ProvTAP implementation - ProvHIPS

F.Bonnarel, CDS on behalf of M.Servillat, M.Louys, M.Nullmeier, M.Sanguillon, L.Michel, G.Mantelet, A.Egner, D.Durand

CENTRE DE DONNÉES ASTRONOMIQUES DE STRASBOURG

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

IVOA

ProvTAP

- ProvTAP specification is mapping IVOA Provenance data model onto the TAP schema.

- The TAP schema defines

  - table and columns names,

  - Datatypes,

  - Units,

  - Ucds

  - and utypes (model attributed id)

  - → for each model feature !

TAP SCHEMA = Entity table

- specification is

currently an internal

IVOA draft : significant

Changes end 2021

IVOA Provenance Table Access Protocol (ProvTAP)

Version 1.0

IVOA Working Draft 2021-12-21

Working group
DM

This version
http://www.ivoa.net/documents/ProvTAP/20211221

Latest version
http://www.ivoa.net/documents/ProvTAP

Previous versions

Author(s)
François Bonnarel, Mireille Louys, Markus Nullmeier, Michèle San-guillon, 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 schema tables and provides guidelines for implementing with TAP 1.1.

Status of this document

This is an IVOA Working Draft for review by IVOA members and other interested parties. It is a draft document and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use IVOA Working Drafts as reference materials or to cite them as other than "work in progress".

A list of current IVOA Recommendations and other technical documents can be found at http://www.ivoa.net/documents/.

ProvHiPS project

- Tracing provenance of HST HiPS and HiPS tiles

  - From HiPS tiles back to

    raw HST images,

  - through « drizzled » images

  - and « calibrated » images
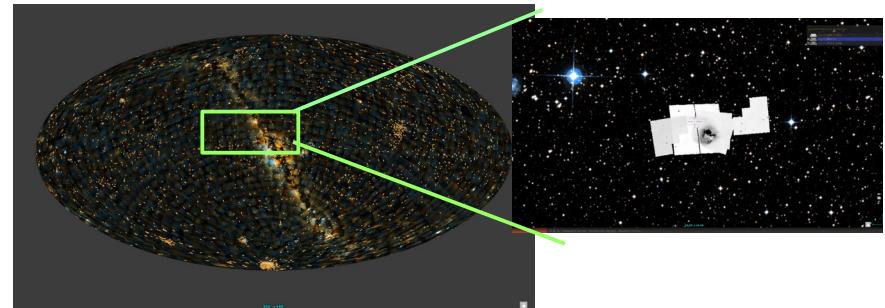
- ProvHiPS is a testbed for « On Top provenance » for image processing in optical astronomy.

  - See examples next slides

ProvHiPS project : prototype access / Try it and test it

Via TapHandle :https://saada.unistra.fr/taphandle/?url=https://saada.unistra.fr/provtap/#

Via new smarter TapHandle:
https://taphandle.astro.unistra.fr/tapcomplex/app/Tap_Handle_MK2/taphandev.html?url=//saada.unistra.fr/provtap

«HiPS» Provenance diagram
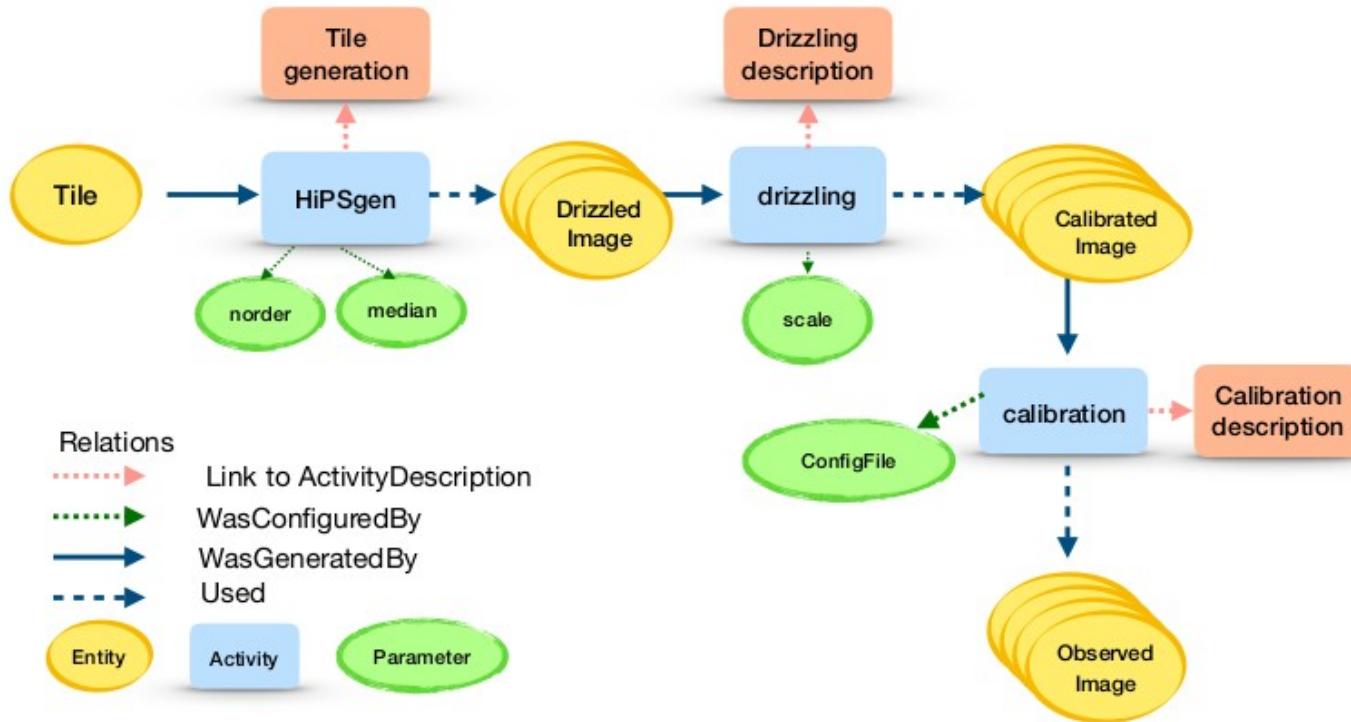
Provenance tracking for Prov-HiPS

«HiPS» tiles at order 10 history
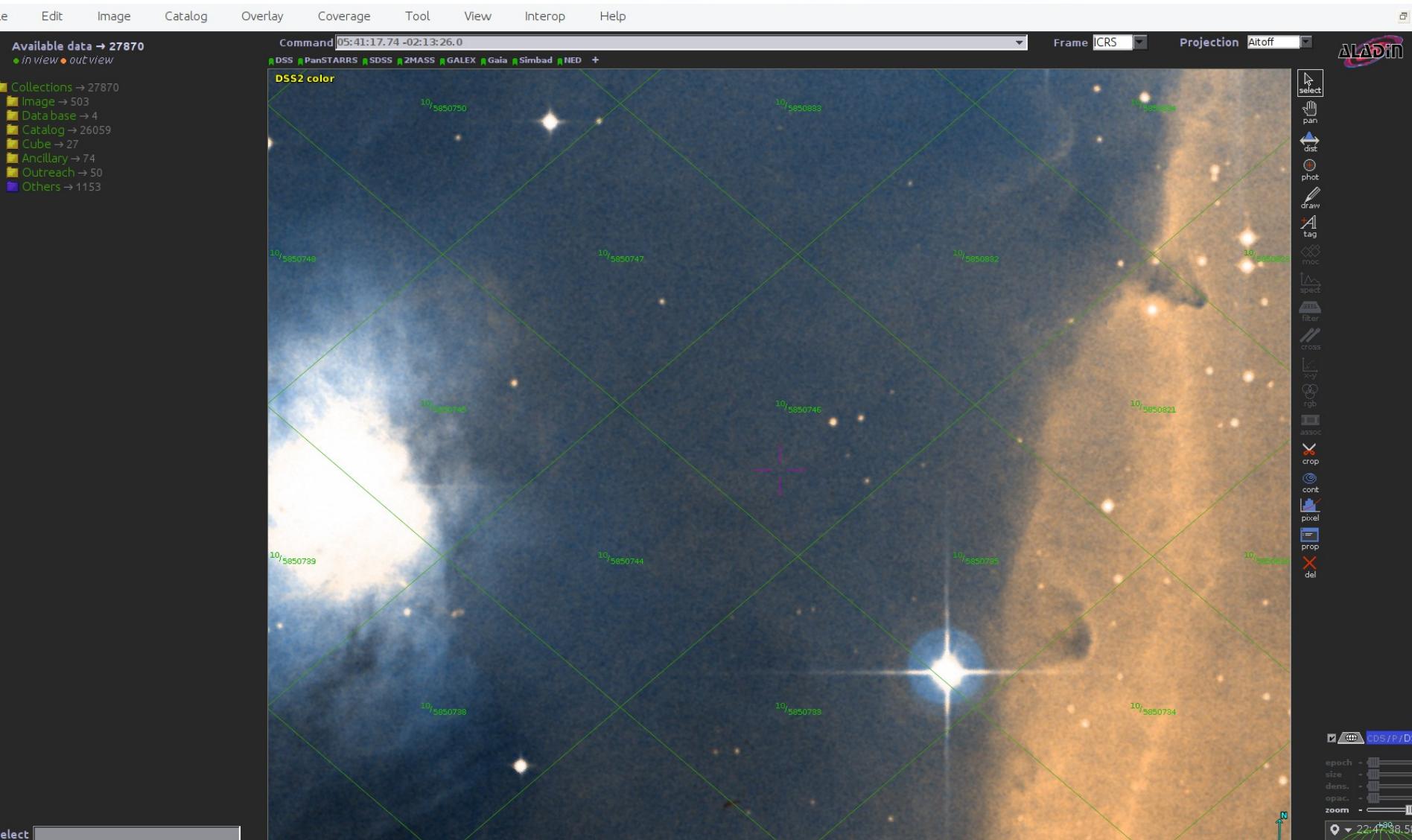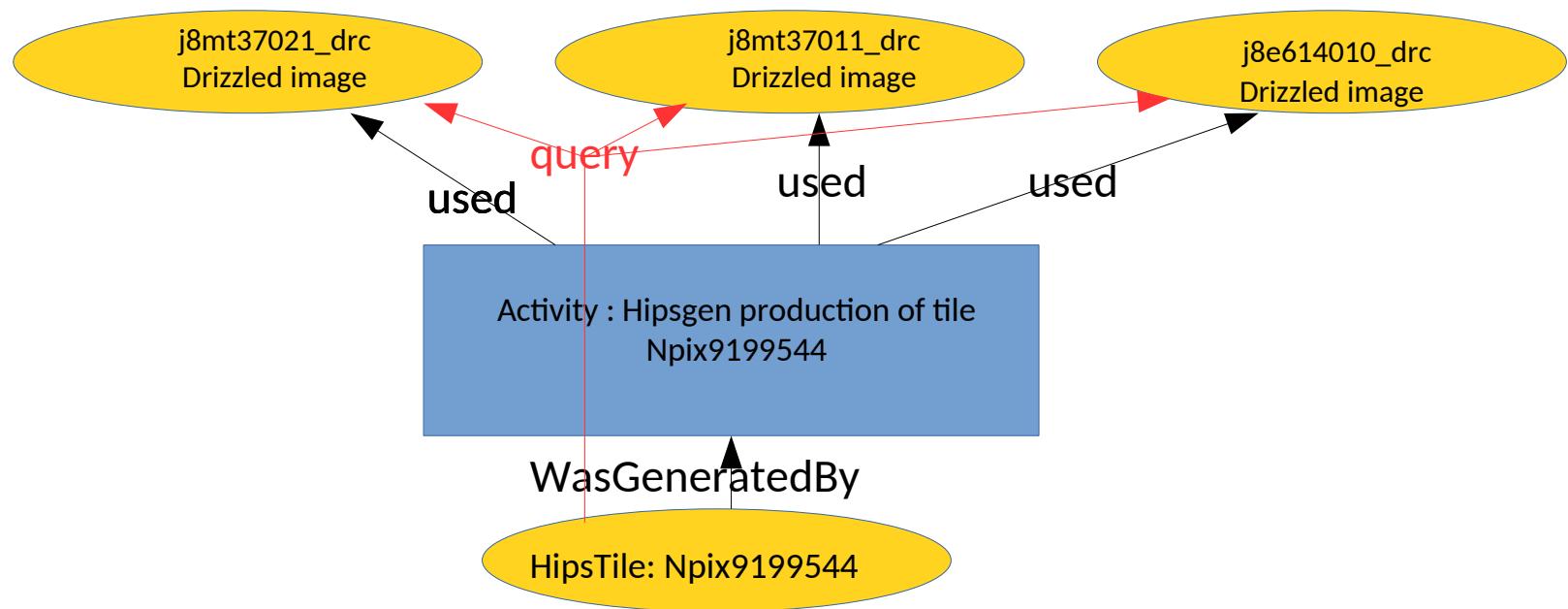
ProvHiPS ADQL query examples :
Finding out drizzled images « progenitors » of a specific HiPS tile.

select tile.e_name, tile.e_comment, hipsgen.a_name,

 hipsgen.a_starttime, hipsgen.a_comment, drizzle.e_name, drizzle.e_comment from
    entity tile

join wasgeneratedby on tile.e_id = wgb_entity

join activity hipsgen on wgb_activity = hipsgen.a_id

join used on hipsgen.a_id = u_activity

join entity drizzle   on drizzle.e_id = u_entity

where tile.e_name like '%Npix9199544'

ProvHiPS ADQL query examples :
## Finding out drizzled images « progenitors » of a specific HiPS tile.

ProvHiPS ADQL query examples :
Displaying some activities documentation for specific activities through their activity description

Calibration
Activity description
-name
- docURL
-etc...

query

Activity : calibration of
Image so and so ... ;

STScI is operating and takes pride in serving the astronomical community. Read **our response** to rapidly changing circumstances due to COVID-19 and stay informed about which STScI events have been **cancelled or rescheduled**.

CALIBRATION

HST Home   About   Proposing   Observing   Instrumentation   Documentation   News & Events

Home > Hubble Space Telescope > Instrumentation > Advanced Camera for Surveys

On This Page

Introduction | Calibration Plans

Introduction

Previously, data requested from MAST were generated using the "on the fly reprocessing" (OTFR) approach. However, with the growing size of the archive and the increasing complexity of calibration software the archive has switched to become a static one. This means that archived data are no longer reprocessed each time users submit requests to MAST. Instead, the data are only reprocessed when significant changes are made to either the calibration software (e.g. CALACS) or or new/improved reference files (e.g. DARKFILE, BIASFILE, etc..) have become available. As soon as these changes are implemented, the affected datasets are reprocessed and the status of their reprocessing is reported on the MAST processing status page ⧉. Once reprocessing has completed, users can download the freshly calibrated data products. For users who wish to reprocess their data locally, the ACS Team has created a Jupyter notebook that outlines this process starting from RAWs and ending with FLTs/FLCs and we strongly recommend you follow this guided example closely. You can access this notebook on the ACS Analysis Tools webpage.

select anyact.a_name, anyact.a_starttime, anyact.a_comment,

 ad_name, ad_description, ad_type, ad_subtype, ad_doculink

from activity anyact

join activitydescription on a_description = ad_id

ProvHiPS ADQL query examples :

Displaying some activities documentation for specific activities through their activity description

Finding agent and activity
to whom entity is attributed
and generating a given entity

- select ag_name,ag_type,ag_comment, e_name, e_generated,e_location, e_comment, a_name, a_comment

from agent

join wasattributedto on ag_id = wat_agent

join entity on wat_entity = e_id

join wasgeneratedby on e_id = wgb_entity

join activity on a_id = wgb_activity

where e_name = 'j8f627010_drz.fits'

Issues

- Table is denormalized : a lot of redundant information

- Loop issue : several occurances of the same triplet (name,utype,ucd) in the same table for different « objects »

- Let's try minimum or last step provenance by creating a standardized view

- 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_list

- → Redundancy may be avoided if we group all used entities ids in a single cell
- → possible Recursivity

- ## 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 ;

- ## View (in postgres)

create view minimum_provenance as select

e.e_id AS entity_id, 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,

 activity.a_name AS generating_activity_name, activity.a_starttime AS generating_activity_starttime,
activity.a_endtime AS generating_activity_endtime, activity.a_comment AS generating_activity_comment,

 wasattributedto.wat_role AS agent_role, agent.ag_name AS agent_name, agent.ag_type AS agent_type,
   agent.ag_affiliation AS agent_affiliation, agent.ag_email AS agent_email, agent.ag_address AS agent_address,
   agent.ag_phone AS agent_phone, agent.ag_comment AS agent_comment,

string_agg(used.u_entity::text, ','::text) AS used_entities_list
FROM entity e

 JOIN wasgeneratedby ON e.e_id::text = wasgeneratedby.wgb_entity::text
 JOIN activity ON wasgeneratedby.wgb_activity::text = activity.a_id::text
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 ; yypo

Solutions

Single step = view query execution

Entity + dataset descriptions

- Finding datasetdescription for entities used by an activity started in november 2018.

SELECT *

FROM datasetdescription

join entity on e_description = dd_id

join used on e_id = u_entity

join activity on a_id = u_activity

where a_starttime like '2018-11%'

Entity + dataset descriptions

full_entity view

```
create view provenance.full_entity AS
select entity.e_id AS full_entity_id,
entity.e_name AS full_entity_name,
entity.e_location AS full_entity_location,
entity.e_generated AS full_entity_generated,
entity.e_invalidated AS full_entity_invalidated,
entity.e_comment AS full_entity_comment,
datasetdescription.dd_name AS full_entity_desc_name,
datasetdescription.dd_type AS full_entity_desc_type,
datasetdescription.dd_subtype AS full_entity_desc_subtype,
datasetdescription.dd_doculink AS full_entity_desc_doculink
FROM entity
 JOIN datasetdescription ON datasetdescription.dd_id = entity.e_description
```

full_entity view

saada.unistra.fr/taphandle/?url=https://saada.unistra.fr/provtap/#

Mettre à jour

Applications   Importés dep...   Sites suggérés   Home   Services de l'É...   www.sweetpa...   MyPCBackup :...   Petit problèm...   driver 1394 - A...   https://docs.g...   Dropbox - Lie...   Téléchargement

Enter a TAP service URL or a keyword

Tap Nodes

- saadaunistrafrprovtap SJAU
  - TAP_SCHEMA
  - ivoa
  - provenance
    - entity
    - datasetdescription
    - full_entity
    - valuedescription
    - activity
    - activitydescription
    - agent
    - parameter
    - parameterdescription
    - used
    - usagedescription
    - wasgeneratedby
    - generationdescription
    - wasassociatedwith
    - wasattributedto
    - wasinformedby
    - wasderivedfrom
    - collection
    - configfile
    - minimum_provenance
    - last_step_provenance
    - configfiledescription
    - wasconfiguredby
    - history
  - Goodies (not used yet)

saadaunistrafrprovtap>provenance>entity>822

Show 25 entries   Showing 1 to 25 of 10,000 entries   Search

full_entity_id	full_entity_name	full_entity_location	full_entity_generated	full_entity_invalidated	full_entity_comment	full_entity_desc_name	full_entity_desc_type	full_entity_desc_su
t_1	HST_B_Norder10_Npix6977 ...	null	2019-01-09T02:34Z	null	Npix6977932 tile of HST ...	HST_B_Norder10	image/fits	hips_tile
t_2	HST_B_Norder10_Npix2924 ...	null	2019-01-09T02:34Z	null	Npix2924791 tile of HST ...	HST_B_Norder10	image/fits	hips_tile
t_3	HST_B_Norder10_Npix5943 ...	null	2019-01-09T02:34Z	null	Npix5943052 tile of HST ...	HST_B_Norder10	image/fits	hips_tile
t_4	HST_B_Norder10_Npix7415 ...	null	2019-01-09T02:34Z	null	Npix7415720 tile of HST ...	HST_B_Norder10	image/fits	hips_tile
t_5	HST_B_Norder10_Npix5145 ...	null	2019-01-09T02:34Z	null	Npix5145589 tile of HST ...	HST_B_Norder10	image/fits	hips_tile
t_6	HST_B_Norder10_Npix6978 ...	null	2019-01-09T02:34Z	null	Npix6978192 tile of HST ...	HST_B_Norder10	image/fits	hips_tile
t_7	HST_B_Norder10_Npix1092 ...	null	2019-01-09T02:34Z	null	Npix10922601 tile of HS ...	HST_B_Norder10	image/fits	hips_tile
t_8	HST_B_Norder10_Npix1555 ...	null	2019-01-09T02:34Z	null	Npix1555068 tile of HST ...	HST_B_Norder10	image/fits	hips_tile
t_9	HST_B_Norder10_Npix2283 ...	null	2019-01-09T02:34Z	null	Npix2283782 tile of HST ...	HST_B_Norder10	image/fits	hips_tile
t_10	HST_B_Norder10_Npix1249 ...	null	2019-01-09T02:34Z	null	Npix12494543 tile of HS ...	HST_B_Norder10	image/fits	hips_tile
t_11	HST_B_Norder10_Npix1204 ...	null	2019-01-09T02:34Z	null	Npix12044279 tile of HS ...	HST_B_Norder10	image/fits	hips_tile
t_12	HST_B_Norder10_Npix7204 ...	null	2019-01-09T02:34Z	null	Npix7204288 tile of HST ...	HST_B_Norder10	image/fits	hips_tile
t_13	HST_B_Norder10_Npix8534 ...	null	2019-01-09T02:34Z	null	Npix8534284 tile of HST ...	HST_B_Norder10	image/fits	hips_tile
drz_484	j8f627010_drz.fits	null	null	null	Drizzled HST image from ...	HST_Drizzled_Images_Des ...	data	coadd_calib
	j8uq10061_drz.fits	null	null	null	Drizzled HST image from ...	HST_Drizzled_Images_Des ...	data	coadd_calib

Hide query

SUBMIT   Select What   Where   Position   Plain Text Query   Job Control

Result Limit
100

select * from full_entity

Widgets do not reflect the query anymore after you modified it directly

History view (provHiPS specific)

create view provenance.history AS
select e.e_name AS tile,
e.e_comment AS tile_comment,
a.a_name AS tile_generation,
a.a_starttime AS tile_generation_time,
ee.e_name AS drizzle_progenitor,
ee.e_comment AS drizzle_comment,
aa.a_name AS drizzle_generation,
aa.a_starttime AS drizzle_generation_time,
eee.e_name AS calibrated_progenitor,
eee.e_comment AS calibrated_comment
FROM entity e
 JOIN wasgeneratedby w ON e.e_id= w.wgb_entity
 JOIN activity a ON a.a_id = w.wgb_activity
 JOIN used u ON u.u_activity = a.a_id
 JOIN entity ee ON ee.e_id = u.u_entity
 JOIN wasgeneratedby ww ON ww.wgb_entity = ee.e_id
 JOIN activity aa ON aa.a_id = ww.wgb_activity
 JOIN used uu ON uu.u_activity = aa.a_id
 JOIN entity eee ON eee.e_id = uu.u_entity

History view (provHiPS specific)

Future work

- Complete the database with more HiPS tiles and original images

- Define a view for workflow as a superactivity

- Add parameters and configuration to the database

- Collaborate for more ProvTAP implementation (CTA ? VizieR?)

- Release ProvTAP WD