VIRTUAL ASTRONOMICAL OBSERVATORY

Utypes: current usages and practices

Matthew J. Graham, Caltech/VAO
and the Utypes Tiger Team

Associated Universities, Inc.

AURA

The VAO is operated by the VAO, LLC.

Overview

- Core part of the IVOA Architecture
- Currently lack a formal definition (only Utypes WD)
- Most general usage is an identifier for a concept defined with an IVOA data model, i.e. a label
- More specific usage is as a pointer (parseable identifier) to a data model concept, semantically equivalent to a URI or XPath in XML
- Also related practices about reuse, inheritance, extensibility, etc.
- A consistent (and formal) definition is required for interoperability and implementation reusability

Data Models - definition

- Spectrum 1.1:
  - Standard identifier for a data model field
  - Case-insensitive
  - Of the form “a.b.c.d” where “dots indicate a ‘has-a’ hierarchy”
  - Enumeration syntax for multiple instances same Utype
- Photometry:
  - “ns.a.b.c.d” form
- ObsCore
  - Camel case form with attributes of a class starting with a lower case letter
- Characterization:
  - “Built from the XML Schema representation of the model which already enforces a hierarchical structure”
  - “a.b.c.d” form “based on instance variable paths in the O-O DM”
- Simulation
  - “A pointer into a data model” that “should allow one to uniquely identify a concept in a data model”
  - Specifies a set of rules for deriving Utypes directly from a UML data model
  - Metamodelling with VO-URP

Data Models – inheritance

- Spectrum 1.1:
  - Utypes have the same form except for leftmost element: “we say that SSA inherits the Spectrum model so ‘SSA.’ Utypes overlap with the ‘Spectrum’ ones”

    => b.c.d concept should be consistent across data models
  - Top level use of Spectrum must be denoted using “spec:” namespace
- Photometry:
  - “Access class defined in ObsTAP and inherited from SSA”: PhotometryFilter.transmissionCurve.Access.*

    => field names must be unique across data models
- Extensibility
  - No specific mechanism described in standards

Data Access Layer

- SSA 1.1:
  - Utypes are “pointers to data model elements” and a mechanism to “flatten a hierarchical data model so that all fields are represented by fixed strings in a flat namespace” but not viable for a complex data model
  - Constructed using “embedded period characters to delimit the fields of the Utype” – implies parsable Utypes
  - Defined within a single namespace: component-name.field-name; however, common concepts between Spectrum DM and SSA do not require “spectrum” prefix
  - Case-insensitive
  - Mechanism for having multiple equal Utypes in same file, e.g., when serializing multiple instance of a component data model as a VOTable
  - 2011 survey of Utype practices in SSA response documents showed rather inconsistent results
- TAP:
  - Utypes are columns in TAP_SCHEMA
  - Can be assigned to schemas, tables, columns and foreign keys

Applications

- VOTable:
  - Utypes are an identifier for something in an external data model with the form: “datamodel_identifer:role_identifier”
  - “xmlns convention which specifies the URI of the data model cited” should be used: stc:AstroCoords.Position2D.Value2.C1
  - Attribute of FIELD, PARAM, FIELDref, PARAMref, RESOURCE, TABLE and GROUP since 1.1
- STC in VOTable (IVOA Note):
  - References to space-time coordinate metadata kept in GROUPs with a specific <prefix>DataModel.URI Utype
  - “Abusing XML namespace declaration for … binding data model names to URIs is explicitly discouraged”
  - Defined through XPaths into STC-X instance documents
- SPLAT:
  - Case-insensitive matches suffix patterns in case-insensitive fashion when processing a SSA response
  - Within a spectrum, uses Utypes to locate flux and spectral axes and errors and units (case-insensitive and suffix rules)
  - Case-sensitive matches TABLE elements with “sed:Segment” Utype for VOTable spectra

Registry

- VODataService 1.1:
  - Utypes are “an identifier for a concept in a data model”
  - Utypes on schemas, tables, columns, and foreign keys (xs:token)
- Registry Interface 2.0:
  - Utypes are relative XPaths generated via XSLT operating on XSD
  - Link VOResource data model with columns in a set of database tables
  - Allow a representation of VOResource extensions without having to explicitly model all aspects of them
  - Requires lower-case on ingestion but is otherwise case-insensitive

And the rest…

- Semantics
  - Used in VOUnits to tie quantities with concepts
- VOEvent
  - Used in time series representations to link to concepts in time series data model
- GWS
  - No usage of Utypes (yet?)

Issues from current usages

- Syntax: ns:a.b.c.d vs. a.b.c.d
- Case-insensitive vs. Camel case
- Uniqueness vs. non-uniqueness of a,b,c,d across data models
- Parsable (pointer) vs. non-parsable (label)
- Where to give a data model URI:
  - xmlns
  - prefix:DataModel.URI
  - Not at all