Integrating Research Data into Digital Preservation Workflows for Libraries & Archives: Mapping the Landscape

**Amber Leahey and Grant Hurley, Scholars Portal, Ontario Council of University Libraries**

Summary: This poster surveys the suite of research data management services and workflows currently in development and production at Scholars Portal, the information technology service provider for the Ontario Council of University Libraries (OCUL) in Canada. OCUL's efforts in this domain have been prompted by consortium member requests for infrastructure and tools to assist researchers in effectively managing research assets during their active phases through to deposit and long-term preservation. Using shared cloud-based storage and open-source technologies, OCUL intends to support researchers with a suite of accessible management and preservation services for research data. We will do so by enabling early interventions in the long-term preservation and reuse of research data through small-scale contributions during active phase data creation and management, such as capturing metadata and documentation and ensuring it matches frameworks for access and preservation. The following workflows illustrate the current models in development and testing and the systems and tools in place, alongside related functions and activities and roles and policies.

Workflows

→ Active phase management in OwnCloud with researcher-driven deposit in Dataverse and preservation using Archivematica with storage in the OLRC.

→ Active phase management in the Open Science Framework with researcher-driven deposit to a Dataverse and preservation using Archivematica with storage in the OLRC.

→ Curation phase deposit by researcher to an institutional repository and automated preservation using Archivematica with storage in the OLRC.

→ Curation phase deposit by researcher to Dataverse and automated preservation using Archivematica with storage in the OLRC

Optional Workflow Connectors

1. Researchers may choose to manage or move files between OwnCloud and the Open Science framework depending on needs.

2. The Open Science Framework allows researchers to upload data to, and pull from, Dataverses.

3. Data may enter Permafrost directly via legacy media stored in archives and libraries. Access copies can be made available via institutional repositories or Dataverse.

Tools
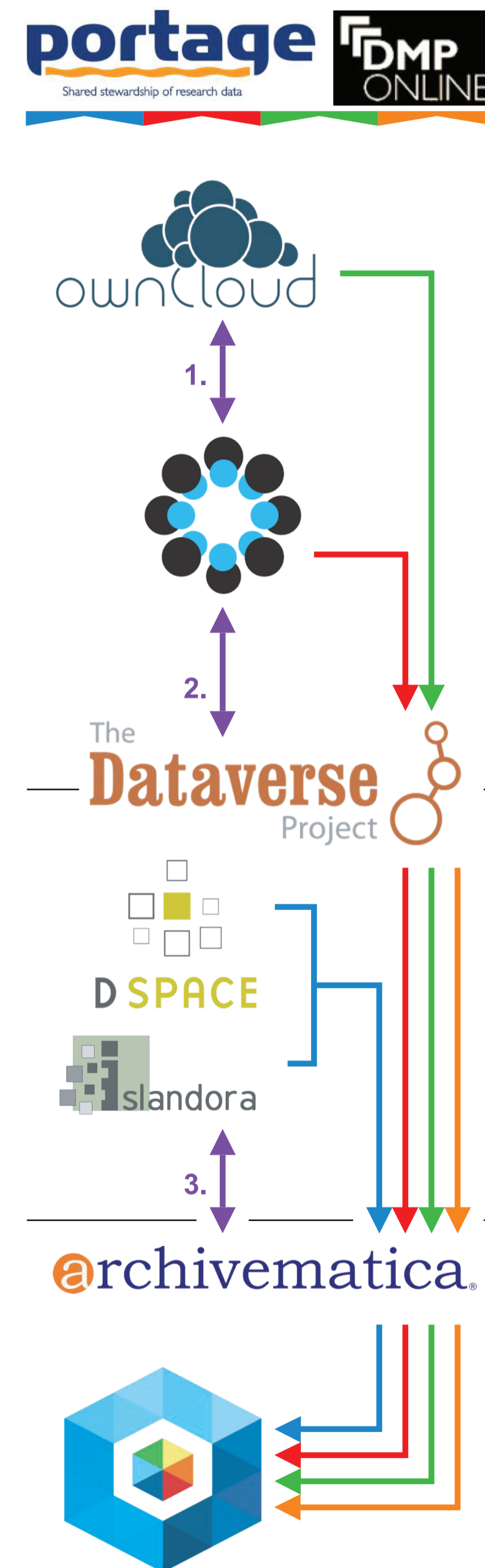
Active Phase Management

DMP Assistant
Tool for preparing research data management plans describing metadata, formats, storage, etc. planning as part of grant process.

OwnCloud
Dropbox-like application using the OLRC for storing and managing active-phase data.

Open Science Framework
Third party hosted platform for research data management and collaboration.

Dataverse
Multi-institutional hosted research data repository for
data deposit, discovery and reuse.

Private settings enable active-phase file management, file permissions, encryption.

Curation & Access

Public settings enable final data files to be accessed, preserved and reused.

Institutional Repositories
Member-hosted repositories where researchers may submit data, such as local DSpace or Islandora installations. May also include data on legacy media donated to archives and special collections.

Preservation Services

Archivematica
OAIS-compliant preservation processing tool.

Permafrost
Cloud hosted preservation service using Archivematica for OCUL libraries and archives.

Ontario Library Research Cloud (OLRC)
Replicated object storage cluster for Ontario academic libraries built using Swift.

Functions & Activities

- Assess data management needs and requirements.
- Manage research data creation, including file versioning, syncing, etc.
- Coordinate and communicate data creation/management functions across individuals involved in project.
- Collect, document, organize, clean and analyze data for publication and presentation.
- Generate basic documentation and related metadata.
- Appraise data and select final versions for deposit, including appropriate final formats and required documentation.

- Assign persistent identifiers and citations, add descriptive metadata, and manage access (restricted vs. open).
- Provide search, discovery and analysis access through open tools (e.g. APIs, discovery interfaces).
- Generate metadata relating to description, provenance, context, fixity, and rights.
- Control versions, retractions, releases.

- Pull data and metadata into transfer folder and arrange for ingest into Archivematica.
- Perform preservation micro-services such as file format identification and normalization, preservation metadata creation, fixity checks, and the creation of access copies (dissemination information packages) and archival copies (archival information packages).
- Store replicated copies of archival packages across 3 nodes. Regular fixity checks monitor integrity and management system monitors for file format obsolescence.

Roles & Policies

- Funder-driven policies.
- Researcher-driven policies, procedures and workflows.
- Guidance on appraisal and file formats from RDM liaison librarians at institutions.
- Institutional policies for sensitive and private data.

- Curation and enhancement, including quality review.
- Rule-based metadata creation and validation.
- Metadata and fixity checks on stored files that can be passed onto preservation services.
- Monitoring by data librarians.

- Appraisal and retention schedules
- Preservation policies and procedures, including format normalization/migration policies.
- Policy-based preservation storage for fixity checks, format migrations, and storage.
- Monitoring by preservation librarian.