EPN-TAP Support at the PPI Node of the PDS

S. Joy[1], I. Moon[1], R. Walker[1], J. Mafi[1], B. Cecconi[2], S. Erard[2]

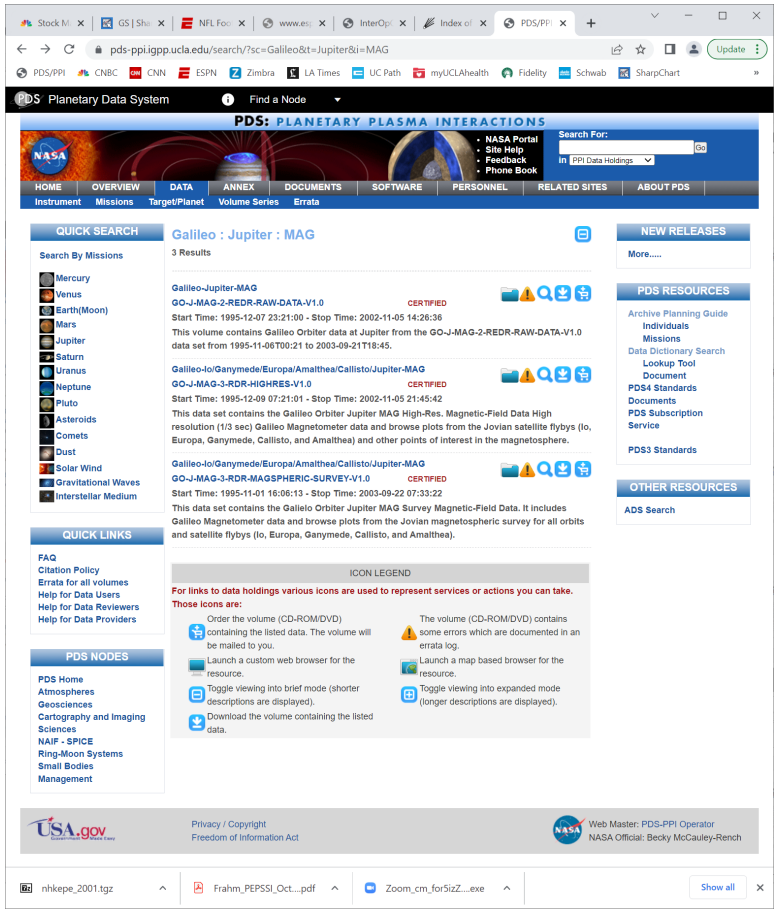
1. Dept. of Earth, Planetary, and Space Sciences – University of California at Los Angeles
2. Observatotoire de Paris - Meudon, France

About PPI and PDS

- The NASA Planetary Data System (PDS) is responsible for archiving data from all NASA missions after 1990, and as many of the earlier missions as possible
  - PDS is a federation of distributed science discipline nodes and support nodes – no central data repository
  - PDS is both an archive and a data system but its primary requirement is to be an archive
    - PDS requires that data are archived in stable formats (ASCII tables, FITS, CDF/A)
    - Data compression is not allowed
    - All of the bytes in a data file must be describable in the PDS metadata – data access is not dependent on any particular software suite
- The Planetary Plasma Interactions (PPI) Node of the PDS archives planetary fields and particles (F&P) data and interplanetary data that are acquired by planetary missions
  - There is no clear division between planetary and interplanetary (Heliophysics) F&P datasets other than proximity to a planet. Both disciplines use similar instruments and data formats.
  - The PDS Small Bodies Node has a similar affinity with the Astrophysics community

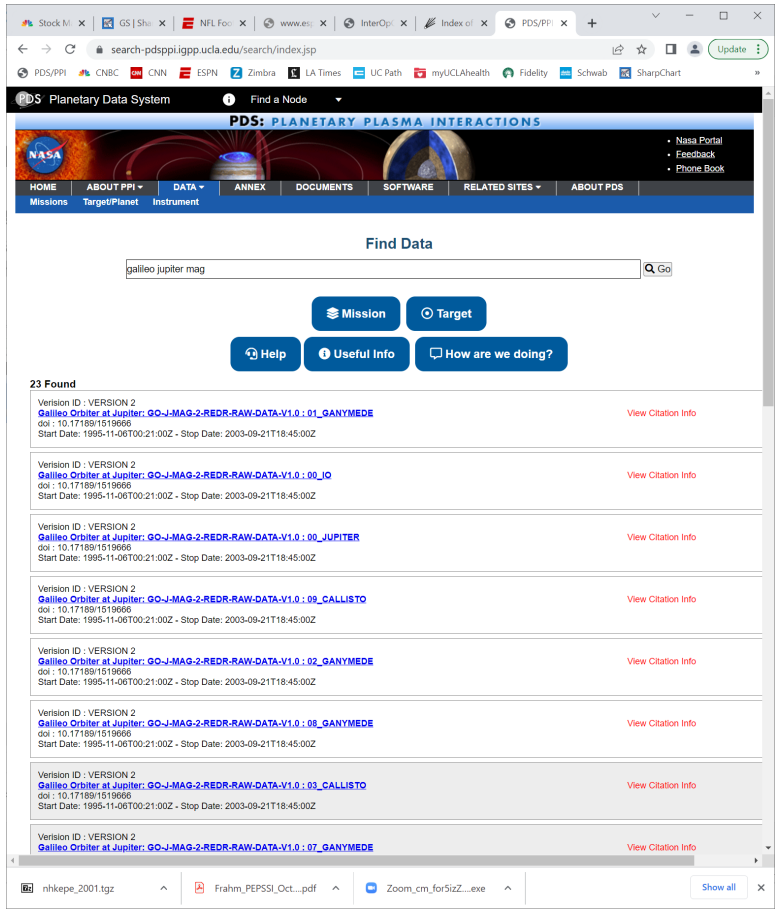
PPI Data Access using the Website

Users can navigate to dataset by selecting targets, spacecraft, and instruments using the Quick Search panel on the left side of the webpage.



https://pds-ppi.igpp.ucla.edu/search/?sc=Galileo&t=Jupiter&i=MAG

Alternatively, users can locate data of interest by using our “Google-like” search interface where keywords are entered and matching datasets are returned

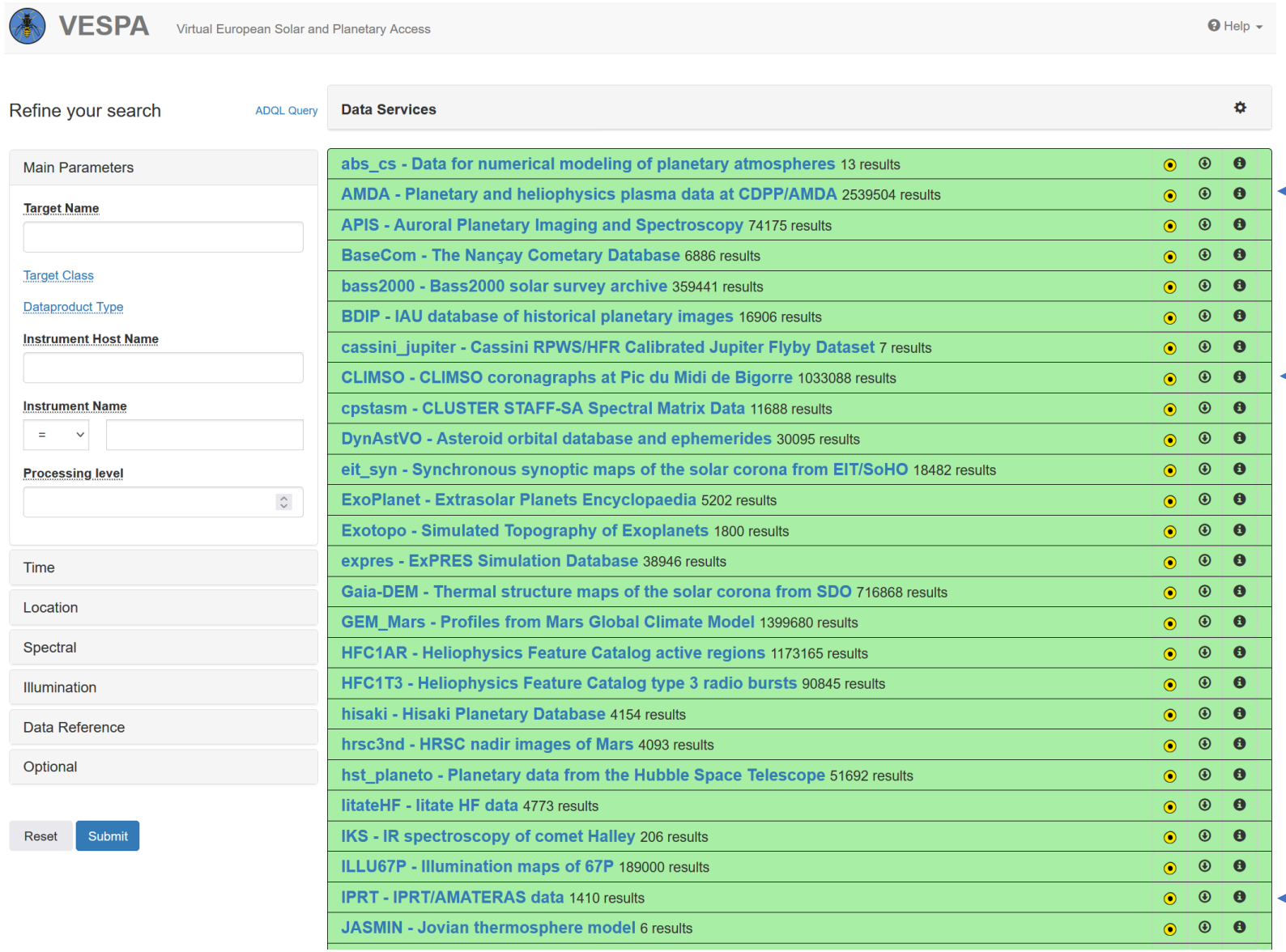


https://search-pdsppi.igpp.ucla.edu/search/index.jsp

Why is PPI using EPN-TAP

- Like all of the PDS Nodes, PPI has a strong commitment to the FAIR principles of data sharing and access
  - Our goal is to make our data findable & accessing to users regardless of how they prefer to locate and access the data.
  - PDS assigns DOIs to datasets which facilitates both access and citation.
- PPI users come from a variety of different communities and are used to using different tools (Autoplot, TOPCAT, VISTA, etc.). Many of these tools work best when the data are streamed using an API or data streaming protocol. PPI does not have the resources to independently develop data visualization and analysis tools so we are trying to make our data compatible with and accessible by existing tools.
  - PPI runs a SAMP (Simple Application Message Protocol) server and a HAPI (Heliophysics API) server although the HAPI server needs some work
  - Starting last December, PPI began making some of it's PDS4 data holdings available to users of EPN-TAP (https://vo-pds-ppi.igpp.ucla.edu)
- EPN-TAP is also being used by the ESA Planetary Science Archive (PSA) and is widely used by the Europlanet community
  - PPI believes that supporting EPN-TAP will improve our FAIRness with the European planetary F&P science community and our interoperability with international data partners.

Europlanet Vespa Portal (https://vespa.obspm.fr/planetary/data/)



VESPA provides access to large repositories like AMDA, CLIMBSO, and the PSA but it also provides access to small data providers like IPRT

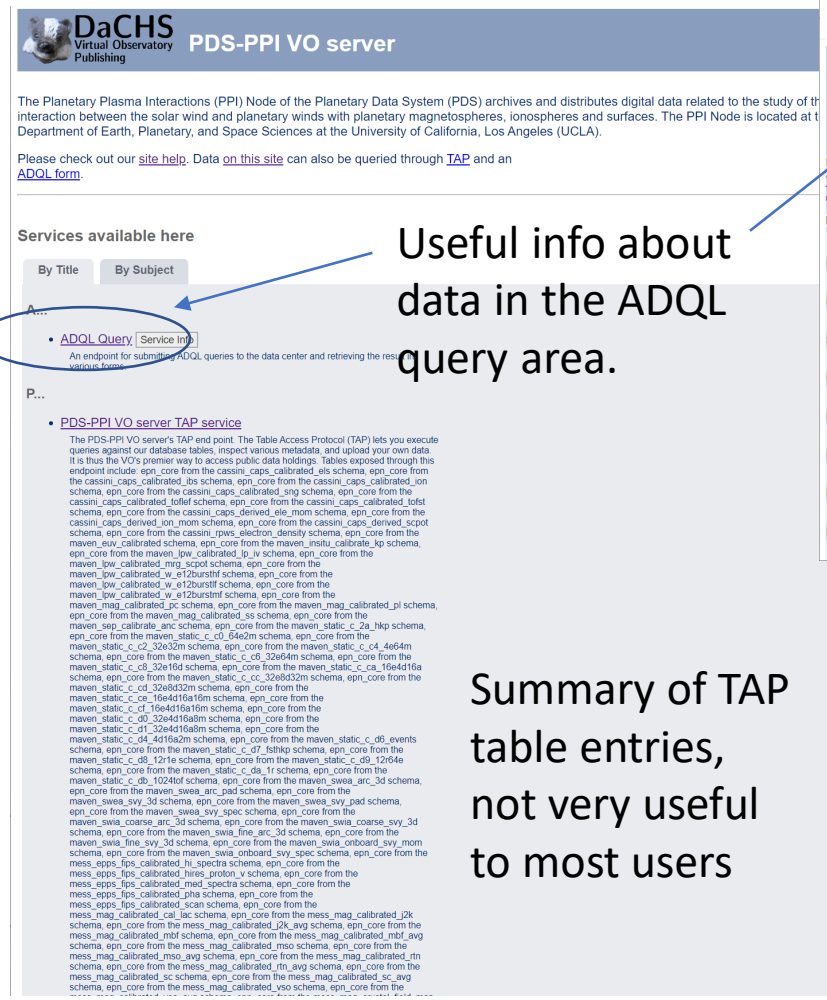
EPN-TAP Implementation at PPI

- PPI is currently producing separate EPN-TAP tables for PDS4 data collections and publishing them using a DaCHS server
  - PPI participated in the VESPA 2021 Workshop to get some initial training and support
  - Not all PDS4 data are currently described in EPN-TAP tables but we're making progress
    - We're starting with calibrated and derived data collections that we feel may be most useful to the European planetary F&P community since EPN-TAP is currently more widely adopted in Europe. This was part of our workshop participation proposal.
  - PPI is in the process of migrating all PDS3 data to the PDS4 standard
- The are pros and cons to the PPI approach to EPN-TAP tables

  Pros
    - All data in a PDS4 data collection at PPI have the same file structure and contents (processing level, reference frames, sample rate, etc.). They are collections of "like" data files. This makes it easy to stream data to tools like TOPCAT.
    - It's easy to implement the EPN-TAP tables when the data files are uniform

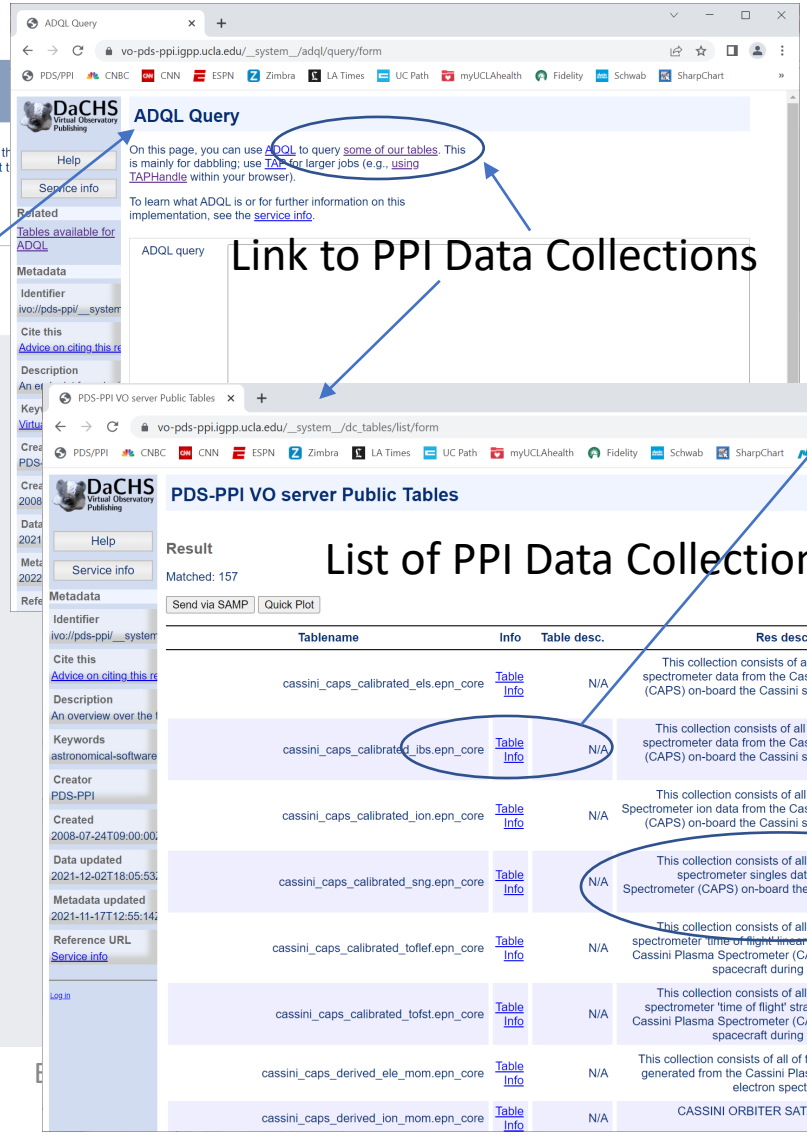
  Cons
    - PPI has hundreds of data collections. Services like the VESPA portal would be overwhelmed if they tried to list all of the PPI data collections
    - Repositories like AMDA and the PSA describe their entire holdings in a single TAP table

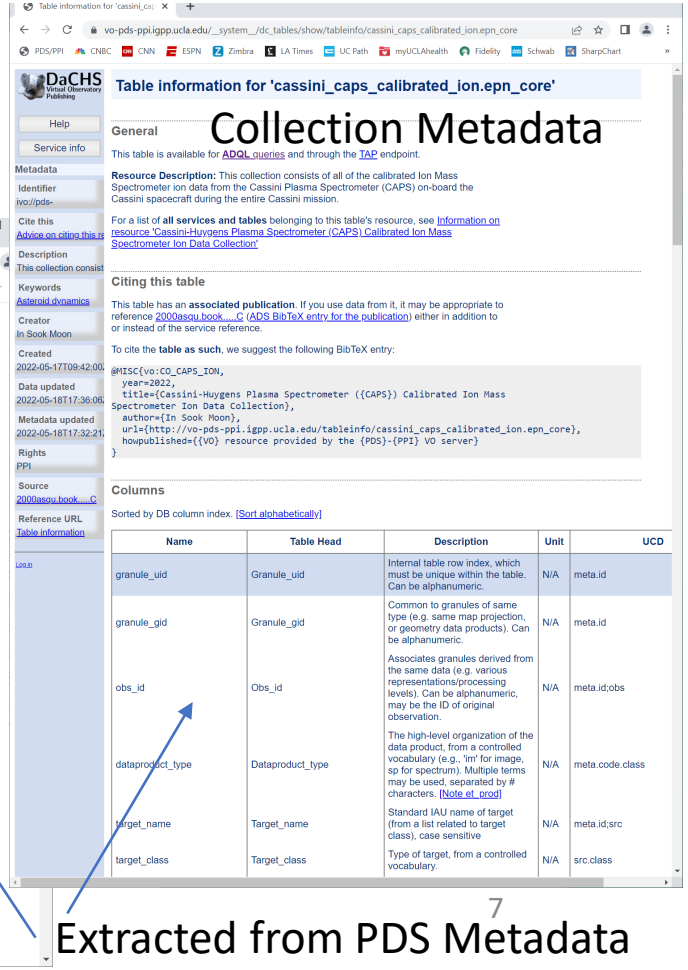
DaCHS Server at https://vo-pds-ppi.igpp.ucla.edu



Useful info about data in the ADQL query area.

Summary of TAP table entries, not very useful to most users





Extracted from PDS Metadata

PPI DaCHS Server Implementation at PPI

Running DaCHS server requires running dachs and a postgres server together. The PPI DaCHS Server is implemented in a docker container.

Data volume organization

The TAP data volume is attached to the ppi_vespa container located under /var/gavo/inputs/

There is a list of data collection folders below that level.

Each collection folder contains a resource descriptor file (ppi.rd) that stores the collection level metadata (title, description, authors, etc.) and a collection data product access listing (data.csv). The data.csv file contains product-level metadata including the access URL for streaming the data.

PPI uses a java program that reads the PDS collection and product-level metadata and translates it into the required EPN-TAP parameters to create the ppi.rd and data.csv files. This code will be in GitHub and can be shared with any interested parties. At present, it focuses on the extraction of PDS metadata relevant to F&P data.

Publishing new data to the DaCHS server requires that the resource file is validated and imported using DaCHS tools:

dachs val ppi.rd to validate a resource file and
dachs imp ppi.rd to import the data into the server

PPI Plans to implement DaCHS pub//services in the near future

Accessing PPI Data in Topcat using EPN-TAP

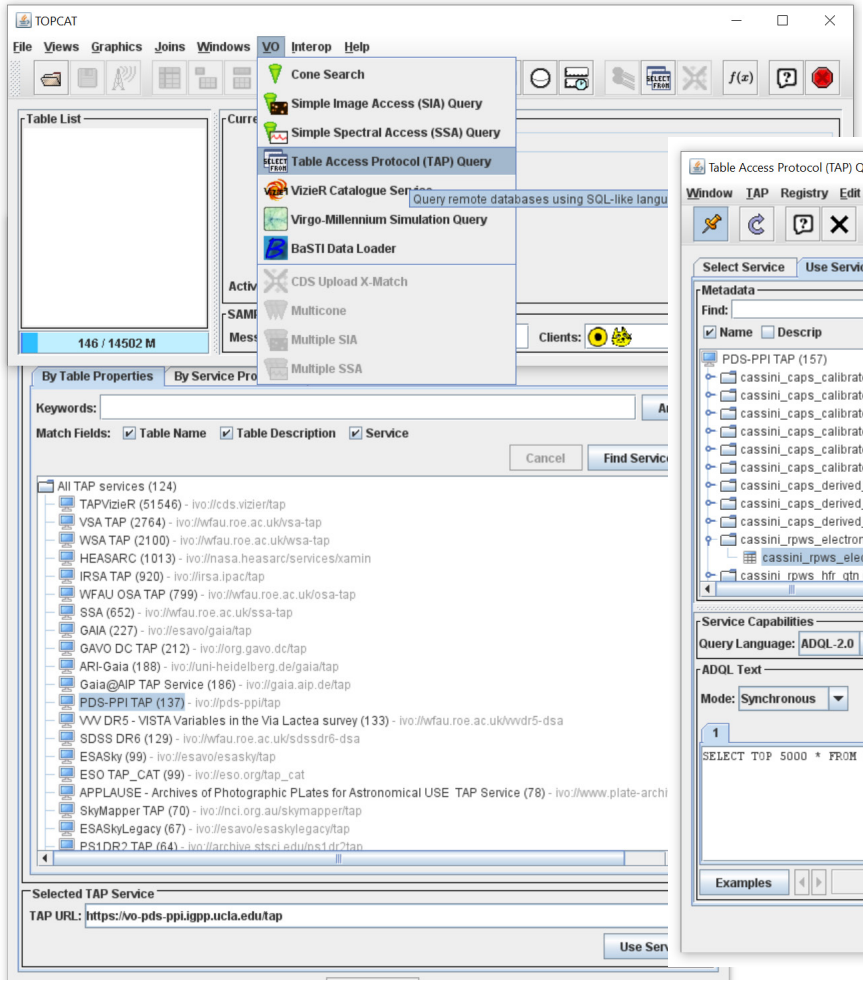
Start TOPCAT

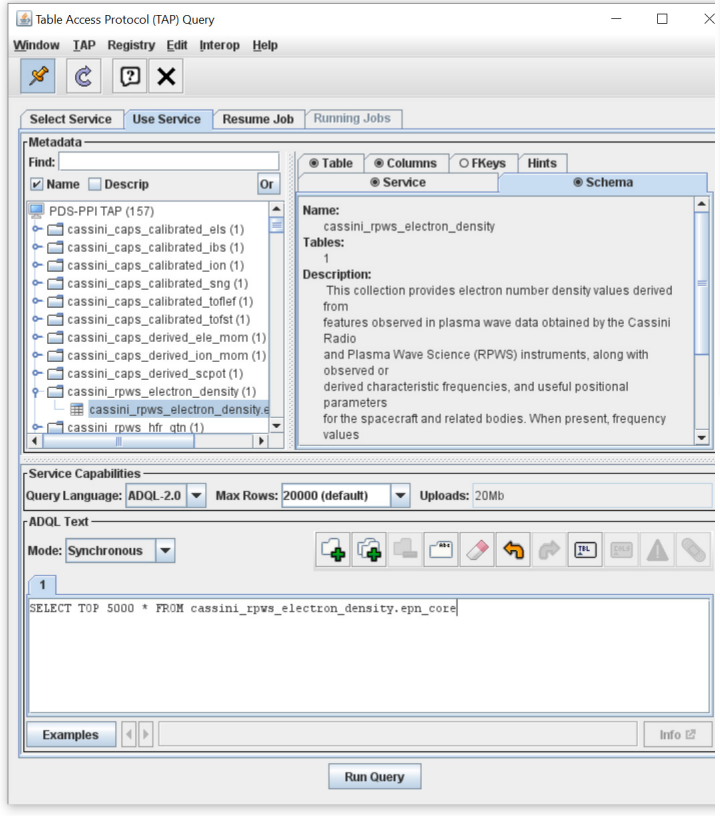
Select VO – TAP server, then PDS-PPI TAP

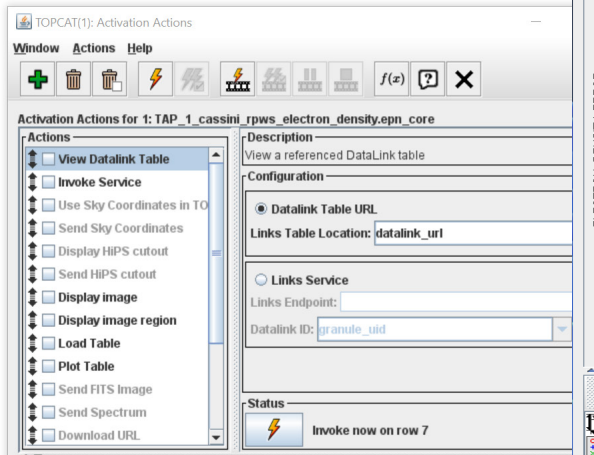
Select a Collection & Run Query

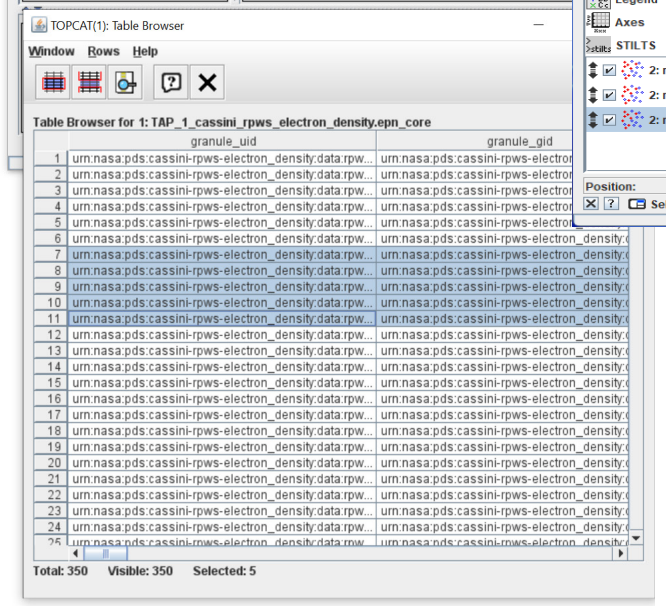
Select Products

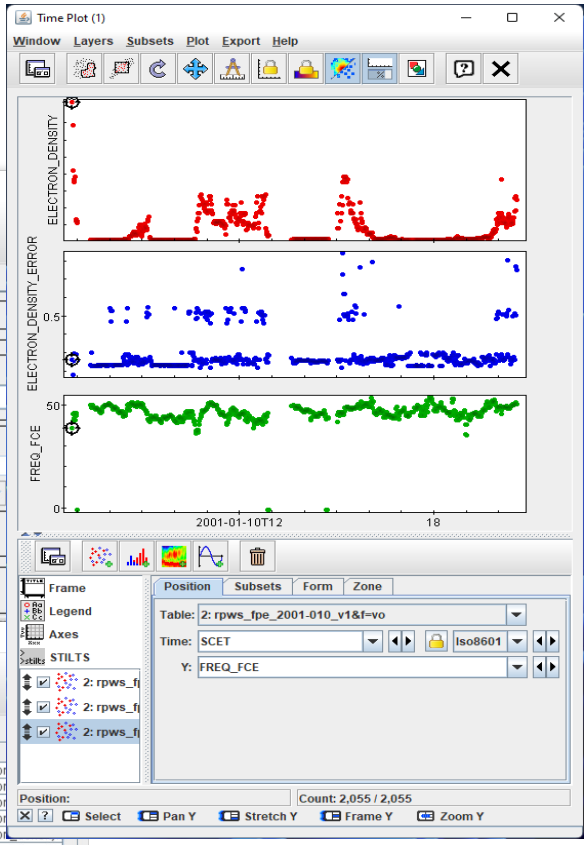
Display











PPI Path Forward

- There is still a lot of work to go in implementing EPN-TAP at PPI
  - We need to complete the descriptions of all PDS4 data collections
  - We need to include creating the EPN-TAP descriptions part of the PDS3 data migration process
  - PPI would love to show up in the VESPA data portal
    - Doing so may require that PPI develop a new/additional view of the EPN-TAP data holdings so that we appear as a single entity like AMDA or the PSA
    - PPI is going to need help making our implementation more VESPA friendly
  - PPI would like to provide our users with access to data at the PSA or other IPDA member locations using EPN-TAP for interoperability. Similarly, PPI would like its data collections to be available to PSA users
    - A multi-repository PDS4 data registry provides an alternative mechanism for interoperability
    - EPN-TAP is available now and is supported by various data display and analysis tools, it's not clear that the PDS4 registry will have the same tool support that EPN-TAP currently enjoys
  - PPI hopes to become a model for other PDS nodes that might choose to support EPN-TAP

Questions / Discussion