Rethinking digital preservation
Looking beyond the repository

CRKN 2020 Conference | 15 October 2020

Corey Davis
Digital Preservation Librarian, University of Victoria
Visiting Program Officer, Canadian Association of Research Libraries

Territorial acknowledgement

CARL ABRC

Overview of the session

- What is digital preservation?
- Why are libraries struggling?
- What are some solutions?
- Discussion

CARL **ABRC**

What is digital preservation?

- "Digital preservation consists of the processes aimed at ensuring the continued accessibility of digital materials. To do this involves finding ways to re-present what was originally presented to users by a combination of software and hardware tools acting on data."

  – UNESCO https://en.unesco.org/themes/information-preservation/digital-heritage/concept-digital-preservation

CARL ABRC

CARL Digital Preservation Survey

- "96% of the respondents indicated that they provide access to digital materials under their care [and] 79% … use a digital object repository or discovery platform."

    – Hurley, G., & Shearer, K. *Final Report of the Survey on Digital Preservation Capacity and Needs at Canadian Memory Institutions, 2017-18*. Canadian Association of Research Libraries.

CARL ABRC

CARL Digital Preservation Survey

- "Only a small proportion of respondents are confident in their abilities to preserve their holdings..."
  - *Ibid.*

CARL✶ABRC

An issue of priorities and attention

- "…memory organizations…have not yet prioritized digital preservation within their operations and do not direct significant resources towards these activities."

    – *Ibid.*

CARL ABRC

Collective action to increase capacity

- Portage Preservation Experts Group
  - "Advise on developing RDM infrastructure and best practices for preserving research data and metadata…"

- CARL Digital Preservation Working Group
  - "Assessing digital preservation capacity, available resources, and funding opportunities within the Canadian research community, and identifying and promote relevant approaches, standards, practices, and technologies…"

CARL ABRC

What we'll be talking about today

1. Its difficult to connect individual repositories to shared preservation services

2. The current repository ecosystem privileges near-term access (via repositories) over long-term access (i.e. preservation)

CARL✦ABRC

COPPUL's WestVault

CASE STUDY

Case study: WestVault

- COPPUL Private LOCKSS Network for distributed preservation storage was underutilized
  - Hard to create repository plug-ins for LOCKSS
  - Each LOCKSS box had to grow to accommodate large deposits
  - Lack of service-level agreements

CARL ABRC

Case study: WestVault

- Reboot!
  - Drag-and-drop deposit
  - Virtual Machines at large data-centres
  - Robust policies, user manuals and training, dedicated staff at COPPUL

CARL ❧ **ABRC**

Case study: WestVault

- Results
  - Getting stuff out of repositories for deposit still requires lots of attention and manual intervention

CARL✦ABRC

Case study: WestVault

- What needs to be done
  - Create capacity locally and through regional consortia and national organizations

  - **Scale repositories like we've scaled preservation services, and connection the two**

  - **Maybe rethink how we do access and preservation!**

CARL ABRC

Looking at exemplar systems for solutions to the problem

SCALING REPOSITORIES

CARL ABRC

BC ELN and COPPUL

EXEMPLAR INFRASTRUCTURE 1

BC ELN and COPPUL

- Arca
  - "…shared platform provides organisations with open-access repositories."
    https://bceln.ca/services/shared-services/arca

- WestVault
  - "…a high-redundancy, independently-administered peer storage network spread across all 4 Western provinces."
    https://bceln.ca/resources/westvault

CARL✦ABRC

Lightweight policy framework…

BC ELN and COPPUL working together: Arca Collaborative Digital Repository and COPPUL Digital Preservation Network

🖨 Printer-friendly version

VANCOUVER, B.C., November 30, 2016 — The British Columbia Electronic Library Network (BC ELN) and the Council of Prairie and Pacific University Libraries (COPPUL) announced today the signing of an agreement that will enable COPPUL members to join the Arca Collaborative Digital Repository while providing existing Arca members access to COPPUL's emerging suite of digital preservation services.

Both BC ELN and COPPUL have a common interest in ensuring that the high-value digital materials of their member institutions are accessible to their communities over the long term. By leveraging their respective expertise and service infrastructures in the areas of institutional repositories and digital preservation, both BC ELN and COPPUL members will benefit from this agreement.

Under this agreement, COPPUL members will be able to participate in the Arca Collaborative Digital Repository, which is based on a shared instance of the open source Islandora Digital Repository Framework. At the same time, Arca participants will have access to an emerging set of digital preservation services being developed as part of the COPPUL Digital Preservation Network (COPPUL DPN), including preservation storage in COPPUL's Private LOCKSS Network, and preservation processing services such as Archivematica-as-a-Service.

CARL ✦ ABRC

E·COLLECTIONS
AT EMILY CARR

Explore Emily Carr University's eCollections

Search Emily Carr University's Institutional Repository. Find scholarly research materials and other digital assets from our archives and the Wosk Masterworks Print Collection.
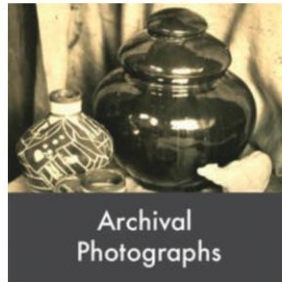
Search All Collections 🔍

Advanced Search

Browse Collections
Browse All Materials

About

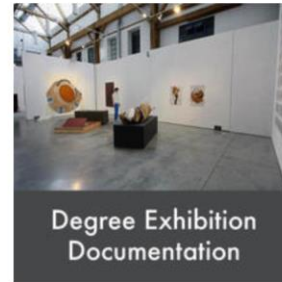Heinz Laffin in ceramics studio, circa late 1950's - early 1960's.

CARL✦ABRC

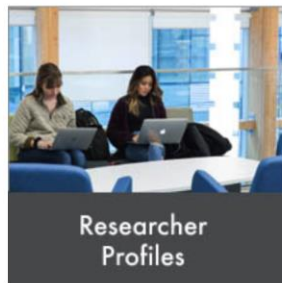Academic Calendar
Archive

Archival
Photographs

Degree Exhibition
Documentation

Graduate Student
Theses
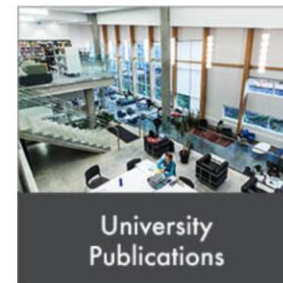
Research + Creation

Researcher
Profiles

Student Publication
Archive

University
Publications

Wosk Masterworks
Print Collection

CARL✳ABRC

Aurora

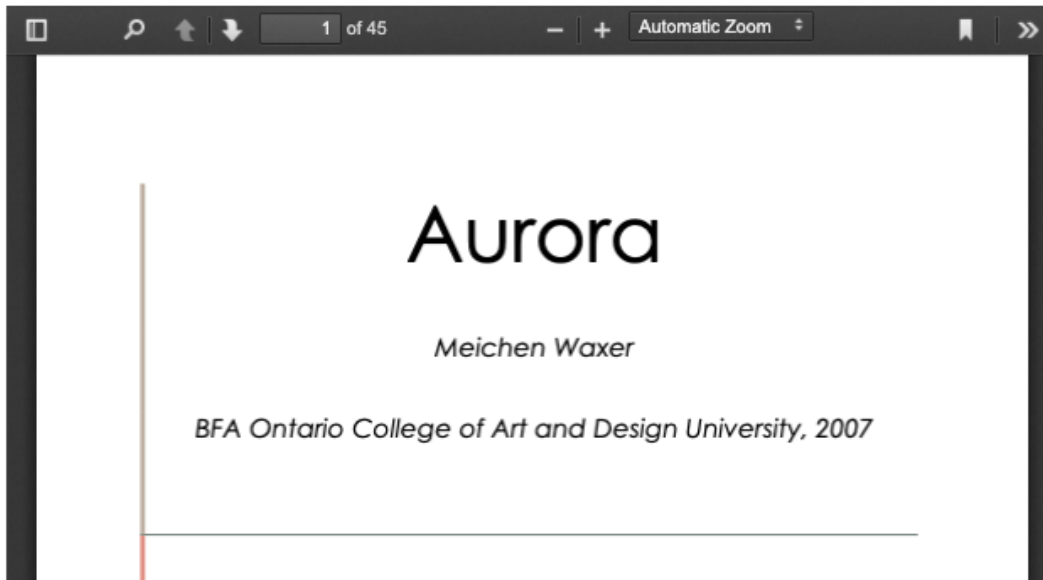| View | IP Embargo | Manage | Edit metadata |

Style

APA ▾

Choose the citation style.

Waxer, M. (2016). Aurora. doi:10.35010/ecuad:8455

▢   🔍   ⬆   ⬇   1   of 45   —   +   Automatic Zoom ▾   🔖   »

Aurora

Meichen Waxer

BFA Ontario College of Art and Design University, 2007

Part of: Aurora (10 objects)
manage parent
Next

Aurora

Untitled (Evenlode)

Bird of paradise, tree of life

CARL◆ABRC

Aurora **COLLECTIONS**

VIEW | IP EMBARGO | MANAGE | EDIT METADATA

Home » Browse Repository » Graduate Theses

Overview | Datastreams | Properties | Compound | Create Bag(s) | Object Policy | Preservation | Solr Document | DOI

☐ Preserve this object?

Checking this box creates a PRESERVATION datastream, which will flag this object for preservation on the next cron run. Un-checking this box will delete the PRESERVATION datastream, which will not delete the preserved objects, but stop the preservation of updates to the object.

Save Configuration

Part of: Aurora (10 objects)

CARL ● ABRC

Overview   Datastreams   Properties   Collection   Collection Permissions   Compound   Create Bag(s)   Object Policy   Owner Proxies

Preservation   Solr Document   DOI

☐ Preserve this object?

Checking this box creates a PRESERVATION datastream, which will flag this object for preservation on the next cron run. Un-checking this box will delete the PRESERVATION datastream, which will not delete the preserved objects, but stop the preservation of updates to the object.

New children of this collection (excludes current children)
✓ All children of this collection (existing and new)
Current children only - new objects will not be preserved   hild collections. Child collections must be configured individually.

Save Configuration

CARL · ABRC

Preservation integration at scale…

- It cost COPPUL $3000 to support BC ELN's development of Islandora WestVault

- Arca = $5100/yr (150Gb allocation)
- WestVault = $200/yr per 100Gb

CARL ABRC

Extending access across the GLAM sector…

BCELN
BRITISH COLUMBIA
ELECTRONIC LIBRARY NETWORK

IRVING K. BARBER LEARNING CENTRE

Hosting Initiative Expands Access to Province's Historical Treasures

Burnaby, B.C., August 24, 2020 – Past, present, and future BC History Digitization Program (BCHDP) grant recipients have access to free, sustainable, standards-based hosting for their online collections into 2025 thanks to an extension of the BCHDP-Arca Hosting Initiative. A renewed 5-year commitment by the British Columbia Electronic Library Network (BC ELN) and the Irving K. Barber Learning Centre (IKBLC) at the University of British Columbia Library builds on an established 3-year partnership that has allowed BCHDP grant recipients to host their digitized material on Arca, BC's collaborative digital repository, at no cost.

Ten GLAM sector (Galleries, Libraries, Archives, and Museums) organizations – groups like the Potters Guild of BC and grunt gallery – have taken part in this initiative over the preliminary 3-year pilot period, and several more are in discussions to add their collections to Arca. They join a growing number of post-secondary institutions and GLAM organizations that are making their local digital assets openly available via the Arca website (arcabc.ca).

IKBLC's continued funding, along with BC ELN's dedicated administration of the initiative and of the award-winning Arca platform, ensures that the BCHDP-Arca Hosting Initiative remains an ongoing and sustainable option for grant recipients. The successful partnership advances both BCHDP's goal of promoting access to British Columbia's unique historical materials and Arca's vision of sharing digital scholarship and heritage and building communities in British Columbia and beyond.

CARL✦ABRC

Extensible to other repository services

📁	includes	Add maximum filesize check	14 months ago
📄	.travis.yml	Change travis php dist	13 months ago
📄	LICENSE.txt	Get started	2 years ago
📄	README.md	Update Readme with config rules	16 months ago
📄	islandora_westvault.drush.inc	Remove debug line	13 months ago
📄	islandora_westvault.info	Setup paged and compound for use with islandora_bagit_c...	16 months ago
📄	islandora_westvault.module	Change compound child preservationto foreach loop	16 months ago

bondjimbond Merge pull request #46 from bondjimbond... ... ✓ 2a5c2cb on Aug 26, 2019 🕐 117 commits

README.md

Islandora Westvault

Introduction

Islandora Westvault lets you flag invididual objects, newspapers, books, or whole collections to be preserved in COPPUL's Westvault LOCKSS service or another OwnCloud-based service.

CARL✦ABRC

Scholars Portal and COPPUL

EXEMPLAR INFRASTRUCTURE 2

CARL ✺ ABRC

Scholars Portal and COPPUL

- Dataverse
  - "…a publicly accessible data repository platform...hosted completely on Canadian servers at the University of Toronto Libraries." https://dataverse.scholarsportal.info/

- Archivematica-as-a-Service
  - "…an open-source digital preservation system [that can be used] without installing and locally managing Archivematica." https://coppul.ca/archivematica

CARL✦ABRC

Working together in Ontario…

OCUL Ontario Council of University Libraries

News   Contact   [   ]   **Search**

About | Community | Services | Resources | Vendors | | Log In

Home » Dataverse-Archivematica Integration Now Available for Testing

Dataverse-Archivematica Integration Now Available for Testing

Posted on: 5 March 2019   **By:** Kaitlin

The Ontario Council of University Libraries (OCUL) and Scholars Portal are pleased to announce that the first version of the Dataverse-Archivematica integration for long-term dataset preservation is now available for community testing and feedback.

For the past several years the Scholars Portal team has been working with Artefactual Systems Inc. to create an integration between the preservation processing tool Archivematica and Dataverse. This technical integration enables users of Archivematica to select and copy datasets from Dataverse instances and process them for long-term access and digital preservation. The Archival Information Packages (AIPs) include metadata and all files stored by Dataverse, and are packaged for long-term storage and management usingArchivematica's standards-compliant workflow. Integration between Archivematica and Dataverse expands capacity to preserve research data, opening up new avenues for librarians and archivists to make valuable research outputs accessible into the future.

CARL ABRC

UVic Research Data Collection

Ocean Networks Canada

CHONe

Canada Health Infoway
Canada Health Infoway

Search this dataverse...   🔍 Find   **Advanced Search**

☑ ⅛ **Dataverses (4)**

☑ 📄 **Datasets (40)**

☐ 🗒 Files (291)

Dataverse Category

Organization or Institution (4)

Publication Year

2019 (21)

2018 (13)

2020 (10)

Author Name

1 to 10 of 44 Results   ⇅ Sort ⌄

Data from: How far have we come? A review of MPA network performance indicators in reaching qualitative elements of Aichi Target 11   📄

Sep 15, 2020 - CHONe

Miller, Mairi; Ban, Natalie; Devillers, Rodolphe; Singh, Gerald; Claudet, Joachim, 2020, "Data from: How far have we come? A review of MPA network performance indicators in reaching qualitative elements of Aichi Target 11", https://doi.org/10.5683/SP2/FPYIVC, Scholars Portal Dataverse, V1, UNF:6:RpVEXjI4EVf/w858NYQ3Gw== [fileUNF]

This dataset is based on a systematic literature review to identify indicators used to assess MPA network effectiveness in achieving the qualitative elements. Web of Science core collection database (1900 to April 2019) and Elsevier's Scopus database (1995 to April 2019) has been...

CARL ABRC

Staging path:

```
/var/archivematica/storage_servi
ce
```

Absolute path to a staging area. Must be UNIX filesystem compatible, preferably on the same filesystem as the path.

Host:

`dataverse.scholarsportal.info` 🔳   Hostname of the Dataverse instance. Eg. apitest.dataverse.org

API key:

`2a4c8992-e710-4c56-ba0c-df12`   API key for Dataverse instance. Eg. b84d6b87-7b1e-4a30-a374-87191dbbbe2d

Agent name:

`University of Victoria Dataverse`   Agent name for premis:agentName in Archivematica

Agent type:

`Dataverse`   Agent type for premis:agentType in Archivematica

Agent identifier:

`https://dataverse.scholarsportal.i`   URI agent identifier for premis:agentIdentifierValue in Archivematica

[Edit Space]

Transfer	UUID	Transfer start time	

Standard — Transfer type

Transfer name

Accession no.

Access system ID

Browse | Start transfer

☑ Approve automatically

Dataverse ▾

Transfer type

Transfer name

Accession no.

Access system ID

Browse | Start transfer

☑ Approve automatically

The University of Victoria Dataverse ▾

📁 72612 (2014 National Survey of Canadian Community Pharmacists: Use of Digital Health Technologies in Practice)
📁 72603 (2014 National Survey of Canadian Nurses)
📁 72614 (2014 National Survey of Caregivers)
📁 72621 (2016 National Survey of Community-Based Pharmacists: Use of Digital Health Technology in Practice)
📁 72617 (2017 National Survey of Canadian Nurses: Use of Digital Health Technology in Practice)
📁 75038 (2018 Canadian Physician Survey)
📁 105311 (Algal lipids in the Canadian arctic)
📁 74061 (B.C. Ministry of Environment Environmental Quality Information System (EQUIS) Database)
📁 71772 (Compiling and mapping global permeability of the unconsolidated and consolidated Earth: GLobal HYdrogeology MaPS 2.0 (GLHYMPS 2.0). [Supporting Data])
📁 77377 (Data from: A climate-associated multi-species cryptic genetic cline in the northwest Atlantic")
📁 93020 (Data from: An hypoxia-tolerant flatfish: Consequences of sustained stress on the slender sole Lyopsetta exilis (Pleuronectidae) in the context of a changing ocean)
📁 96763 (Data from: Biodiversity and habitat assessment of coastal benthic communities in a sub-Arctic industrial harbour area)
📁 72090 (Data from: Composition and functional diversity of macrofaunal assemblages on vertical walls of a deep northeast Pacific fjord)
📁 111204 (Data from: Continental margin sediments underlying the NE Pacific oxygen minimum zone are a source of nitrous oxide to the water column)
📁 93083 (Data from: Food and initial size influence overwinter survival and condition of a juvenile marine fish (age-0 Atlantic cod))

Add

Transfer	UUID	Transfer start time	⊖

Test_Dataset_for_UVic_Dataverse Archival Information Package

UUID	cd64bee3-0fe8-47c1-bd95-ae36956f1f59
Size	12.03 MB
Date stored	2020-05-12 17:57
Status	Stored
Encrypted	False
Location	Download [...]/Test_Dataset_for_UVic_Dataverse-cd64bee3-0fe8-47c1-bd95-ae36956f1f59.7z
Pointer file	View

Actions

Upload DIP   Re-ingest   Delete

Metadata-only DIP upload to AtoM

Insert slug *

Only compatible with AtoM 2.4 or newer.

Upload

CARL ABRC

Portage Preservation Expert Group…

- Recommendations for a national preservation strategy

CARL✦ABRC

Its not just about preservation!

OTHER BENEFITS TO SCALE

Encouraging investments in infrastructure

- When was the last time you got grant funding to enhance your repository?

CANARIE Awards up to $3.2M in Funding to 9 Research Teams to Develop Research Data Management Software Tools

Posted on **November 27, 2018** in **Press Releases**

The ability to reuse research data helps accelerate discovery, allows for reproducibility of scientific results, and maximizes return on investment of research funding

Funded software tools aim to strengthen Canada's capacity to manage research data
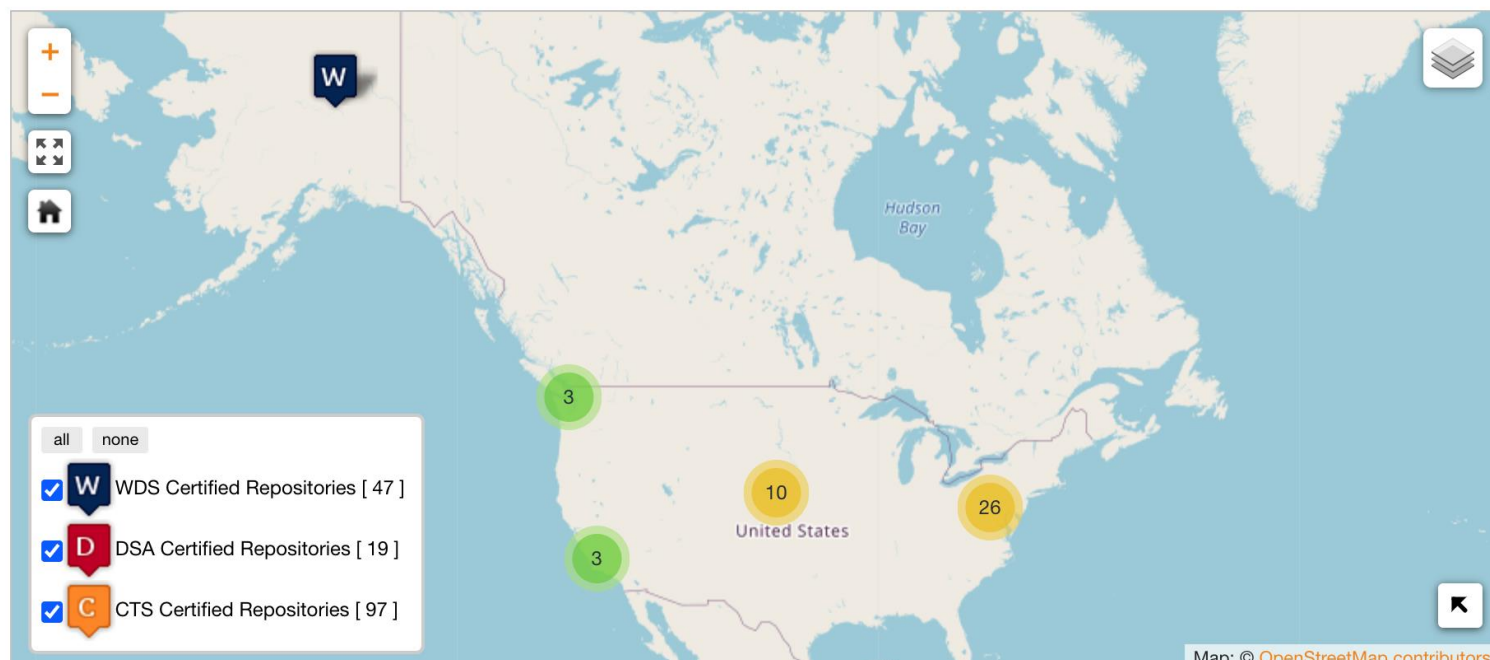
Building 'Trust' with research communities

- "Repositories must earn the trust of the communities they intend to serve and demonstrate that they are reliable and capable of appropriately managing the data they hold."

  - Lin, D., Crabtree, J., Dillo, I. *et al.* The TRUST Principles for digital repositories. *Sci Data* **7,** 144 (2020). https://doi.org/10.1038/s41597-020-0486-7

CARL✦ABRC

Core Certified Repositories

Applications are made public only once certification of a data repository has been approved by the CoreTrustSeal Board. Certification is against the version of the Core Trustworthy Data Repositories Requirements named in the link to the public application (e.g., 2017–2019). The CoreTrustSeal for Data Repositories is valid for three years from the certification date listed within the public application.

W

3

10

26

United States

3

Hudson Bay

all   none

☑ **W** WDS Certified Repositories [47]

☑ **D** DSA Certified Repositories [19]

☑ **C** CTS Certified Repositories [97]

Map: © OpenStreetMap contributors

CARL ABRC

Pooled developer resources mean more agility

- "We must prepare for the rapidly evolving technological landscape needed to support scholarly communication. Agility is essential."

  - Arlitsch, K., & Grant, C. (2018). Why so many repositories? Examining the limitations and possibilities of the institutional repositories landscape. *Journal of Library Administration*, *58*(3), 264-281.

CARL ABRC

And more…

- Enhanced discovery through more consistent metadata

- Integration w/ other systems at scale (e.g. the PID graph, ORCHID, etc.)
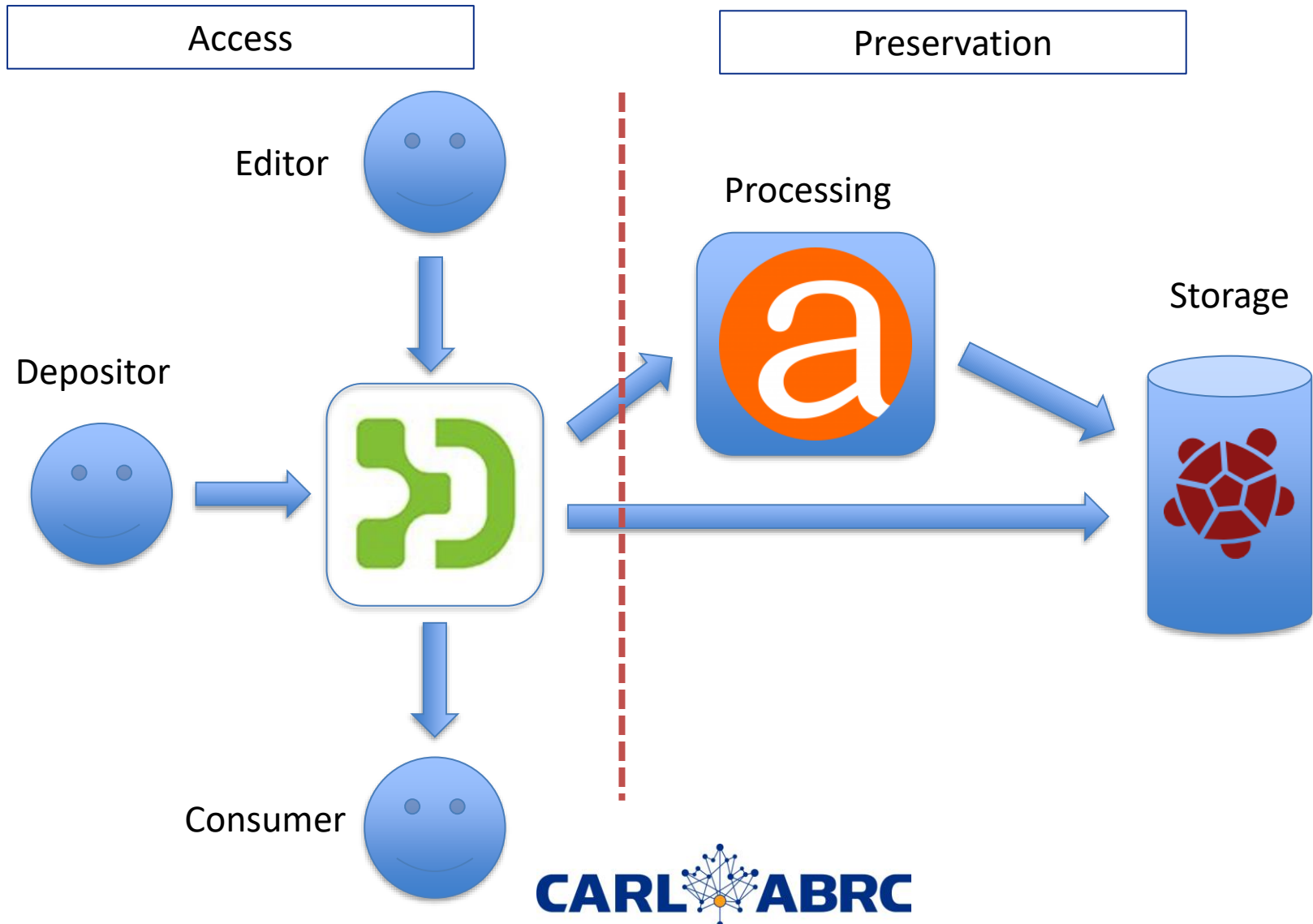
- Network effects amplify impact

Thinking outside the repository box
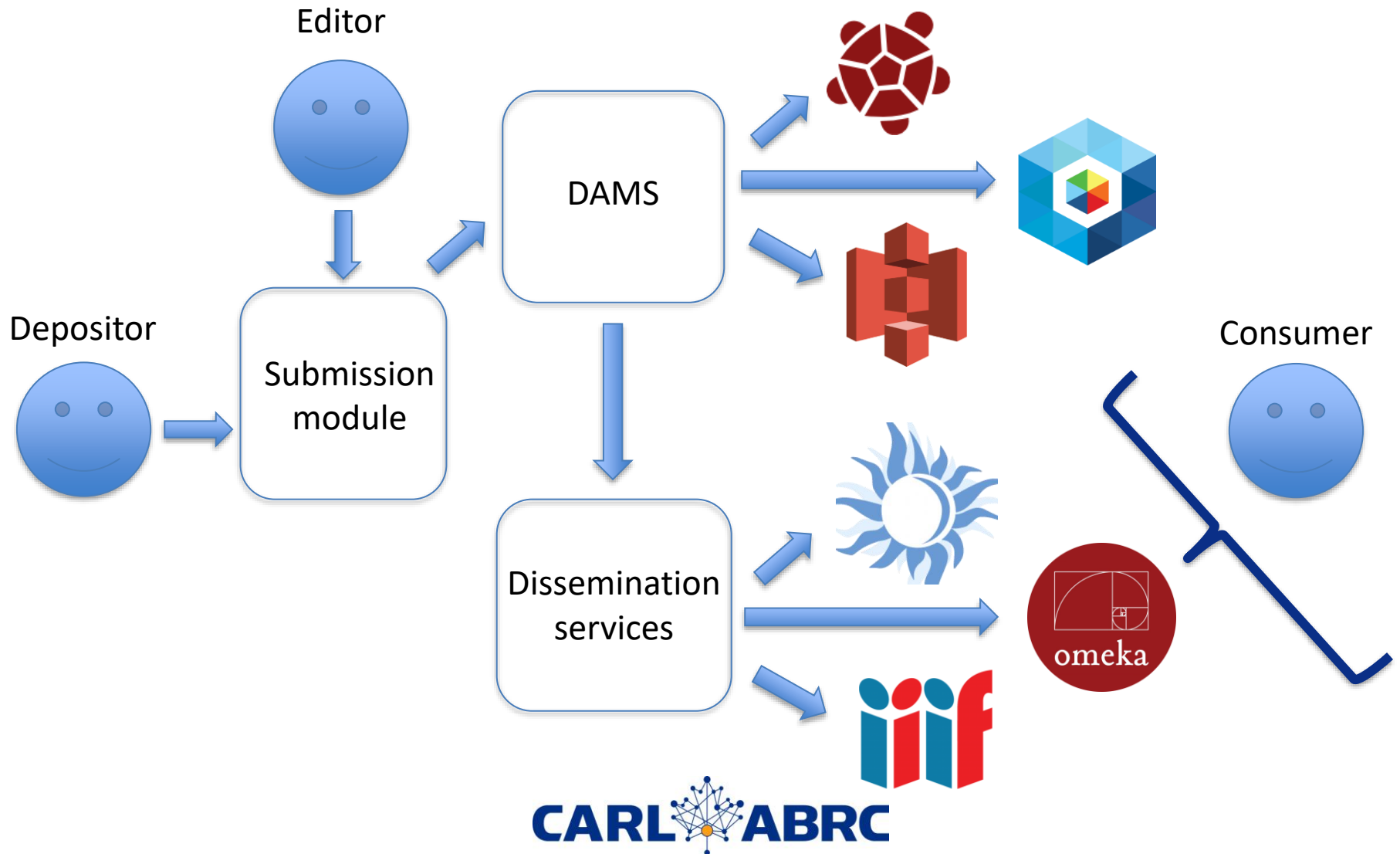
RETHINKING ACCESS AND PRESERVATION

Inspired by the folks at U of T and Scholars Portal, UBC, and others…

- What if we rethought our workflow and put preservation first!

CARL✺ABRC

Traditional repository preservation workflow

Thinking beyond the repository

Editor

Depositor

Consumer

Submission module

DAMS

Dissemination services
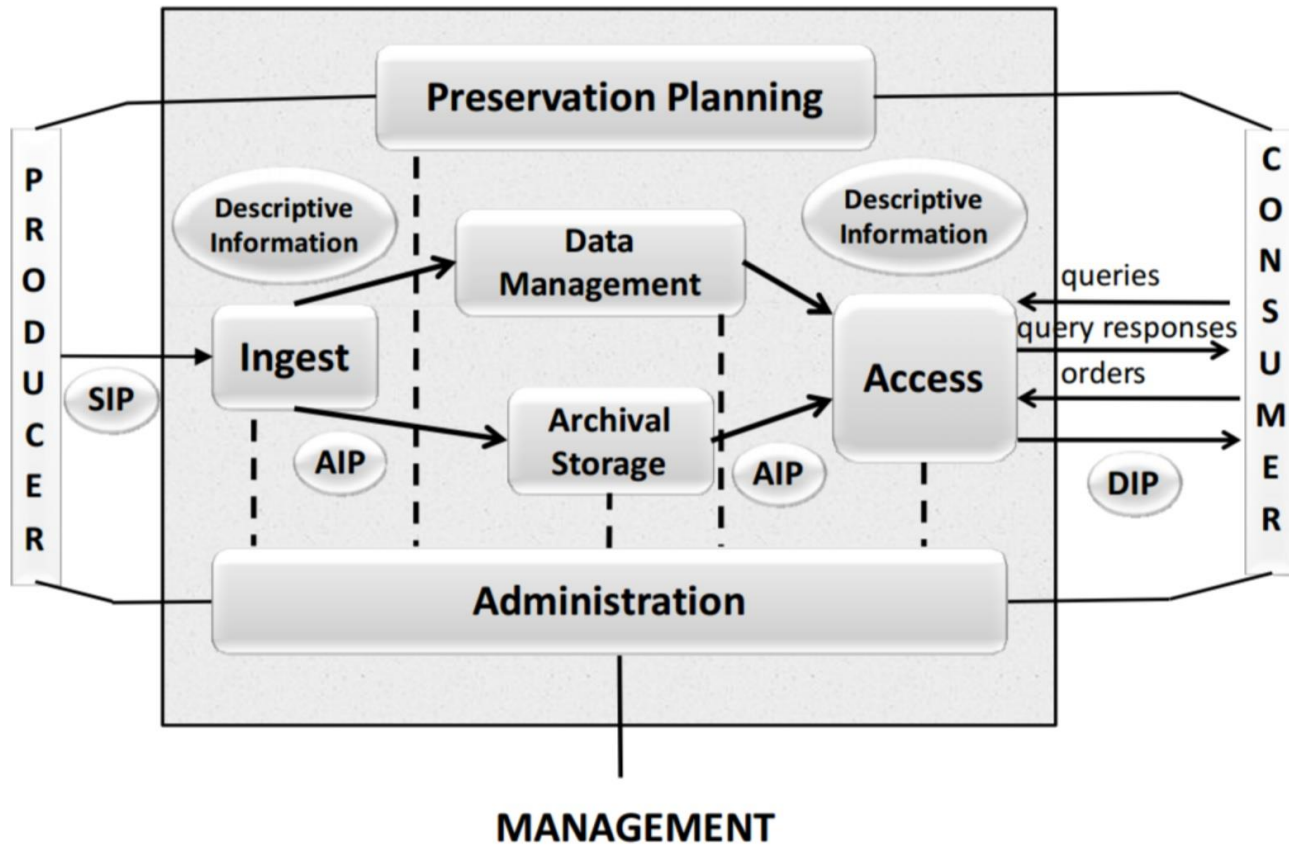
CARL✷ABRC

Open Archival Information System

Figure 4-1: OAIS Functional Entities

The archival advantage

- Libraries have always privileged access
- Archives have made preservation a part of access (and were decades ahead of libraries in terms of digital preservation)
- Library thinking privileges repositories at the expense of preservation

OAIS

- OAIS provides a path to a more service-based, researcher-focused, preservation-friendly approach to access
- Librarians and archivists need to work together

DISCUSSION

Questions and comments

- Diversity vs. scale: What's the right scale? Institution, region, nation?

- How unique are local repository needs?

- What are the switching costs?

CARL✱ABRC

Thanks!

- coreyd@uvic.ca

CARL ABRC