PPI DATA SEARCH AND USABILITY

PDS DATA SERVICE WORKSHOP

NOVEMBER 5, 2019

R. WALKER, S. JOY, J. MAFI

A USER-CENTERED APPROACH

What users want:

- **Google** – Most users begin data searches from Google.
- **Search Services** – Users want services that make data search more accurate, and download more efficient.
- **Useable Data** – Users want data that are readily accessible using generally available data visualization and manipulation tools.

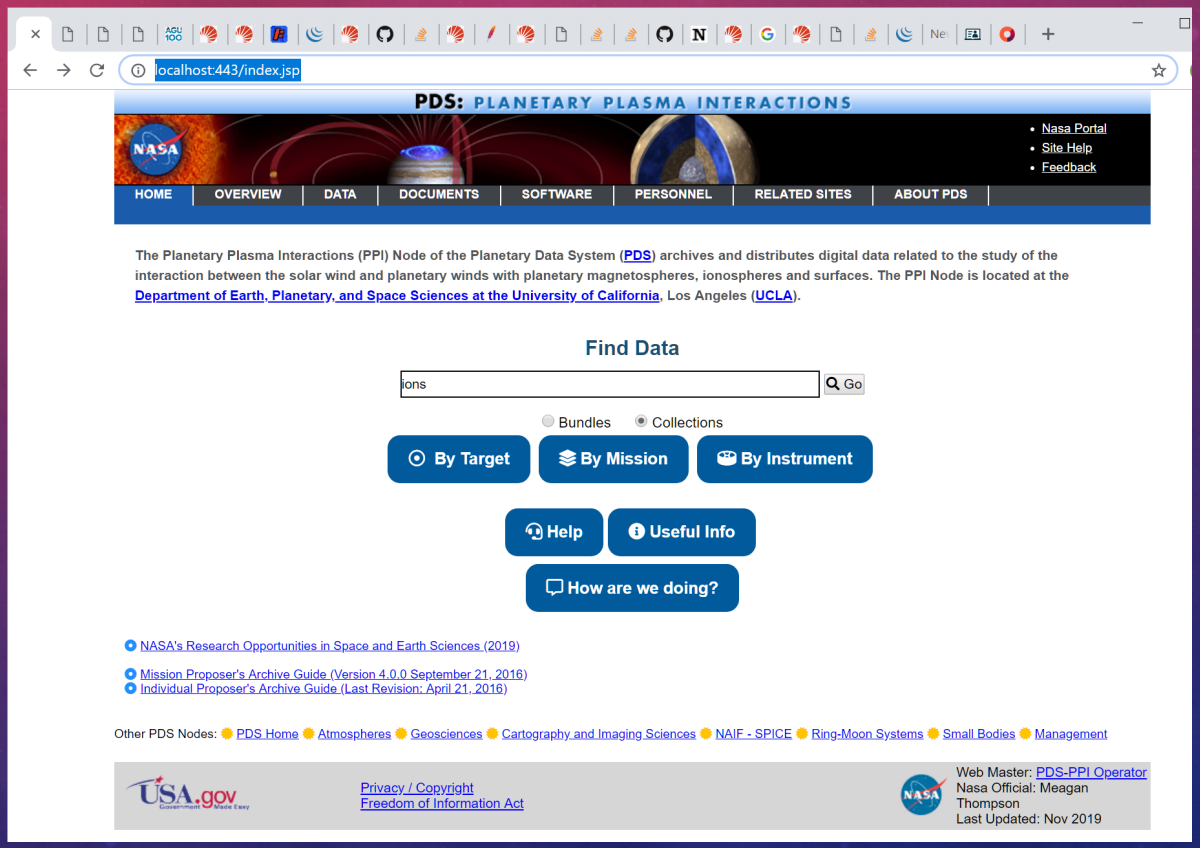
GOOGLE SEARCH TESTING

- PPI conducted internal testing comparing the results from identical searches using a PPI text search and an identical search in Google.
  - PPI searches generally did fairly well, though some data sets were missing.
    - Reasonable variations in search terms (e.g. “mag” instead of “magnetometer” or “fgm”)
    - Minor spelling variations (e.g. “ion” vs. “ions”, “magnetic field” vs. “magnetic-field”, etc.)
  - Google searches were less consistently reliable.
    - Search results must be in the first screen/page to be considered successful.
    - Successful results depended very much on selecting the right search terms.
    - Search terms of high public interest often push PDS holdings far down in search results.
  - Hierarchical mission and target pages were more reliable in finding all available data.

BETTER SUPPORT FOR WORD SEARCHES

- PDS metadata need to be better conditioned to support text search:
  - More consistent metadata
  - Accommodation for reasonable variations of search terms
  - Support for associating reasonable alternative search terms with the canonical terms used found in the metadata
  - Metadata support for terms which will move PDS data up in relevant Google searches

NEW PPI WEB INTERFACE

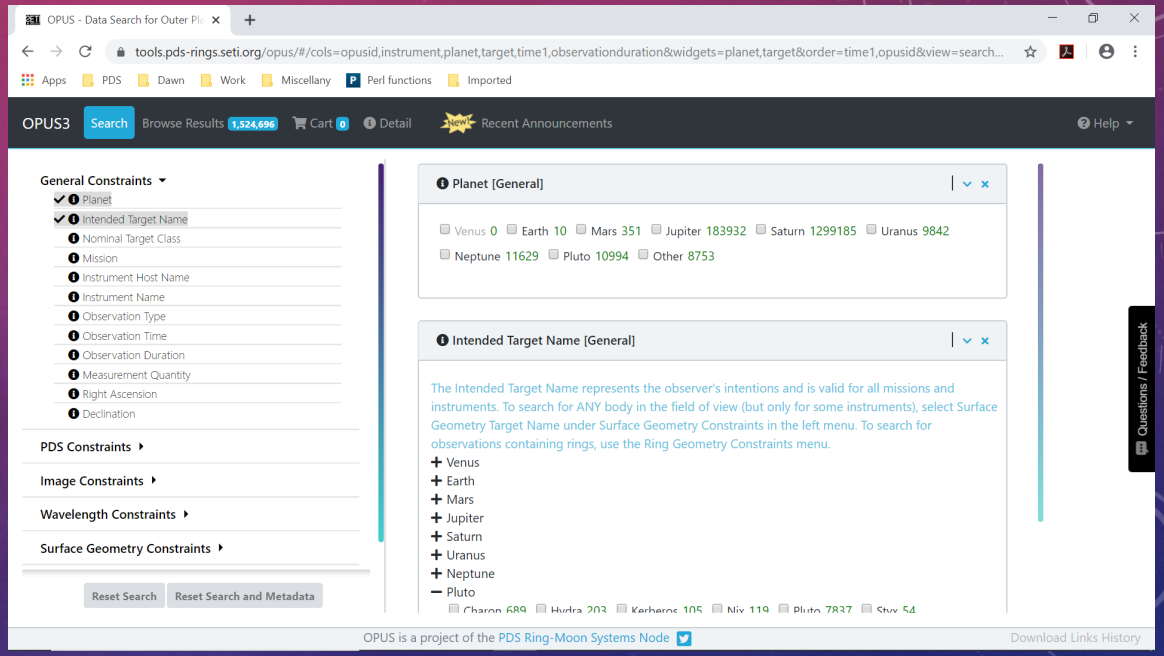


PPI DATA SEARCH SERVICE CONCEPT

- Ring-Moon Systems OPUS-style search interface
- PPI-style results list (by bundle/data set)
- Geosciences ODE-style checkout

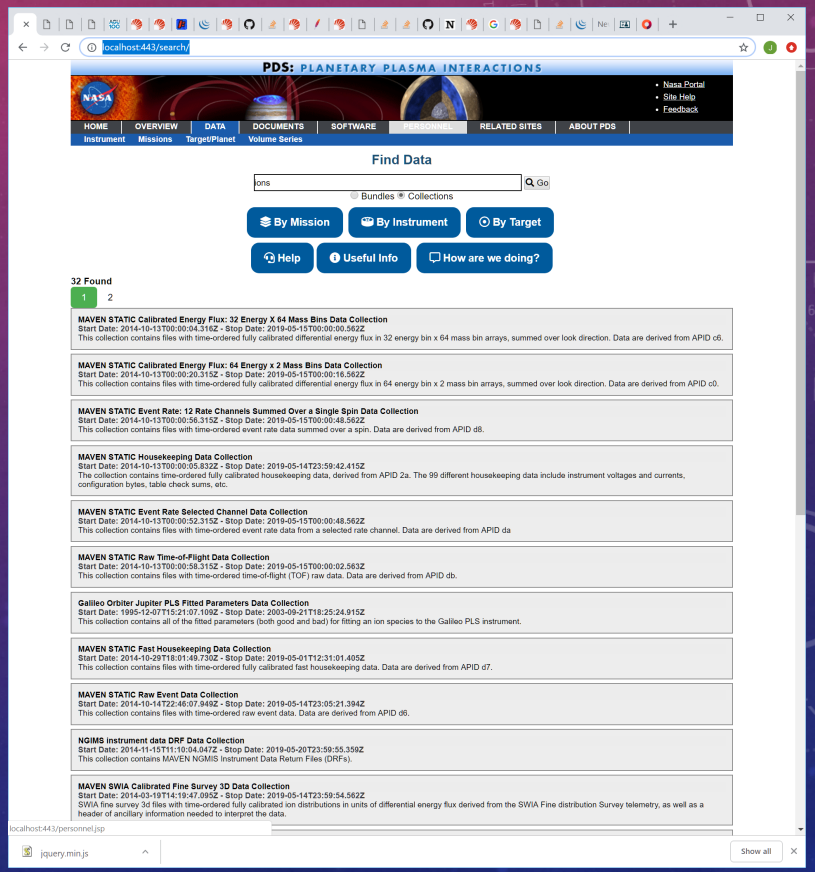
OPUS-STYLE SEARCH INTERFACE

- OPUS (RMS)
  - Clean, intuitive interface
    - Well grouped, collapsible facet lists
  - Realtime selection criteria updates
  - Realtime product counts
- Add PPI-specific selection criteria
  - Primary Results



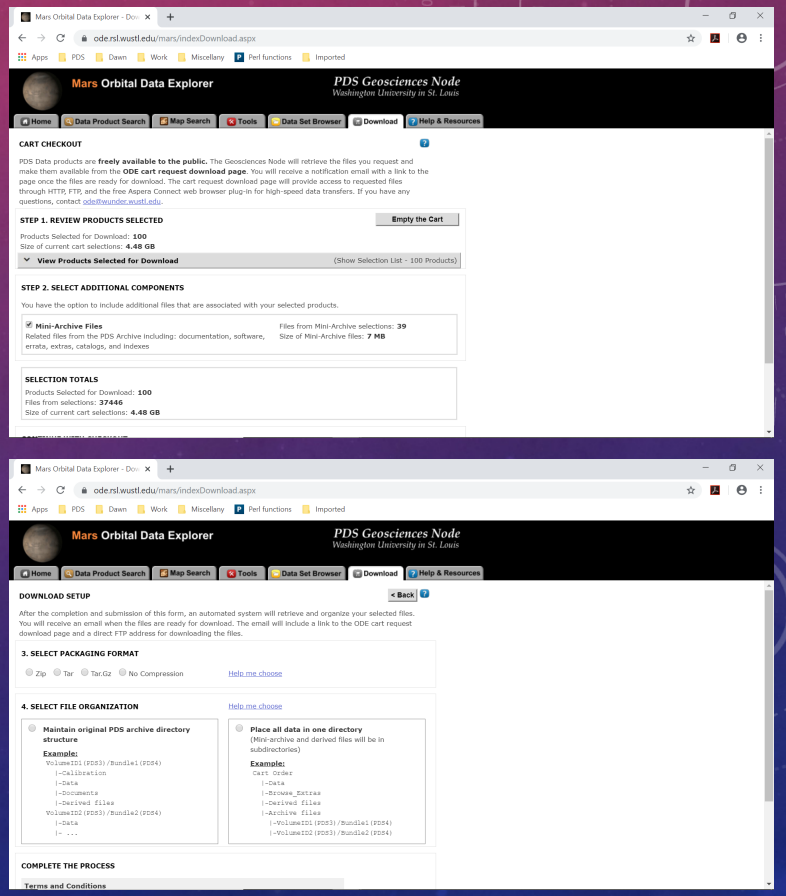
PPI-STYLE SEARCH RESULTS

- PPI
  - Results listed by bundle
    - Add product counts/volume for each data set
    - Add checkbox to allow selection of multiple data sets



ODE STYLE CHECKOUT

- Orbital Data Explorer (GEO)
  - Option to include associated documentation
  - Option to maintain original multiple volume structures, or merge into single volume
  - Compression (or no compression) options
  - Warnings/time estimate for large downloads

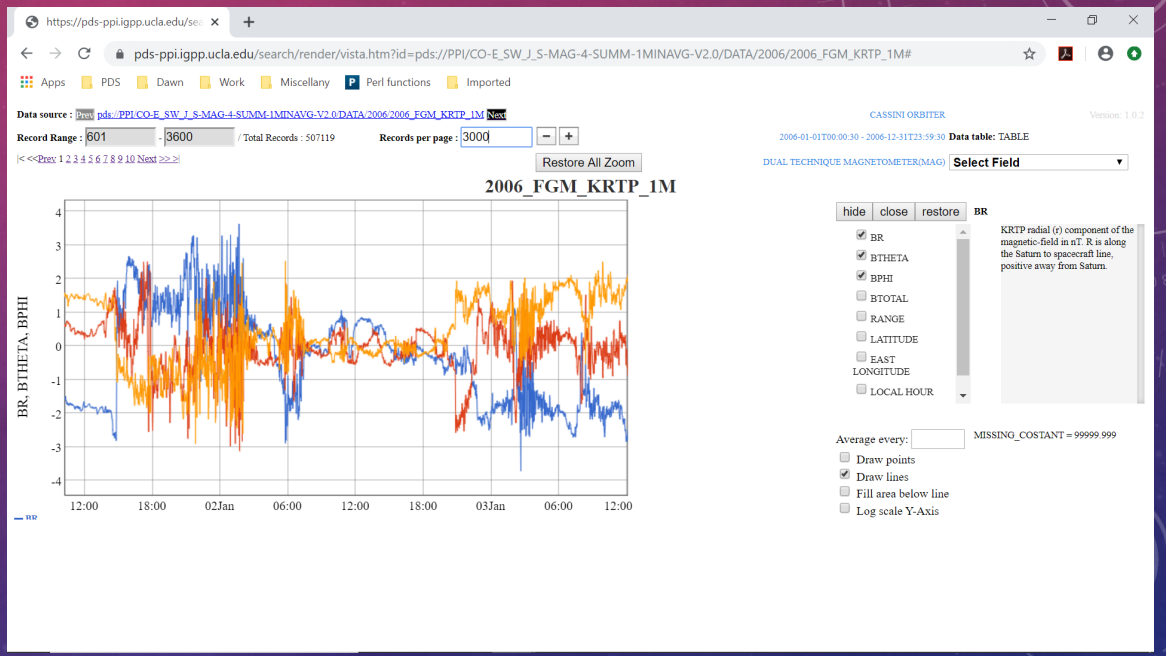


PPI USER SUPPORT TOOLS

- PPI users are frequently looking for tools to support their research.
  - Web based visualization tools can help users to select data to download.
  - Analysis tools can support data users' research.

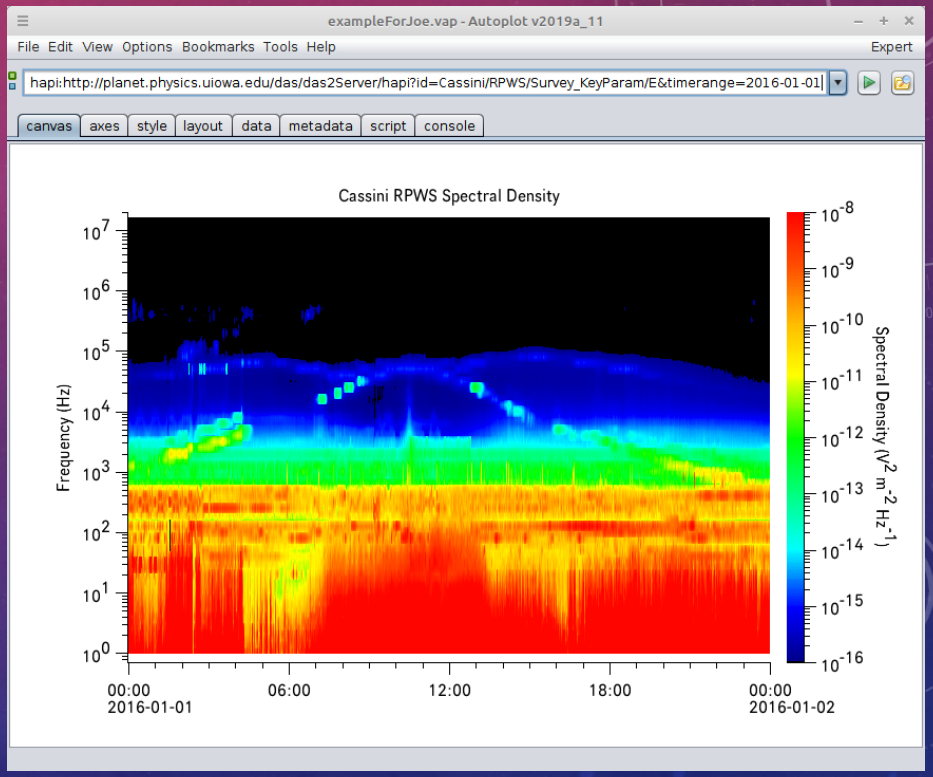
VISTA

- Simple line plotter
- Supports: PDS labelled ASCII, binary, and delimited (CSV) tables



AUTOPLOT

- Spectral image plotter
- Supports: ASCII tables, CDF, FITS, and other formats



PPI VISUALIZATION SUPPORT

- PPI has been working to expand visualization support for its data holdings.
  - VISTA has been adapted to support previously unsupported data sets.
  - Autoplot config files will be added to enable visualization of appropriate data sets.
  - Metadata will be stored in the server to indicate the appropriate visualization tool for each data set.

SUPPORT FOR PARTICLE DATA

- Many particle data sets are in formats not supported by any visualization or analysis tools.
  - Data are multi-dimensional arrays flattened into two-dimensional tables.
  - Flattening has been performed differently from data set to data set.
- PPI is exploring alternatives for improving tool support for these data sets:
  - Expanding PDS4 metadata to more accurately describe array flattening
  - Converting flattened arrays to CDF

EXPANDING PDS4 METADATA

Pros

- Relatively minor effort

Cons

- May not solve the problem for all data sets.
- Existing tools would have to be adapted, or new tools developed to work with the expanded metadata.

CONVERSION TO CDF

Pros

- CDF supported by existing tools (IDL, MATLAB, Autoplot, etc.)
- SPARTA (developed by ESA for Rosetta) – tool to convert PDS3 labeled data to CDF

Cons

- Major effort
- Converted data sets would have to be peer reviewed in order to become part of the permanent archive (could be maintained as a “contemporary format”).

BACKUP SLIDES

PPI DATA SEARCH DATABASE PARAMETERS

- LID
- Title
- Description
- start time/stop time
- Science_Facets
- Investigation_Area/name and type
- Observing_System_Component/name and type for spacecraft and instrument
- Target

MULTI-DIMENSIONAL ARRAY FLATTENING

- “Road Kill” method
  - Each record contains the whole array: header followed all of the data rows arranged end-to-end.
    - In some cases ITEMS and/or CONTAINER objects have been used in an attempt to represent the multi-dimensional structure of these data.
- “Flat Stanley” method
  - Each record contains header information, followed by a single row of the array.
    - Header data (e.g. time stamps, etc.) are repeated.

MULTI-DIMENSIONAL ARRAY FLATTENING

- “Off with its head!” method
  - The first record of the array contains either the header information by itself, or the header followed by a single data row. Subsequent records contain subsequent rows of data (without the header) through the end of the array.
    - This method only works by the creative use of PDS3 containers.
    - Data flattened using this method are often reformatted.
- “This is a what?!” method
  - Each record contains a header followed by a cell ID and data value for that cell.
    - Header data (e.g. time stamps, etc.) are repeated.