ProvTAP evolution after first community feedback

F.Bonnarel, CDS
on behalf of M.Servillat, M.Louys, M.Nullmeier, M.Sanguillon, L.Michel

CENTRE DE DONNÉES ASTRONOMIQUES DE STRASBOURG

ESCAPE
European Science Cluster of Astronomy & Particle physics ESFRI research Infrastructures

IVOA

Why a ProvTAP specification ?

- Provenance information can be attached to data in various ways :
  - Embedded in the data « header » itself
  - Linked to the data record via DataLink or URL
  - Retrievable via ProvSAP via data id.
- In addition to that , ProvTAP allows to discover « data » by constraining Provenance features.
  - It's a « reverse » mechanism.

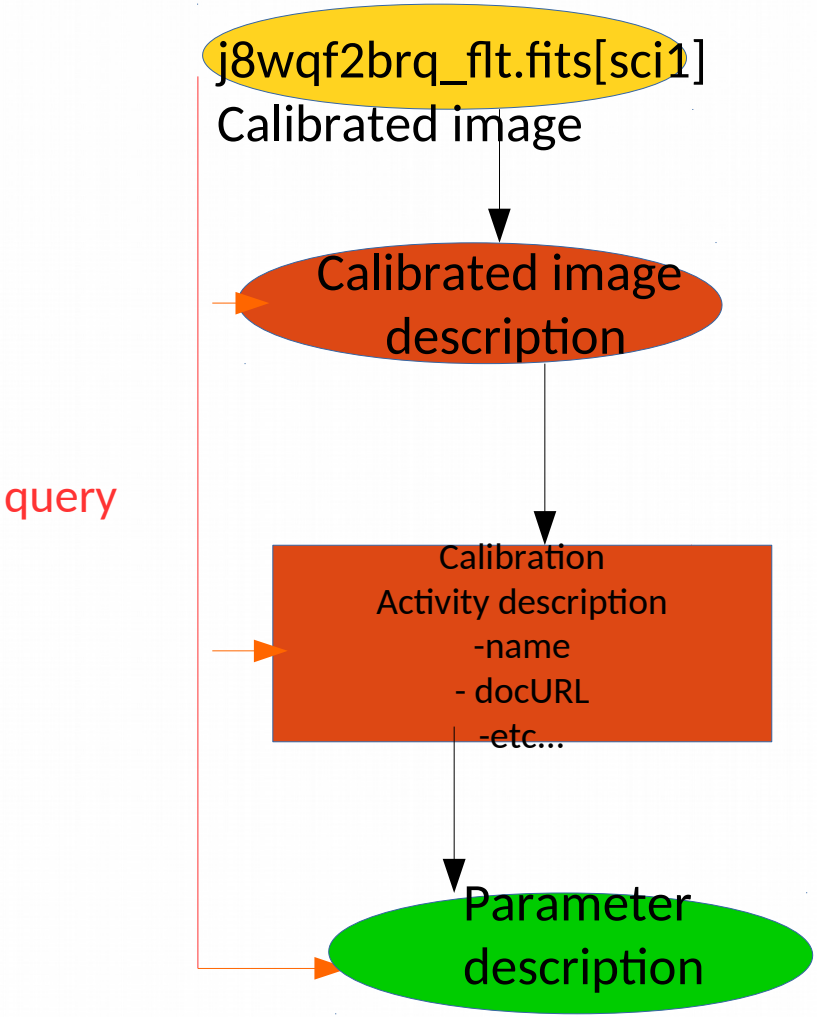
Why a ProvTAP specification ?

Examples :

- Discovering data produced with the same version of a given software.
- Discovering data produced with some specific value of a software configuration parameter :
  - Known by its name
  - Known by its ucd
  - ….
- Discovering a « family » of data : datasets produced by ancestor activities of a given datasets
- Discovering data or activities related to some agent with a given role (operator, editor, author, etc…)

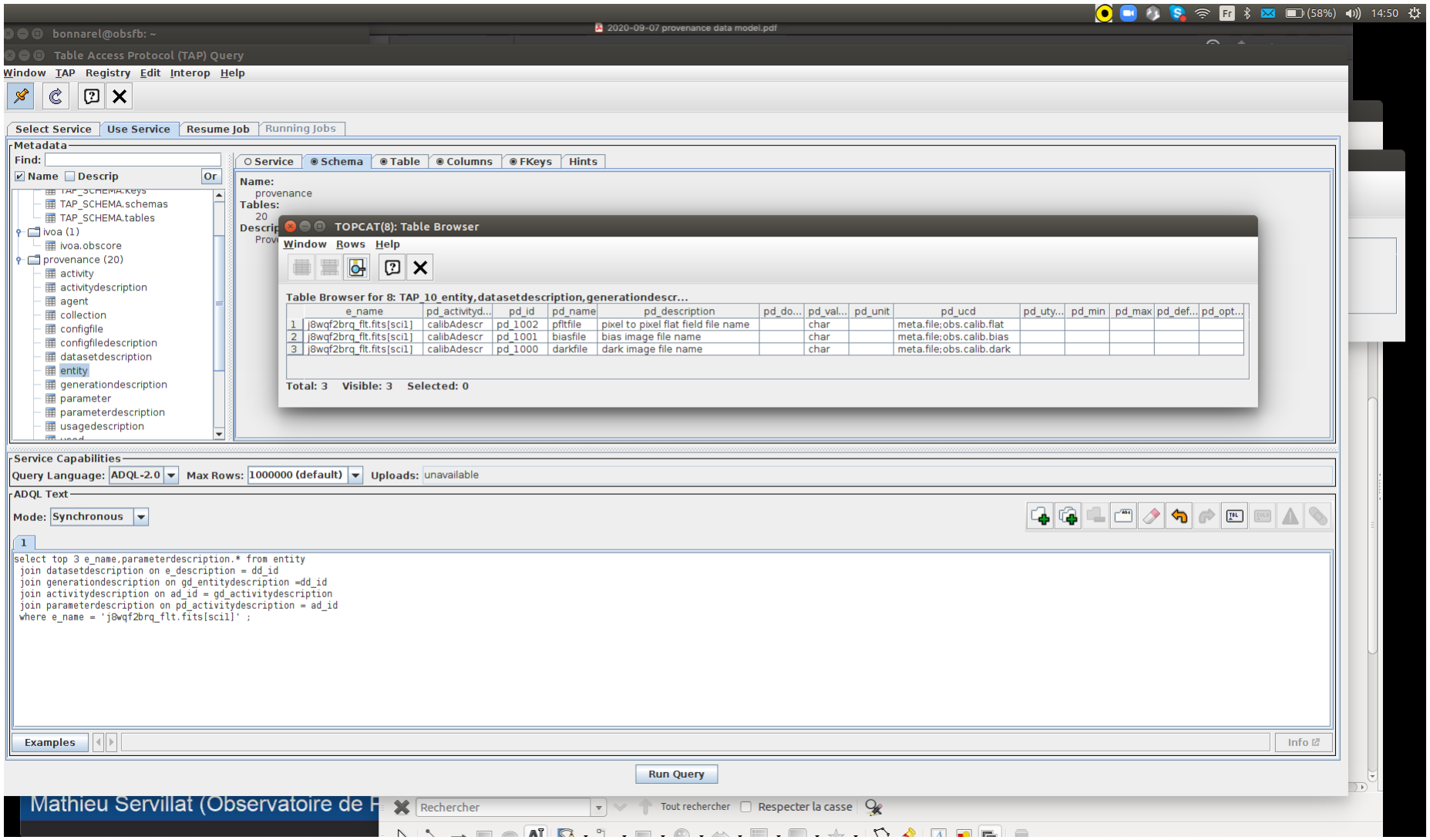
ProvHiPS ADQL query examples :

Find out parameter descriptions of parameter used to generate a calibrated file



ProvHiPS ADQL query examples :

Find out parameter descriptions of parameter used to generate a calibrated file



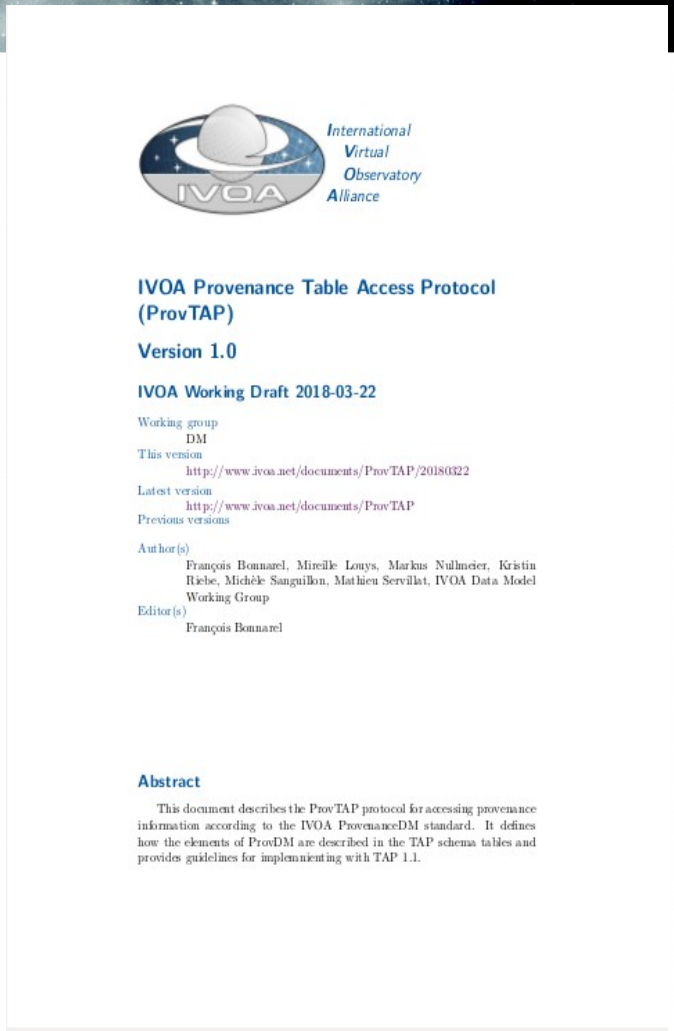
ProvHiPS ADQL query examples :

Find out parameter descriptions of parameter used to generate a calibrated file

```
select top 3 * from entity
join datasetdescription on e_description = dd_id
join generationdescription on gd_entitydescription =dd_id
join activitydescription on ad_id = gd_activitydescription
join parameterdescription on pd_activitydescription = ad_id
where e_name = 'j8wqf2brq_flt.fits[sci1]' ;
```

ProvTAP = where are we ?

- There is an internal draft on the IVOA DAL pages
- Was not a WD by lack of discussion
- TAP schema mapping classes as tables
- ProvHiPS (provenance of HiPS and HiPS tiles) is an implementation prototype
- Discussion among authors on various points (see DAL running meeting slides)



International Virtual Observatory Alliance

IVOA Provenance Table Access Protocol (ProvTAP)

Version 1.0

IVOA Working Draft 2018-03-22

Working group
DM
This version
http://www.ivoa.net/documents/ProvTAP/20180322
Latest version
http://www.ivoa.net/documents/ProvTAP
Previous versions

Author(s)
François Bonnarel, Mireille Louys, Markus Nullmeier, Kristin Riebe, Michèle Sanguillon, Mathieu Servillat, IVOA Data Model Working Group
Editor(s)
François Bonnarel

Abstract

This document describes the ProvTAP protocol for accessing provenance information according to the IVOA ProvenanceDM standard. It defines how the elements of ProvDM are described in the TAP schema tables and provides guidelines for implemnienting with TAP 1.1.

ProvTAP = where are we ?
External point of view

- DAL chairs
  - 1 to 1 class/table mapping too ambitious. Need for simplification/ denormalization
- ESFRI projects (within ESCAPE)
  - Looking for simplified/partial views
    - Tracing the last step ?
    - Concept of « ProvCore » : minimal model attributes ?
    - Depends from project requirements and uses cases

1 table per Class ?

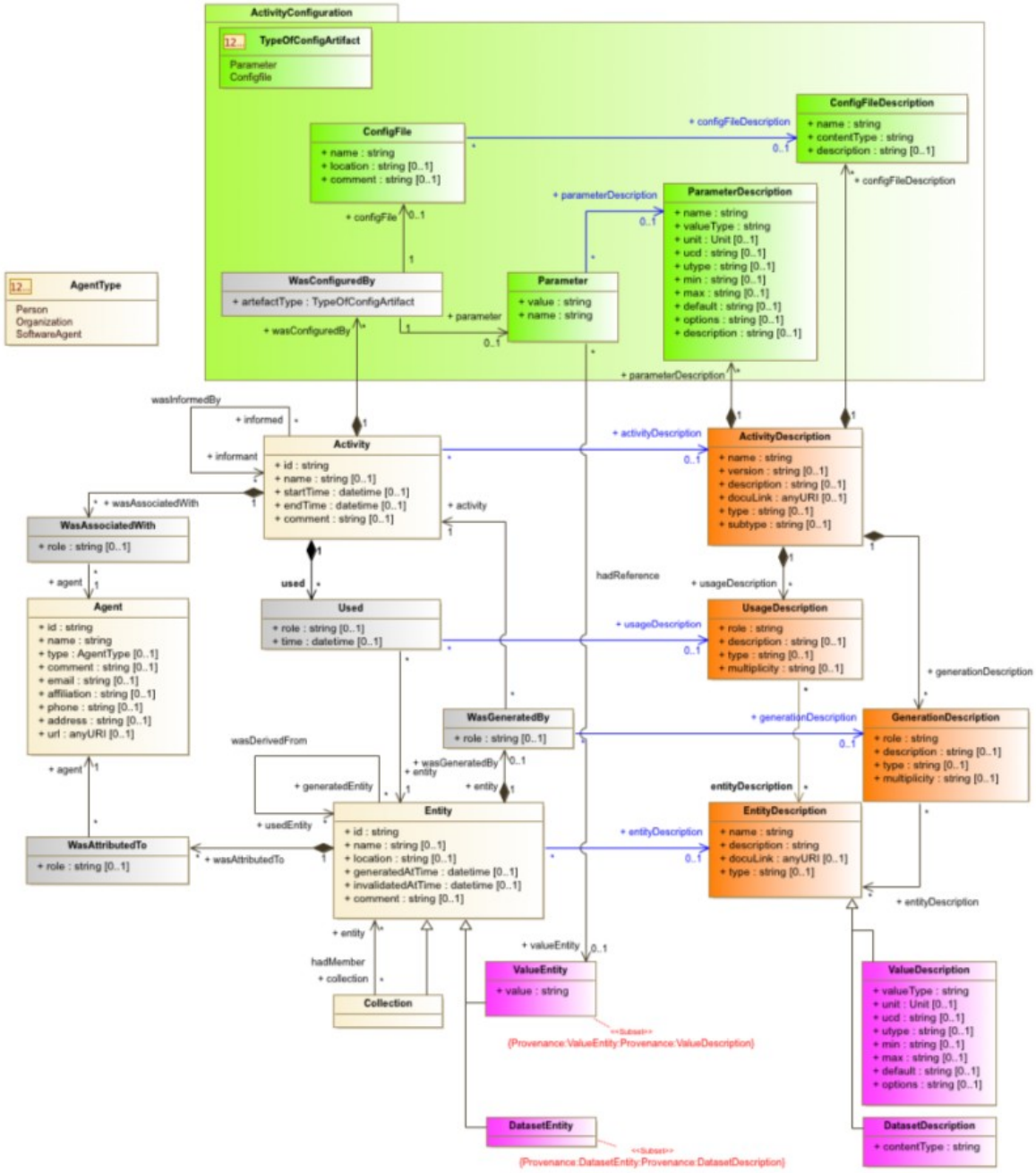


Figure 8: Full class diagram of the IVOA Provenance Data Model.

Solutions
-1 Single step = single table (= join)

- The join is a permanent view described in the TAP schema
- Columns :

  entity_name, entity_location, entity_comment, …

  generating_activity_name, generating_activity_starttime, ….

  agent_role, agent_name, ….

  used_entity_name
- → Redundancy (several lines for a single entity/activity).
- → possible Recursivity

Solutions
-1 Single step = single table (= join)

- View (in postgres)

create view last_step_provenance as select

e.e_name as entity_name, e.e_location as entity_location, e.e_generated as entity_generated, e.e_invalidated as entity_invalidated, e.e_comment as entity_comment,

a_name as generating_activity_name,a_starttime as generating_activity_starttime, a_endtime as generating_activity_endtime, a_comment as generating_activity_comment,

wat_role as agent_role, ag_name as agent_name, ag_type as agent_type ,ag_affiliation as agent_affiliation, ag_email as agent_email, ag_address as agent_address, ag_phone as agent_phone, ag_comment as agent_comment,

ee.e_name as used_entity_name from entity as e

join wasgeneratedby on wgb_entity = e.e_id
join activity on a_id = wgb_activity
join used on u_activity = a_id
join entity as ee on ee.e_id = u_entity
join wasattributedto on wat_entity = e.e_id
join agent on ag_id = wat_agent ;

Solutions

-1 Single step = single table (= join)

Solutions

-1 Single step = single table (= join)

Table Browser for 6: TAP_6_last_step_provenance

	entity_name	entity_location	entity_generated	...	entity_comment	generating_activity_name	generating_activity_starttime	generating_activity_end.
1	j90x37020_drc	https://www.cadc-ccda.hia-iha.nrc-cnrc.gc.ca/...	2018-12-05T02:19Z		Context image of Drizzled HST image from ACS...	j90x37020_drc_DrizzleGeneration	2018-12-05T02:19Z	2018-12-05T03:19Z
2	j90x37020_drc	https://www.cadc-ccda.hia-iha.nrc-cnrc.gc.ca/...	2018-12-05T02:19Z		Context image of Drizzled HST image from ACS...	j90x37020_drc_DrizzleGeneration	2018-12-05T02:19Z	2018-12-05T03:19Z
3	j90x37020_drc	https://www.cadc-ccda.hia-iha.nrc-cnrc.gc.ca/...	2018-12-05T02:19Z		Context image of Drizzled HST image from ACS...	j90x37020_drc_DrizzleGeneration	2018-12-05T02:19Z	2018-12-05T03:19Z
4	j90x37020_drc	https://www.cadc-ccda.hia-iha.nrc-cnrc.gc.ca/...	2018-12-05T02:19Z		Context image of Drizzled HST image from ACS...	j90x37020_drc_DrizzleGeneration	2018-12-05T02:19Z	2018-12-05T03:19Z
5	j90x37020_drc	https://www.cadc-ccda.hia-iha.nrc-cnrc.gc.ca/...	2018-12-05T02:19Z		Context image of Drizzled HST image from ACS...	j90x37020_drc_DrizzleGeneration	2018-12-05T02:19Z	2018-12-05T03:19Z
6	j90x37020_drc	https://www.cadc-ccda.hia-iha.nrc-cnrc.gc.ca/...	2018-12-05T02:19Z		Context image of Drizzled HST image from ACS...	j90x37020_drc_DrizzleGeneration	2018-12-05T02:19Z	2018-12-05T03:19Z
7	j90x37020_drc	https://www.cadc-ccda.hia-iha.nrc-cnrc.gc.ca/...	2018-12-05T02:19Z		Context image of Drizzled HST image from ACS...	j90x37020_drc_DrizzleGeneration	2018-12-05T02:19Z	2018-12-05T03:19Z
8	j90x37020_drc	https://www.cadc-ccda.hia-iha.nrc-cnrc.gc.ca/...	2018-12-05T02:19Z		Context image of Drizzled HST image from ACS...	j90x37020_drc_DrizzleGeneration	2018-12-05T02:19Z	2018-12-05T03:19Z
9	j90x37020_drc	https://www.cadc-ccda.hia-iha.nrc-cnrc.gc.ca/...	2018-12-05T02:19Z		Context image of Drizzled HST image from ACS...	j90x37020_drc_DrizzleGeneration	2018-12-05T02:19Z	2018-12-05T03:19Z
10	j90x37020_drc	https://www.cadc-ccda.hia-iha.nrc-cnrc.gc.ca/...	2018-12-05T02:19Z		Context image of Drizzled HST image from ACS...	j90x37020_drc_DrizzleGeneration	2018-12-05T02:19Z	2018-12-05T03:19Z
11	j90x37020_drc	https://www.cadc-ccda.hia-iha.nrc-cnrc.gc.ca/...	2018-12-05T02:19Z		Context image of Drizzled HST image from ACS...	j90x37020_drc_DrizzleGeneration	2018-12-05T02:19Z	2018-12-05T03:19Z
12	j90x37020_drc	https://www.cadc-ccda.hia-iha.nrc-cnrc.gc.ca/...	2018-12-05T02:19Z		Context image of Drizzled HST image from ACS...	j90x37020_drc_DrizzleGeneration	2018-12-05T02:19Z	2018-12-05T03:19Z

Table Browser for 6: TAP_6_last_step_provenance

	generating_activity_comment	agent_name	agent_role	agent_type	agent_affiliation	agent_email	agent_address
1	calibrated ACS HST images j90x37hpq_flc.fits[sci1] j90x37hpq_flc.fits[sci2] j90x37hrq_flc.fits[sci1] j90x37hrq_flc.fits[sci2]...	STScI	creator	Organization	AURA	stsci@stsci.edu	Martin Drive Baltimore
2	calibrated ACS HST images j90x37hpq_flc.fits[sci1] j90x37hpq_flc.fits[sci2] j90x37hrq_flc.fits[sci1] j90x37hrq_flc.fits[sci2]...	STScI	creator	Organization	AURA	stsci@stsci.edu	Martin Drive Baltimore
3	calibrated ACS HST images j90x37hpq_flc.fits[sci1] j90x37hpq_flc.fits[sci2] j90x37hrq_flc.fits[sci1] j90x37hrq_flc.fits[sci2]...	STScI	creator	Organization	AURA	stsci@stsci.edu	Martin Drive Baltimore
4	calibrated ACS HST images j90x37hpq_flc.fits[sci1] j90x37hpq_flc.fits[sci2] j90x37hrq_flc.fits[sci1] j90x37hrq_flc.fits[sci2]...	STScI	creator	Organization	AURA	stsci@stsci.edu	Martin Drive Baltimore
5	calibrated ACS HST images j90x37hpq_flc.fits[sci1] j90x37hpq_flc.fits[sci2] j90x37hrq_flc.fits[sci1] j90x37hrq_flc.fits[sci2]...	STScI	creator	Organization	AURA	stsci@stsci.edu	Martin Drive Baltimore
6	calibrated ACS HST images j90x37hpq_flc.fits[sci1] j90x37hpq_flc.fits[sci2] j90x37hrq_flc.fits[sci1] j90x37hrq_flc.fits[sci2]...	STScI	creator	Organization	AURA	stsci@stsci.edu	Martin Drive Baltimore
7	calibrated ACS HST images j90x37hpq_flc.fits[sci1] j90x37hpq_flc.fits[sci2] j90x37hrq_flc.fits[sci1] j90x37hrq_flc.fits[sci2]...	STScI	creator	Organization	AURA	stsci@stsci.edu	Martin Drive Baltimore
8	calibrated ACS HST images j90x37hpq_flc.fits[sci1] j90x37hpq_flc.fits[sci2] j90x37hrq_flc.fits[sci1] j90x37hrq_flc.fits[sci2]...	STScI	creator	Organization	AURA	stsci@stsci.edu	Martin Drive Baltimore
9	calibrated ACS HST images j90x37hpq_flc.fits[sci1] j90x37hpq_flc.fits[sci2] j90x37hrq_flc.fits[sci1] j90x37hrq_flc.fits[sci2]...	STScI	creator	Organization	AURA	stsci@stsci.edu	Martin Drive Baltimore
10	calibrated ACS HST images j90x37hpq_flc.fits[sci1] j90x37hpq_flc.fits[sci2] j90x37hrq_flc.fits[sci1] j90x37hrq_flc.fits[sci2]...	STScI	creator	Organization	AURA	stsci@stsci.edu	Martin Drive Baltimore
11	calibrated ACS HST images j90x37hpq_flc.fits[sci1] j90x37hpq_flc.fits[sci2] j90x37hrq_flc.fits[sci1] j90x37hrq_flc.fits[sci2]...	STScI	creator	Organization	AURA	stsci@stsci.edu	Martin Drive Baltimore
12	calibrated ACS HST images j90x37hpq_flc.fits[sci1] j90x37hpq_flc.fits[sci2] j90x37hrq_flc.fits[sci1] j90x37hrq_flc.fits[sci2]...	STScI	creator	Organization	AURA	stsci@stsci.edu	Martin Drive Baltimore

Table Browser for 6: TAP_6_last_step_provenance

	entity_name	entity_location	...	ag...	...	agent_comment	used_entity_name
1	j90x37020_drc	https://www.cadc-ccda.hia-iha.nrc-cnrc.gc.ca/data/pub/MAST/HST/product/j90x37020_drc.fits		A...		Space Telescope Science Institute	j90x37hpq_flc.fits[sci1]
2	j90x37020_drc	https://www.cadc-ccda.hia-iha.nrc-cnrc.gc.ca/data/pub/MAST/HST/product/j90x37020_drc.fits		A...		Space Telescope Science Institute	j90x37hpq_flc.fits[sci2]
3	j90x37020_drc	https://www.cadc-ccda.hia-iha.nrc-cnrc.gc.ca/data/pub/MAST/HST/product/j90x37020_drc.fits		A...		Space Telescope Science Institute	j90x37hrq_flc.fits[sci1]
4	j90x37020_drc	https://www.cadc-ccda.hia-iha.nrc-cnrc.gc.ca/data/pub/MAST/HST/product/j90x37020_drc.fits		A...		Space Telescope Science Institute	j90x37hrq_flc.fits[sci2]
5	j90x37020_drc	https://www.cadc-ccda.hia-iha.nrc-cnrc.gc.ca/data/pub/MAST/HST/product/j90x37020_drc.fits		A...		Space Telescope Science Institute	j90x37hpq_sci1_final_mask.fits
6	j90x37020_drc	https://www.cadc-ccda.hia-iha.nrc-cnrc.gc.ca/data/pub/MAST/HST/product/j90x37020_drc.fits		A...		Space Telescope Science Institute	j90x37hpq_sci2_final_mask.fits
7	j90x37020_drc	https://www.cadc-ccda.hia-iha.nrc-cnrc.gc.ca/data/pub/MAST/HST/product/j90x37020_drc.fits		A...		Space Telescope Science Institute	j90x37hrq_sci1_final_mask.fits
8	j90x37020_drc	https://www.cadc-ccda.hia-iha.nrc-cnrc.gc.ca/data/pub/MAST/HST/product/j90x37020_drc.fits		A...		Space Telescope Science Institute	
9	j90x37020_drc	https://www.cadc-ccda.hia-iha.nrc-cnrc.gc.ca/data/pub/MAST/HST/product/j90x37020_drc.fits		A...		Space Telescope Science Institute	
10	j90x37020_drc	https://www.cadc-ccda.hia-iha.nrc-cnrc.gc.ca/data/pub/MAST/HST/product/j90x37020_drc.fits		A...		Space Telescope Science Institute	j90x37hrq_sci1_final_mask.fits
11	j90x37020_drc	https://www.cadc-ccda.hia-iha.nrc-cnrc.gc.ca/data/pub/MAST/HST/product/j90x37020_drc.fits		A...		Space Telescope Science Institute	j90x37hpq_sci2_final_mask.fits
12	j90x37020_drc	https://www.cadc-ccda.hia-iha.nrc-cnrc.gc.ca/data/pub/MAST/HST/product/j90x37020_drc.fits		A...		Space Telescope Science Institute	j90x37hpq_sci1_final_mask.fits

Solutions
2- denormalization

- Adding « description » classes attributes to the « execution » classes.
  - This is another permanent view
  - A lot of redundancy.

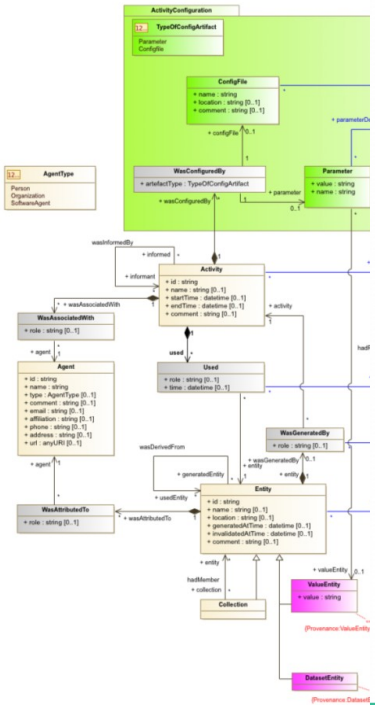


Figure 8: Full class diagram of the IVO

Solutions
2- denormalization

- Full entity :
  - entity_name
  - entity_location
  - entity_comment
  - entity_generated
  - entity_invalidated
  - entitydescription_content
  - entitydescription_type
  - entitydescription_description
  - entitydescription_docurl

Solutions
2- denormalization

- Full activity :
  - activity_name
  - activity_comment
  - activity_starttime
  - activity_endtime
  - activitydescription_description
  - activitydescription_type
  - activitydescription_docurl
- And parameter/parameterdescription

Solutions

3 - Simplification of « descriptions » « linkage »

- Suppress UsageDescription and GenerationDescription ?

- May introduce other difficulties ?

How to go on ?

- Keep full ProvTAP schema and propose views.
- Make these views become simplified TAP schemata part of the standard.
- Limited or full services possibility (implementing only views or the whole TAP schema)
- Upgrade the specification and open prototype