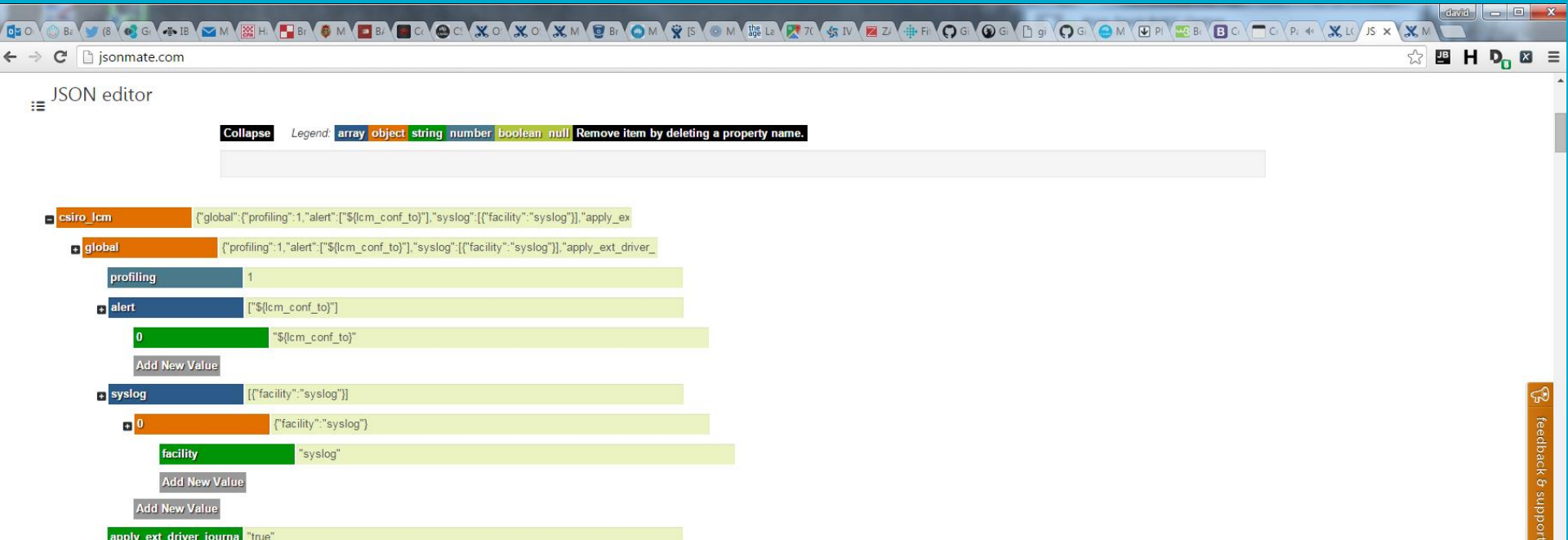


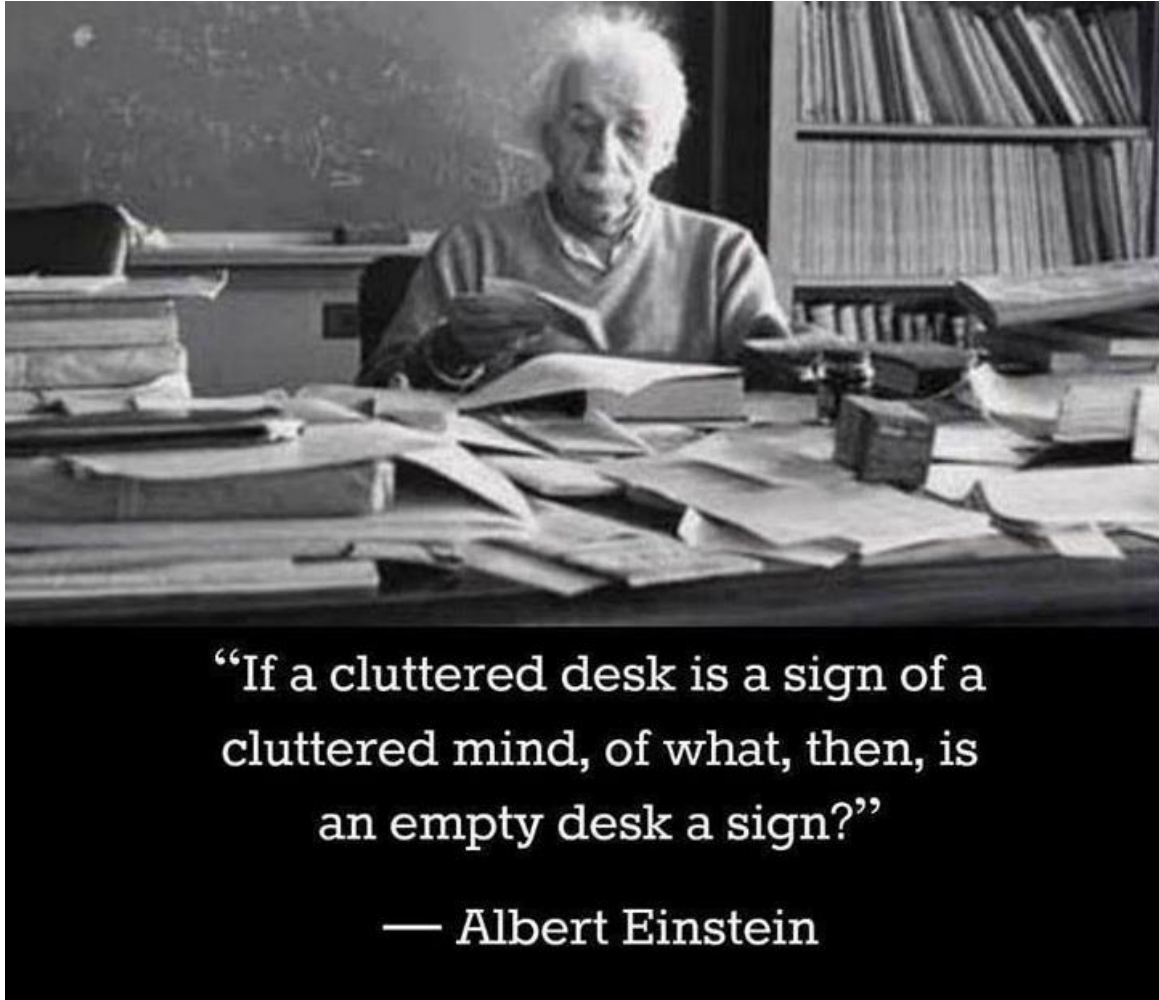


LCM (Logical Collection Manager)

And DMF

Andrew Smith | Data Storage and Data Centre Services Specialist
August 2019

Logical Collection Manager



Logical Collection Manager

Example Data sources:

- Enterprise File Server
- External/USB hard drive
- Lab/Instrument eg: DNA sequencer, Electron Microscope
- Boat – RV Investigator
- Radio Astronomy
- Unstructured data
- Large collections

Logical Collection Manager

**LAW NUMBER XVII: SOFTWARE IS LIKE
ENTROPY. IT IS DIFFICULT TO GRASP,
WEIGHS NOTHING, AND OBEYS THE SECOND
LAW OF THERMODYNAMICS; I.E., IT ALWAYS
INCREASES.**

- NORMAN R. AUGUSTINE -

Logical Collection Manager

Data Access Portal



CSIRO's Data Access Portal **data.csiro.au**

Why manage your data?

- To maximise research impact
- To enable collaboration
- To support reproducibility and reliability
- To facilitate citation and reward

What does the Data Access Portal mean for you?

It allows you to:

- **Deposit** - self serve or automated
- **Discover and reuse** - search for data, software or web service
- **Manage** - version control, access and licensing
- **Share** - internally, nationally or internationally or just with a select few
- **Link data to publications**

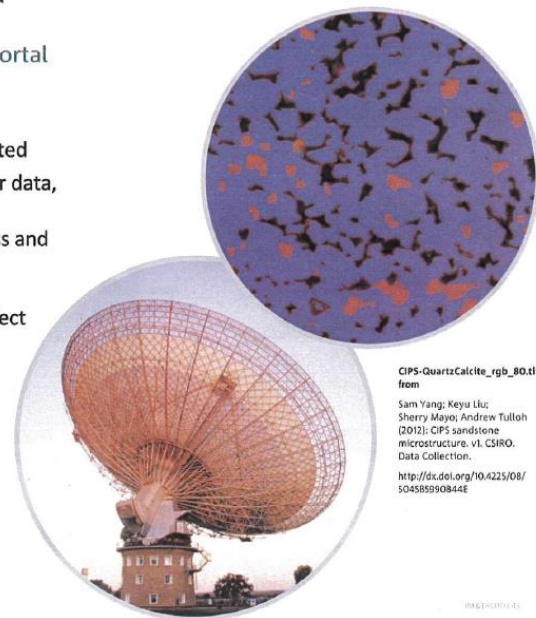
It allows **Data Citation** similar to publications. Published data will get a DOI and will be citable by yourself and others.

It is supported by the IM&T Storage and Compute Cloud platform.

Research Data Service Support

The support team is able to help you with:

- Training and demonstrations
- Describing your data
- Data organisation and transformations
- Data management planning
- Data citation



Logical Collection Manager

Acknowledgements

FUNDING



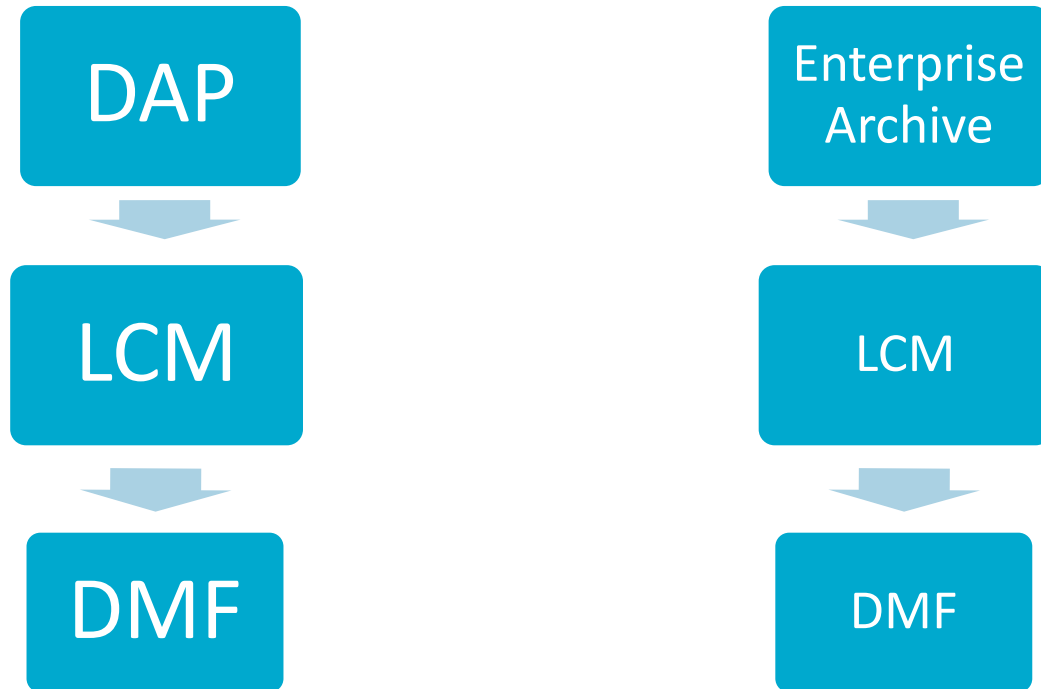
This project is supported by the Australian National Data Service (ANDS), now the Australian Research Data Commons (ARDC), through the National Collaborative Research Infrastructure Strategy Program and the Education Investment Fund (EIF) Super Science Initiative.

DEVELOPMENT

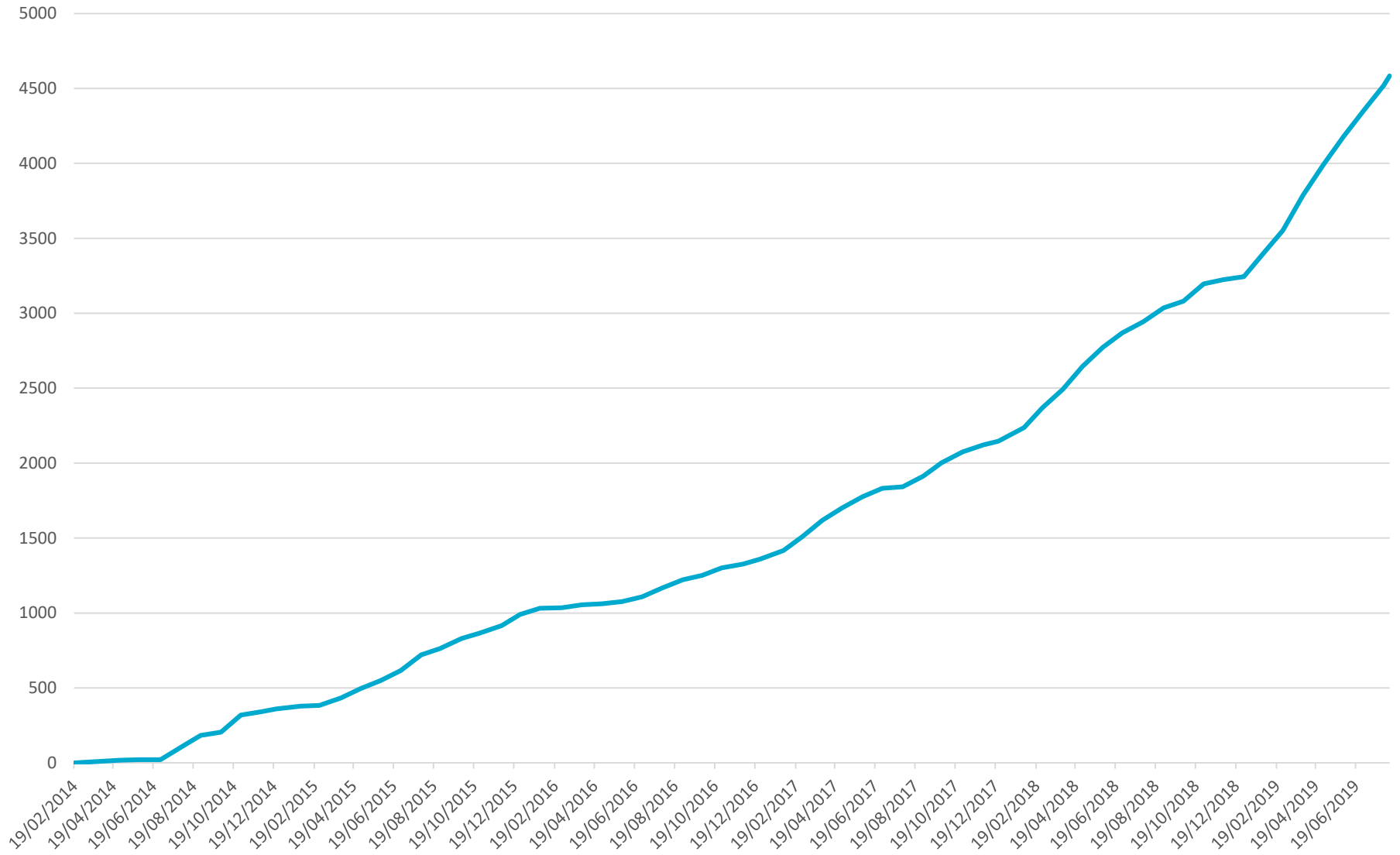
This site is produced by CSIRO Information Management and Technology. With thanks to all CSIRO staff past and present for their input into the development of the Data Access Portal.

Open Source & 3rd-party components in the Data Access Portal

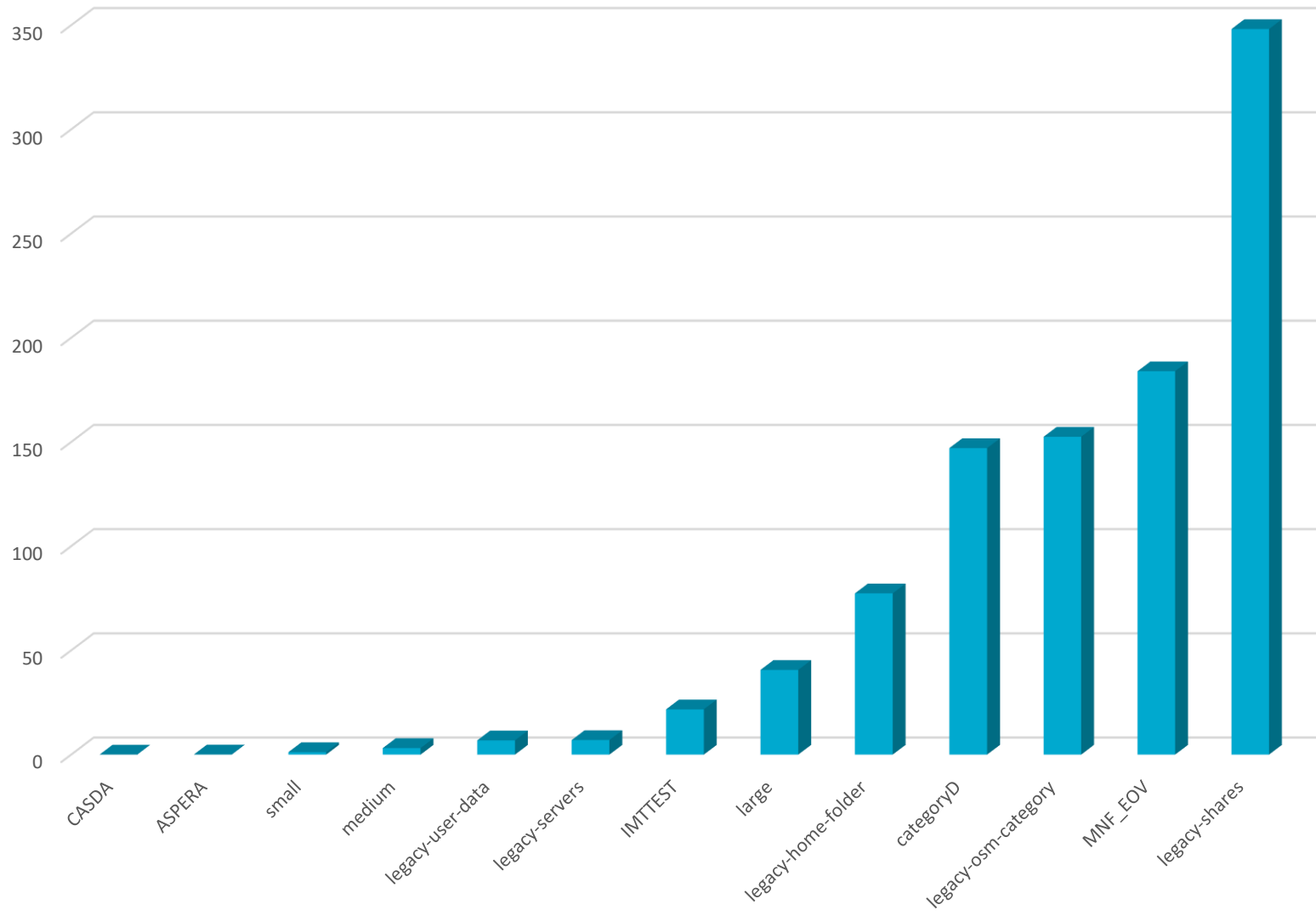
Logical Collection Manager

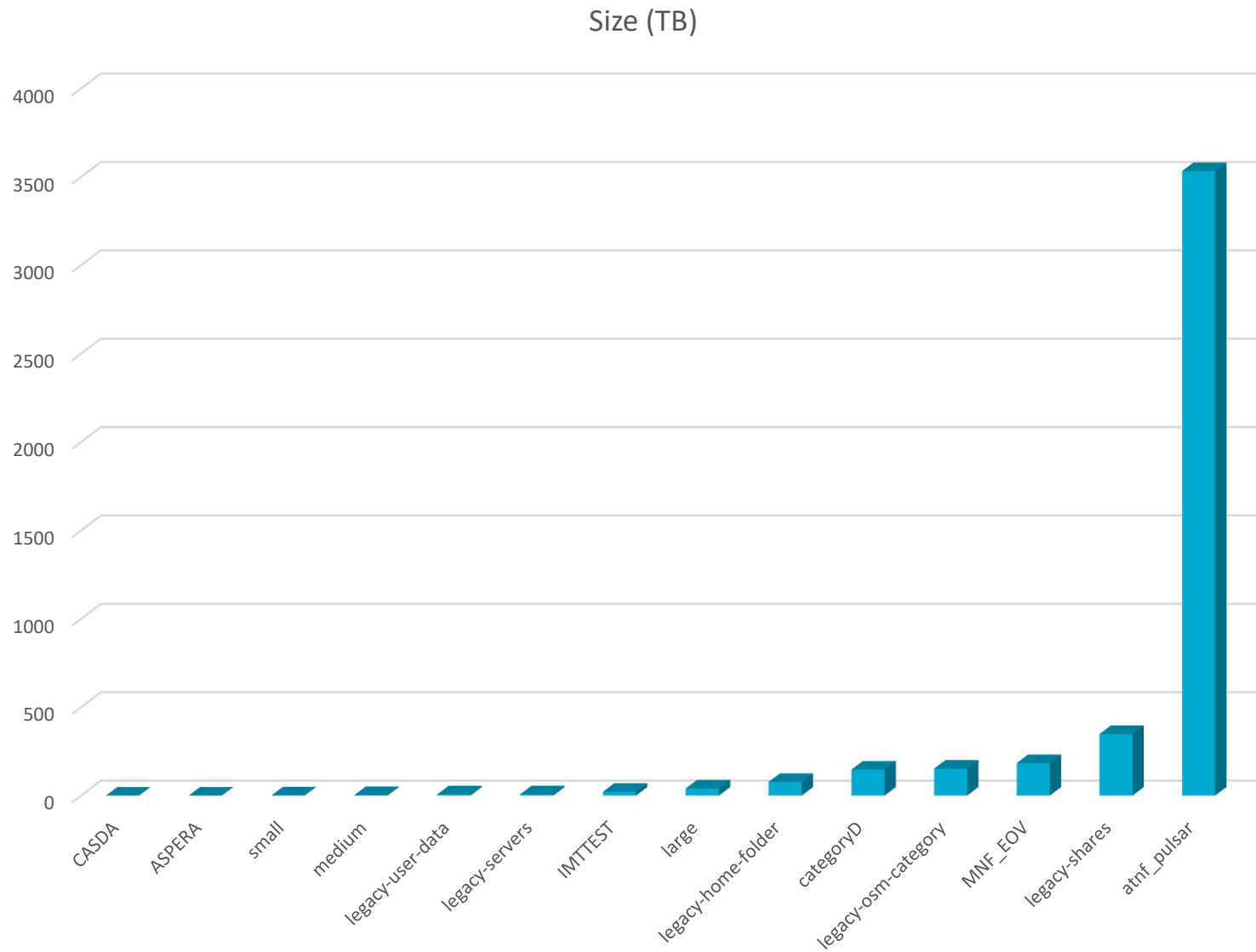


Cumulative Size (TB)



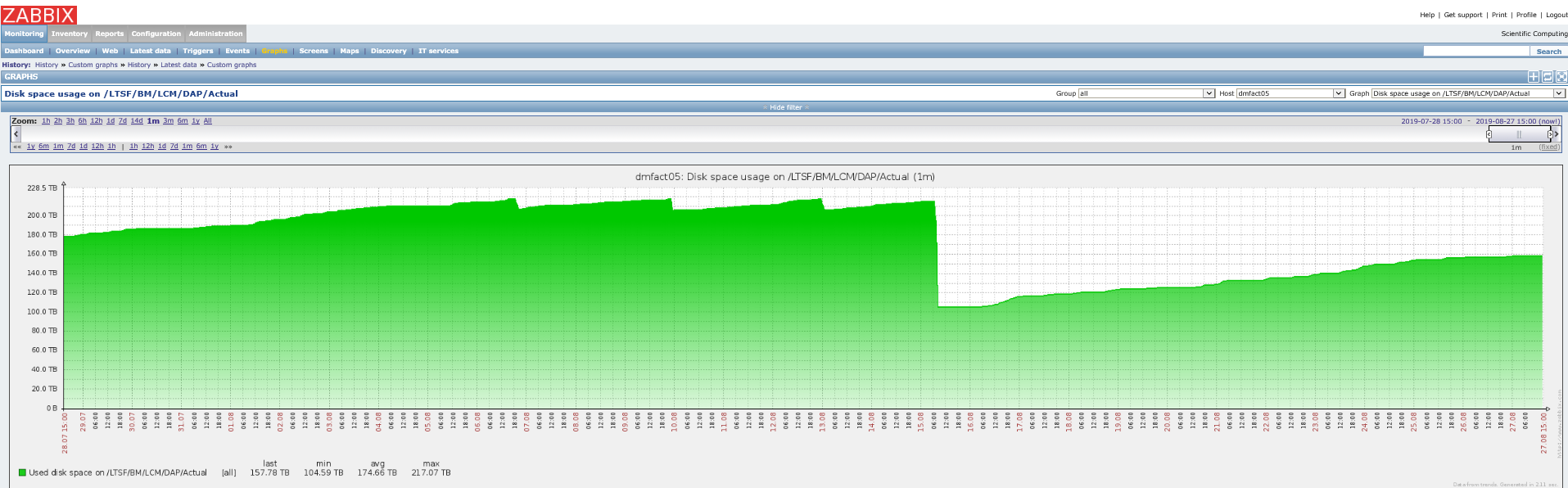
Size (TB)





Logical Collection Manager

DAP DMF usage



Logical Collection Manager

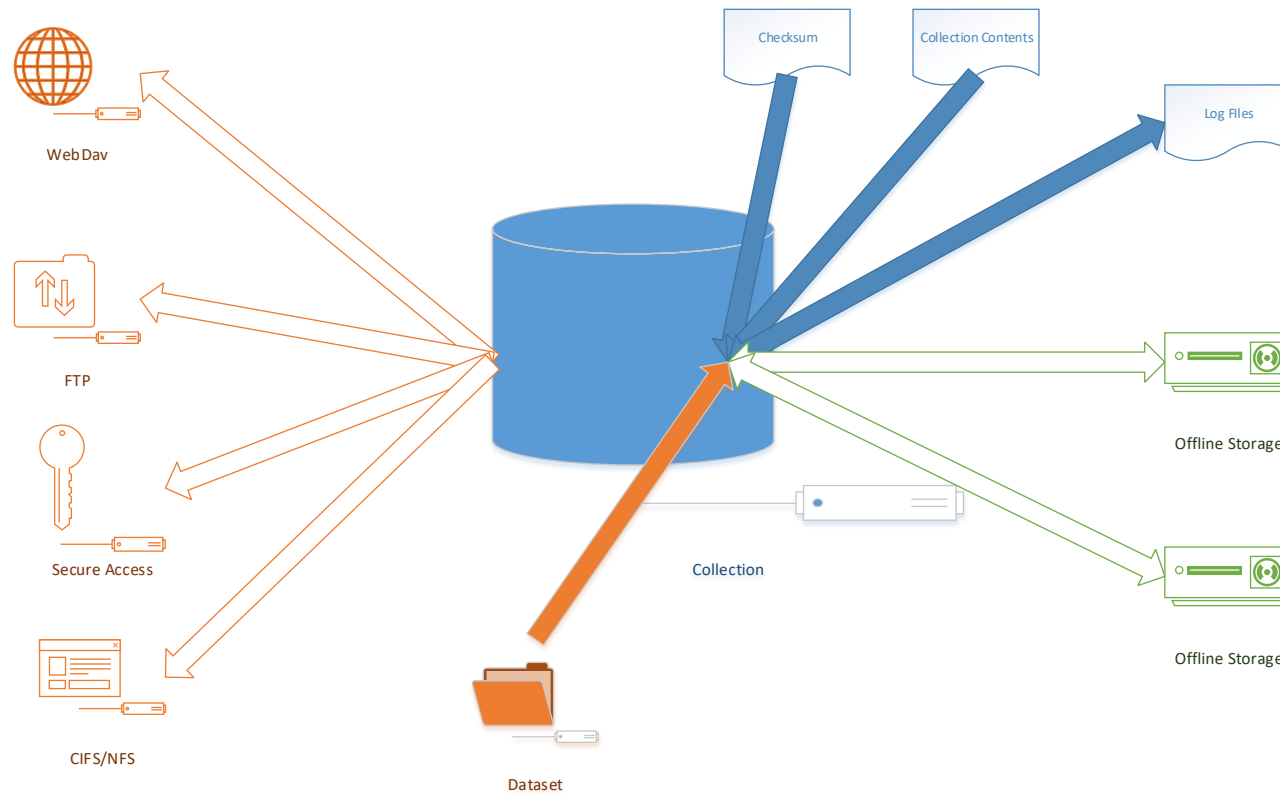
Core Features

- Manage Large Sets of Unstructured Data
- Data Protection
- Data Providence
- Data Security
- HSM Friendly
- Efficient Usage of Storage
- Quick Access to Contents
- Expose Contents via Common Protocols

Logical Collection Manager

- Data stored in a single file
 - File system inside a file system
 - Exposes collection through various protocols (CIFS, NFS, ftp, etc)
- Efficient tape usage in replacing millions of files by one
- Check summed to ensure no data loss or corruption
- File Permissions
- Surrounded by various metadata files
 - Checksums
 - Logs/Audits
 - Content list
- Maintains multiple geographic copies
- Imposes naming convention
- Unix/Linux Command line

Logical Collection Manager



Logical Collection Manager

- Command Line

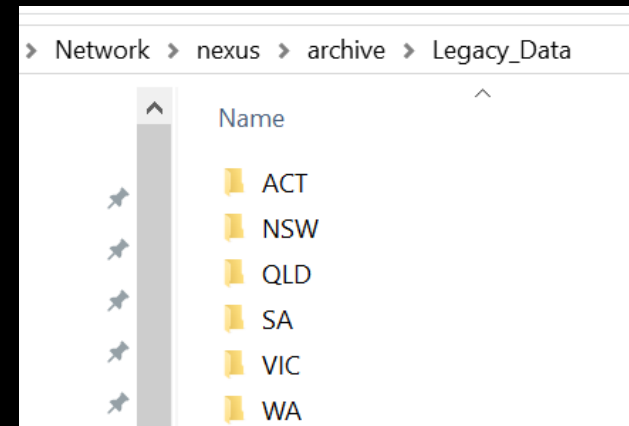
eg: root@entlcm-01-mel: /# ./encapsulate_legacy_home andrew 1 //fsvic/Home/andrew

- Profile, Environment and Category
- *Hosts (optional, override default)*
- *Formats (optional, override default)*
- Collection Name
- Action
 - Create/Refresh
 - Extract
 - Copy
 - Verify
 - Mount/Unmount
 - List
- Output Variables as name+value pair– Used by DAP

Logical Collection Manager

Enterprise Archive

```
=====
Menu ---Legacy Enterprise Data Archive-----
=====
Enter U to Unmount Archive
Enter C to Unmount All Archive based on Max Mount Date
Enter L to List Mounted Archives
Enter A to List All Available Archives
Enter M to Mount Required Archive
Enter H for Usage Info
Enter Q to quit
Enter your selection here and hit <return> .. c
█
```



Logical Collection Manager

Command Statistics

Location	Command	Count
BNE	create	242
BNE	finalise	30
BNE	mount	315
BNE	umount	231
BNE	verify	49
CDC	create	1833
CDC	extract	91
CDC	finalise	23
CDC	mount	7195
CDC	refresh	23
CDC	umount	144218
CDC	verify	1778

Location	Command	Count
MEL	create	3559
MEL	finalise	267
MEL	mount	35252
MEL	refresh	3
MEL	umount	54638
MEL	verify	32123
PER	create	237
PER	credentials	1
PER	finalise	15
PER	mount	306
PER	umount	180
PER	verify	18

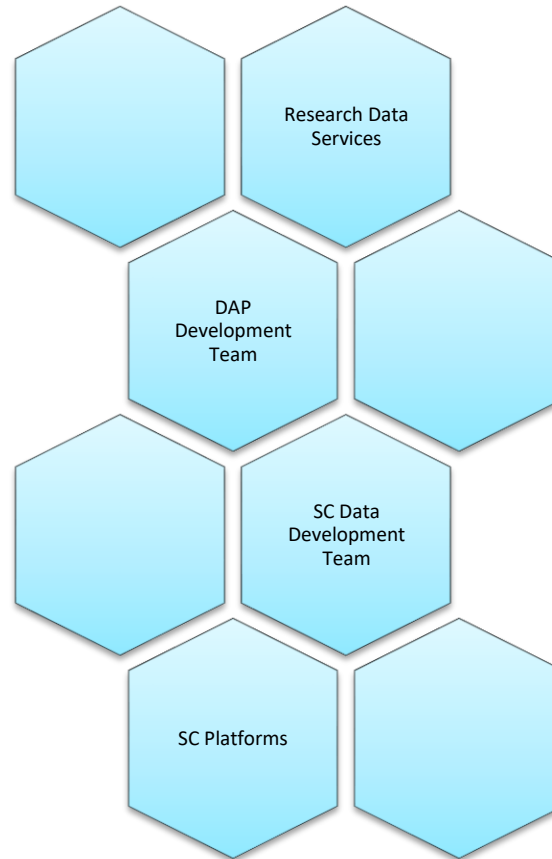
Logical Collection Manager

Data Access Portal

The screenshot shows the CSIRO Data Access Portal website. The browser address bar displays <https://data.csiro.au/collections/>. The page header includes the site title "CSIRO Data Access Portal" and navigation links for "Contact", "Help", and "API". A large banner with a bokeh background contains the text: "Access research data, software and other digital assets across a range of disciplines." Below the banner, there are tabs for "Search" and "Browse", and a "Sign in" link. The search section features three options: "Search by keyword", "Search by location", and "Search by domain". A search input field contains the placeholder text "Search CSIRO collections" and a search button. Below the input field, a note states: "Each word is searched with results returned for one of the words, any combination or all of the words." The "Featured collections" section displays two items:

- Bird survey data from the Eucalyptus salubris chronosequence 2013** (02 Aug 2019)
Elizabeth Fox, Carl Gosper, Allan Burbidge, Michael Craig, Tegan Douglas, James Fitzsimons, Shapelle McNea, Nick Nicholls, James O'Connor, Suzanne Prober, David Watson, Simon Watson, Colin Yates
Dataset of bird survey results at the multi-century Eucalyptus salubris (gimlet) chronosequence in the Great Western Woodlands, south-western Australia. This data has been used to describe responses of bird species, functional groups and community composition to time since fire (Gosper et al. 2019 Biol Cons 230, ...)
- Australian vegetated coastal ecosystems as global hotspots for climate change mitigation** (26 Jul 2019)
Oscar Serrano, Catherine Lovelock, Trisha Atwood, Peter Macreadie, Robert Canto, Stuart Phinn, Ariane Arias-Ortiz, Le Bai, Jeff Baldock, Camila Bedulli, Paul Carnell, Rod Connolly, Paul Donaldson, Alba Esteban, Carolyn J. Ewers Lewis, Brad Eyre, Matthew A. Hayes, Pierre Horwitz, Lindsay Hutley, Christopher Kavazos, ...
Data on coastal and near-coastal water in Australia that includes mangrove forests and seagrass

Logical Collection Manager



Logical Collection Manager

For DevOps

- Access Logging (Providence)
- Read Only
- DAP Friendly
- Geographically diverse copies
- Contents are verified on mount (every ~6mo)
- “Storage Friendly”
- Exposed with multiple protocols (CIFS, NFS, WebDav)
- Mature
- Fits the “Managed Data” Model

Logical Collection Manager

Future changes

- Extra Protocols (S3)
- Interface Upgrade - U.I.
- API (Webservice)
- Orthogonal Access
- 64 TB Collections
- Performance and Access Metrics
- Manifest Support
- Translate collection into new format

Logical Collection Manager



Thank you

Information Management & Technology

Andrew Smith

Data Storage and Data Centre Services Specialist

t +61 3 9545 8674

e Andrew.Smith@csiro.au

w www.csiro.au/Support-Services/IMT.aspx

INFORMATION MANAGEMENT AND TECHNOLOGY

www.csiro.au

