



ProvSAP prototypes

Michèle Sanguillon

Kristin Riebe

Provenance Working Group



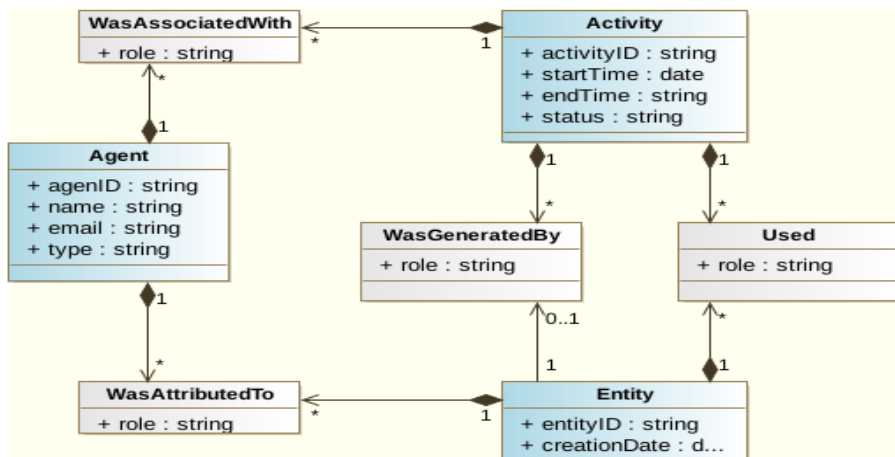
Provenance Group



Data Model (proposed recommendation)

Access Protocols (drafts)

Core classes



- **ProvSAP:** retrieve provenance information of one or several items (entity, activity)

3 implementations:
Pollux database, RAVE experiment & MuseWise

- **ProvTAP:** discovery of datasets based on provenance criteria

Cf F. Bonnarel's talk – ADASS



ProvSAP draft



- Why ProvSAP?
 - Allows users to retrieve serializations for a given item
 - Offers W3C compatible serializations that can be processed by tools that understand W3C provenance metadata
 - For existing projects with their own provenance data structure, gives the opportunity to share the provenance data in a common serialization format with other tools
 - Returns valid IVOA or W3C serializations, which can be processed or uploaded with other tools, or also placed into FITS-headers
 - Easy to implement: by a Web service (GET request)
- Draft : need to be updated (ProvenanceDM last version: model and serialization)



ProvSAP parameters



Parameter	Values	Description
ID	qualified ID	a valid qualified identifier for an entity, activity or agent (can occur multiple times)
DEPTH	0,1,2,..., ALL	number of relations to be followed or ALL for everything, independent of the relation type
RESPONSEFORMAT	PROV-N, PROV-JSON, PROV-XML, PROV-VOTABLE	serialisation format of the response
DIRECTION	<u>BACK</u> , FORTH	BACK = track the provenance history, FORTH = explore the results of activities and where entities have been used
MEMBERS	true (1) or <u>false</u> (0)	if true/1, retrieve and track members of collections
STEPS	true (1) or <u>false</u> (0)	if true/1, retrieve and track steps of activityFlows
AGENT	true (1) or <u>false</u> (0)	if true/1, explore all relations for agents, i.e. find out what an agent is responsible for
MODEL	<u>IVOA</u> or W3C	compatibility of the serialization to the IVOA or W3C provenance data model

Mandatory

Optional

ID: ongoing discussion.

Do we have to restrict to IDs of entities?

DEPTH: Last discussions:

1 = 1 processing step

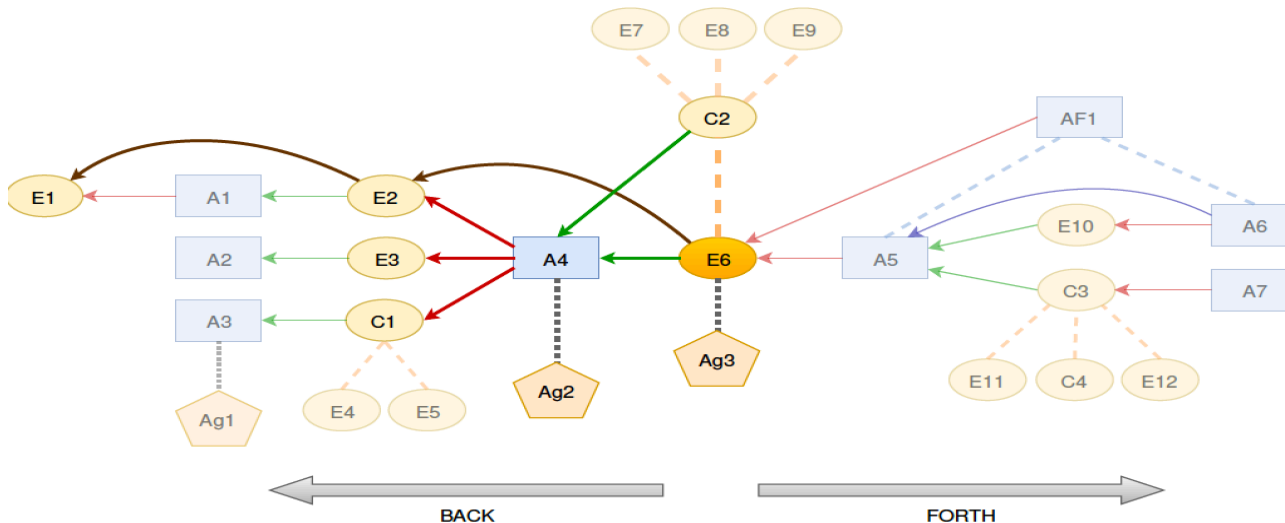
STEPS: no workflow in the DM 1.0

AGENT: Last discussions: won't track further
=> display or not agents

MODEL: ongoing discussion. Only W3C or IVOA serialization? If IVOA, other parameters?



ProvSAP draft



A	Activity	<i>Processing relations</i>	<i>Hierarchical relations</i>	
AF	ActivityFlow			
E	Entity			
C	Collection			
Ag	Agent	<i>Responsibility relations</i>		
	used			hadStep
	wasGeneratedBy			hadMember
	wasDerivedFrom		wasAssociatedWith, wasAttributedTo	
	wasInformedBy			



The context of Pollux

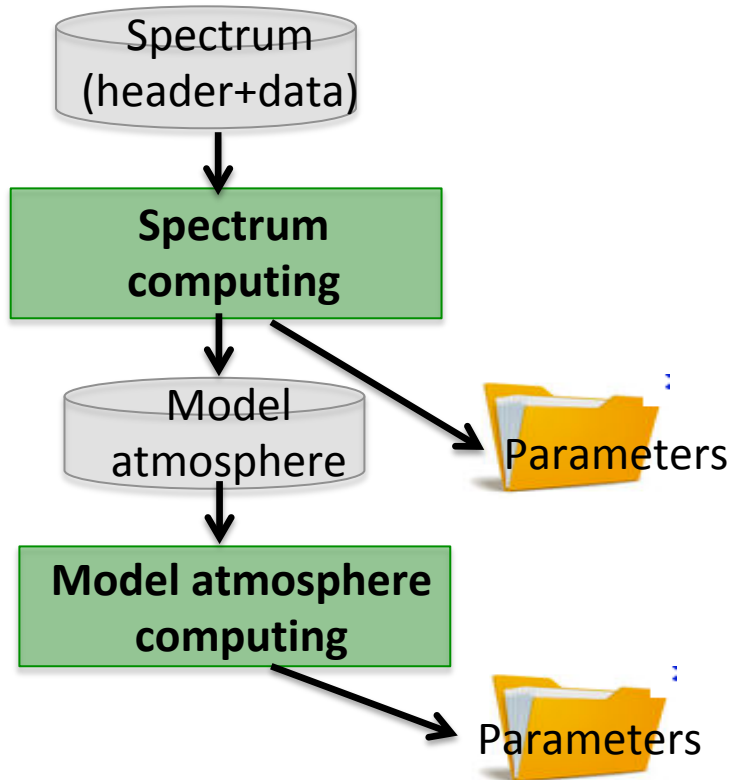


- **Pollux: Synthetic stellar spectra database**
- Access via a website (<http://pollux.oreme.org>) or via the SSA protocol (<ivo://ov-gso/ssap/pollux>)
- Spectra computed at very high resolution (computing done by the producers and not on the fly)
- Spectra provided in different formats (Flat tables, FITS, VOTable, ...)
- Has its own provenance data structure provided in a header file or in a specific HDU in FITS files
- Wish to provide provenance information in a standardized format

=> Implementation of ProvSAP



Pollux spectra history



Only the spectra are hosted at LUPM and available on the VO
⇒ We decide
to provide only the provenance of the spectra



Implementation of the parameters



Parameter	Values	Description
ID	qualified ID	a valid qualified identifier for an entity, activity or agent (can occur multiple times)
DEPTH	0, <u>1</u> ,2,..., ALL	number of relations to be followed or ALL for everything, independent of the relation type
RESPONSEFORMAT	PROV-N, <u>PROV-JSON</u> , PROV-XML, PROV-VOTABLE	serialisation format of the response
DIRECTION	<u>BACK</u> , FORTH	BACK = track the provenance history, FORTH = explore the results of activities and where entities have been used
MEMBERS	true (1) or <u>false</u> (0)	if true/1, retrieve and track members of collections
STEPS	true (1) or <u>false</u> (0)	if true/1, retrieve and track steps of activityFlows
AGENT	true (1) or <u>false</u> (0)	if true/1, explore all relations for agents, i.e. find out what an agent is responsible for
MODEL	<u>IVOA</u> or W3C	compatibility of the serialization to the IVOA or W3C provenance data model

ID: Multiple Ids spectra ids only

DEPTH: 0,1,...,ALL

RESPONSEFORMAT: PROV-N, PROV-JSON, PROV-XML, PROV-VOTABLE, SVG, PNG, PDF

DIRECTION: BACK only

MEMBERS: no coll defined

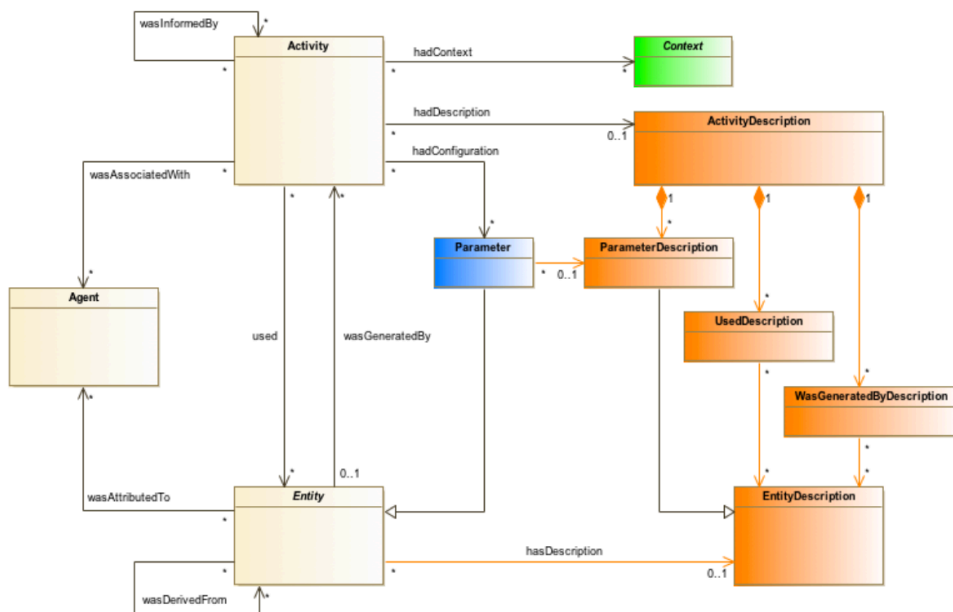
STEPS: no workflow defined

AGENT: true or false (to display or not)

MODEL: W3C, IVOA in progress



Mapping IVOA-W3C used



Non W3C classes:

- mapped to entities
- `prov:type=voprov:<class_name>`

Non W3C relations:

- mapped to `wasInfluencedBy`
- `prov:type=voprov:<relation_name>`



ProvSAP Web service



Web service

- <http://pollux.oreme.org/vo/provsap?...>
- Python
 - prov module (1.5.0 version) developed by Trung Dong Huynh (Southampton University)
 - implementing W3C provenance data model
 - Output formats of serialized data : PROV-N, JSON
 - Graphics output fomats: PDF, PNG, SVG
 - voprov:
 - Adapted to IVOA provenance data model (Flow class, hadStep relationship)
 - Addition of the output serialized format: VOTable
 - Beta version: <https://github.com/sanguillon/voprov>

Request →

← Response

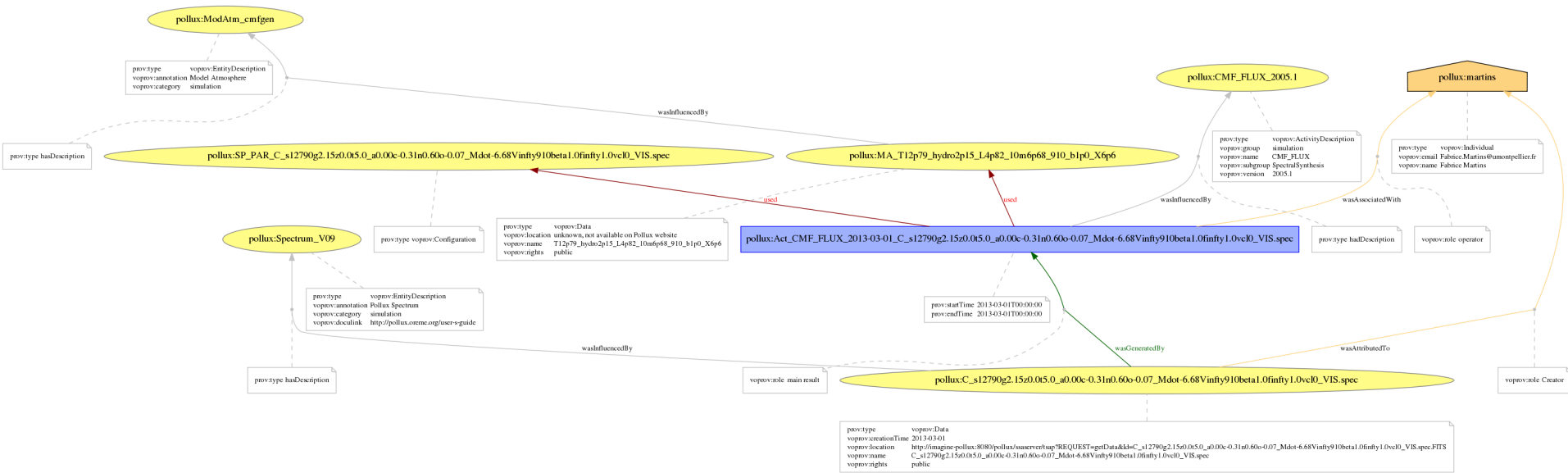
← Metadata database



ProvSAP query



http://pollux.oreme.org/vo/provsap?id=C_s12790g2.15z0.0t5.0_a0.00c-0.31n0.60o-0.07_Mdot-6.68Vinfy910beta1.0finfty1.0vcl0_VIS.spec.FITS&RESPONSEFORMAT=PNG&DEPTH=1&AGENT=true&model=W3C



ProvSAP prototype for RAVE

- RAVE = Radial Velocity Experiments; data for half a million stars
- Prototype web application using Django framework (Python) (incl. description classes; however, currently not reflecting the latest model)
 - Live demo: <https://escience.aip.de/provenance-rave> (only a few test data sets available)
 - GitHub: <https://github.com/kristinriebe/provenance-rave>
- ProvSAP interface supporting all parameters of the current draft
- Returns PROV-N, PROV-JSON or PROV-XML
- IVOA serialization: using voprov-prefix and IVOA extensions and IVOA class/attribute names
- W3C serialization: attributes etc. renamed and mapped to create W3C compatible serialization

Parameter	Values
ID	qualified ID
DEPTH	0,1,2,..., ALL
RESPONSEFORMAT	PROV-N, PROV-JSON, PROV-XML, PROV-VOTABLE
DIRECTION	<u>BACK</u> , FORTH
MEMBERS	true (1) or <u>false</u> (0)
STEPS	true (1) or <u>false</u> (0)
AGENT	true (1) or <u>false</u> (0)
MODEL	<u>IVOA</u> or W3C

ProvSAP webform

mandatory parameters

The screenshot shows a webform with the following fields and options:

- Identifier:** A text input field containing the value `rave:20121220_0752m38_089`. Below it is a note: "Please enter the identifier for an entity (e.g. rave:20030411_1507m23_001 or rave:20121220_0752m38_089) or an activity (e.g. rave:act_iraftReduction)".
- Depth:** A dropdown menu set to `1`. Below it is the text: "Specify number of relations to be tracked".
- Direction:** Radio buttons for `back` (selected) and `forth`. Below it is the text: "Choose the tracking direction".
- Data model:** Radio buttons for `IVOA` (selected) and `W3C`. Below it is the text: "Choose W3C for W3C Prov-DM compliant serialization".
- Response format:** Radio buttons for `PROV-N`, `PROV-JSON` (selected), and `Graphics`. Below it is the text: "Format of returned provenance record".
- Members:** A checkbox. Below it is the text: "Also find and track members of collections".
- Activity steps:** A checkbox. Below it is the text: "Also find and track steps of activityFlows".
- Agent:** A checkbox. Below it is the text: "Also find and track other entities and activities that the found agents are responsible for".

Submit

additional option

Automatically generates the ProvSAP GET request URL: https://escience.aip.de/provenance-rave/provapp/provdal/?ID=rave:20121220_0752m38_089&DEPTH=1&RESPONSEFORMAT=PROV-JSON&DIRECTION=BACK&MODEL=IVOA&MEMBERS=false&STEPS=false&AGENT=false



ProvSAP



Thank you for your attention