperature

 ▼ narrowerThan

 ✎ Oceans--Ocean_Temperature

sea_surface_temperature

label : sea_surface_temperature

standard_name : sea_surface_temperature

hasCanonical_Units : K

description : Sea surface temperature is usually abbreviated as "SST". It is the temperature of sea water near the surface (including the part under sea-ice, if any), and not the skin temperature, whose standard name is surface_temperature. For the temperature of sea water at a particular depth or layer, a data variable of sea_water_temperature with a vertical coordinate axis should be used.

type : Standard_Name

Map | Mapping Results

Click on a vocabulary to select it

all none cf **gcmd**

Search for: use REGEX

water temp find

All (Found: 3 Checked: 1)

- Atmosphere--Atmospheric_Water_V
- Hydrosphere--Water_Quality_Water_
- Oceans--Ocean_Temperature--Wat

▼ Oceans--Ocean_Temperature--Water_T

 ▼ broaderThan

 ✎ sea_surface_temperature

Oceans--Ocean_Temperature--Water_Temperature

label : Oceans > Ocean Temperature > Water Temperature

type : TOPIC-TERM-VARIABLE



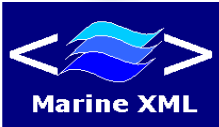
Conclusions



- **Significant scientific benefits result from sharing distributed data and information resources**
- **Semantic interoperability work to date in oceanographic data management has given us:**
 - **A firm foundation of lessons learned from our past efforts**
 - **Technology and vocabulary infrastructure that now deliver the means to build new systems that are semantically interoperable**
 - **The challenge of liberating the vast information resource locked away in isolated stovepipe systems**

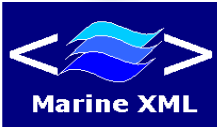


Conclusions



- **Meeting this challenge requires:**
 - **The technology, guidance and community building provided by MMI**
 - **The combined efforts of the oceanographic scientific community to build the Semantic Web resources that will deliver effective semantic interoperability to the benefit of all**

The End



**Thank you for your
attention**

Questions?

