8 Years of the Parkes Pulsar Data Archive

James Dempsey (CSIRO Information Management and Technology), George Hobbs, Lawrence Toomey (CSIRO Astronomy and Space Science)

The Parkes Pulsar Data Archive (PSRDA) has been operating since 2011, providing access to science-ready data of pulsar observations taken at CSIRO's Parkes radio telescope. The PSRDA is part of CSIRO's Data Access Portal, and provides both web and Virtual Observatory access to the data.

The Parkes Pulsar Data Archive

More than half of the known pulsars have been discovered using the Parkes radio telescope. It continues to be a highly productive instrument for studying pulsars. The PSRDA holds the majority of pulsar observations taken since 2011, and historical observations are being added progressively. The archive holds ~1.8 PB of pulsar data currently and is growing rapidly (Fig. 2).

All of the data in the archive are in PSRFITS format, meaning the data can be processed with tools such as PSRCHIVE [1], DSPSR [2], PRESTO [3], and TEMPO2 [4]. The data are grouped into collections based on the observing program and semester, along with observing runs. Data are published within 30 days of each observing run. Each collection is issued with a DOI to allow data citation.

The PSRDA has been developed as a collaboration between the science, application and infrastructure teams within CSIRO.

Data Access

The PSRDA web site, part of CSIRO's Data Access Portal (DAP), allows pulsar observations to be queried by position, pulsar name, project and selected observing parameters. Matching data files are listed with preview images and metadata (Fig. 1). Online data can be accessed as individual data files, and collections can be brought online and made available for download or bulk access via WebDAV.

Observations may also be queried through our Table Access Protocol (TAP) service. This allows discovery based on position, time, receiver, backend etc using SQL like queries. If the data are online they can then be directly downloaded from the result.

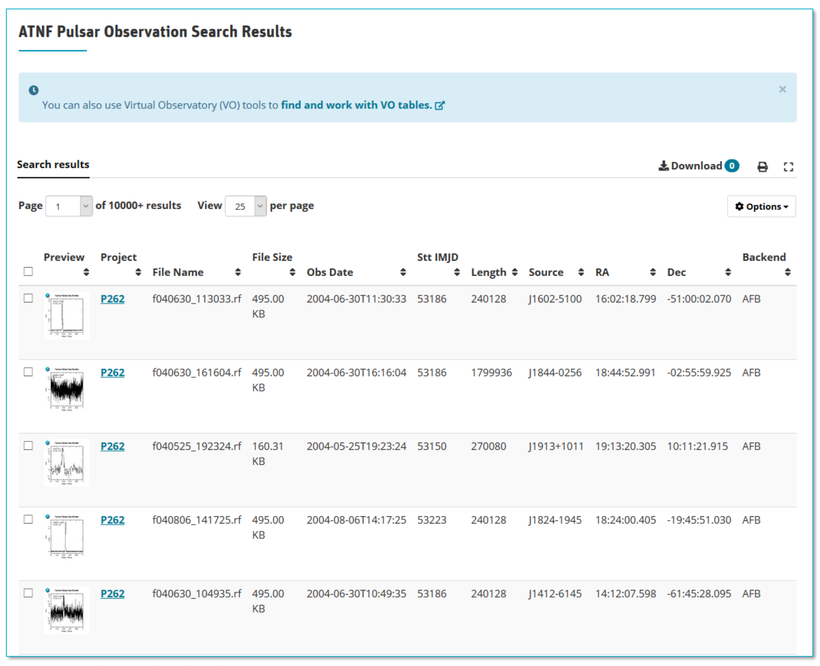

Figure 1: PSRDA search results display

Architecture

PSRDA is a Java web application using a PostgreSQL database for observation. The metadata is indexed in ElasticSearch to power the search functionality. Collection descriptions are sourced from the ATNF observing proposals system (OPAL).

Raw data products are generated by the backends at Parkes. These are transferred to a processing server in Canberra where derived data products, including preview images, are produced. All data products are grouped up into collections for every block of observing for a project and transferred to DAP. In DAP each collection is created automatically prior to ingest. DAP then extracts observational metadata from the PSRFITS headers, adds that to the metadata database, saves the data to archival storage, mints a DOI for the collection and indexes the metadata in ElasticSearch.

Science Enabled

- Discovery of new Fast Radio Bursts in archival data [5]
- Collections for studies of gravitational wave limits [6, 7, 8]
- Data for calibrating pulsar observations [9]
- Data from pulsar polarimetry studies [10]
- Providing a history of observations at Parkes radio observatory (Fig. 3)

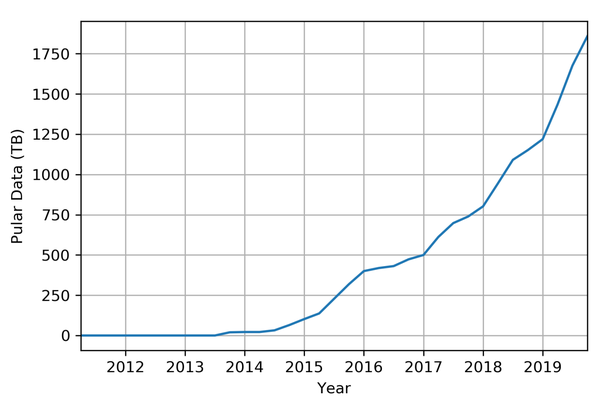

Figure 2: PSRDA data holding growth since 2011.

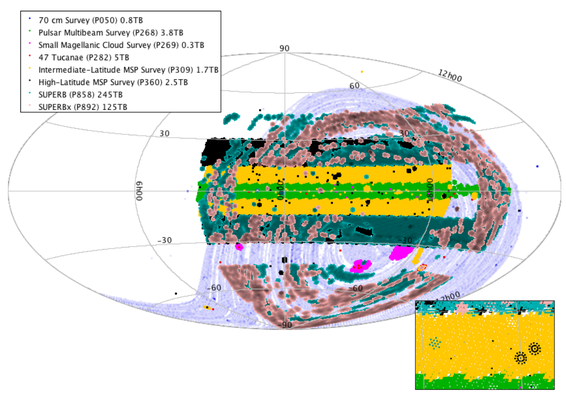

Figure 3: Sky coverage of the pulsar observations in the PSRDA, coloured by observing project (to 2018).

Future Enhancements

Planned future improvements include:

- Automating the deposit process to support higher deposit data rates
- Support data from other surveys or observatories
- Improved access speed and support for higher volume access methods

As Australia's national science agency and innovation catalyst, CSIRO is solving the greatest challenges through innovative science and technology.

CSIRO. Unlocking a better future for everyone.

FOR FURTHER INFORMATION

James Dempsey
CSIRO Information Management & Technology
James.dempsey@csiro.au
https://data.csiro.au/

VISIT PSRDA

REFERENCES

1. http://psrchive.sourceforge.net
2. http://dspsr.sourceforge.net
3. https://www.cv.nrao.edu/~sransom/presto/
4. https://bitbucket.org/psrsoft/tempo2
5. https://doi.org/10.1093/mnrasl/slz023
6. http://doi.org/10.4225/08/534CC21379C12
7. http://doi.org/10.4225/08/561EFD72D0409
8. http://doi.org/10.4225/08/54F3990BDF3F1
9. http://doi.org/10.4225/08/58363e853a58b
10. http://doi.org/10.4225/08/59952c840ae35

ACKNOWLEDGEMENTS

- CSIRO Information Services Applications team
- CSIRO Research Data Service support team
- CSIRO Scientific Computing support teams