
ProvTAP access to provenance

Metadata :

specification and ProvHiPS reference
implementation

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



TAP

- ProvTAP is a TAP service → what is TAP ?
- Most of astronomical archives and services rely on relational databases
 - Too much various DBMS remote interfaces and SQL-like languages
 - an issue for interoperability
 - TAP :
 - specification defining an interoperable interface layer on top of DBMS.
 - Using an universal SQL-like language adapted to astronomy : ADQL
 - TAP schema = describes available tables and columns
 - Example : Gaia catalog query



TAP service distributing Gaia at ESAVO

The image shows a screenshot of the TOPCAT software interface, which is used for accessing and querying astronomical data. The interface is divided into several windows:

- Table Browser for 2: TAP_2_gaiadr2.vari_long_period_variable**: This window displays a table with columns for solution_id, source_id, abs_mag, abs_mag_b, rsg_flag, bolometri., bolometri., frequency, and frequency_error. The table contains 55 rows of data.
- Table List**: This window shows a list of tables, including 'TAP_1_gaiadr2.vari_long' and 'TAP_2_gaiadr2.vari_long'.
- Current Table Properties**: This window displays properties for the selected table, such as Label, Location, Name, Rows, and Columns.
- Table Columns for 1: TAP_1_gaiadr2.vari_long_period_variable**: This window shows a list of columns with their indices, visibility, names, IDs, classes, units, and descriptions.
- Table Access Protocol (TAP) Query**: This window shows the query interface, including a 'Select Service' dropdown, a 'Query Language' dropdown (set to ADQL-2.0), and a text area for the query. The query is: `SELECT TOP 1000 * FROM gaiadr2.vari_long_period_variable where frequency > 0.01`.

The status bar at the bottom of the Table Browser window indicates: Total: 1 000, Visible: 1 000, Selected: 0.

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

The screenshot displays the TOPCAT software interface. The main window shows a metadata browser with a tree view on the left and a details pane on the right. The tree view is expanded to show the 'entity' table under the 'provenance' schema. The details pane shows the 'Name: provenance' and 'Description: Provenance schema'.

A 'Table Browser' window is open, displaying a table with 11 columns and 11 rows. The table is titled 'Table Browser for 3: TAP_3_TAP_SCHEMA.columns'. The columns are: column..., table..., column_name, datatype, arraysize, size, descri..., utype, unit, ucd, indexed, principal, and std. The rows contain data for each column, including values like 'entity', 'e_id', 'VARCHAR', and 'voprov:Entity.id'.

The 'Service Capabilities' section shows 'Query Language: ADQL-2.0', 'Max Rows: 1000000 (default)', and 'Uploads: unavailable'. The 'ADQL Text' section shows the query: 'select * from TAP_SCHEMA.columns where table_name = 'entity''. The 'Mode' is set to 'Synchronous'.

The 'Run Query' button is visible at the bottom of the interface.

ProvTAP specification

for IVOA provenance datamodel serialisation and
metadata service

- specification is currently an internal IVOA draft → WD Before interop



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)

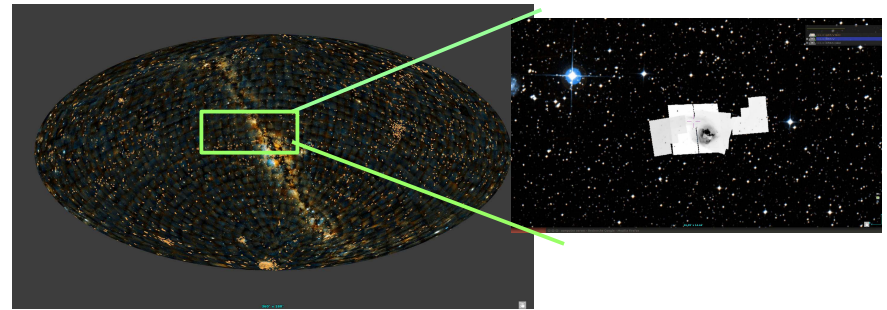
François Bonnarel

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.

ProvHiPS project

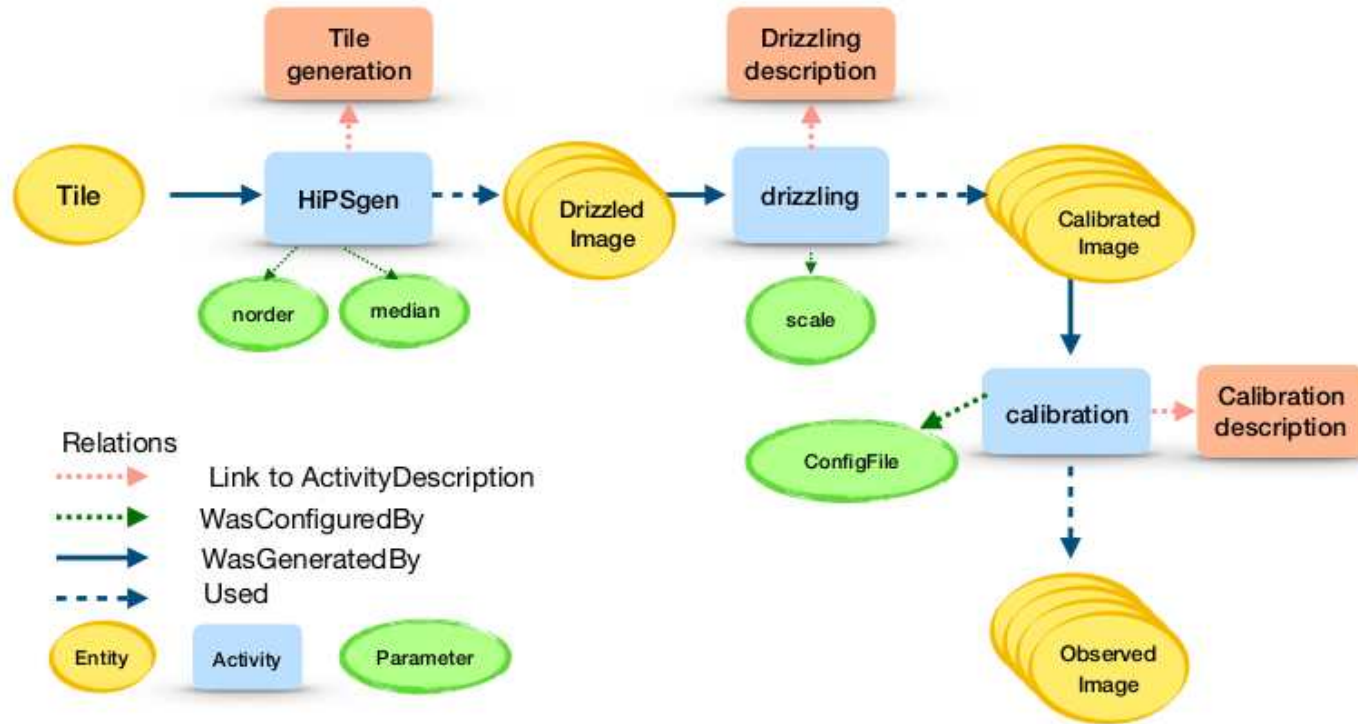
- Tracing **provenance of HiPS and HiPS tiles**
 - From **HiPS tiles** back to **raw HST images**,
 - through « **drizzled** » images
 - and « **calibrated** » images



- ProvHiPS is a testbed for « **On Top provenance** » for image processing in optical astronomy.
 - See examples next slides



«HiPS» Provenance diagram



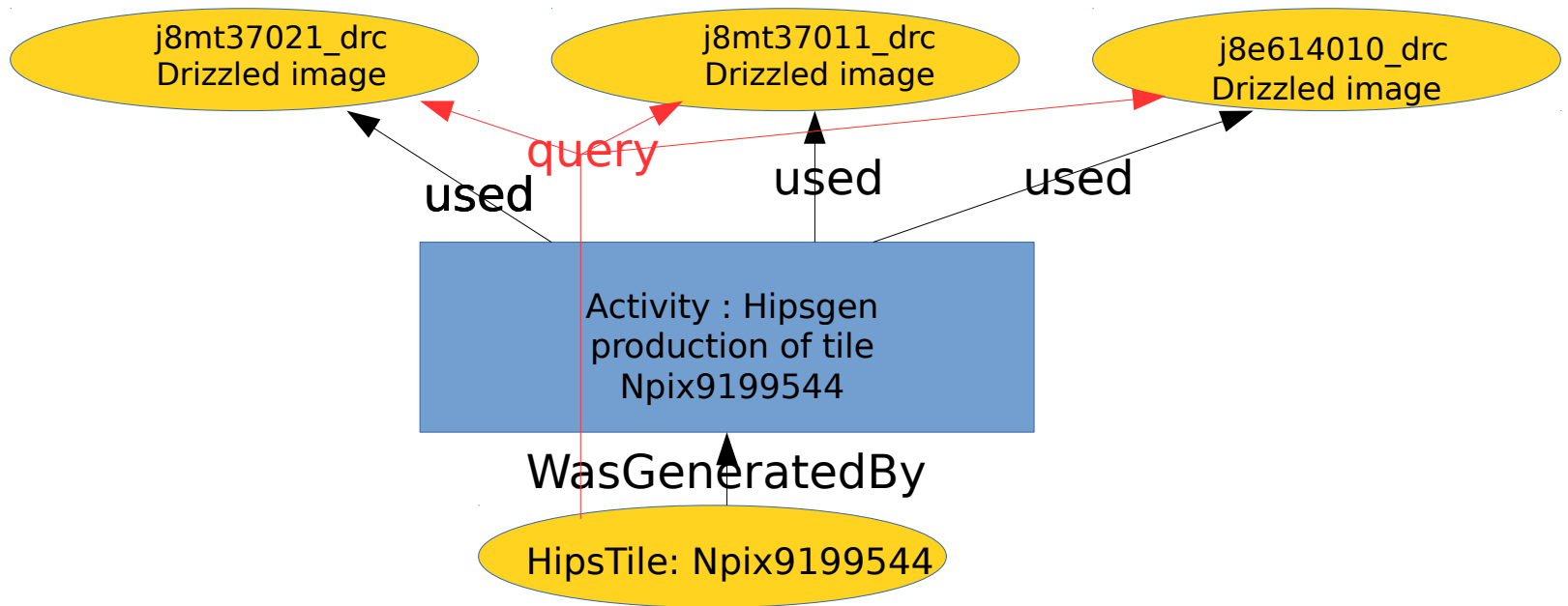
Itiples

Provenance tracking for Prov-HiPS



ProvHiPS ADQL query examples :

Finding out drizzled images « progenitors » of a specific HiPS tile.



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.

Window TAP Registry Edit Interop Help

TOPCAT(48): Table Browser

Table Browser for 48: TAP_54_entity,wasgeneratedby,activity,used,entit...

	e_name	e_comment	a_name	a_starttime	a_comment	e_name	e_comment
1	HST_V_Order10_Npix9199544	Npix9199544 tile of HST-V HiPS of size 12.25 a...	HST_V_Order10_Npix9199544 Generation	2019-01-09T02:34Z	hipsGEN version 10.101 generation of Npix91...	j8e614010_drc	Drizzled HST image from ACS centered on 53.0...
2	HST_V_Order10_Npix9199544	Npix9199544 tile of HST-V HiPS of size 12.25 a...	HST_V_Order10_Npix9199544 Generation	2019-01-09T02:34Z	hipsGEN version 10.101 generation of Npix91...	j8mt37011_drc	Drizzled HST image from ACS centered on 53.0...
3	HST_V_Order10_Npix9199544	Npix9199544 tile of HST-V HiPS of size 12.25 a...	HST_V_Order10_Npix9199544 Generation	2019-01-09T02:34Z	hipsGEN version 10.101 generation of Npix91...	j8mt37021_drc	Drizzled HST image from ACS centered on 52.9...
4	HST_V_Order10_Npix9199544	Npix9199544 tile of HST-V HiPS of size 12.25 a...	HST_V_Order10_Npix9199544 Generation	2019-01-09T02:34Z	hipsGEN version 10.101 generation of Npix91...	j8mt37031_drc	Drizzled HST image from ACS centered on 52.9...
5	HST_V_Order10_Npix9199544	Npix9199544 tile of HST-V HiPS of size 12.25 a...	HST_V_Order10_Npix9199544 Generation	2019-01-09T02:34Z	hipsGEN version 10.101 generation of Npix91...	j8mt38011_drc	Drizzled HST image from ACS centered on 53.0...
6	HST_V_Order10_Npix9199544	Npix9199544 tile of HST-V HiPS of size 12.25 a...	HST_V_Order10_Npix9199544 Generation	2019-01-09T02:34Z	hipsGEN version 10.101 generation of Npix91...	j8mt38021_drc	Drizzled HST image from ACS centered on 53.0...
7	HST_V_Order10_Npix9199544	Npix9199544 tile of HST-V HiPS of size 12.25 a...	HST_V_Order10_Npix9199544 Generation	2019-01-09T02:34Z	hipsGEN version 10.101 generation of Npix91...	j8mt38031_drc	Drizzled HST image from ACS centered on 52.9...
8	HST_V_Order10_Npix9199544	Npix9199544 tile of HST-V HiPS of size 12.25 a...	HST_V_Order10_Npix9199544 Generation	2019-01-09T02:34Z	hipsGEN version 10.101 generation of Npix91...	j8mt39011_drc	Drizzled HST image from ACS centered on 53.0...
9	HST_V_Order10_Npix9199544	Npix9199544 tile of HST-V HiPS of size 12.25 a...	HST_V_Order10_Npix9199544 Generation	2019-01-09T02:34Z	hipsGEN version 10.101 generation of Npix91...	j8mt39021_drc	Drizzled HST image from ACS centered on 53.0...
10	HST_V_Order10_Npix9199544	Npix9199544 tile of HST-V HiPS of size 12.25 a...	HST_V_Order10_Npix9199544 Generation	2019-01-09T02:34Z	hipsGEN version 10.101 generation of Npix91...	j8mt39031_drc	Drizzled HST image from ACS centered on 53.0...
11	HST_V_Order10_Npix9199544	Npix9199544 tile of HST-V HiPS of size 12.25 a...	HST_V_Order10_Npix9199544 Generation	2019-01-09T02:34Z	hipsGEN version 10.101 generation of Npix91...	j8mt39041_drc	Drizzled HST image from ACS centered on 53.0...
12	HST_V_Order10_Npix9199544	Npix9199544 tile of HST-V HiPS of size 12.25 a...	HST_V_Order10_Npix9199544 Generation	2019-01-09T02:34Z	hipsGEN version 10.101 generation of Npix91...	j8mt39051_drc	Drizzled HST image from ACS centered on 53.0...
13	HST_V_Order10_Npix9199544	Npix9199544 tile of HST-V HiPS of size 12.25 a...	HST_V_Order10_Npix9199544 Generation	2019-01-09T02:34Z	hipsGEN version 10.101 generation of Npix91...	j8mt39061_drc	Drizzled HST image from ACS centered on 53.0...
14	HST_V_Order10_Npix9199544	Npix9199544 tile of HST-V HiPS of size 12.25 a...	HST_V_Order10_Npix9199544 Generation	2019-01-09T02:34Z	hipsGEN version 10.101 generation of Npix91...	j8mt39071_drc	Drizzled HST image from ACS centered on 53.0...
15	HST_V_Order10_Npix9199544	Npix9199544 tile of HST-V HiPS of size 12.25 a...	HST_V_Order10_Npix9199544 Generation	2019-01-09T02:34Z	hipsGEN version 10.101 generation of Npix91...	j8mt39081_drc	Drizzled HST image from ACS centered on 53.0...
16	HST_V_Order10_Npix9199544	Npix9199544 tile of HST-V HiPS of size 12.25 a...	HST_V_Order10_Npix9199544 Generation	2019-01-09T02:34Z	hipsGEN version 10.101 generation of Npix91...	j8mt39091_drc	Drizzled HST image from ACS centered on 53.0...
17	HST_V_Order10_Npix9199544	Npix9199544 tile of HST-V HiPS of size 12.25 a...	HST_V_Order10_Npix9199544 Generation	2019-01-09T02:34Z	hipsGEN version 10.101 generation of Npix91...	j8mt39101_drc	Drizzled HST image from ACS centered on 52.9...
18	HST_V_Order10_Npix9199544	Npix9199544 tile of HST-V HiPS of size 12.25 a...	HST_V_Order10_Npix9199544 Generation	2019-01-09T02:34Z	hipsGEN version 10.101 generation of Npix91...	j8mt39111_drc	Drizzled HST image from ACS centered on 52.9...
19	HST_V_Order10_Npix9199544	Npix9199544 tile of HST-V HiPS of size 12.25 a...	HST_V_Order10_Npix9199544 Generation	2019-01-09T02:34Z	hipsGEN version 10.101 generation of Npix91...	j8mt39121_drc	Drizzled HST image from ACS centered on 52.9...

Total: 36 Visible: 36 Selected: 0

used
valuedescription
wasassociatedwith
unattributed

Service Capabilities

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

ADQL Text

Mode: Synchronous

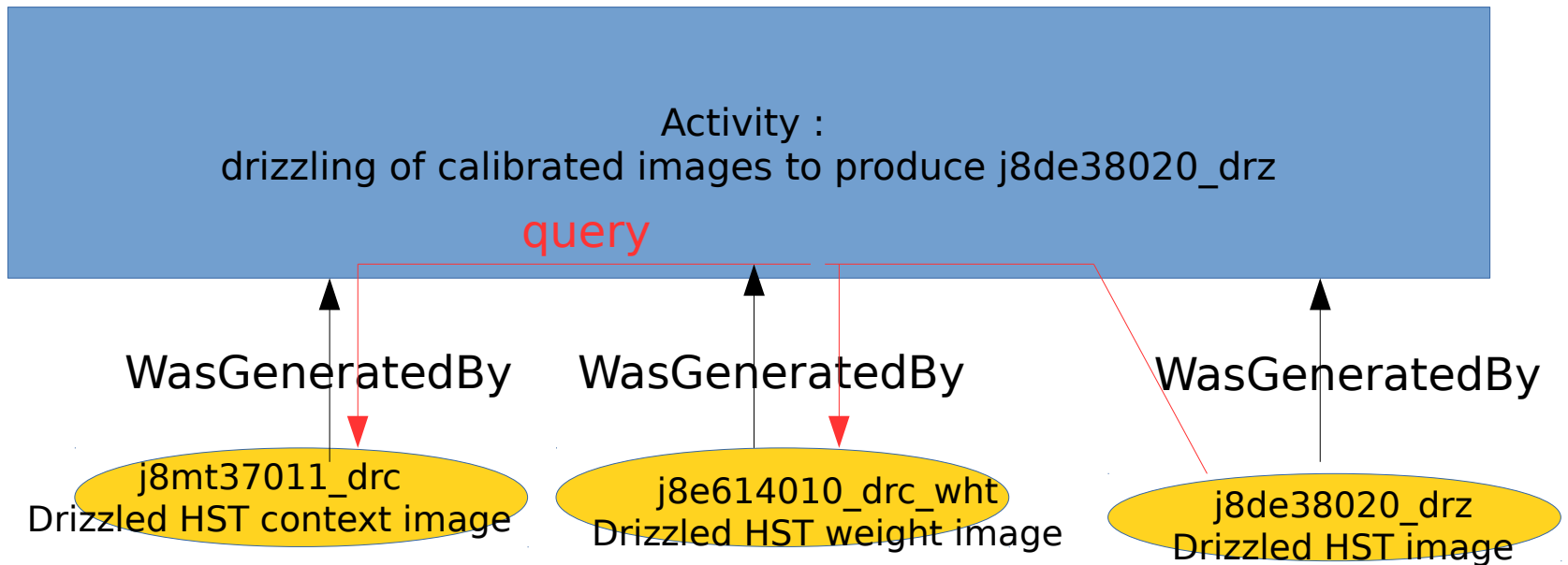
```

select ee.e_name, ee.e_comment, a_name, a_starttime, a_comment, ee.e_name, ee.e_comment from entity e
join wasgeneratedby on e.e_id = wgb_entity
join activity on wgb_activity = a_id
join used on a_id = u_activity
join entity ee on ee.e_id = u_entity
where e.e_name like '%Npix9199544'
    
```



ProvHiPS ADQL query examples :

Finding out other products of a drizzling activity which generated a specific drizzled image.



ProvHiPS ADQL query examples :

Finding out other products of a drizzling activity which generated a specific drizzled image

```
select drizzle.e_name, drizzle.e_comment,  
       drizzling.a_name, drizzling.a_starttime,  
       drizzleaux.e_type, drizzleaux.e_name, drizzleaux.e_comment  
from entity drizzle  
join wasgeneratedby w on drizzle.e_id = w.wgb_entity  
join activity drizzling on w.wgb_activity = drizzling.a_id  
join wasgeneratedby ww on drizzling.a_id = ww.wgb_activity join  
  entity drizzleaux on drizzleaux.e_id = ww.wgb_entity  
where drizzle.e_name = 'j8de38020_drz'  
and drizzleaux.e_id != drizzle.e_id
```



ProvHiPS ADQL query examples :

Finding out other products of a drizzling activity which generated a specific drizzled image

The screenshot shows the TOPCAT software interface. On the left, the 'Metadata' browser displays a tree view of 'TAP Service (25)' and 'provenance (20)'. The 'provenance' section is expanded, showing various entities like 'activity', 'agent', 'collection', etc. The main window shows the 'Table Browser' for the table 'TAP_58_entity,wasgeneratedby,activity,wasgenerat...'. The table contains two rows of data, both with 'e_name' 'j8de38020_drz' and 'e_comment' 'Drizzled HST image from ACS centered on 53.0...'. The table browser also shows columns for 'a_name', 'a_s...', 'e_type', and 'e_name'. The status bar at the bottom indicates 'Total: 2 Visible: 2 Selected: 0'. Below the table browser, the 'Service Capabilities' section shows 'Query Language: ADQL-2.0', 'Max Rows: 1000000 (default)', and 'Uploads: unavailable'. The 'ADQL Text' section shows a query:

```
1
select e.e_name, e.e_comment, a_name, a_starttime, ee.e_type, ee.e_name, ee.e_comment from entity e
join wasgeneratedby w on e.e_id = w.wgb_entity
join activity on w.wgb_activity = a_id
join wasgeneratedby ww on a_id = ww.wgb_activity
join entity ee on ee.e_id = ww.wgb_entity
where e.e_name = 'j8de38020_drz' and ee.e_id != e.e_id
```



ProvHiPS ADQL query examples :

Displaying some activities documentation for specific activities through their activity description

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

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

query

STScl 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 STScl 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 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 FLT/FLCs and we strongly recommend you follow this guided example closely. You can access this notebook on the [ACS Analysis Tools](#) webpage.



ProvHiPS ADQL query examples :

Displaying some activities documentation for specific activities through their activity description

```
select anyact.a_name, anyact.a_starttime,  
       anyact.a_comment,  
       ad_name, ad_description, ad_type, ad_subtype,  
       ad_docurl  
from activity anyact  
join activitydescription on a_description = anycat.ad_id
```



ProvHiPS ADQL query examples :

Displaying some activities documentation for specific activities through their activity description

TOPCAT(56): Table Browser

Window Rows Help

Table Browser for 56: TAP_64_activity,activitydescription

	a_name	a_starttime	a_comment	ad_name	ad_description	ad_type	ad_subtype	ad_doculink
1	jcx15r1q_fit.fits_Calibration	2018-10-26T00:00:00	Calibration of image jcx15r1q_fit.fits obtaine...	HST_CALACS_Activity	HST_ACS_Calibration activity	Calibration	PhotometricCalibration	http://www.stsci.edu/hst/instrumentation/acs/calibration
2	j8mbncsfq_fit.fits_Calibration	2018-05-31T00:00:00	Calibration of image j8mbncsfq_fit.fits obtaine...	HST_CALACS_Activity				
3	jdgq04c7q_fit.fits_Calibration	2018-10-27T00:00:00	Calibration of image jdgq04c7q_fit.fits obtaine...	HST_CALACS_Activity				
4	jca531heq_fit.fits_Calibration	2018-10-22T00:00:00	Calibration of image jca531heq_fit.fits obtaine...	HST_CALACS_Activity				
5	j8f0a1b4q_fit.fits_Calibration	2017-11-02T00:00:00	Calibration of image j8f0a1b4q_fit.fits obtaine...	HST_CALACS_Activity				
6	j90znag1q_fit.fits_Calibration	2018-06-03T00:00:00	Calibration of image j90znag1q_fit.fits obtaine...	HST_CALACS_Activity				
7	j91y01p7s_fit.fits_Calibration	2018-06-04T00:00:00	Calibration of image j91y01p7s_fit.fits obtaine...	HST_CALACS_Activity				
8	jd851ohq_fit.fits_Calibration	2018-07-02T00:00:00	Calibration of image jd851ohq_fit.fits obtaine...	HST_CALACS_Activity				
9	jbevelv4q_fit.fits_Calibration	2018-10-15T00:00:00	Calibration of image jbevelv4q_fit.fits obtaine...	HST_CALACS_Activity				
10	j6mi10c7q_fit.fits_Calibration	2017-11-01T00:00:00	Calibration of image j6mi10c7q_fit.fits obtaine...	HST_CALACS_Activity				
11	jcat11bnq_fit.fits_Calibration	2018-10-22T00:00:00	Calibration of image jcat11bnq_fit.fits obtaine...	HST_CALACS_Activity				
12	jcc210e1q_fit.fits_Calibration	2018-10-22T00:00:00	Calibration of image jcc210e1q_fit.fits obtaine...	HST_CALACS_Activity				
13	j96012q2q_fit.fits_Calibration	2018-06-05T00:00:00	Calibration of image j96012q2q_fit.fits obtaine...	HST_CALACS_Activity				
14	02c14512j_npl.fits_Calibration	2018-06-04T00:00:00	Calibration of image 02c14512j_npl.fits obtaine...	HST_CALACS_Activity				
15	j90o48b9q_fit.fits_Calibration	2018-06-03T00:00:00	Calibration of image j90o48b9q_fit.fits obtaine...	HST_CALACS_Activity				
16	j9h522zq_fit.fits_Calibration	2018-10-12T00:00:00	Calibration of image j9h522zq_fit.fits obtained...	HST_CALACS_Activity				
17	jdkb13fcq_fit.fits_Calibration	2018-10-28T00:00:00	Calibration of image jdkb13fcq_fit.fits obtaine...	HST_CALACS_Activity				
18	jbwjccbeq_fit.fits_Calibration	2018-10-19T00:00:00	Calibration of image jbwjccbeq_fit.fits obtaine...	HST_CALACS_Activity				
19	j8y2leqz_fit.fits_Calibration	2017-11-05T00:00:00	Calibration of image j8y2leqz_fit.fits obtained...	HST_CALACS_Activity				

Total: 10 373 Visible: 10 373 Selected: 1

TAP_SCHEMA.schemas
TAP_SCHEMA.tables
provenance (20)
activity
activitydescription
agent
collection
configfile
configfiledescription

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

ADQL Text
Mode: Synchronous

```
select a_name, a_starttime, a_comment, ad_name, ad_description, ad_type, ad_subtype, ad_doculink from activity
join activitydescription on a_description = ad_id
```

Examples

Calibration - Mozilla Firefox

Calibration

https://www.stsci.edu/hst/instrumentation/acs/calibration

STScI SPACE TELESCOPE SCIENCE INSTITUTE

STScI is operating and takes pride in serving the astronomical community. Read [our response](#) to rapidly changing circumstances. **rescheduled**

TOPCAT(56): Activation Actions

Window Actions Help

Activation Actions for 56: TAP_64_activity,activitydescription

Actions

- Send HiPS cutout
- Display image
- Display image region
- Load Table
- Plot Table
- Send FITS Image
- Send Spectrum
- Download URL
- View in Web Browser
- Delay
- Execute code
- Run system command
- Send row index

Description
Load an associated resource into a web browser

Configuration
Resource Identifier: ad_doculink

Identifier Type: auto
Browser: system browser

Status
Invoke now on row 1

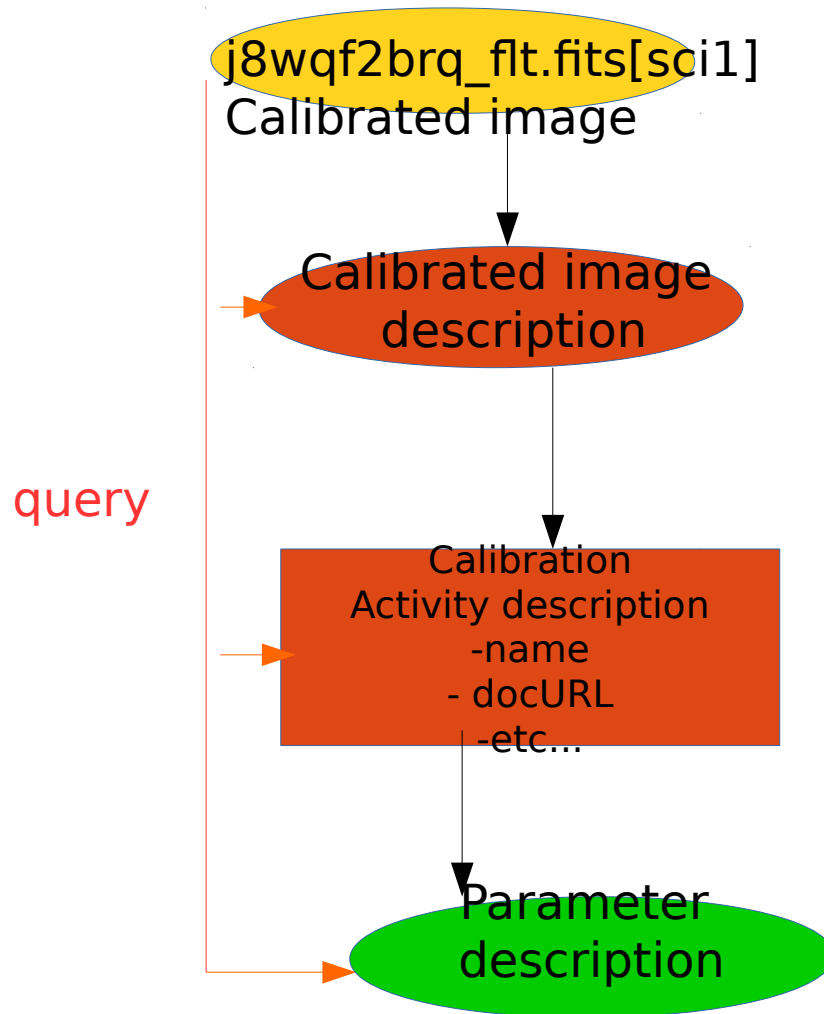
Results

Seq	Row	Status	Message
1	1	OK	http://www.stsci.edu/hst/instrumentation/acs/calibration
2	1	OK	http://www.stsci.edu/hst/instrumentation/acs/calibration
3	1	OK	http://www.stsci.edu/hst/instrumentation/acs/calibration
4	1	OK	http://www.stsci.edu/hst/instrumentation/acs/calibration
5	1	OK	http://www.stsci.edu/hst/instrumentation/acs/calibration

On This Page
Introduction | Calibration Pl

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_fit.fits[sci1]' ;
```



ProvHiPS ADQL query examples :

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

The screenshot shows the TOPCAT software interface. The main window displays a metadata browser with a tree view on the left and a table browser window on the right. The table browser window shows a table with 10 columns and 3 rows of data.

Table Browser for 8: TAP_10_entity,datasetdescription,generationdescr...

	e_name	pd_activityd...	pd_id	pd_name	pd_description	pd_do...	pd_val...	pd_unit	pd_ucd	pd_uty...	pd_min	pd_max	pd_def...	pd_opt...
1	j8wqf2brqflt.fits[sc1]	calibAdeschr	pd_1002	pftfile	pixel to pixel flat field file name		char		meta.file:obs.calib.flat					
2	j8wqf2brqflt.fits[sc1]	calibAdeschr	pd_1001	biasfile	bias image file name		char		meta.file:obs.calib.bias					
3	j8wqf2brqflt.fits[sc1]	calibAdeschr	pd_1000	darkfile	dark image file name		char		meta.file:obs.calib.dark					

ADQL Text

Mode: Synchronous

```
select top 3 e_name,parameterdescription.* 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 = 'j8wqf2brqflt.fits[sc1]';
```

Run Query