Implementation feedback of the IVOA Provenance data model



Goal

Following the IVOA Provenance data model (ivoa.net/Documents/DM/ Provenance) we have implemented a PROV-TAP service tracing the provenance of HiPS representations of the HST data products based on the metadata and knowledge provided by the Canadian Astronomical Data Center (CADC).

CDS Prov-HiPS Database

A PostGres database tracing the processing of HST image data sets has been set-up. Entities as defined in the IVOA DM instantiate various HST data products: raw images, calibrated then drizzled images and finally the 10 order HiPS tiles within the HST's HiPS collection. Corresponding transformations of data are stored in the Activity table and bound to their ActivityDescription metainformation explaining which kind of computation was performed. The Parameter values used to run these Activities are also available and queriable from the Parameter table.

We present the building blocks of this service and illustrate how queries can be submitted and how results can be interpreted.

Due to the complexity of queries with ADQL in the relational database we tested various tree-based solutions with collaborators: TripleStore, Common Table Extensions, and an adhoc 'graph upon table' mapping.



Provenance metadata have been extracted from the FITS Headers of HST data products and compiled in order to rebuild the historical track of data production steps in details.

The database is accessed through a TAP/ADQL query interface, based on the PROV-TAP XML schema (https://wiki.ivoa.net/twiki/bin/view/IVOA/ProvTAP).



Provenance tracking for Prov-HiPS

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'



Query response as VOTable in TAPHandle

| | | | | | 1 | | | (|
|-------------------------|--------------------------|----------------|--------------------|-------------------------|--------------------------|----------------------|-------------------------|-------------------------|
| 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 | چ اللہ 1 | 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] | i 🛃 📜 🥌 | 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 |
| | | drizzleo | d progenitors | • | calibrate | d images | | |



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

- This work highlights how provenance metadata can be • extracted from existing data collections to a posteriori build provenance services
- represented, named and structured following the IVOA Provenance DM
- queried and accessed through the Prov-TAP protocol interpreted with VO applications (TapHandle, Topcat, TapLibrary native web interface, Aladin)

Finally, with IVOA PROV-DM definitions and the TAP protocol we have been able to:

• track back execution flows run in a former project to generate data • factorize the description of various processing steps and structure processing knowledge within ActivityDescription templates convey configuration parameters and their description in order to foster reproducibility



Mireille Louys, F. Bonnarel, D. Durand, A. Egner

mireille.louys@astro.unistra.fr







Université