Federation of Earth Science Information Partners
MAKING DATA MATTER

www.esipfed.org

Datatype and Service Ontology

Prepared for: ESIP Semantic Web Cluster
Prepared by: Peter Fox, Cluster lead, Rob Raskin, Luis Bermudez, Howard Burrows, Rahul Ramachandran, Liping Di and Brian Wilson

February 13, 2007, last revision May 2, 2007
White paper number: SemWeb-001

Executive Summary

Objective

To develop a community-based data-type and service-type ontology for use in the application domain of Earth and space sciences.

Goals

To develop formal specification and encodings of these ontologies and publish them along with demonstration end-points and tools.

Solution

Adopt best practices within Earth and Space science community to date and adhere to existing or emerging standards (formal, ad-hoc and community) both generally and within the broader Earth and Space sciences community.

Relevant Organization and standards

- Syntactic efforts
- Schematic efforts
- Convention efforts
- Semantic efforts

W3C: OWL, OWL-S, SWSO, WSMO

OGC: WFS, WMS, WCS

CGI: GML, GeoSciML

ISO (19115, 19119, etc.),

Climate community: CF, COARDS

Data community: OPeNDAP, CDM,

Scope of this effort and terms of reference

Earth Science problems that need to bring together diverse data sources, with processing and analysis programs are increasingly relying on technical approaches such as service oriented architectures (for e.g. web services, often multiple ones in succession) ingesting and outputting a variety of data types (for e,g. floating float numbers from netCDF files as an image onto a geo-spatially referenced local area map). As service and data providers are enabling new and more rapid science progress they are realizing that there are limitations to an approach where data sources and outputs, service inputs and outputs must be matched up manually and checked often the consistency and validity - an approach which quickly becomes unwieldy as data and service diversity grows.

A complementary technical approach is to utilize a semantic ...

We can add a sentence at the end, of the form: This is how this effort may fit in with other science areas.

The need to specific this level of detail for services and data/data-type is built on and arises out of the existence of emerging community ontology efforts for scientific terms, concepts, physical processes, phenomena, domains of applicability, instruments, platforms, data archives, and so on from projects such as SWEET, GEON, VSTO, MMI, NOESIS and GeoBRAIN. These projects have advance our ability to built semantically-enabled smart search, service discovery and data integration.

Data type ontology

What is a data ontology? (Peter Fox contributing)

Ontologies in descriptions that represent agreed domain semantics. Unlike data models, the fundamental asset of ontologies is their relative independence of particular applications, i.e. an ontology consists of relatively generic knowledge that can be reused by different kinds of applications/tasks.

A data model, on the contrary, represents the structure and integrity of the data elements of the, in principle “single”, specific enterprise application(s) by which it will be used. An ontology in general an ontology contains the vocabulary (terms or labels) and the definition of the concepts and their relationships for a given domain. In many cases, the instances of the application (domain) are included in the ontology as well as domain rules (e.g. identity, mandatoriness, rigidity, etc.) that are implied by the intended meanings of the concepts. Domain rules restrict the semantics of concepts and conceptual relationships in a specific conceptualisation of a particular application domain. These rules must be satisfied by all applications that want to use – or “commit to” an interpretation of – an ontology.

Thus, a data-type ontology is a formal set of classes, sub-classes, properties and relation representing data-types as they are used in scientific and engineering applications (representing data and information in files, programming codes and scripts and dynamically generated by internet services).

In general data-types can be split into two broad categories: Atomic (Base) and Derived/Compound (Constructor) which are built on the Atomic types.

From: http://www.w3.org/TR/xmlschema-2/#typesystem

In this specification, a datatype is a 3-tuple, consisting of a) a set of distinct values, called its ·value space·, b) a set of lexical representations, called its ·lexical space·, and c) a set of ·facet·s that characterize properties of the ·value space·, individual values or lexical items.

Value space

[Definition:] A value space is the set of values for a given datatype. Each value in the value space of a datatype is denoted by one or more literals in its ·lexical space·.

The ·value space· of a given datatype can be defined in one of the following ways:

- defined axiomatically from fundamental notions (intensional definition) [see ·primitive·]
- enumerated outright (extensional definition) [see ·enumeration·]
- defined by restricting the ·value space· of an already defined datatype to a particular subset with a given set of properties [see ·derived·]
- defined as a combination of values from one or more already defined ·value space·(s) by a specific construction procedure [see ·list· and ·union·]

value space·s have certain properties. For example, they always have the property of ·cardinality·, some definition of equality and might be ·ordered·, by which individual values within the ·value space· can be compared to one another. The properties of ·value space·s that are recognized by this specification are defined in Fundamental facets (§2.4.1).

Lexical space

In addition to its ·value space·, each datatype also has a lexical space.

[Definition:] A lexical space is the set of valid literals for a datatype.

For example, "100" and "1.0E2" are two different literals from the ·lexical space· of float which both denote the same value. The type system defined in this specification provides a mechanism for schema designers to control the set of values and the corresponding set of acceptable literals of those values for a datatype.

Note: The literals in the ·lexical space·s defined in this specification have the following characteristics:

Interoperability:

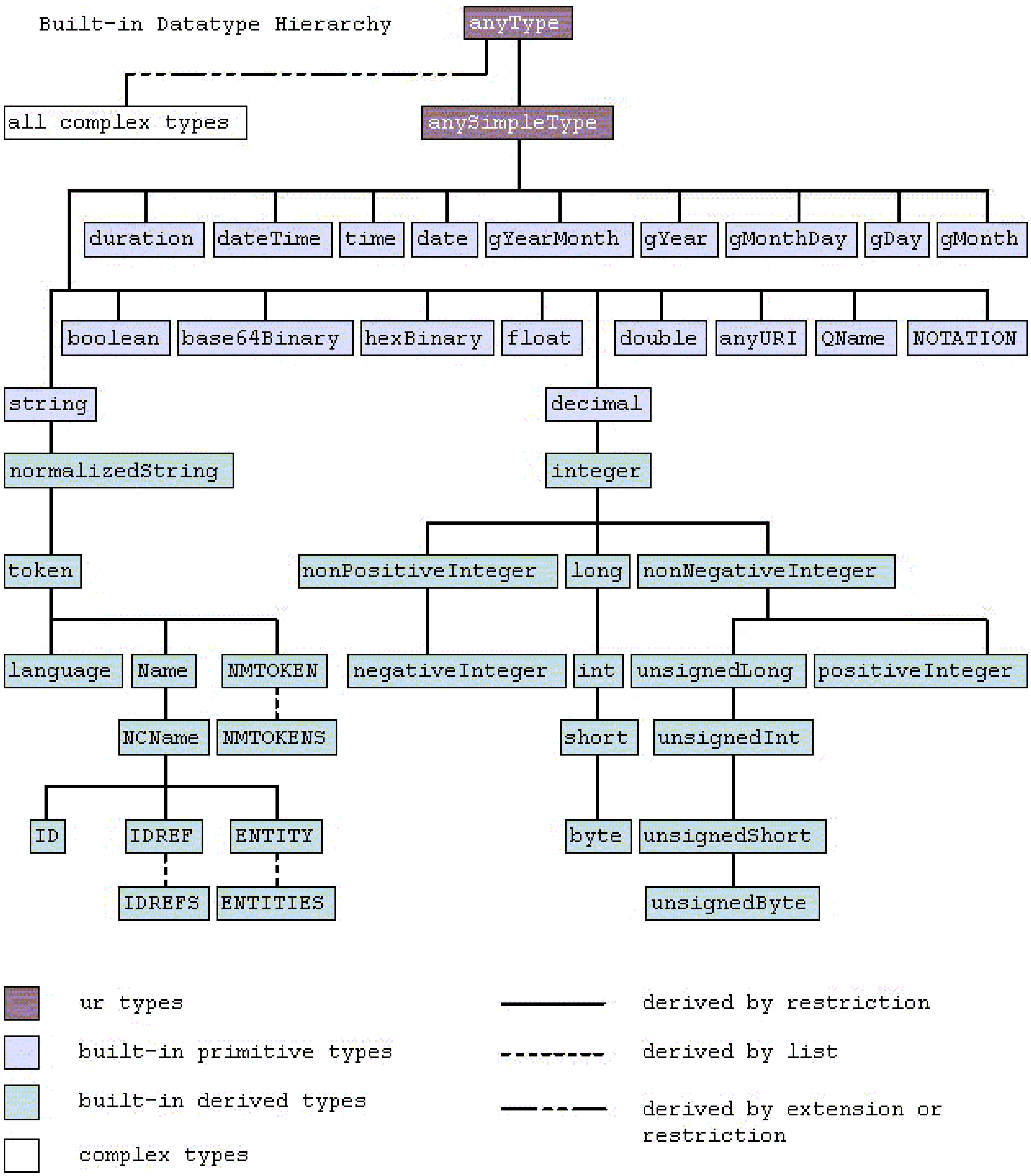
The number of literals for each value has been kept small; for many datatypes there is a one-to-one mapping between literals and values. This makes it easy to exchange the values between different systems. In many cases, conversion from locale-dependent representations will be required on both the originator and the recipient side, both for computer processing and for interaction with humans.

Basic readability:

Textual, rather than binary, literals are used. This makes hand editing, debugging, and similar activities possible.

Ease of parsing and serializing:

Where possible, literals correspond to those found in common programming languages and libraries.



Built-in data-types.
Can also have User-derived datatypes are those ·derived· datatypes that are defined by individual schema designers.

Links to existing work:

see Meersman paper

http://www.w3.org/TR/owl-ref/#Datatype

Links to the watchdog strategies: http://marinemetadata.org/wd, where CF transformation is done.

link to realtime CF ontology:

http://marinemetadata.org/cf

Mediation services at MMI, to query RDF triples and infer statements (ran using SESAME and a MySQL back-end):

http://marinemetadata.org/ontws

What Exists in General (Luis Bermudez contributing)

There are four basic types of ontologies [6]: top-level, domain, task and application ontology. Top level ontologies provide abstract concepts. For example they may define that an event is a temporal thing. Examples that contain general top level concepts are: OpenCyc [3], SENSUS[1], SUMO [12], DOLCE[8], and WORDNET [14]. Domain and task ontologies are a specialization of the top-level ontology. Domain ontologies represented concepts on a specific domain (e.g. earth science). Tasks ontology represented activities (e.g. model-running, interpolating, visualizing).

Application ontologies ties together a domain ontology with providers of information. For example, and ocean information system will internally has it own data base and schemas that need to be represented in an ontology and mapped with a domain ontology. One type of mapping is stating that a concept form the application ontology “is a” type of concept of a domain ontology; often called subsumption relation. Example references of domain and application ontologies are provided in the next section.

What has been done in Earth and Space Science?

SWEET[11] has the most extensive set of concepts, that could qualified as a top-level-domain ontology for earth and space sciences. SWEET includes terms defining earth realms, physical phenomena, physical processes properties, substances, numerics, time space etc. Several initiatives in the Earth and Space domain have created their own ontologies which could be consider as micro-domain ontologies, since they are very project centric. They define top level concepts to represent the knowledge in those projects and facilitate solving specific problems.

In outer space domain the VSTO [7] has developed ontologies describing observatories their instruments and the capabilities of the instruments, and the INAF-OAT Trieste has created concept maps describing a basic foundation for Solar Space Weather [9]. In the Hydrosphere, the Ordnance Survey of Great Britain has an ontology for topographic features [13] and a top hydrologic ontology was design for the CUAHSI project [2] . In the environmental domain Ecoinformatics has a working group to address ontology issues, ECOTERM [4]. In the marine domain the MMI initiative has created ontologies from already existing marine controlled vocabularies and it has been working on sensors and platforms ontologies.[10]. GEON[5] has developed ontologies related to Geology, including properties, time scales and rock classifications.

Problem

Most of them are not extending any upper domain nor any top level ontology. For example none of these ontologies extend SWEET and SWEET doesn't extend any upper level ontology. Need to: build more community to really create a domain ontology, facilitate the use of SWEET. Need to refractor and present the top concepts for earth science. This will reduce the size, but will help reusability and will leave the details for the different subdomains. Improve and develop tools to facilitate collaboration for creating, editing , mapping, versioning and reviewing of concepts and the relations within ontologies.

Another 'technology' to discuss in this context is GRDDL:

- Gleaning Resource Descriptions from Dialects of Languages (GRDDL) - http://www.w3.org/2004/01/rdxh/spec

[1] Bateman, J. A., Kasper, R. T., Moore, J. D. and Whitney, R. A., A General Organization of Knowledge for Natural Language Processing: The Penman Upper Model, USC/Information Sciences Institute, Marina del Rey, CA., 1989.

[2] Bermudez, L. E., ONTOMET: Ontology Metadata Framework, Drexel University, Philadelphia, Pennsylvania, 2004.

[3] Cycorp, OpenCyc Selected Vocabulary and Upper Ontology, 2008, http://www.cyc.com/cycdoc/upperont-diagram.html.

[4] Ecoinformatics Initiative, ECOTERM - Environmental Thesaurus and terminology working group, 2008, http://www.eea.europa.eu/cooperations/eco-informatics.

[5] GEON, 2007, http://www.geongrid.org/.

[6] Guarino, N., Formal Ontology and Information Systems, FOIS'98, Amsterdam, IOS Press, Tento, Italy, 1998, 3-15.

[7] High Altitude Observatory and Scientific Computing Division of the National Center for Atmospheric Research, Virtual Solar-Terrestrial Observatory (VSTO) 2007, http://vsto.hao.ucar.edu.

[8] Masolo, C., Borgo, S., Gangemi, A., Guarino, N. and Oltramari, A., The WonderWeb Deliverable D18, WonderWeb, 2003, http://wonderweb.semanticweb.org/deliverables/documents/D18.pdf.

[9] Messerotti, M., Building a Foundation Ontology for Solar Space Weather 2005, http://www.iono.noa.gr/cost724/Documents/WG1/M_Messerotti.pdf.

[10] MMI, Ontology Team at MMI, 2007, http://marinemetadata.org/ontteam.

[11] NASA Jet Propulsion Laboratory, Semantic Web for Earth and Environmental Terminology (SWEET), 2007, http://sweet.jpl.nasa.gov.

[12] Niles, I. and Pease, A., Towards a Standard Upper Ontology, in C. Welty and B. Smith, eds., 2nd International Conference on Formal Ontology in Information Systems (FOIS-2001), Ogunquit, Maine, 2001, http://www.ontologyportal.org/.

[13] Ordnance Survey, Ordnance Survey Ontologies, 2007, http://www.ordnancesurvey.co.uk/oswebsite/ontology/.

[14] Princeton University, Wordnet: A lexical database for the English Language, 2008, http://wordnet.princeton.edu/.

Service ontology

- review of existing implementations, detail of what is represented and what is needed

Service types:

- dependent variable transform
- independent variable transform
- analog to digital and digital to analog
- sub-setting
- aggregation
- statistics
- algorithms
- visualization

Feature types:

- need to fill this in

Check this link for recent docs: http://www.ai.sri.com/daml/services/owl-s/1.2/

see services docs.

Submissions to W3C

- OWL-S - http://www.w3.org/Submission/OWL-S
- SWSO/F/L - Semantic Web Services Ontology/Framework/Language - http://www.w3.org/Submission/SWSF/
- WSMO/X/L - Web Services Modeling Ontology/Exection/Language - http://www.w3.org/Submission/WSMX/ www.wsmo.org, www.wsmx.org
- SAWSDL - Semantic Annotations for Web Services Description Language - http://www.w3.org/2002/ws/sawsdl/

Schedule

Date	Task
March 2007	Outline of white paper with initial section structure and statement of scope
April 2007	Review of white paper with initial contributions - moved to May
May 2007	Review of white paper with initial contributions
June 2007	Second draft with examples
July 2007	Presentation of status at ESIP Federation meeting, technical review
August 2007	
September 2007	
October 2007	
November 2007	
December 2007	