

Registry Ideas

Paul Harrison (JBO/UKSRC)

VOA Interop Northern Autumn 2025



SKAO Regional Centre **United Kingdom**



Introduction

- ✦ Have created a new Publishing Registry Implementation
 - ✦ Uses XML database (BaseX) as the document store
 - ✦ enables XQuery as a query language + makes it easy to store arbitrary XML
 - ✦ Interfaces
 - ✦ Has OAI-PMH interface (obviously)
 - ✦ Management interface to add resources as XML
 - ✦ protected with HTTP basic authentication
 - ✦ can check against Registry schema
 - ✦ GUI generated as a Swagger-UI over the API
 - ✦ Dockerized distribution, configurable with environment variables
- ✦ Will produce a separate fully searchable variant

Possible new standard interfaces

- Search interfaces
 - let's just stick with RegTAP
- retrieval interfaces
 - Something that returns all or part of a record in json, given the ivoid
 - plus related records.
 - perhaps return all ivoids for a particular category
- admin interfaces
 - something to trigger an “immediate” harvest
 - standard “upload” endpoint
- Any other things that people have implemented that might be worthy of standardisation?

DM concept duplication

Concept

DM

Observatory (Facility)

VOResource, ProposalDM

Instrument
Service

VODataService, ProposalDM, CAOM
VOResource, Mango, Datalink

Data Set

DataSet, VODataService

Observation

ObsCore, ProposalDM

Creator/Curator

VOResource, DataSet

Parameter

PDL, VOTable, UWS, Mango

- Not guaranteed to be an accurate/complete survey, but illustrative
- Sometimes the concept might be intended as a reference
- Note that data models are created 'outside' the DM WG

Other approaches

- ✦ SPASE (Space Physics Archive Search and Extract)
- ✦ NASA PDS
- ✦ These are easier to query - the whole “Domain” is in the one model (or set of non-overlapping models)
 - ✦ No possible ambiguities in concepts
 - ✦ Don’t have to try to join together services as with the IVOA
 - ✦ we have evolved in a different way
 - ✦ but it would be nice to try some rationalisation

DM rationalization/refactor

- There is some slightly unfortunate 'silo'ing in DM, Registry and Semantics WGs when there is clearly much potential common interest.
- Sometimes as with ProposalDM which *needs* concepts like Observatory, Telescope and Instrument, but does not *want* to be the central repository of instances of such things - it *feels* like the registry would be natural place for them to live.
 - Note there are facility and instrument fields in the registry without specific types to point to.
- Have started a pilot that unifies the VOResource data model with other models by defining it in VO-DML
 - <https://github.com/ivoa/RegistryDM>
 - initial aim is to try to make resulting schema validate current resource instances
 - but then introduce some new types

DataCollection vs DataResource

- DataCollection was there to model the “many services serve same dataset use case”
 - Unfortunately it was given a TableSet - there are other types of data
 - Vizier broke that model too
- DataResource “isA” Service so is expected to have Capabilities
 - though Capabilities are optional - so it could occupy the role that DataCollection was supposed to...
- Neither alternative is ideal
 - It would be nice to have a Registry V2.0 that did not have to deal with legacy “mistakes”

Conclusions

- ✦ There is a new Publishing Registry Implementation
 - ✦ intended to be easy to deploy - feedback welcome
- ✦ Would like to add some new content types to Registry