

Data Model Working Group

Closing Remarks

Chair: Mark Cresitello-Dittmar
Vice-Chair: Laurent Michel

Oct 27, 2017

IVOA Interop – Santiago



Summary – TDIG

- TDIG/DM/DAL
 - Generating/documenting use cases
 - Assessing current models
 - Developing requirements for needs
 - Active communication between groups



Summary – TDIG

- TDIG/DM/DAL
 - Generating/documenting use cases
 - Assessing current models
 - Developing requirements for needs
 - Active communication between groups



Summary - DM1

- STC 2.0
 - General model(s) stable – looking great
 - Exercising in several threads
 - Working 'shortcuts' for common cases
 - “make simple things simple”
 - **Questions**
 - Repercussions on STC-S serialization
 - Standard serialization → astroPy
 - Generation of XSD schema (VO-DML?)



Summary - DM1

- Provenance

- Nice presentations on model state and usage in SVOM ground segment
- Document maturing
- Good discussion on WD

- Goal: PR ready by end of year.

Next evolutions to be discussed

1 Introduction

- 1.1 Goal of the provenance model
- 1.2 Minimum requirements for provenance
- 1.3 Role within the VO architecture
- 1.4 Previous efforts

2 The provenance data model

- 2.1 Overview: Conceptual UML class diagram and introduction to core classes
- 2.2 Model description
 - 2.2.1 Class diagram and VO-DML compatibility
 - 2.2.2 Entity and EntityDescription
 - 2.2.3 Collection
 - 2.2.4 Activity and ActivityDescription
 - 2.2.5 ActivityFlow
 - 2.2.6 Entity-Activity relations → WasDerivedFrom? WasInformedBy?
 - 2.2.7 Parameters
 - 2.2.8 Agent

3 Links to other data models

- 3.1 Links with Dataset/ObsCore Model
- 3.2 Links with Simulation Data Model

Moved to the Appendices

4 Serialization of the provenance data model

- 4.1 Introduction
- 4.2 Serialization formats: PROV-N, PROV-JSON and PROV-XML
- ~~4.3 PROV-XML format~~ comes with ProvTAP
- 4.4 Serialization of description classes in the data processing context
- 4.5 W3C PROV-DM compatible serializations

5 Accessing provenance information

- 5.1 Access protocols
- 5.2 ProvDAL
 - 5.2.1 ProvDAL example use cases
- 5.3 ProvTAP
- 5.4 VOSI availability and capabilities

→ DAL document

6 Use cases – applying the data model

- 6.1 How to use the data model
- 6.2 voprov Python package
- 6.3 Provenance of RAVE database tables
- 6.4 Provenance for CTA
- 6.5 Provenance for the POLLUX database
- 6.6 Provenance of HiPS datasets

Appendices → Implementation note

+ consistent VO vocabulary, map with external ID



Summary - DM1

- Usage of DMs by INAF Radio Data Archive
 - Archive model is complex while scientific users are not interested in this complexity
 - Common theme bridging model → users
 - Tar file datasets
 - Conclusions
 - ObsCore satisfies discovery requirements
 - CAOM may preserve inherent data complexity
 - Q: what metadata is needed to bridge the gap from data access to data processing?



Summary - DM1

- Source model

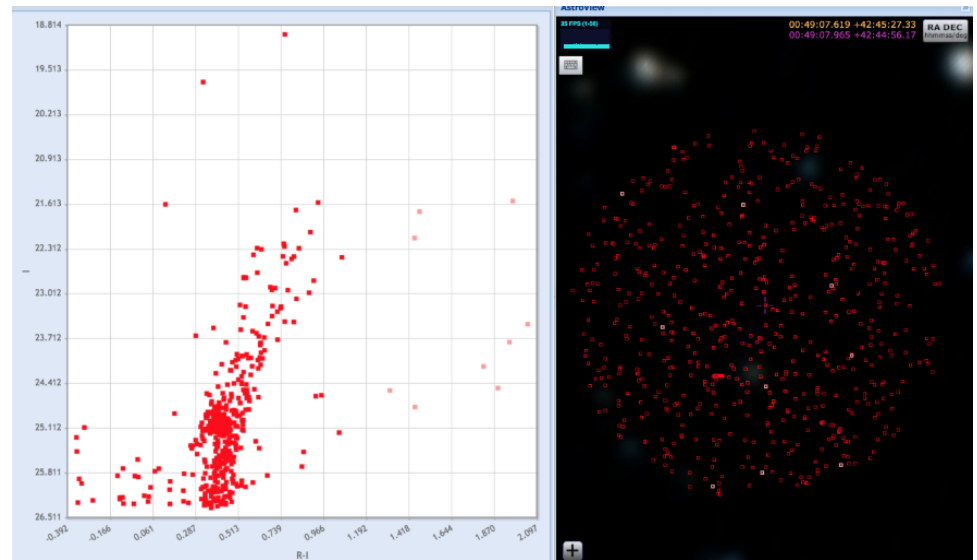
- Interest in re-opening this project
- A simple model could be useful for many cases
- Complexity comes in provenance..
 - How was this parameter computed?
- Thoughts
 - Current Source work takes much of this into account and is vo-dml compliant
 - With provenance maturing, maybe can revisit
 - Needs to factor in more catalogues



Summary – DM/Apps

- DM/Apps joint session

- Very nice presentations on current syntax application to various models
- 'Demo' of simple science thread on annotated source catalogue response
- No consensus on discussion topics
- Need to engage WGs more directly to allow informed feedback.



Roadmap - Victoria

- Sorry.. no racetrack yet
- VO-DML: REC
- Provenance: PR by end of year
- STC-2: REC
- DS/Cube: REC
- Mapping: Involve broader community to form consensus on content and usage
- SourceDM: form focus team

