ProvTAP:

A TAP service for providing IVOA provenance metadata

F. Bonnarel
on behalf of the « provenance datamodel » author team of the IVOA
What is ProvTAP for?

- Distributing provenance metadata for astronomical datasets
- Selecting datasets by provenance
- ProvTAP is a specification for services serializing IVOA provenance metadata model
Provenance data model

Let's speak first of the IVOA provenance data model:

A lot of definitions are possible. Look at W3C one.
W3C PROV (PROV-DM, 2013)

Provenance is defined as a record that describes the people, institutions, entities, and activities involved in producing, influencing, or delivering a piece of data or a thing.

In particular, the provenance of information is crucial in deciding whether information is to be trusted, how it should be integrated with other diverse information sources, and how to give credit to its originators when reusing it.

Core concepts from the W3C PROV recommendations:

- **Entity** - **Activity** - **Agent**
- **Relations** and roles: e.g. generation, usage, influence, association, attribution, derivation, information
- W3C PROV has more relations (see components and extensions)
- IVOA Provenance connected to **VO concepts** and **astronomy needs**
A: Tracking the production history

Find out which steps were taken to produce a dataset and list the methods/tools/software that was involved.

B: Attribution and contact information

Find the people involved in the production of a dataset, that need to be cited or can be asked for more information.

C: Locate error sources

Find the location of possible error sources in the generation of a dataset.

D: Quality assessment

Judge the quality of an observation, production step or dataset.

E: Search in structured provenance metadata

This would allow one to also do a “forward search”, i.e. locate derived datasets or outputs.
IVOA model Extends the W3C one

Core model (W3C)
Provenance data model

Hot Topic!

5 presentations during IVOA interoperability meeting last week.

2 Posters at this conference:

« data model » (P11.6) and « provenance in triplestore » (P11.5)
Serialisation and services:

- ProvSAP exists

A parameter based service to get provenance information for a dataset in several formats including graphical format.
1) ProvTAP is ... TAP
2) mapping of the model classes/attributes to the relational view.
3) specification is currently an internal IVOA draft
A specification which defines:

- Interoperable table services, with relational view
- Queriable via a SQL-oriented language: ADQL
- Lot of tap services in many datacenters and big projects archives.

DataModels can be mapped in TAP via the "TAP schema" (the database schema) using object/relational mapping guidelines.
• A TAP schema has been defined
  – All classes and attributes of the model are mapped onto tables and columns of the schema
• A Prototype has been recently developed at CDS
  → screenshots in next slides
• CTA/HESS implementation in development in collaboration with CDS
CDS prototype content : HiPS and progenitors

• HiPS
  – Multiresolution all sky view, hierarchical, based on healpix cells at all orders
  – needs processing of « original images » to be generated
  – It's a VO standard.

• Tools exist to generate and read it

• Progenitors are some time available

• Metainformation on the HiPS has been transferred in a relational database underlying the ProvTAP service
HiPS
Goals of the prototype

• Create a first ProvtaP implementation
• Integrate information on HiPS as well as classical images in the same design
• Full integration of provenance searches in the VO framework
Simple queries to browse the content

- Entities
- Activities
- Agents
- Select parameters with associated ParameterDescriptions and activities to which they are related
first query in the html interface provided with the TAP library (G.Mantelet) : select * from entity
VOTable response

-- VOTABLE version="1.3" xsi:schemaLocation="http://www.ivoa.net/xml/VOTable/v1.3 http://www.ivoa.net/xml/VOTable/v1.3"
-- RESOURCE type="results"
    <INFO name="QUERY_STATUS" value="OK"/>
    <INFO name="PROVIDER" value="CDS"/>
    <INFO name="QUERY" value="SELECT * FROM entity;"/>
-- TABLE name="result_S15420304441455"
    <FIELD arraysize="1" datatype="char" name="e_id" ucd="meta.id" utype="vprov:EntityId"/>
    <FIELD arraysize="1" datatype="char" name="e_name" ucd="meta.title" utype="vprov:EntityName"/>
    <FIELD arraysize="1" datatype="char" name="e_type" ucd="meta.code.class" utype="vprov:EntityType"/>
    <FIELD arraysize="1" datatype="char" name="e_rights" ucd="meta.code.class" utype="vprov:EntityRights"/>
    <FIELD arraysize="1" datatype="char" name="e_annotation" ucd="meta.description" utype="vprov:EntityAnnotation"/>
    <FIELD arraysize="1" datatype="char" name="e_description" ucd="meta.id" utype="vprov:EntityName"/>
-- TABLEDATA
  -- TR
    <TD>ivo://CDS/P/2MASS/H</TD>
    <TD>2MASS H (1.66um) Hips</TD>
    <TD>data</TD>
    <TD>public</TD>
    <TD>hipsdata</TD>
  -- TR
    <TD>ivo://CDS/P/2MASS/J</TD>
    <TD>2MASS J (1.23um) Hips</TD>
    <TD>data</TD>
    <TD>public</TD>
    <TD>origimages</TD>
  -- TR
    <TD>ivo://CDS/P/2MASS/H</TD>
    <TD>2MASS H (1.66um) original data</TD>
    <TD>data</TD>
    <TD>public</TD>
    <TD>origimages</TD>
  -- TR
    <TD>ivo://CDS/P/2MASS/J</TD>
    <TD>2MASS J (1.23um) Hips</TD>
    <TD>data</TD>
    <TD>public</TD>
    -- TR
    2MASS has uniformly scanned the entire sky in three near-infrared bands to detect and characterize point sources brighter than about 1 mJy in each band, with signal-to-noise ratio (SNR) greater than 10, using a pixel size of 2.0". This has achieved an 80,000-fold improvement in sensitivity relative to earlier surveys. 2MASS used two highly-automated 1.3 m telescopes, one at Mt. Hopkins, AZ, and one at CTIO, Chile. Each telescope was equipped with a three-channel camera, each channel consisting of a 256x256 array of HgCdTe detectors, capable of observing the sky simultaneously at J (1.25 microns), H (1.65 microns), and Ks (2.17 microns). The University of Massachusetts (UMass) was responsible for the overall management of the project, and for developing the infrared cameras and on-site computing systems at both facilities. The Infrared Processing and Analysis Center (IPAC) is responsible for all data processing through the Production Pipeline, and construction and distribution of the data products. Funding is provided primarily by NASA and the NSF.
  -- TR
    <TD>hipsdata</TD>
  -- TR
    <TD>origimages</TD>
    <TD>2MASS J (1.23um) original data</TD>
    <TD>data</TD>
    <TD>public</TD>
    <TD>origimages</TD>
  -- TR
    <TD>origimages</TD>
    <TD>2MASS J (1.23um) original data</TD>
    <TD>data</TD>
    <TD>public</TD>
    <TD>origimages</TD>
01/10/2015
SELECT * FROM ACTIVITY

JSON Response
<table>
<thead>
<tr>
<th>ag_id</th>
<th>ag_name</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;noagent&quot;</td>
<td>&quot;noname&quot;</td>
</tr>
<tr>
<td>&quot;agent_1_277&quot;</td>
<td>&quot;1.8&quot;</td>
</tr>
<tr>
<td>&quot;agent_1_328&quot;</td>
<td>&quot;Pierre Fernique [CDS]&quot;</td>
</tr>
<tr>
<td>&quot;agent_1_537&quot;</td>
<td>&quot;L. Michel [Observatoire de Strasbourg]&quot;</td>
</tr>
<tr>
<td>&quot;agent_1_222&quot;</td>
<td>&quot;P. Fernique [CDS]&quot;</td>
</tr>
<tr>
<td>&quot;agent_1_190&quot;</td>
<td>&quot;P. Fernique (CDS)&quot;</td>
</tr>
<tr>
<td>&quot;agent_1_378&quot;</td>
<td>&quot;ESA (ESDC &amp; Planck Science Office)&quot;</td>
</tr>
<tr>
<td>&quot;agent_1_5&quot;</td>
<td>&quot;CDS (T. Boch)&quot;</td>
</tr>
<tr>
<td>&quot;agent_1_318&quot;</td>
<td>&quot;Stefan Meingast (Institute for Astrophysics, University of Vienna)&quot;</td>
</tr>
<tr>
<td>&quot;agent_1_371&quot;</td>
<td>&quot;ESA/ESDC&quot;</td>
</tr>
<tr>
<td>&quot;agent_1_191&quot;</td>
<td>&quot;CDS (Pierre Fernique)&quot;</td>
</tr>
<tr>
<td>&quot;agent_1_432&quot;</td>
<td>&quot;D. Paradis (IRAP/CADE)&quot;</td>
</tr>
<tr>
<td>&quot;agent_1_330&quot;</td>
<td>&quot;Thomas Boch [CDS]&quot;</td>
</tr>
<tr>
<td>&quot;agent_1_33&quot;</td>
<td>&quot;CDS (Thomas Boch)&quot;</td>
</tr>
<tr>
<td>&quot;agent_1_407&quot;</td>
<td>&quot;Guilherme Soares&quot;</td>
</tr>
<tr>
<td>&quot;agent_1_36&quot;</td>
<td>&quot;Thomas Boch&quot;</td>
</tr>
<tr>
<td>&quot;agent_1_99&quot;</td>
<td>&quot;CDS (A.Oberto, P. Fernique)&quot;</td>
</tr>
<tr>
<td>&quot;agent_1_97&quot;</td>
<td>&quot;CDS (P. Fernique)&quot;</td>
</tr>
<tr>
<td>&quot;agent_1_8&quot;</td>
<td>&quot;CDS [P. Fernique]&quot;</td>
</tr>
<tr>
<td>&quot;agent_1_44&quot;</td>
<td>&quot;T. Boch&quot;</td>
</tr>
<tr>
<td>&quot;agent_1_7&quot;</td>
<td>&quot;CDS&quot;</td>
</tr>
<tr>
<td>&quot;agent_1_352&quot;</td>
<td>&quot;ESA (ESDC &amp; Herschel SOC)&quot;</td>
</tr>
<tr>
<td>&quot;agent_1_342&quot;</td>
<td>&quot;China-VO&quot;</td>
</tr>
<tr>
<td>&quot;agent_1_130&quot;</td>
<td>&quot;CADC (Daniel Durand)&quot;</td>
</tr>
<tr>
<td>&quot;agent_1_409&quot;</td>
<td>&quot;NASA/HEASARC&quot;</td>
</tr>
<tr>
<td>&quot;agent_1_9&quot;</td>
<td>&quot;P. Fernique [CDS]&quot;</td>
</tr>
<tr>
<td>&quot;agent_1_14&quot;</td>
<td>&quot;M. Buca [CDS]&quot;</td>
</tr>
<tr>
<td>&quot;agent_1_354&quot;</td>
<td>&quot;ESA (ESDC &amp; Herschel Science Centre)&quot;</td>
</tr>
<tr>
<td>&quot;agent_1_16&quot;</td>
<td>&quot;P. Fernique [CDS]&quot;</td>
</tr>
<tr>
<td>&quot;agent_1_536&quot;</td>
<td>&quot;WFAU, Institute for Astronomy, University of Edinburgh&quot;</td>
</tr>
<tr>
<td>&quot;agent_1_126&quot;</td>
<td>&quot;Christoph Deil, Axel Donath, Pierre Fernique&quot;</td>
</tr>
<tr>
<td>&quot;agent_1_1&quot;</td>
<td>&quot;CDS (A. Oberto)&quot;</td>
</tr>
<tr>
<td>&quot;agent_2_229&quot;</td>
<td>&quot;Axel Mellinger&quot;</td>
</tr>
<tr>
<td>&quot;agent_2_227&quot;</td>
<td>&quot;JPL/Photojournal&quot;</td>
</tr>
<tr>
<td>&quot;agent_2_535&quot;</td>
<td>&quot;SVO, CAB (INTA-CSIC)&quot;</td>
</tr>
<tr>
<td>&quot;agent_2_221&quot;</td>
<td>&quot;Arizona State University&quot;</td>
</tr>
<tr>
<td>&quot;agent_2_350&quot;</td>
<td>&quot;<a href="http://archives.esac.esa.int/hsa/whsa/">http://archives.esac.esa.int/hsa/whsa/</a>&quot;</td>
</tr>
<tr>
<td>&quot;agent_2_36&quot;</td>
<td>&quot;<a href="http://portal.nersc.gov/project/cosmo/data/decaps/drl/coadd/">http://portal.nersc.gov/project/cosmo/data/decaps/drl/coadd/</a>&quot;</td>
</tr>
<tr>
<td>&quot;agent_2_232&quot;</td>
<td>&quot;USGS Astrogeology Science Center from Arizona State University&quot;</td>
</tr>
<tr>
<td>&quot;agent_2_170&quot;</td>
<td>&quot;MAST archives&quot;</td>
</tr>
<tr>
<td>&quot;agent_2_114&quot;</td>
<td>&quot;NASA s Earth Observatory&quot;</td>
</tr>
<tr>
<td>&quot;agent_2_34&quot;</td>
<td>&quot;<a href="http://portal.nersc.gov/project/cosmo/data/legacysurvey/dr5/coadd/">http://portal.nersc.gov/project/cosmo/data/legacysurvey/dr5/coadd/</a>&quot;</td>
</tr>
<tr>
<td>&quot;agent_2_216&quot;</td>
<td>&quot;<a href="https://photojournal.jpl.nasa.gov/catalog/PIA209284">https://photojournal.jpl.nasa.gov/catalog/PIA209284</a>&quot;</td>
</tr>
<tr>
<td>&quot;agent_2_377&quot;</td>
<td>&quot;<a href="http://iso.esac.esa.int/idax/">http://iso.esac.esa.int/idax/</a>&quot;</td>
</tr>
<tr>
<td>&quot;agent_2_17&quot;</td>
<td>&quot;CFTP&quot;</td>
</tr>
</tbody>
</table>
```sql
SELECT p_isaparamof, pd_name, pd_ucd, pd_unit, p_value
FROM parameter INNER JOIN parameterdescription
ON parameter.p_parameterdescription = parameterdescription.pd_id;
```

<table>
<thead>
<tr>
<th>p_isaparamof</th>
<th>pd_name</th>
<th>pd_ucd</th>
<th>pd_unit</th>
<th>p_value</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;act:CDS/P/2MASS/H&quot;</td>
<td>&quot;hips_order&quot;</td>
<td>&quot;meta.id&quot;</td>
<td>&quot;px&quot;</td>
<td>&quot;equatorial&quot;</td>
</tr>
<tr>
<td>&quot;act:CDS/P/2MASS/H&quot;</td>
<td>&quot;hips_frame&quot;</td>
<td>&quot;meta.id&quot;</td>
<td>&quot;px&quot;</td>
<td>&quot;ivo://CDS/P/2MASS/H&quot;</td>
</tr>
<tr>
<td>&quot;act:CDS/P/2MASS/K&quot;</td>
<td>&quot;hips_order&quot;</td>
<td>&quot;meta.id&quot;</td>
<td>&quot;px&quot;</td>
<td>&quot;equatorial&quot;</td>
</tr>
<tr>
<td>&quot;act:CDS/P/2MASS/K&quot;</td>
<td>&quot;hips_frame&quot;</td>
<td>&quot;meta.id&quot;</td>
<td>&quot;px&quot;</td>
<td>&quot;ivo://CDS/P/2MASS/K&quot;</td>
</tr>
<tr>
<td>&quot;act:CDS/P/2MASS/J&quot;</td>
<td>&quot;hips_order&quot;</td>
<td>&quot;meta.id&quot;</td>
<td>&quot;px&quot;</td>
<td>&quot;equatorial&quot;</td>
</tr>
<tr>
<td>&quot;act:CDS/P/2MASS/J&quot;</td>
<td>&quot;hips_frame&quot;</td>
<td>&quot;meta.id&quot;</td>
<td>&quot;px&quot;</td>
<td>&quot;ivo://CDS/P/2MASS/J&quot;</td>
</tr>
<tr>
<td>&quot;act:CDS/P/2MASS/S&quot;</td>
<td>&quot;hips_order&quot;</td>
<td>&quot;meta.id&quot;</td>
<td>&quot;px&quot;</td>
<td>&quot;equatorial&quot;</td>
</tr>
<tr>
<td>&quot;act:CDS/P/2MASS/S&quot;</td>
<td>&quot;hips_frame&quot;</td>
<td>&quot;meta.id&quot;</td>
<td>&quot;px&quot;</td>
<td>&quot;ivo://CDS/P/2MASS/S&quot;</td>
</tr>
</tbody>
</table>

Configuration parameters with their description (name, ucd, unit) and associated activity.
Real-life queries:
To select HiPS activities or entities via criteria

- Select activities which have been attributed to a given « Agent »
- Select activities described by the same ActivityDescription ( = here, running the same software)
- Select activities from some configuration parameters values
- Select entities and display them in Aladin (HiPS or classical images)
Select activities which have been attributed to a given « Agent »
(here « CADC (Danie Dürand) »)
select activities described by the same ActivityDescription

( = here, running the same hipsgen software)
Select activities from some configuration parameters values
(Here « created only in jpeg »)
select activities from some configuration parameters values
(here selected by ucd and « created in galactic frame)
Select entities and display them in Aladin (HiPS or classical images) (here « public » entities)
select entities and display them in Aladin (HiPS or classical images)

(Here progenitors centers overlay - ready to be loaded)
Functions for complex queries
(M.Nulmeier Heidelberg)

• ADQL queries on a database with 14 or more tables may rapidly become difficult to write
• Graph query technology required
• Implementation experimented via predefined functions, recursive CTE, etc..
• ProvSAP functionalities can be reproduced
Conclusion/future work

• Add provenance information for HiPS progenitors
  – Schmidt plate digitization
  – Raw data if available
• Enrich HiPS description in the service
• Cross combine information with HESS/CTA database