# ProvTAP et provenance minimale
## Grâce aux vues sur la BDD

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

# 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.

# 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 implementing 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 ?
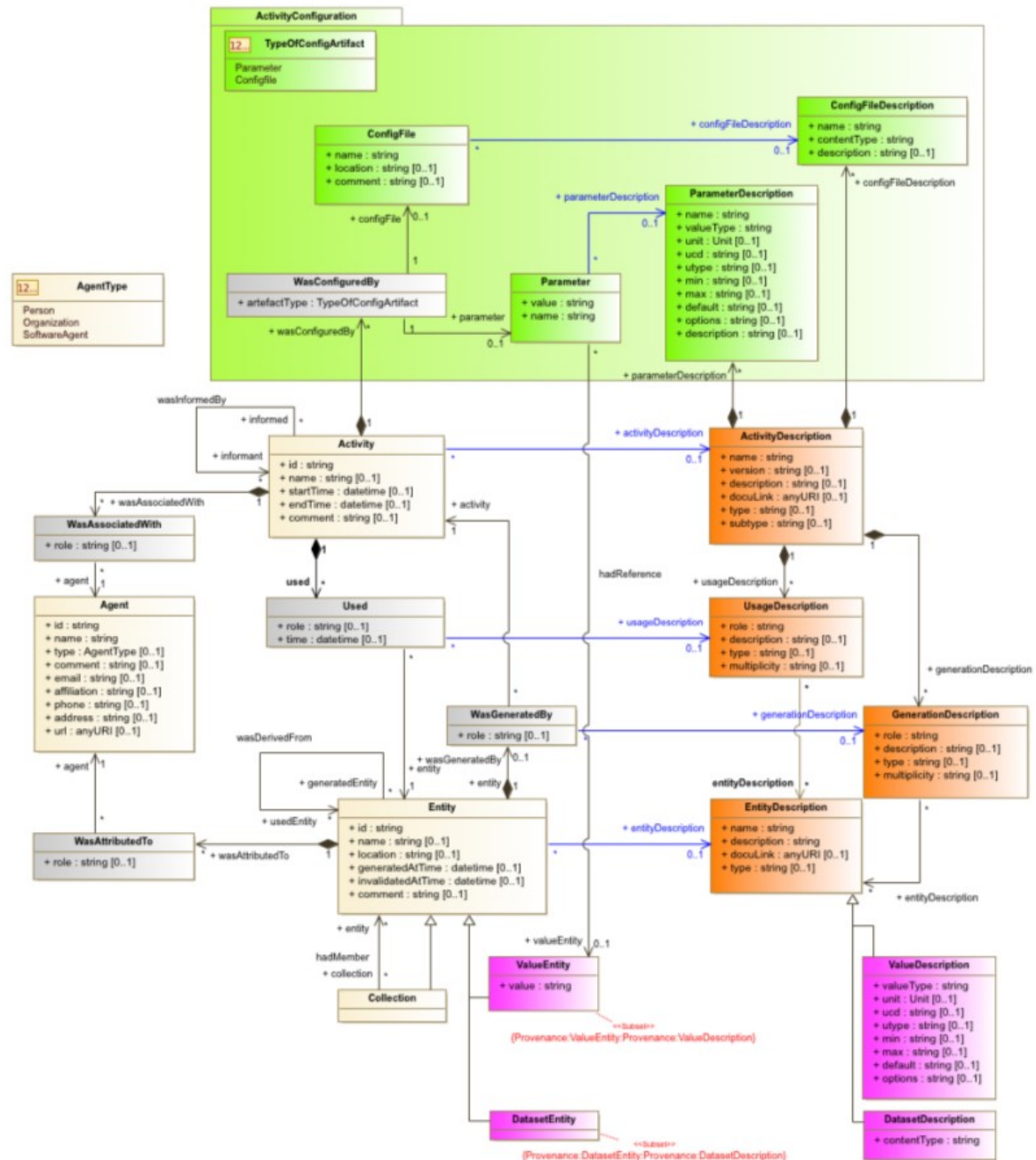    - Depends from project requirements and uses cases

# 1 table per Class ?



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

| Select What | Where | Position | Plain Text Query | Job Control |
|---|---|---|---|---|

```
SELECT e.e_name as "Tile", a.a_name as "Tile generation activity",
ee.e_name as "drizzled image", ee.e_location as "drizzled image URL", aa.a_name as "Drizzling activity",
eee.e_name as "calibrated image",eee.e_location as "calibrated image URL", aaa.a_name as "calibration",
eeee.e_name as "raw image"
FROM provenance.entity e
join provenance.wasgeneratedby w on w.wgb_entity = e.e_id
join provenance.activity a on a.a_id = w.wgb_activity
join provenance.used u on u.u_activity = a.a_id
join provenance.entity ee on ee.e_id = u.u_entity
join provenance.wasgeneratedby ww on ww.wgb_entity = ee.e_id
join provenance.activity aa on aa.a_id = ww.wgb_activity
join provenance.used uu on uu.u_activity = aa.a_id
join provenance.entity eee on eee.e_id = uu.u_entity
join provenance.wasgeneratedby www on www.wgb_entity = eee.e_id
join provenance.activity aaa on aaa.a_id = www.wgb_activity
join provenance.used uuu on uuu.u_activity = aaa.a_id
join provenance.entity eeee on eeee.e_id = uuu.u_entity
where e.e_name = 'HST_V_Order10_Npix8530404'
```

## Tile to Obs scenario
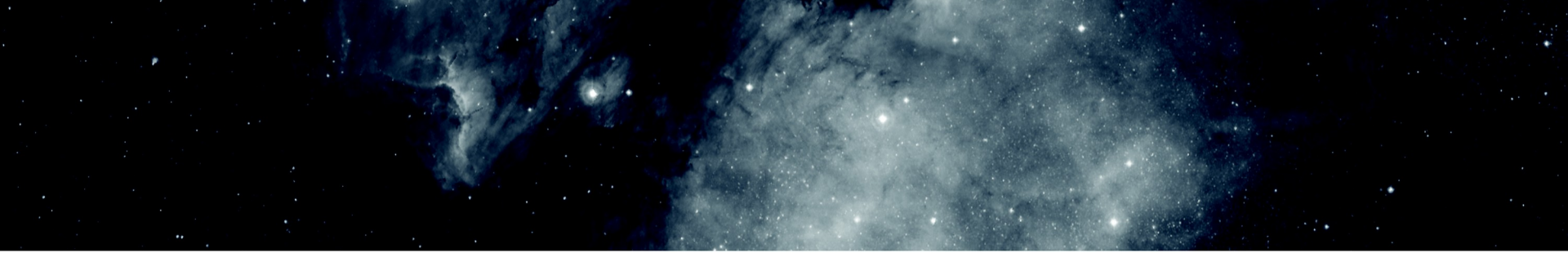
| Tile | Tile generation activity | drizzled image | drizzled image URL | Drizzling activity | calibrated image | calibrated image URL | calibration | raw image |
|---|---|---|---|---|---|---|---|---|
| HST_V_Order10_Npix85304 ... | HST_V_Order10_Npix85304 ... | j8uq70031_drz | | j8uq70031_drz_DrizzleGe ... | j8uq70qnq_flt.fits[sci1] | | j8uq70qnq_flt.fits_Cali ... | j8uq70qnq_flt.fits[sci1 ... |
| HST_V_Order10_Npix85304 ... | HST_V_Order10_Npix85304 ... | j8uq70031_drz | | j8uq70031_drz_DrizzleGe ... | j8uq70qoq_flt.fits[sci1] | | j8uq70qoq_flt.fits_Cali ... | j8uq70qoq_flt.fits[sci1 ... |
| HST_V_Order10_Npix85304 ... | HST_V_Order10_Npix85304 ... | j8uq70011_drz | | j8uq70011_drz_DrizzleGe ... | j8uq70qkq_flt.fits[sci1] | | j8uq70qkq_flt.fits_Cali ... | j8uq70qkq_flt.fits[sci1 ... |

drizzled progenitors

calibrated images
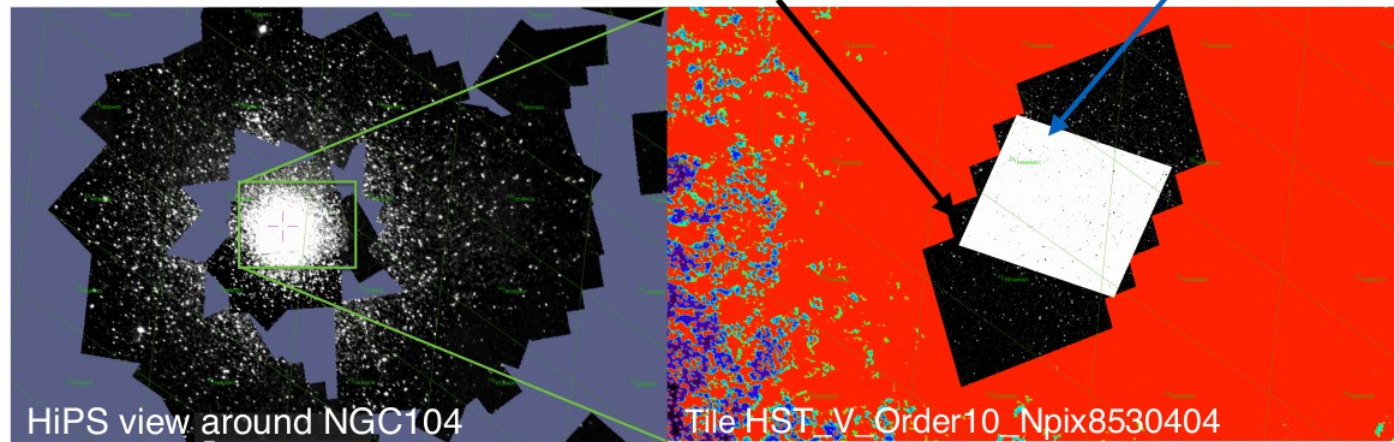
HiPS view around NGC104

Tile HST_V_Order10_Npix8530404

a_name=j8uq70qnq_flt.fits_Calibration
a_comment="obtained with HST ACS at target NGC104 with filters POL120UV and F330W using configuration profile j8uq70qnq_flt.fits.prof"
a_startTime=2018-06-02T00:00:00
a_endTime=2018-06-02T00:00:00

to its ActivityDescription

ad_name =HST_CALACS_Activity
ad_type=Calibration
ad_subtype=Photometric Calibration
ad_description=HSTACS calibration activity
ad_doculink= www.stsci.edu/hst/instrumentation/acs/calibration

# 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
## 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 |

# • View (in postgres)

create view minimum_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,
 string_agg(u_entity,',') as used_entities_list 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 ;*

**GROUP BY  e.e_id, e.e_name, e.e_location, e.e_generated, e.e_invalidated, e.e_comment, activity.a_name, activity.a_starttime, activity.a_endtime, activity.a_comment, wasattributedto.wat_role, agent.ag_name, agent.ag_type, agent.ag_affiliation, agent.ag_email, agent.ag_address, agent.ag_phone , agent.ag_comment**

Window  TAP  Registry  Edit  Interop  Help

Select Service | Use Service | Resume Job | Running Jobs

**Metadata**

Find:

☑ Name  ☐ Descrip  Or

○ Service  ○ Schema  ○ Table  ○ Columns  ○ FKeys  Hints

- last_step_proven
- minimum_proven
- parameter
- parameterdescri
- usagedescription
- used
- valuedescription
- wasassociatedwi
- wasattributedto
- wasconfiguredby
- wasderivedfrom
- wasgeneratedby
- wasinformedby

Name:

Tables:

Description:

**Service Capabilities**

Query Language: ADQL-2.0 ▾  Max Rows: 1000000 (default) ▾  Uploads: unavailable

**ADQL Text**

Mode: Synchronous ▾

| 1 |

```
select entity_name, entity_comment, generating_activity_name,
agent_role,agent_name,used_entities_list  from minimum_provenance
```

Examples    Info

Run Query

---

TOPCAT(3): Table Browser

Window  Rows  Help

Table Browser for 3: TAP_3_minimum_provenance

| | entity_name | entity_comment | generating_activity_name | agent_r... | agent_... | used_entities_list |
|---|---|---|---|---|---|---|
| 1 | j90x37020_drc | Context image of Drizzled HST image from ACS with filter CLEAR1L and F435W centered on 273.56883363507 , 34.2090407699622 , with a position angle of 1... | j90x37020_drc_DrizzleGeneration | creator | STScI | flc_225341,flcw_225340,flcw_225342,flcw_225... |
| 2 | jcqt99020_drc | Context image of Drizzled HST image from ACS with filter CLEAR1L and F435W centered on 273.56883363507 , 34.2090407699622 , with a position angle of 1... | jcqt99020_drc_DrizzleGeneration | creator | STScI | flcw_227732,flcw_227732,flcw_227740,flc_227... |
| 3 | jca536010_drc | Context image of Drizzled HST image from ACS with filter CLEAR1L and F435W centered on 273.56883363507 , 34.2090407699622 , with a position angle of 1... | jca536010_drc_DrizzleGeneration | creator | STScI | flcw_230234,flc_230217,flc_230219,flc_23022... |
| 4 | jbyq07020_drc | Context image of Drizzled HST image from ACS with filter CLEAR1L and F435W centered on 273.56883363507 , 34.2090407699622 , with a position angle of 1... | jbyq07020_drc_DrizzleGeneration | creator | STScI | flcw_232870,flcw_232868,flcw_232870,flcw_23... |
| 5 | jbyp02020_drc | Context image of Drizzled HST image from ACS with filter CLEAR1L and F435W centered on 273.56883363507 , 34.2090407699622 , with a position angle of 1... | jbyp02020_drc_DrizzleGeneration | creator | STScI | flcw_233104,flcw_233106,flcw_233108,flcw_23... |
| 6 | j92qibgcq_drc | Context image of Drizzled HST image from ACS with filter CLEAR1L and F435W centered on 273.56883363507 , 34.2090407699622 , with a position angle of 1... | j92qibgcq_drc_DrizzleGeneration | creator | STScI | flc_237143,flcw_237140,flcw_237142,flcw_237... |
| 7 | jd9r22010_drc | Context image of Drizzled HST image from ACS with filter CLEAR1L and F435W centered on 273.56883363507 , 34.2090407699622 , with a position angle of 1... | jd9r22010_drc_DrizzleGeneration | creator | STScI | flcw_237526,flcw_237530,flcw_237524,flcw_23... |

Total: 7    Visible: 7    Selected: 0

**Table Access Protocol (TAP) Query**

Window  TAP  Registry  Edit  Interop  Help

Select Service | Use Service | Resume Job | Running Jobs

**Metadata**

Find:

☑ Name  ☐ Descrip  Or

○ Service  ● Schema  ○ Table  ○ Columns  ○ FKeys  Hints

- last_step_proven
- minimum_proven
- parameter
- parameterdescri
- usagedescription

Name:

Tables:

Description:

**Service Capabilities**

Query Language: ADQL-2.0 ▾   Max Rows: 1000000 (default) ▾   Uploads: unavailable

**ADQL Text**

Mode: Synchronous ▾

1

`select * from entity where e_id = 'flc_225341'`

Examples   Info ⧉

**Run Query**

---

**TOPCAT(2): Table Browser**

Window  Rows  Help

Table Browser for 2: TAP_2_entity

| e_id | e_name | e_type | e_rights | e_location | e_generated | ... | ... | e_classt... | e_value | e_descrip... |
|------|--------|--------|----------|------------|-------------|-----|-----|-------------|---------|--------------|
| 1 | flc_225341 | j90x37hpq_flc.fits[sci1] | Science_image | public | https://www.cadc-ccda.hia-iha.nrc-cnrc.gc.ca/data/pub/MAST/HST/product/j90x37hpq_flc.fi... | 2018-10-08T00:00:00 | ... | ... | dataset | | sccdescr |

Total: 1    Visible: 1    Selected: 0

---

**TOPCAT(2): Table Browser**

Window  Rows  Help

Table Browser for 2: TAP_2_entity

| e_id | e_name | e_type | e_rights | ... | e_generated | ... | e_comment | e_classt... | e_value | e |
|------|--------|--------|----------|-----|-------------|-----|-----------|-------------|---------|---|
| 1 | flc_225341 | j90x37hpq_flc.fits[sci1] | Science_image | public | ... | 2018-10-08T00:00:00 | ... | Raw image obtained with HST ACS and filters  and  around position 50.58159985383 -37.41356521926 at observation time 2004-12-18T04:15:53 | dataset | | s |

Total: 1    Visible: 1    Selected: 0