Michele.Sanguillon@umontpellier.fr

Provenance Tools for Astronomy

Michèle Sanguillon[1], François Bonnarel[2], Mireille Louys[2,3], Markus Nullmeier[4], Kristin Riebe[5], Mathieu Servillat[6]

[1] LUPM-France, [2] CDS-France, [3] ICUBE-France, [4] ZAH-Germany, [5] AIP-Germany, [6] LUTH-France

Abstract

We present the current status of our developments of libraries and tools, mainly open source, which implement the IVOA Provenance data model in order to produce, serve, load and visualize provenance information. These implementations are also needed to validate and adjust the data model and the standard definitions for accessing provenance. The Provenance tools developed and created into the W3C framework are reused and extended when possible to tackle the domain of astronomical data.

IVOA PROVENANCE DATA MODEL

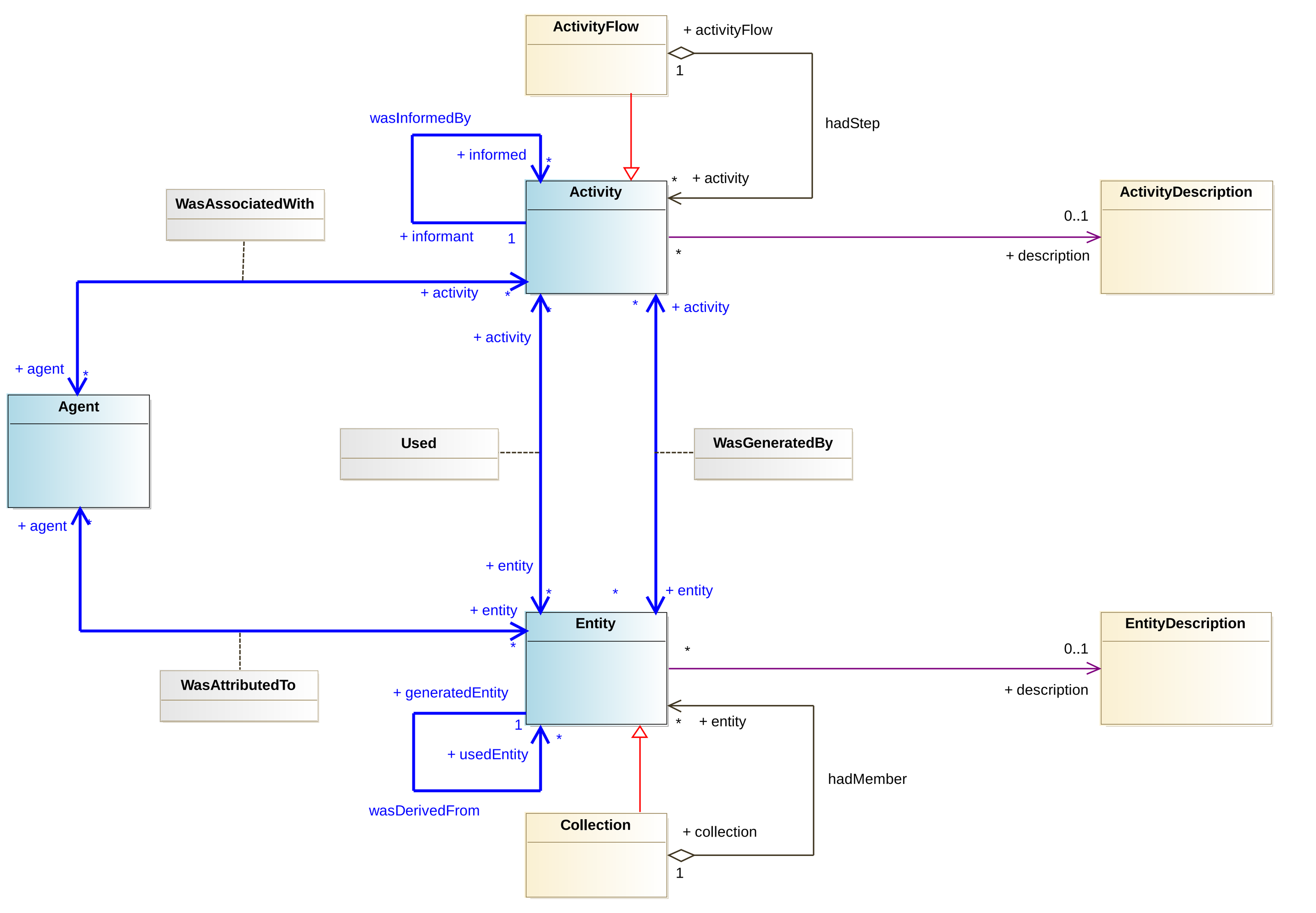
for astronomical data and processing steps

In this diagram, the blue classes are **core elements**: Entity, Activity and Agent.

An **Entity** is a thing in a certain state (ex: image, calibration data). It can represent either a single Entity or a collection of them.

An **Activity** is an action occuring over a period of time (ex: simulation, calibration). It can be composed of a set of Activities and have type ActivityFlow.

An **Agent** is a person or an organization who executes or controls an activity, who is responsible for an activity or an entity (ex: telescope astronomer, pipeline operator).



This figure shows the conceptual UML class diagram for the Provenance Data Model.

There are a number of many-to-many relationships which attached association classes (grey): wasGeneratedBy, used, wasAssociatedWith, wasAttributedTo, etc. They appear as extra classes because they can have additional attributes.

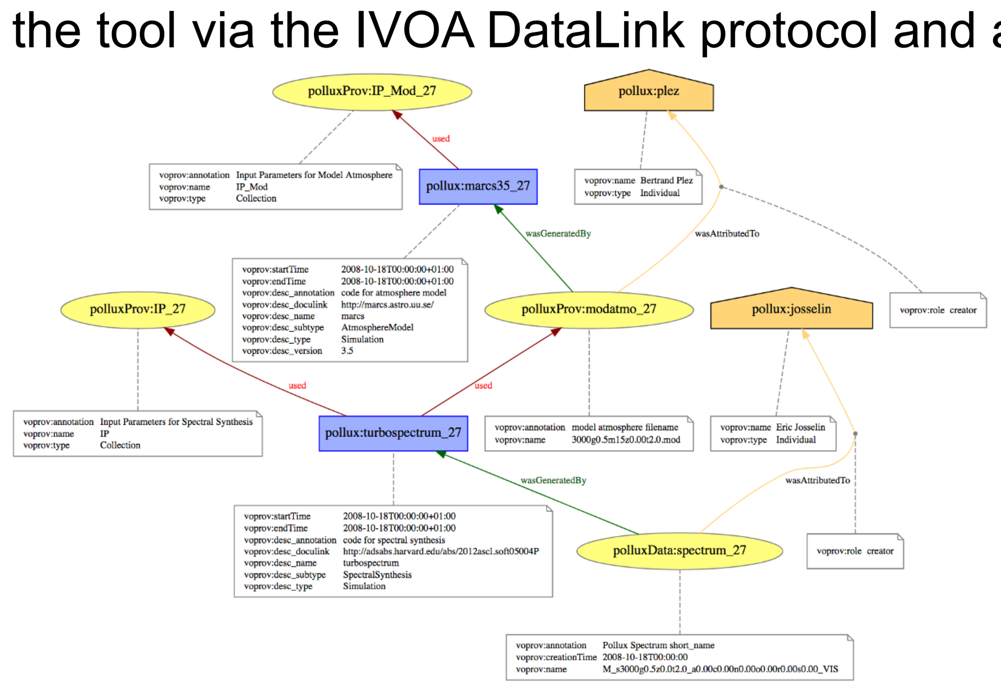
In the domain of astronomy, certain processes and steps are repeated over and over again, using different parameters. They use and generate Entities that have a common description. That means the Activity can be separated from its description (idem for the Entity).

Detailed description at http://www.ivoa.net/documents/ProvenanceDM/

VOPROV LIBRARY

The voprov package is an open source Python library which allows developers to serialize their provenance information in different formats: PROV-N, JSON, XML, VOTable or in graphical ones: PNG, SVG,PDF. This information can then be transmitted to the end user or to the tool via the IVOA DataLink protocol and a Web service.

(Cf https://github.com/sanguillon/voprov/)



This package is used in the context of **Pollux**.

Pollux is a stellar spectra database proposing access to high resolution synthetic spectra computed using the best available models of atmosphere and efficient spectral synthesis codes.

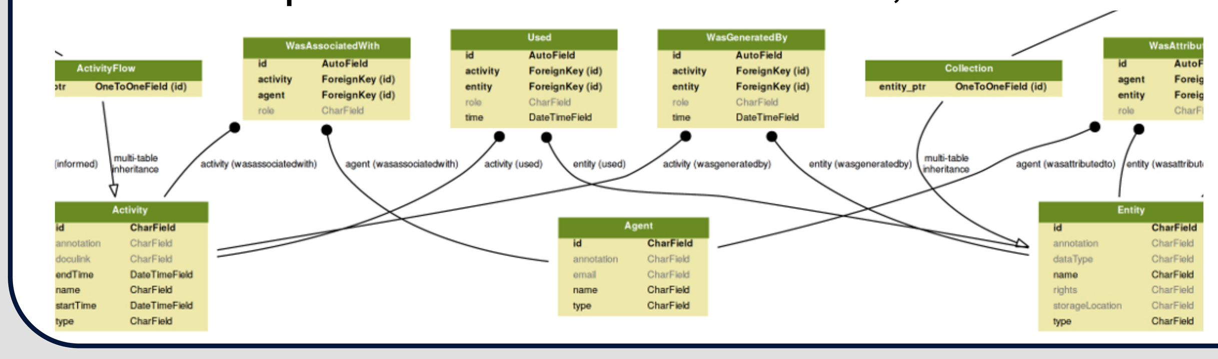
DJANGO PACKAGE

The Django provenance package is an open source Python package that can be reused in Django web applications for serving provenance information via a ProvDAL and a REST interface. The data model classes are directly mapped to tables in a relational database. It supports IVOA as well as W3C serializations into PROV-JSON and PROV-N formats.

(Cf https://github.com/kristinriebe/django-prov_vo)

This package is used in the context of **RAVE**.

The RAVE (RAdial Velocity Experiment) is a survey that observed the spectra of half a million stars from the southern hemisphere. In a pipeline of several steps the data were calibrated, reduced and stellar properties were determined, which were then released in the form of star catalogues.

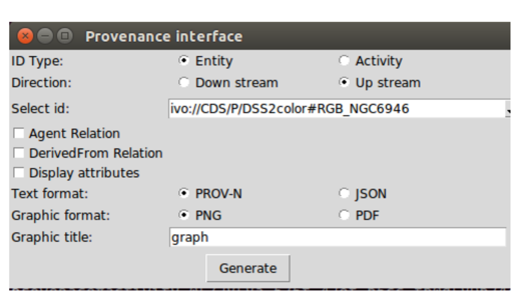


Prototype PostGreSQL database at CDS

In the CDS prototype we implemented a PostGreSQL database for Provenance information attached to image datasets. A database schema has been designed from the IVOA Provenance DM and implemented.

A set of images, together with their digitization and extraction steps, RGB color composition and HiPS generation activities are fed to the database. Various scenarii for querying and displaying the Provenance information have been tested. PROV-N, PROV-Json and PROV-VOTable formats are provided for the query response.

A simple user interface allowing to select the main types of requests and to display the responses via W3C Prov software has been designed. It allows querying for various combinations of Provenance relationships in the database.



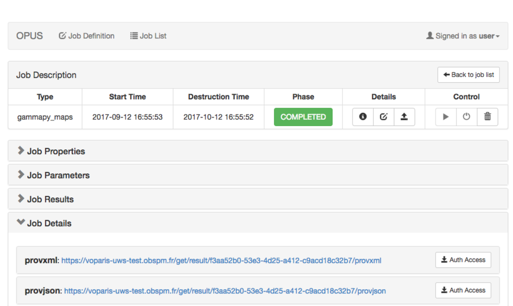
UWS Server at Observatoire de Paris

OPUS (Observatoire de Paris UWS System) is an open source job control system based on the IVOA UWS pattern.

It is developed in the context of the Cherenkov Telescope Array (CTA) project to test the execution of CTA data analysis tools on a work cluster.

It implements the concept of ActivityDescription files and provides the serialized provenance information as files for each executed job (see also ADASS Poster p3822).

(Cf https://github.com/mservillat/OPUS)



The **CTA** is the next generation ground-based very high energy gamma-ray instrument. It will serve as an open observatory providing data to a wide astrophysics community, with the requirement to propose self-described data products to users with detailed provenance information.

Acknowledgements: This work was partially funded by the Federal Ministry of Education and Research in Germany and by the ASTERICS project (http://www.asterics2020.eu/). Additional funding was provided by the INSU (Action Spécifique Observatoire Virtuel, ASOV), the Grand-Sud-Ouest Data Centre, the Paris Astronomical Data Centre and the Observatoire Astronomique de Strasbourg.