University of Alberta Dataverse: A journey from standalone to a hosted community platform

Guanwen Zhang[1], John Huck[1],
James Doiron[1] and Leah Vanderjagt[1]

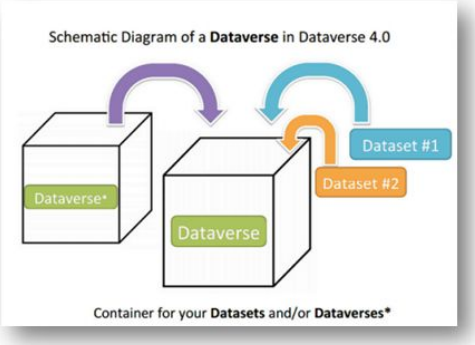
[1] *University of Alberta Library*

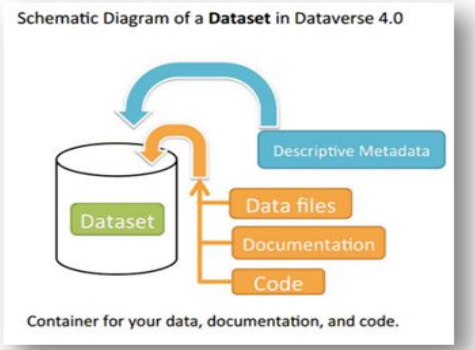
Agenda

1. Overview of Dataverse
2. Dataverse Instance at UAL
3. Scholarsportal's Hosting Platform for Dataverse
4. Migration from UA to Scholarsportal
5. Benefits and Expectations

Overview of Dataverse

- Open-source
- web based application
- by IQSS, Harvard University
- For Research Data
  - publishing
  - discovery
  - reuse and citation
  - preservation
  - FAIR

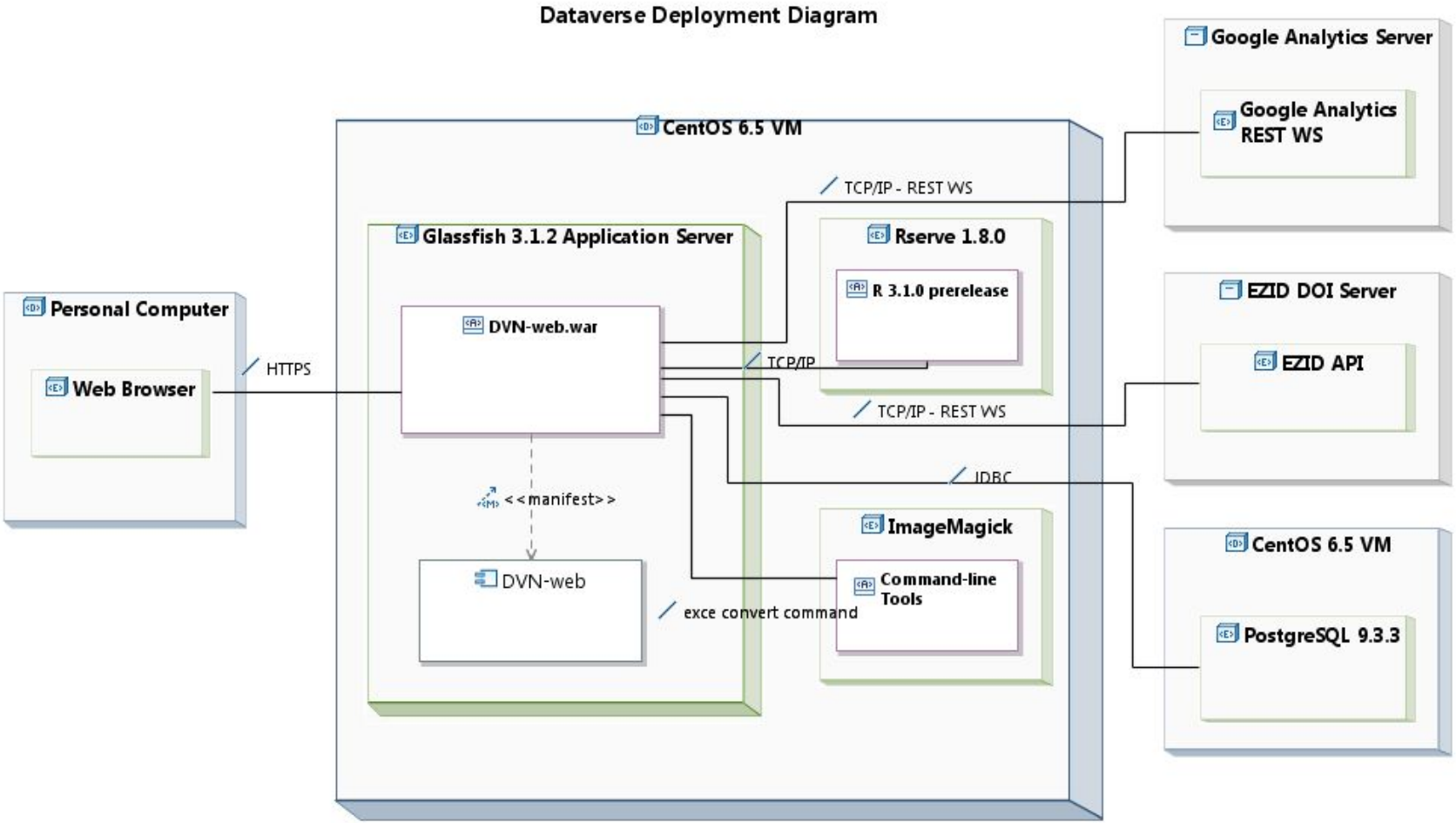




- 79 installations
- 6 continents
- 11.7K DV
- 137K datasets
- 2M data files
- 53.4M downloads

University of Alberta Dataverse

Deployment Diagrams

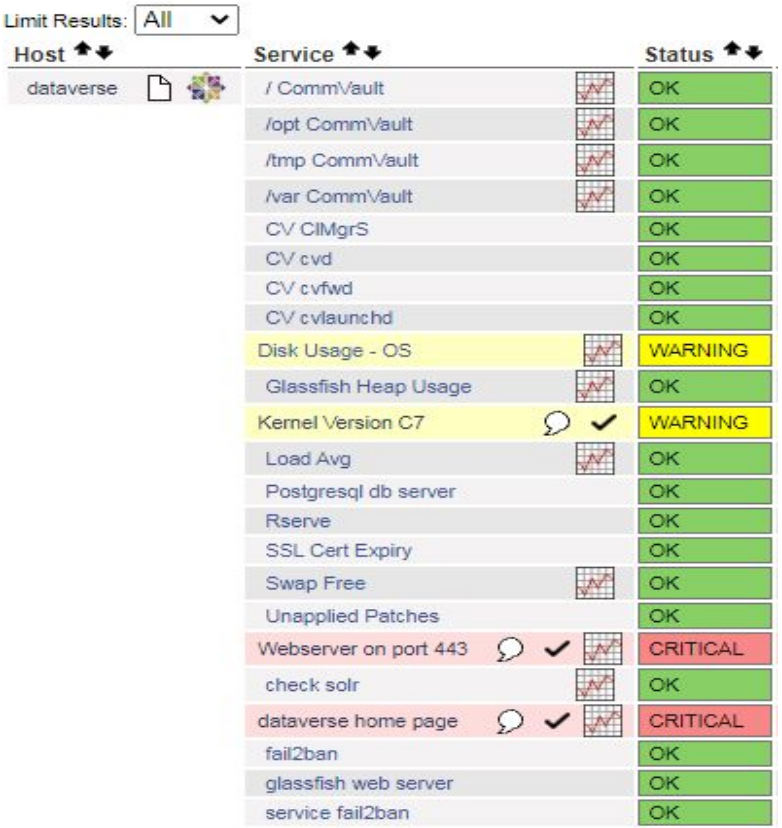


Dataverse Servers at UAL

	Development	Staging	Production
RAM	4GB	16GB	16GB
CPU	2	4	4
Disk	20GB	250GB	250GB

- First deployed in 2014
- Experienced frequent upgrading from v3.x to v4.20 till 2021

Monitoring

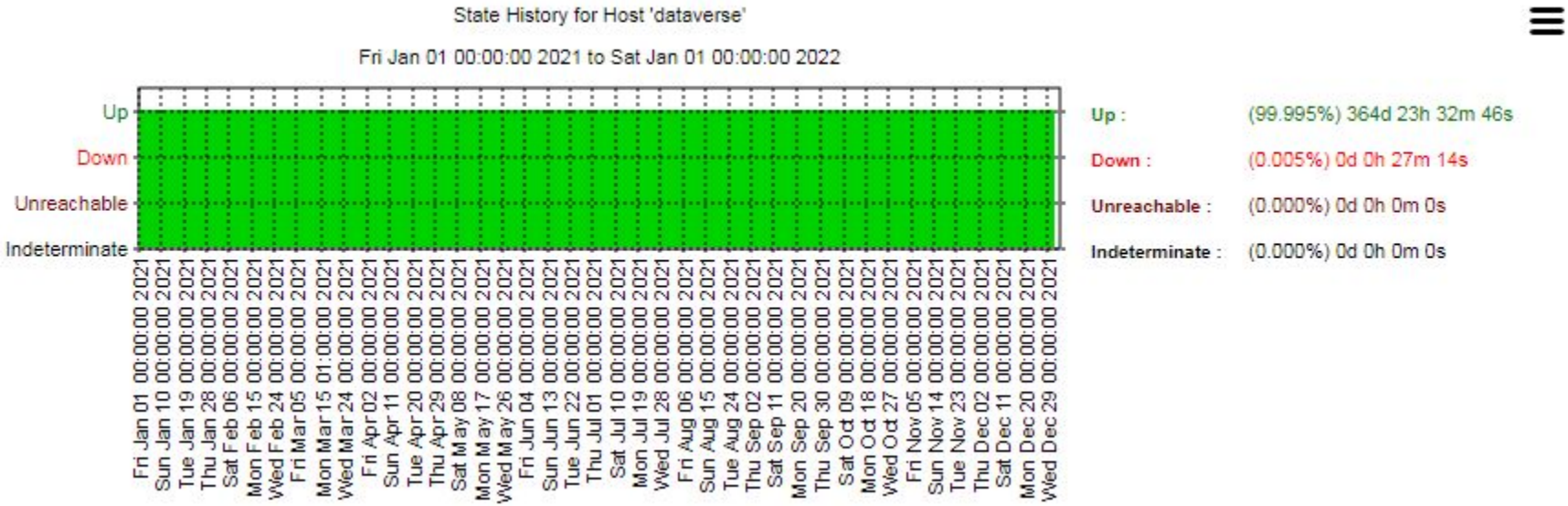


Limit Results: All

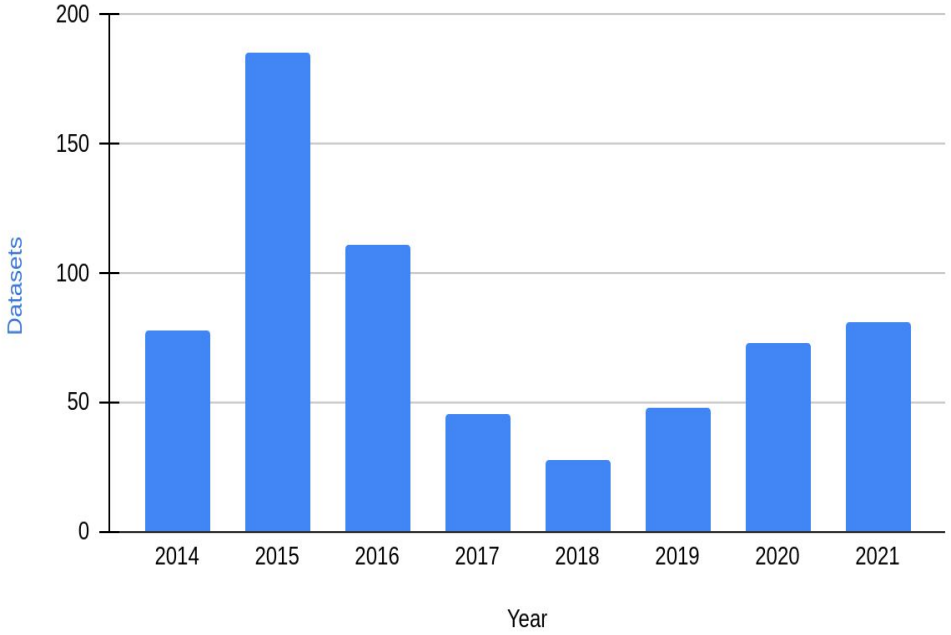
Host	Service	Status
dataverse	/ CommVault	OK
	/opt CommVault	OK
	/tmp CommVault	OK
	/var CommVault	OK
	CV ClMgrS	OK
	CV cvd	OK
	CV cvfwd	OK
	CV cvlaunchd	OK
	Disk Usage - OS	WARNING
	Glassfish Heap Usage	OK
	Kernel Version C7	WARNING
	Load Avg	OK
	Postgresql db server	OK
	Rserve	OK
	SSL Cert Expiry	OK
	Swap Free	OK
	Unapplied Patches	OK
	Webserver on port 443	CRITICAL
	check solr	OK
	dataverse home page	CRITICAL
	fail2ban	OK
	glassfish web server	OK
	service fail2ban	OK

- Is the OS update to date?
- Are the computer resources enough?
- Is the Glassfish server in a good state?
- Is the database server working well?
- Is the Solr search engine working well?
- Has the system being backed up?

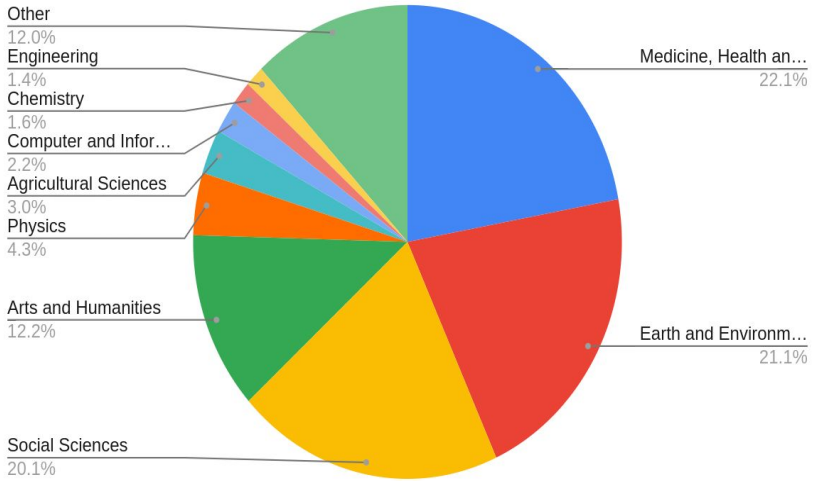
Service Level Agreement (SLA)



Number of datasets vs Year



Subject Distributions



Dataverses	128	**Data Files**	12005
Datasets	522	**Downloads**	77544

Challenges of sustaining UALD and Solutions

- Frequent and ad-hoc releases
- From v3.6.2 (2014) to present, 53 releases.
  - Scheduling updates become very challenging
  - Updating scripts developed were used only once. Not cost-effective in terms of time and efforts
- Financial unpredictability: infrastructure becomes obsolete needs to be upgraded. Money!
- Centralization of IT infrastructure and resources

-- Community based collaborative solutions --

Scholars Portal's hosting platform for Dataverse

- Hosted by Univ of Toronto Library
- Stored on the Ontario Library Research Cloud (OLRC)
- Geographically distributed: UT, UG, UO, YU, QU
- NFS and tape backup

- CentOS/Ubuntu
- OpenStack Swift/AWS
- Solr Search Engine
- Postgresql DB
- Haproxy
- Virtualization with KVM
- Java Runtime

- Archivematica
- AWS S3
- Dropbox
- JHove

Participating Institutions:

- More than 60 universities
- Across all Canada: East, West, Ontario regions

Migration of UALD to Scholars Portal

Migration Team

- University of Alberta
  - Service Managers (2)
  - Metadata Librarian
  - Systems Administrator
  - Other staff
- Scholars Portal
  - Dataverse Service Manager
  - Digital Projects Librarian
  - Programmer
  - Other staff

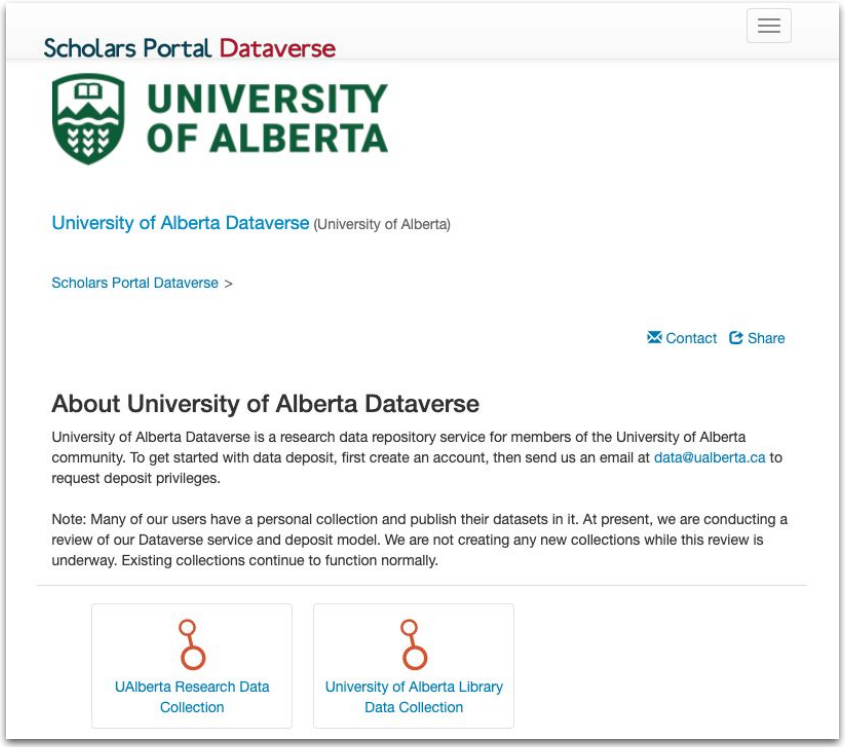
Migration Timeline

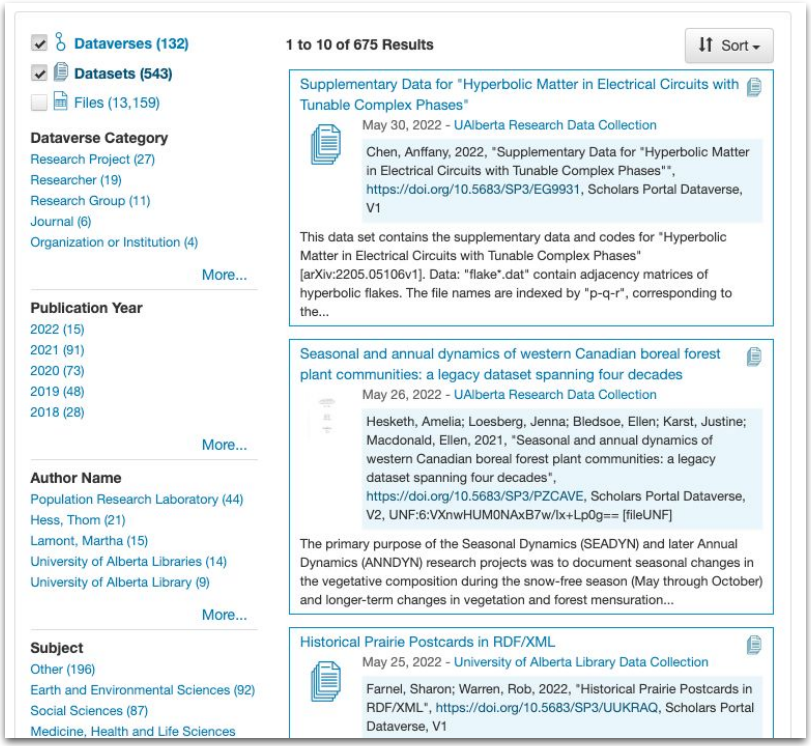
November, 2020	Project kickoff meeting
January–February, 2021	Set up technical environment and tools for team
March–June	Develop migration strategy—Create scripts
July	Small-batch migration test
August	Develop additional scripts
September	Finalize timeline—Prepare data—User communications
October	Full migration test—DOI transfer test—Communications
November	Migration completed

Migration Details

- Datasets, collections, user accounts, and permissions were migrated.
- Python scripts developed by Scholars Portal used Dataverse APIs to copy and recreate data.
- Custom SQL scripts recreated original dates of publication in database.
- Original DOIs were retained—Ownership transferred to SP via DataCite.
- Shibboleth authentication was enabled to support institutional credentials.
- Some data not migrated, like download metrics.
- Fixity checks were performed post-migration.

Migrated Dataverse Repository





Challenges & Outcomes

- Configuration of permissions on Scholars Portal Dataverse was different—Required adjustment of U of A policies.
- Initial scripts did not retain original dataset publication dates—Potential impact on data citation made this issue a priority.
- Testing uncovered a time-zone issue affecting publication dates.
- For the migration, we needed to restrict access to UAL Dataverse for users, but allow it for Scholars Portal's scripts.
- Data went "live" immediately because target repository already in production.
- No significant user issues were reported.

Benefits and Expectations

Direct Cost Reduction on HR

- System Administrator
- IT Manager
- Security Manager
- Programmer Analyst
- DI Librarian/Technical Manager
- Metadata Librarian
- RDM Services Coordinator

- DI Librarian/Technical Manager
- Metadata Librarian
- RDM Services Coordinator

Shared Infrastructure & Technical Resources

- Infrastructure is shared by more than 50 institutions
- Better usage of infrastructure: cost effectiveness
- Enhanced sustainability of infrastructure.
- Shared technical pools, and better use of expertises.
- Better positioned to integrate with other tools:
  - Cloud storage
  - Archivematica
- Increased visibility of research datasets, and hopefully increased usage of the research datasets.

Thank You To Scholars Portal's Migration Team

- Amber Leahey (Data & GIS Librarian)
- Kaitlin Newson (Digital Projects Librarian)
- Victoria Lubitch (Programmer/Analyst)

References

Crosas, M. (2011). “The Dataverse Network: An Open-Source Application for Sharing, Discovering and Preserving Data.” *D-Lib Magazine 17 (1)*. https://doi.org/10.1045/january2011-crosas.

King, G. (2007). “An Introduction to the Dataverse Network as an Infrastructure for Data Sharing.” *Sociological Methods and Research*, 36, Pp. 173–199.

Scholars Portal Annual Report (2019). https://ocul.on.ca/sites/default/files/Scholars%20Portal%20Annual%20Report%20May%202019.pdf, Last accessed on June 1, 2022.

Scholars Portal Dataverse. https://learn.scholarsportal.info/all-guides/dataverse/, Last accessed on June 1, 2022.

University of Alberta Dataverse. https://dataverse.scholarsportal.info/dataverse/ualberta, Last accessed on June 1, 2022.