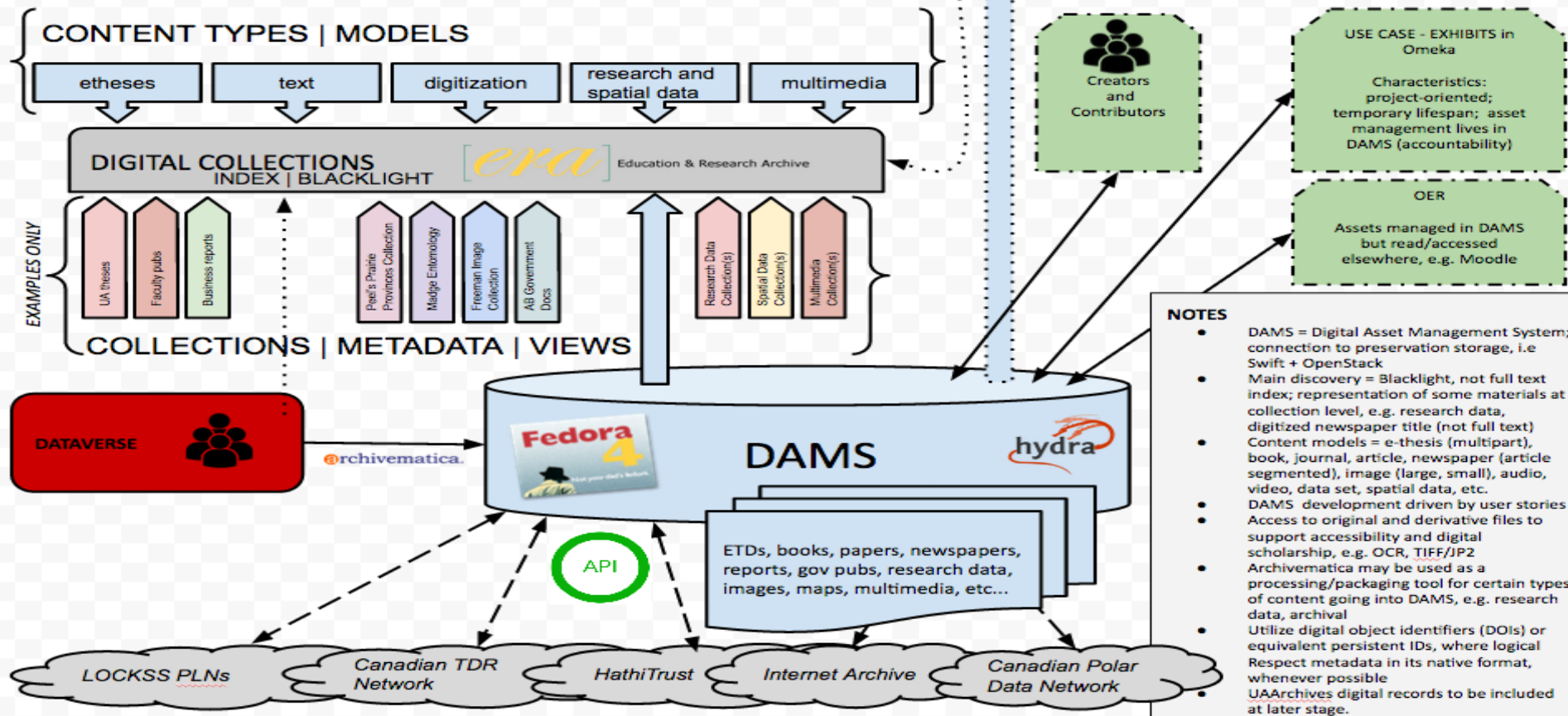


Dataverse at the University of Alberta

Larry Laliberte and Chuck Humphrey
University of Alberta Libraries
2015

The University of Alberta Context

- Research data management services is part of Digital Initiatives, which also includes the institutional repository, digitization production & collections, digital preservation, and scholarly communications.
- Digital Initiatives has ten librarians, one post-doc, three programmers, three student assistants.
- Support from two metadata librarians, one discovery librarian, and two system administrators.



Dataverse & Researcher Needs

- Back-up storage for research data.
- An easy way to share data with a research team that does not involve the use of email or Dropbox.
- A way of tracking which dataset was used to produce a table in a publication or of tracking other version issues.
- A way to satisfy a publisher's requirements, e.g. a DOI for a dataset associated with an article or providing access to a dataset for reviewers.

Dataverse & Curator Needs

- A simplified approach to capture project-level metadata that can be used repeatedly with other applications.
- A one-step method of minting PIDs for data files.
- A service that helps organize data for preservation processing.

Use Cases

- Humanities uptake
 - Linguistic collections
 - Audio recordings of aboriginal elders
 - Audio recordings of aboriginal accents
 - Translation of Spanish texts in the public domain
 - Image collection of texts and current translations
 - Gamegate
 - Social media web-scrapings on gamegate discussions

Use Cases

- Physical Education & Recreation
 - PLOS publication requiring the researcher to share her data for anonymous review
- Canadian Century Research Infrastructure
 - Dissemination of spatial data for the 1911, 1921, 1931, 1941, and 1951 Censuses
- Biology (in need of a quick DOI)
 - Grizzly bear habitat

Next Steps

- Install the new version (4.0) of Dataverse.
- Tailor Dataverse workshops for graduate students.
 - Promote the sharing of thesis data through Dataverse.
 - Provide experience using dataset DOIs with dissertations.
 - Promote the rewards of data sharing.

The Canadian Context

- Research Data Canada (RDC) and the Canadian Association of Research Libraries (CARL) federated pilot program
 - Emerged from the January 28-29, 2014 Digital Infrastructure Summit in Ottawa
 - Prove that existing infrastructure can be coordinated to support (i) research council data management plan policies and (ii) a national research data preservation service

Approved repositories will:

- be able to transmit data to the Interoperability Layer
- provide a crosswalk from any internally used DMD formats to a standard "lowest common denominator" DMD standard, e.g., DataCite, DC
- be able to link deposited data to grant #

Funded researcher
 Researcher deposits data in any approved repository in which they are comfortable.

- TC3+ Approved Repositories
- Hydra Repos
 - Islandora VREs
 - Dataverse Networks
 - CADC
 - ATLAS
 - CBRAIN

Creates a Standard Data Interoperability Package

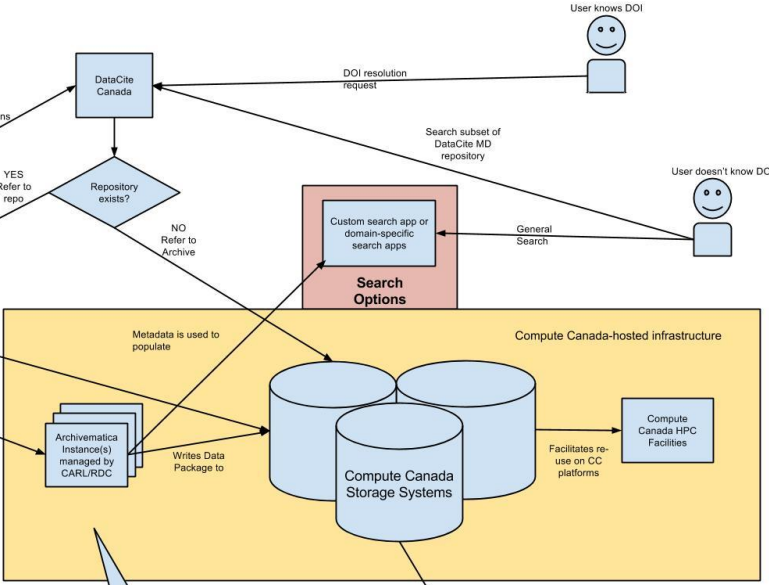
Deposit is made through existing CC endpoints such as Globus and FTP access.
 Deposit should include:

- DOI (if assigned already)
- CASRAI ID / Grant ID
- Domain-specific descriptive metadata that is crosswalkable per the repo's provided crosswalk

Standard Data Interoperability package includes:

- Data as provided
- Data as transformed for reuse/preservation purposes, if necessary
- Domain specific DMD in native format
- DMD in transformed, "lowest common denominator" format (DataCite, DC?)


Resources for transformation include Archivematica Format Policy Registry (FPR), possible other sources.



Approved repositories will:

- be able to transmit data to the Interoperability Layer
- provide a crosswalk from any internally used DMD formats to a standard "lowest common denominator" DMD standard, e.g., DataCite, DC
- be able to link deposited data to grant #

PROJECT OR RESEARCH PROGRAM LEVEL



Funded researcher

Researcher deposits data in any approved repository in which they are comfortable.

- TIC3+ Approved Repositories
- Hydra Repos
 - Islandora VREs
 - Dataverse Networks
 - CADC
 - ATLAS
 - CBRAIN

Creates a Standard Data Interoperability Package

Deposit is made through existing CC endpoints such as Globus and FTP access.

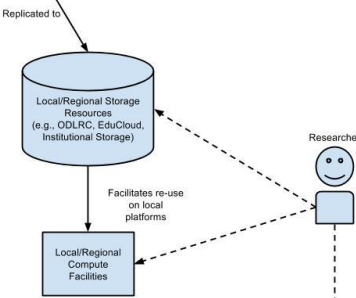
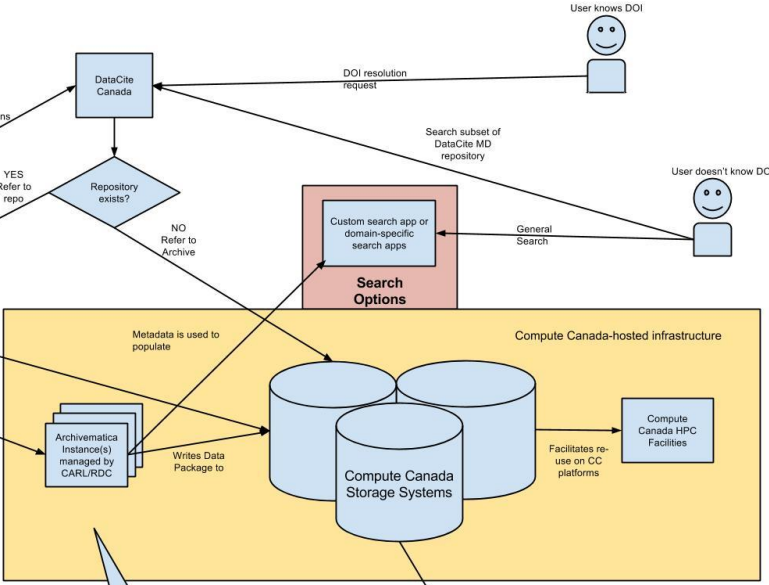
Deposit should include:

- DOI (if assigned already)
- CASRAI ID / Grant ID
- Domain-specific descriptive metadata that is crosswalkable per the repo's provided crosswalk

Standard Data Interoperability package includes:

- Data as provided
- Data as transformed for reuse/preservation purposes, if necessary
- Domain specific DMD in native format
- DMD in transformed, "lowest common denominator" format (DataCite, DC?)


Resources for transformation include Archivematica Format Policy Registry (FPR), possible other sources.



Approved repositories will:

- be able to transmit data to the Interoperability Layer
- provide a crosswalk from any internally used DMD formats to a standard "lowest common denominator" DMD standard, e.g., DataCite, DC
- be able to link deposited data to grant #

PROJECT OR RESEARCH PROGRAM LEVEL



Funded researcher

Researcher deposits data in any approved repository in which they are comfortable.

EXCHANGE

DATA REPOSITORY LEVEL

TC3+ Approved Repositories

- Hydra Repos
- Islandora VREs
- Dataverse Networks
- CADC
- ATLAS
- CBRAIN

Standard Data Interoperability package includes:

- Data as provided
- Data as transformed for reuse/preservation purposes, if necessary
- Domain specific DMD in native format
- DMD in transformed, "lowest common denominator" format (DataCite, DC?)

Resources for transformation include Archivematica Format Policy Registry (FPR), possible other sources.

Deposit is made through existing CC endpoints such as Globus and FTP access.

Deposit should include:

- DOI (if assigned already)
- CASRAI ID / Grant ID
- Domain-specific descriptive metadata that is crosswalkable per the repo's provided crosswalk

Creates a Standard Data Interoperability Package

DataCite Canada


DOI resolution request



Search subset of DataCite MD repository

General Search

User doesn't know DOI



Canada systems

Compute Canada-hosted infrastructure

Facilitates re-use on CC platforms


Compute Canada HPC Facilities

Replicated to

Local/Regional Storage Resources (e.g., ODLRC, EduCloud, Institutional Storage)

Facilitates re-use on local platforms

Local/Regional Compute Facilities




Accesses/processes data for replication/reuse. If reusing, deposits data back into approved repository.

DOI Transactions

- Approved repositories will:
- be able to transmit data to the Interoperability Layer
 - provide a crosswalk from any internally used DMD formats to a standard "lowest common denominator" DMD standard, e.g., DataCite, DC
 - be able to link deposited data to grant #

PROJECT OR RESEARCH PROGRAM LEVEL



Funded researcher

Researcher deposits data in any approved repository in which they are comfortable.

EXCHANGE

DATA REPOSITORY LEVEL

TC3+ Approved Repositories

- Hydra Repos
- Islandora VREs
- Dataverse Networks
- CADC
- ATLAS
- CBRAIN

Dataverse Networks

Search subset of DataCite MD repository

User doesn't know DOI

General Search

Compute Canada-hosted infrastructure

Canada systems

Facilitates re-use on CC platforms

Compute Canada HPC Facilities

Creates a Standard Data Interoperability Package

- Deposit is made through existing CC endpoints such as Globus and FTP access.
- Deposit should include:
- DOI (if assigned already)
 - CASRAI ID / Grant ID
 - Domain-specific descriptive metadata that is crosswalkable per the repo's provided crosswalk

- Standard Data Interoperability package includes:
- Data as provided
 - Data as transformed for reuse/preservation purposes, if necessary
 - Domain specific DMD in native format
 - DMD in transformed, "lowest common denominator" format (DataCite, DC?)
- Resources for transformation include Archivematica Format Policy Registry (FPR), possible other sources.

Replicated to

Local/Regional Storage Resources (e.g., ODLRC, EduCloud, Institutional Storage)

Facilitates re-use on local platforms

Local/Regional Compute Facilities

Researcher

Accesses/processes data for replication/reuse. If reusing, deposits data back into approved repository.

DOI Transactions

DataCite Canada

DOI resolution request

User knows DOI

Search subset of DataCite MD repository

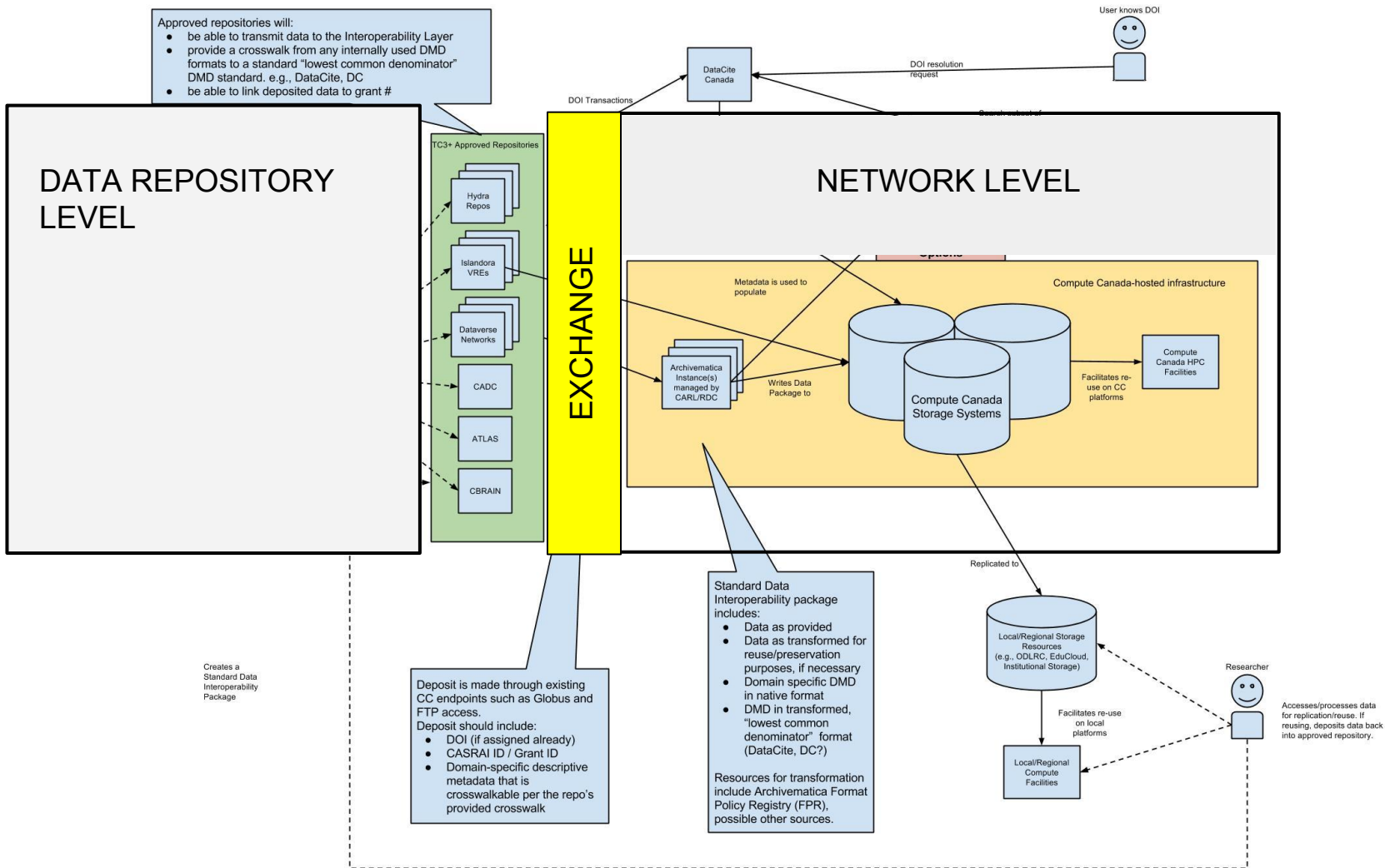
User doesn't know DOI

Compute Canada HPC Facilities

Local/Regional Compute Facilities

Scholars Portal Project

- Scholars Portal of the Ontario Council of University Libraries is working with IQSS and Artefactual to move studies from Dataverse to Archivematica for the production of Archival Information Packages (AIPs)



Next Steps

- Complete the RDC-CARL Federated Pilot with the Dataverse-Archivematica software stack.
- Find ways for Canadian Dataverse Networks to collaborate.