CSIRO Data Access Portal: 1 PB and beyond

Use and impact
Katie Hannan, Dom Hogan, Sue Cook, CSIRO IMT.

The CSIRO Data Access Portal (DAP) provides access to data collections published by CSIRO across a wide range of research domains. The DAP was publicly released in 2011 and storage growth rose from an average of 18 GB a week during Sept 2012-Feb 2013, to an average of 8 TB a week during Sept 2017-Feb 2018. During the first week of March 2018 the DAP passed the 1 Petabyte (PB) mark. This significant milestone means that 1 PB of CSIRO data and software files are preserved together with their metadata for long-term discovery.

Download Growth per Collection

Here we count one IP address downloading one collection any number of times on a given date as constituting one download (this has the flaw that users behind a proxy will be counted as the same individual). Collections that had promotional events on their release (e.g. Soil and Landscape Grid, Workspace), or that received media coverage (e.g. Income Inequality) show strong early growth.

Landing Page Hits

The landing page hits chart shows web traffic visiting the DAP on a daily basis. Large peaks in web traffic correspond to communications referencing data collections in the DAP.

DAP Citations

Finding data citations can be tricky, and the number of citations is quite small compared to the number of collections published. While the rates of publication and self-citation are somewhat linear, one thing to monitor in future will be to see if the rate of 3rd party citations is increasing.

Research Impact

Demonstrating research impact is a complex issue. After analysing the DAP logs we have discovered that there is a relationship between the active promotion of scientific scholarship through publication, data publication, best practice data citation and downloads of data collections on the DAP. As the culture of data citation matures it will become easier to demonstrate research impact via this method.