

Prototyping a Provenance model for astronomical data in the IVOA

Mireille Louys¹, Matthieu Servillat², Michèle Sanguillon³, Johan Bregeon³, Kristin Riebe⁴, François Bonnarel¹, Catherine Boisson²

¹ CDS, Observatoire de Strasbourg

² LUTH, Observatoire de Paris Meudon

³ LUPM, Montpellier

⁴ AIP, Postdam

1. Context

The IVOA Modeling effort has been focused on the representation of metadata necessary to describe the content of datasets searched and used by astronomers for their science work.

Most data are shared in the astronomical community, and distributed on-line via archive centers or desktop services. Simulations of the astronomical objects composition and behavior are available and observations from various campaigns are delivered by astronomical data centers.

The IVOA modeling covers two types of distributed data: simulations and observations results.

The Simulation Data Model ¹ and the micro-simulation specifications describe how to represent and characterize simulated data.

ObsCore DM ² and more generally DataSetMetadata DM describe the main metadata necessary to select a dataset such as:

- Dataset properties: VO identifier, type of observation, date, data format, rights, etc.
- Physical properties of the content of a dataset (of the measurements), based on the description of physical axes observed for a dataset (Characterisation DM³).

These models focus on the description of the data product as the result of an observation process.

However raw data are combined and reprocessed in order to lead to a final science-ready data product, so information about acquisition and processing in the production pipeline, would help the user to evaluate data features and quality and adjust at best his/her selection criteria.

This is the reason why the IVOA Data Model WG investigates how to model the Provenance of a data set, both in terms of observing configuration and of data processing.

W3C Provenance data models and tools are interesting to check and compare to our needs. Current status of the IVOA Provenance DM effort is available at <http://wiki.ivoa.net/wiki/bin/view/IVOA/ObservationProvenanceDataModel>.

2. Requirements and use-cases

Requirements are currently discussed in the CTA project⁴ and in the GAVO (German Virtual Observatory) project ⁵.

3. Assessment of the W3C Provenance model and attached tools for astronomical data

Simulation DM provided a first pattern to highlight the method used to produce data and describe precisely the parameter used in and out. The W3C provenance Data model offers a pattern formalizing the logical links we found out in our test situations like for CTA multilevel data products.

¹ <http://www.ivoa.net/documents/SimDM/20120503/index.html>

² <http://www.ivoa.net/documents/ObsCore/20111028/index.html>

³ <http://www.ivoa.net/Documents/latest/CharacterisationDM.html>

⁴ <https://portal.cta-observatory.org/Pages/Home.aspx>

⁵ <https://volute.googlecode.com/svn/trunk/projects/dm/provenance/ProvenanceDMRequirements.html>

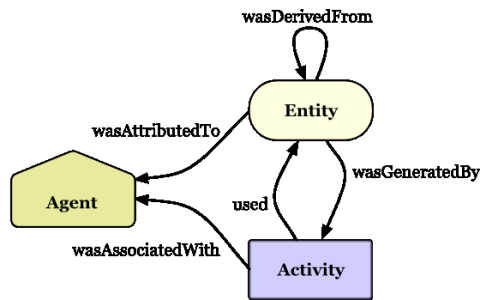


Figure 1 Agent/Activity/Entity Pattern from Prov-DM primer Document, W3C

4. Expressing Provenance metadata in astronomy context

PROV-N and W3C tools: testing the expressiveness and advantages of the W3C PROV-N tools

a. RAVE Prototype

RAVE use-case encoding with PROV-N: GAVO/IAP (Kristin Riebe)

b. CTA Prototype

Prototype for data products links between Level 3 to Level 4 of the CTA data products

UML design with import of Activity/Entity/Agent pattern

Tests for PROV-N representation

Implementations tests in Python (cf Karl Kozack)?

Connections to existing IVOA data models