CARL ABRC

CANADIAN ASSOCIATION OF RESEARCH LIBRARIES
ASSOCIATION DES BIBLIOTHÈQUES DE RECHERCHE DU CANADA

compute canada

BUILDING A SCALEABLE NATIONAL REPOSITORY INFRASTRUCTURE FOR CANADA

Todd Trann, University of Saskatchewan
Alex Garnett, Simon Fraser University

CANHEIT June, 2016

What Is A Data Repository?

Main functions: **digital storage, retrieval and preservation** of research data

Institutional Repositories are run by colleges and universities

- cIRcle (Run by UBC)

Regional Repositories are run by groups of universities

- Scholars Portal (Run by OCUL)

Domain Specific Repositories are run by organizations or communities of researchers

- BioGRID (Biology; run by the BioGRID team)

There are thousands of research repositories in the world (see: re3data.org)

YARR

- Current repositories don’t scale to very large data sets
- Hard to discover data sets in existing repositories
- Some researchers don’t have access to a suitable repository

Project Overview

Designing infrastructure for a national research data repository

- Partnerships between Compute Canada, Globus and CARL (Canadian Association of Research Libraries)
- Using Compute Canada hardware

Designing discovery platform

- Indexing metadata of existing repositories
- Sharing back a search API
- Improving cross-repository discovery
- Facilitating domain specific search development

CARL ABRC
CANADIAN ASSOCIATION OF RESEARCH LIBRARIES
ASSOCIATION DES BIBLIOTHÈQUES DE RECHERCHE DU CANADA

compute canada | calcul canada

Planned Features

- **Federated Storage Model:** Storage and repositories can be distributed, owned and operated by organizations / institutions

- **Scalable:** Many files; large files; large total amounts of storage

- **National Data Discovery:** Single search to discover data, regardless of location

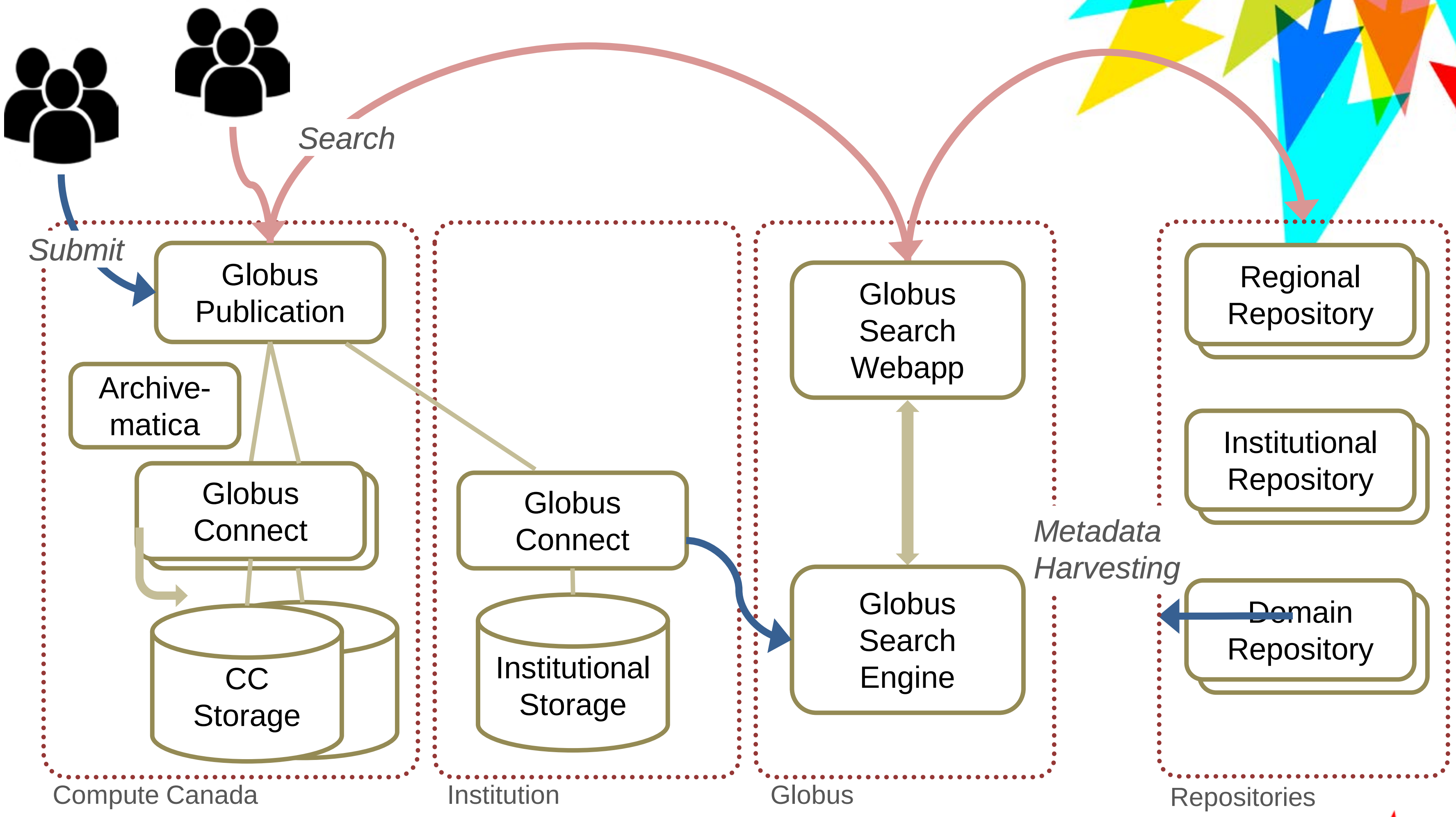
Planned Features (cont'd)

- **Data Preservation:** Use Archivematica to automatically preserve data

- **Suitable for broad range of data types**

- **Automatic geographic data replication:** We intend to leverage Compute Canada national data infrastructure being deployed in 2016

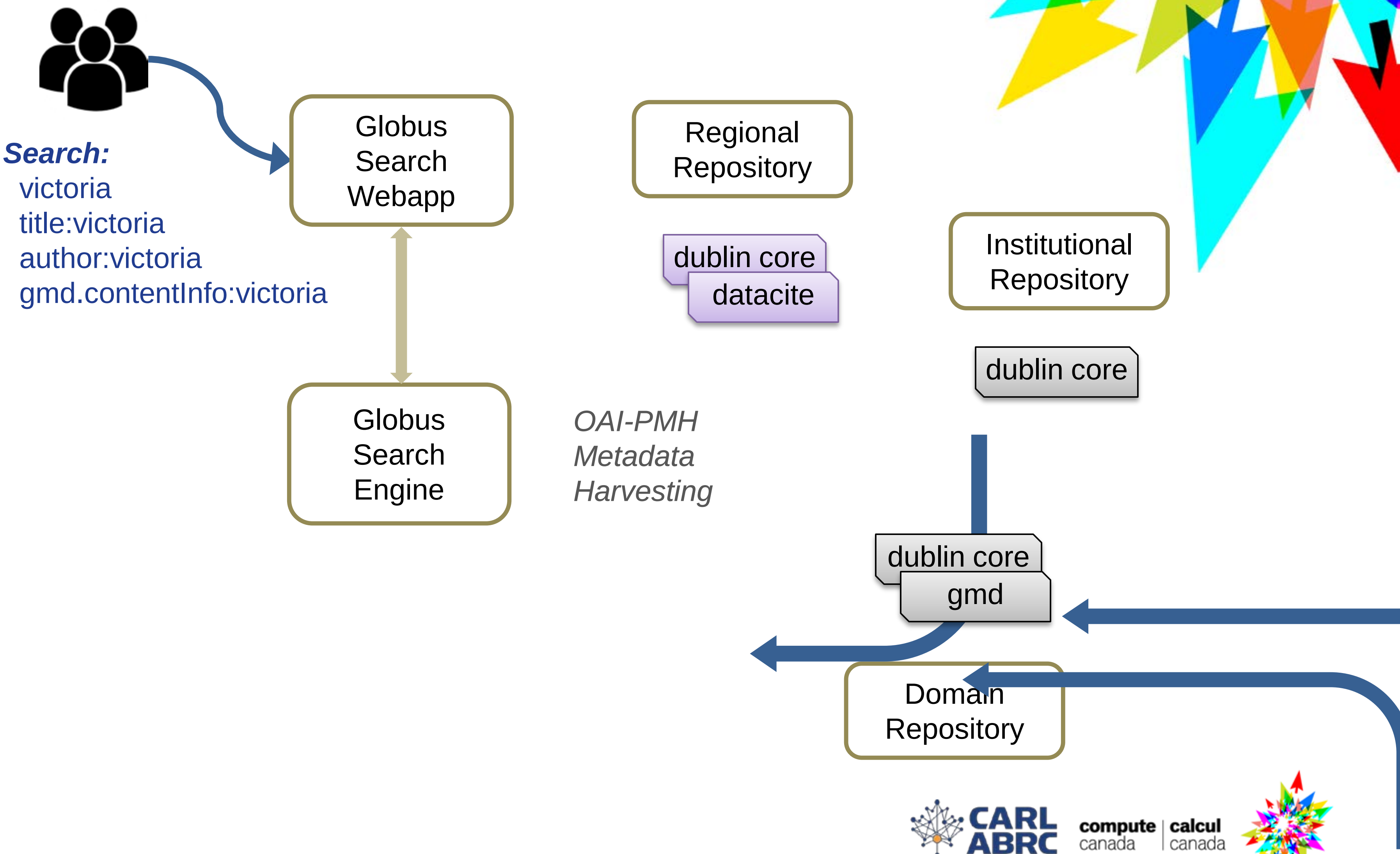
System Diagram



CARL ABRC
CANADIAN ASSOCIATION OF RESEARCH LIBRARIES
ASSOCIATION DES BIBLIOTHÈQUES DE RECHERCHE DU CANADA

compute canada | calcul canada

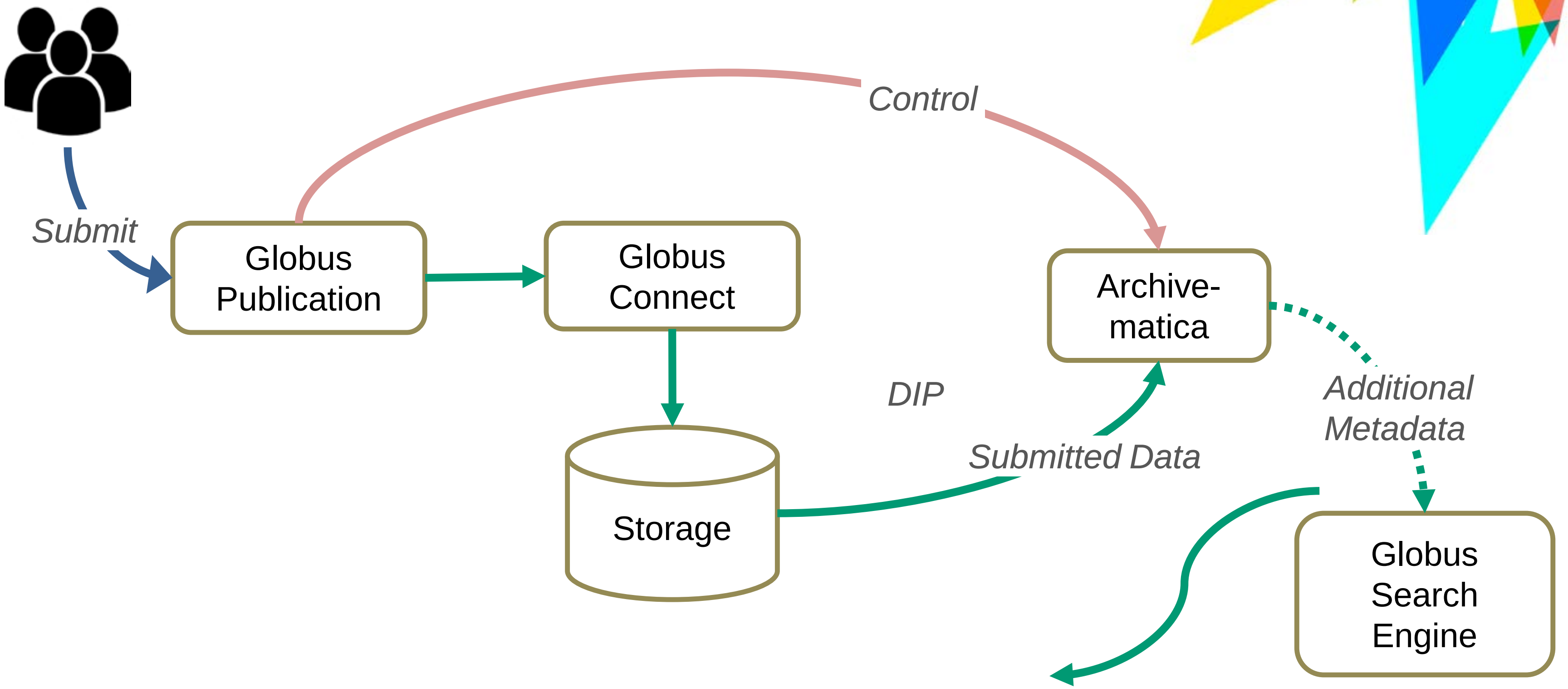
Metadata Overlay



Search:

victoria
title:victoria
author:victoria
gmd.contentInfo:victoria

Preservation Automation



Submit

Globus Publication

Control

Globus Connect

Archive-matica

Storage

DIP

Submitted Data

Additional Metadata

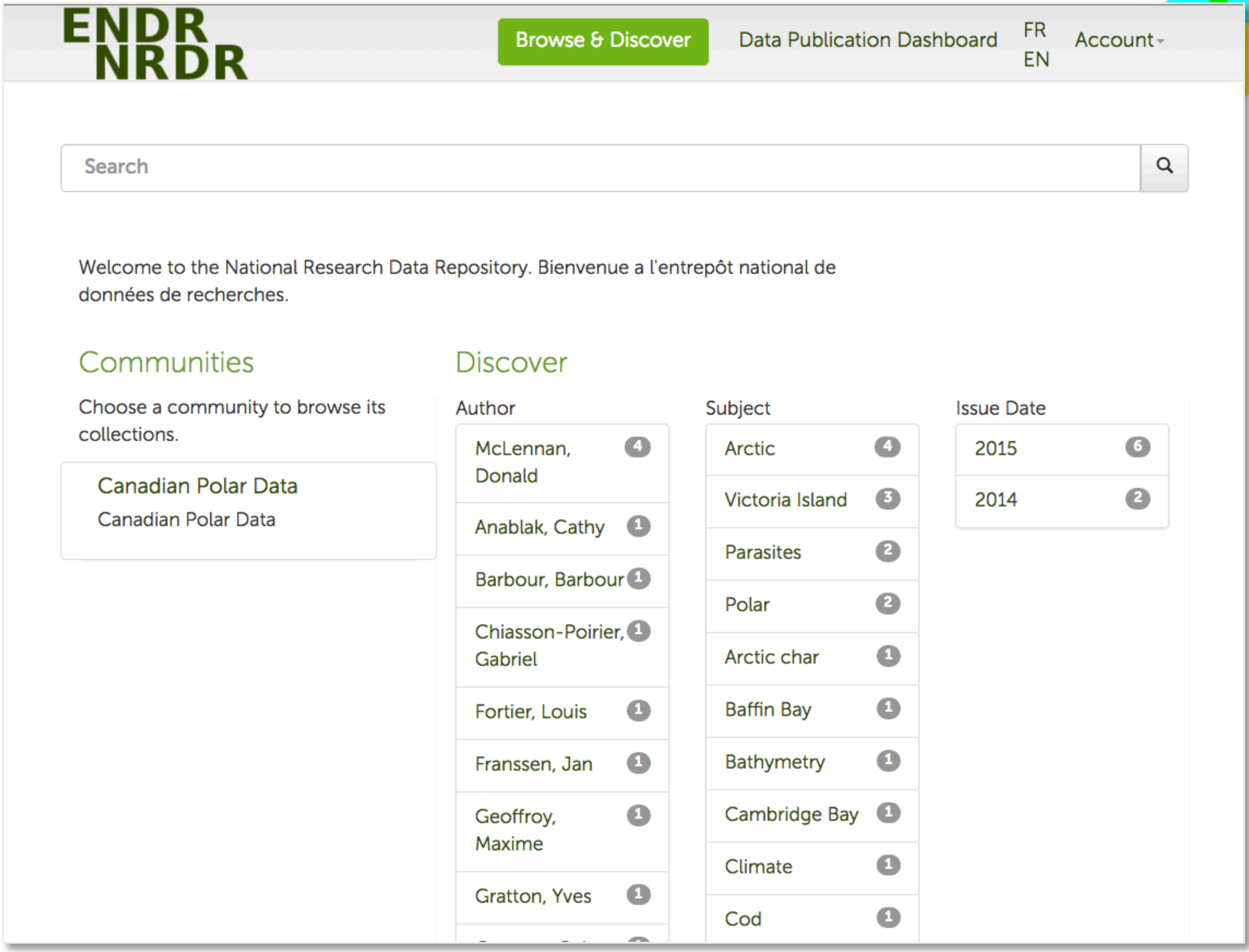
Globus Search Engine

Discovery Landscape

- A University of British Columbia/CARL-Portage group (led by Eugene Barsky) is doing an environmental scan of data discovery systems, and considering options for the design of a coordinated, national discovery service for research data in Canada

- The white paper from this scan will inform finalization of the discovery interface

Landing Page



Welcome to the National Research Data Repository. Bienvenue a l'entrepôt national de données de recherches.

Communities

Choose a community to browse its collections.

Canadian Polar Data

Canadian Polar Data

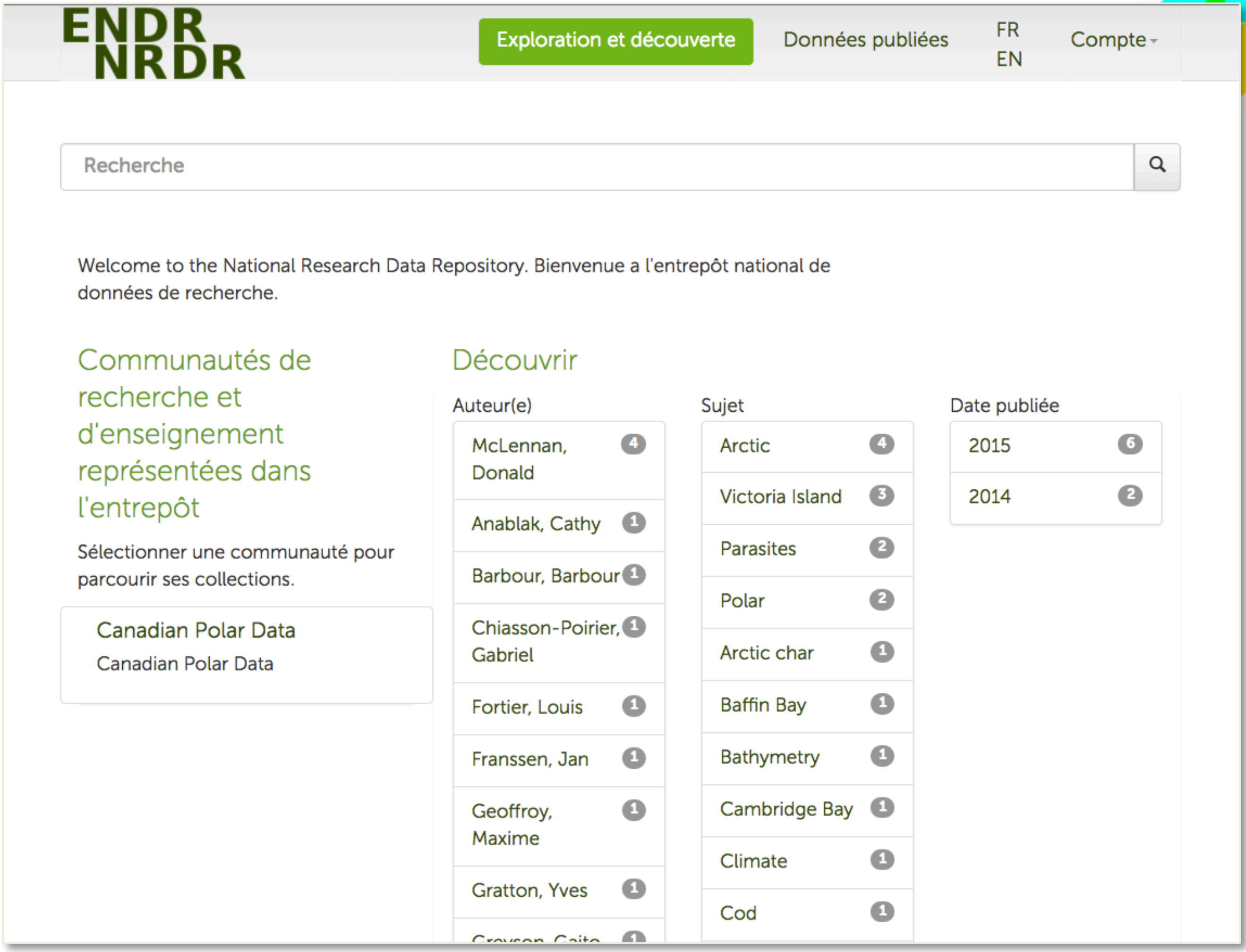
Discover

Author	
McLennan, Donald	4
Anablak, Cathy	1
Barbour, Barbour	1
Chiasson-Poirier, Gabriel	1
Fortier, Louis	1
Franssen, Jan	1
Geoffroy, Maxime	1
Gratton, Yves	1

Subject	
Arctic	4
Victoria Island	3
Parasites	2
Polar	2
Arctic char	1
Baffin Bay	1
Bathymetry	1
Cambridge Bay	1
Climate	1
Cod	1

Issue Date	
2015	6
2014	2

Bilingual Interface



Welcome to the National Research Data Repository. Bienvenue a l'entrepôt national de données de recherche.

Communautés de recherche et d'enseignement représentées dans l'entrepôt

Sélectionner une communauté pour parcourir ses collections.

Canadian Polar Data

Canadian Polar Data

Découvrir

Auteur(e)	
McLennan, Donald	4
Anablak, Cathy	1
Barbour, Barbour	1
Chiasson-Poirier, Gabriel	1
Fortier, Louis	1
Franssen, Jan	1
Geoffroy, Maxime	1
Gratton, Yves	1
Greyson, Gaite	1

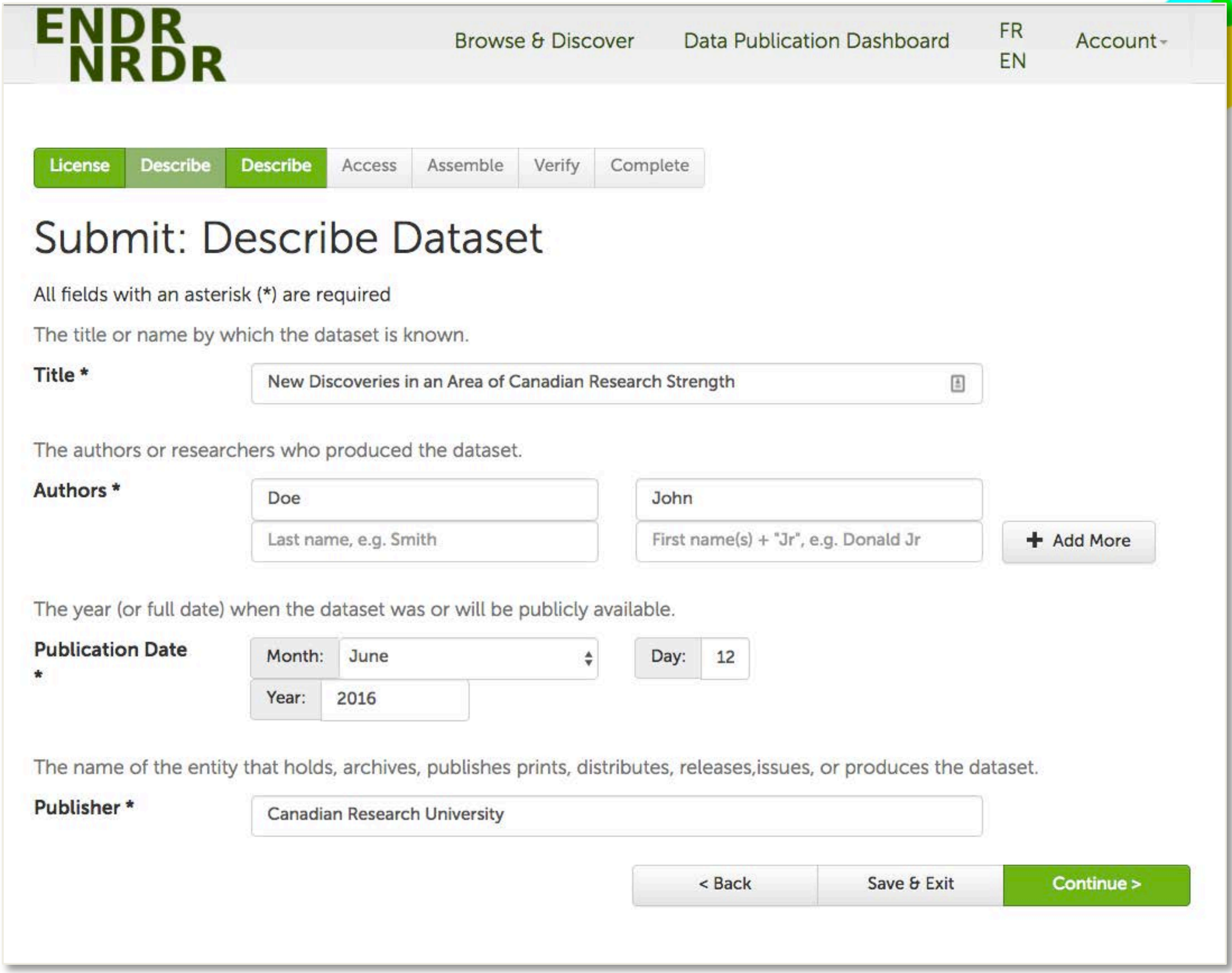
Sujet	
Arctic	4
Victoria Island	3
Parasites	2
Polar	2
Arctic char	1
Baffin Bay	1
Bathymetry	1
Cambridge Bay	1
Climate	1
Cod	1

Date publiée	
2015	6
2014	2

CARL ABRC
CANADIAN ASSOCIATION OF RESEARCH LIBRARIES
ASSOCIATION DES BIBLIOTHÈQUES DE RECHERCHE DU CANADA

compute canada | calcul canada

Submitting An Item



Submit: Describe Dataset

All fields with an asterisk (*) are required

The title or name by which the dataset is known.

Title * New Discoveries in an Area of Canadian Research Strength

The authors or researchers who produced the dataset.

Authors * Doe | John

Last name, e.g. Smith | First name(s) + "Jr", e.g. Donald Jr

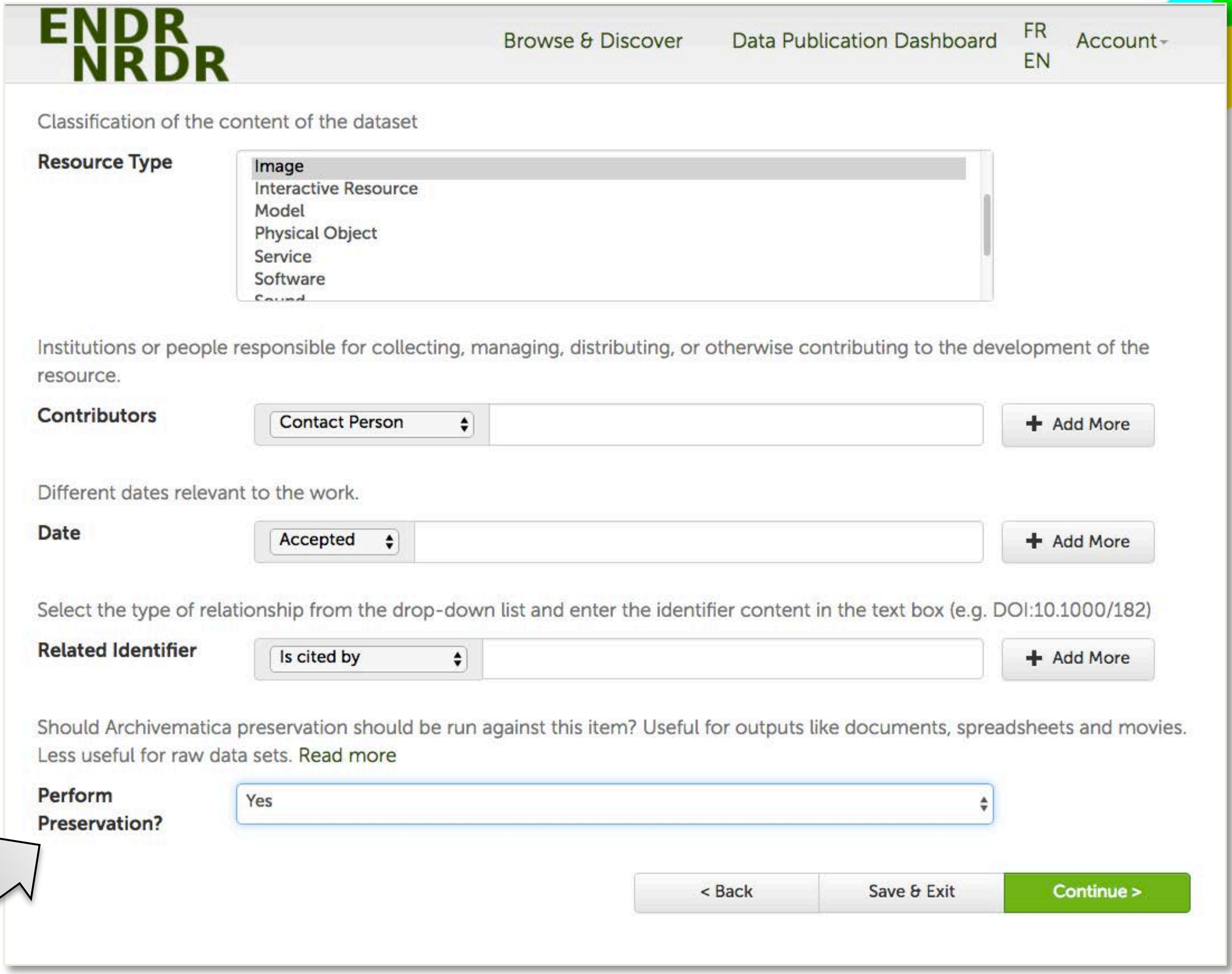
The year (or full date) when the dataset was or will be publicly available.

Publication Date * Month: June Day: 12 Year: 2016

The name of the entity that holds, archives, publishes prints, distributes, releases,issues, or produces the dataset.

Publisher * Canadian Research University

Requesting Preservation



Classification of the content of the dataset

Resource Type

- Image
- Interactive Resource
- Model
- Physical Object
- Service
- Software
- Sound

Institutions or people responsible for collecting, managing, distributing, or otherwise contributing to the development of the resource.

Contributors Contact Person

Different dates relevant to the work.

Date Accepted

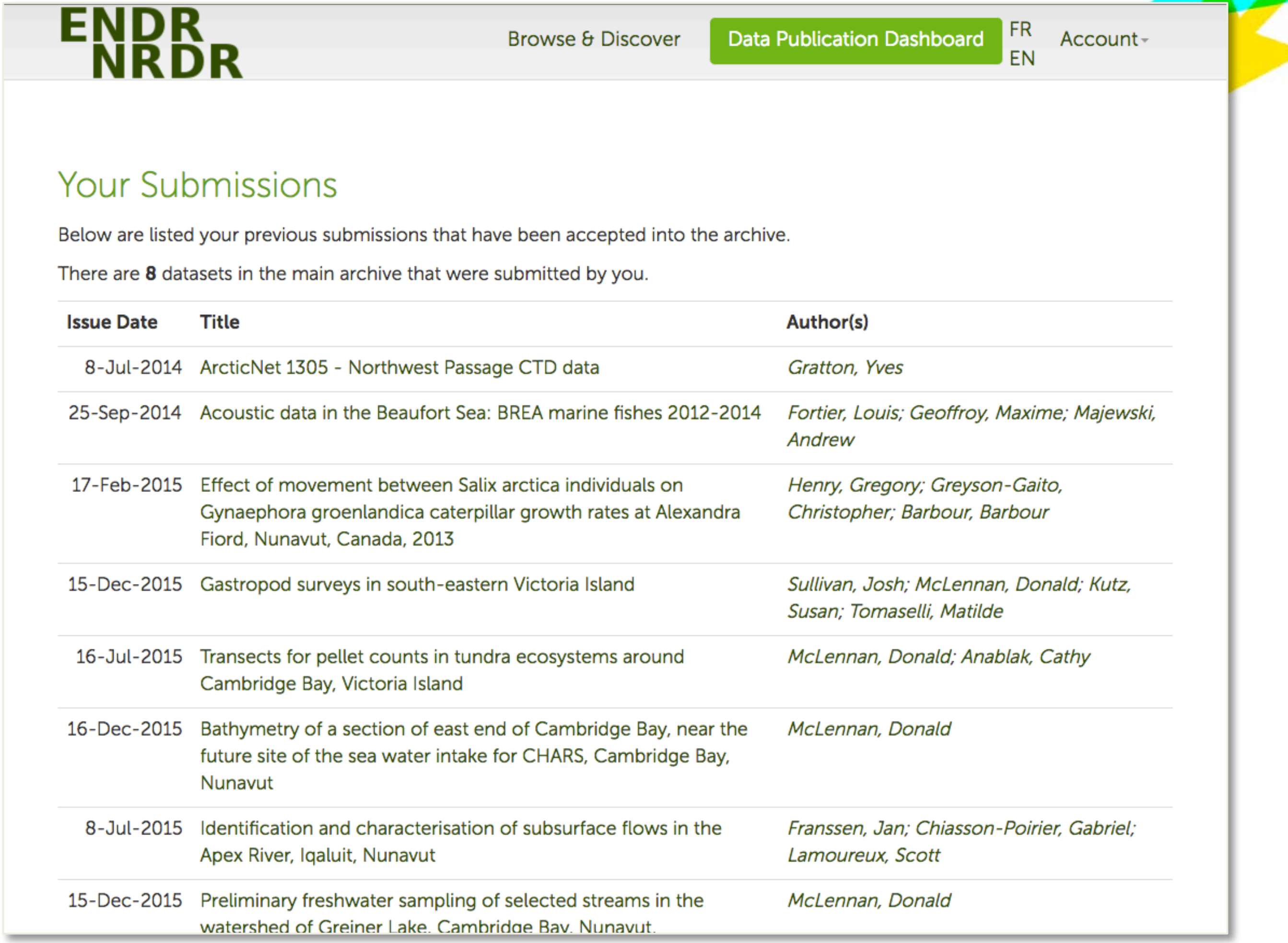
Select the type of relationship from the drop-down list and enter the identifier content in the text box (e.g. DOI:10.1000/182)

Related Identifier Is cited by

Should Archivematica preservation should be run against this item? Useful for outputs like documents, spreadsheets and movies. Less useful for raw data sets. Read more

Perform Preservation? Yes

Submission List



Your Submissions

Below are listed your previous submissions that have been accepted into the archive.

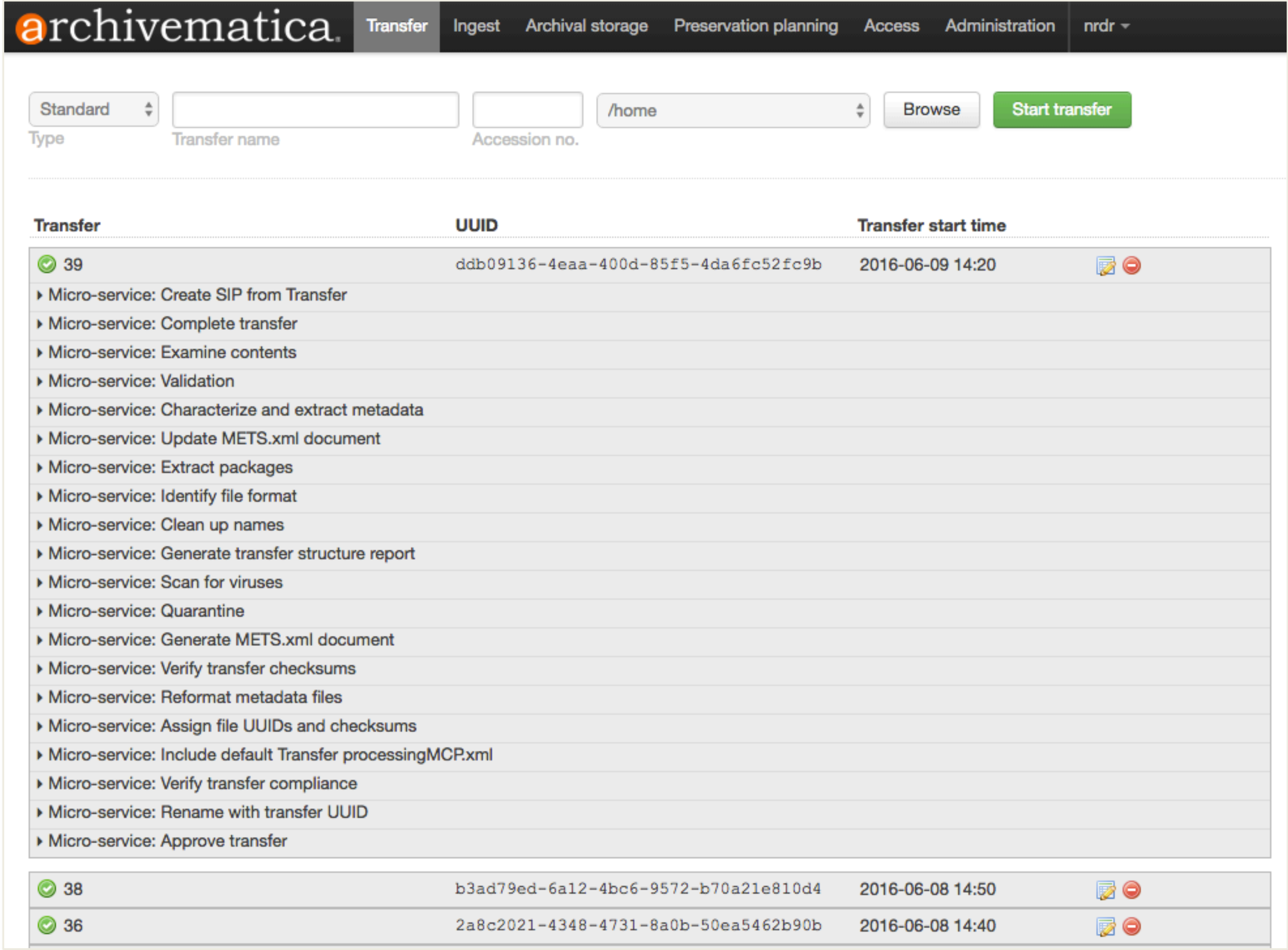
There are **8** datasets in the main archive that were submitted by you.

Issue Date	Title	Author(s)
8-Jul-2014	ArcticNet 1305 - Northwest Passage CTD data	*Gratton, Yves*
25-Sep-2014	Acoustic data in the Beaufort Sea: BREA marine fishes 2012-2014	*Fortier, Louis; Geoffroy, Maxime; Majewski, Andrew*
17-Feb-2015	Effect of movement between Salix arctica individuals on Gynaephora groenlandica caterpillar growth rates at Alexandra Fiord, Nunavut, Canada, 2013	*Henry, Gregory; Greyson-Gaito, Christopher; Barbour, Barbour*
15-Dec-2015	Gastropod surveys in south-eastern Victoria Island	*Sullivan, Josh; McLennan, Donald; Kutz, Susan; Tomaselli, Matilde*
16-Jul-2015	Transects for pellet counts in tundra ecosystems around Cambridge Bay, Victoria Island	*McLennan, Donald; Anablak, Cathy*
16-Dec-2015	Bathymetry of a section of east end of Cambridge Bay, near the future site of the sea water intake for CHARS, Cambridge Bay, Nunavut	*McLennan, Donald*
8-Jul-2015	Identification and characterisation of subsurface flows in the Apex River, Iqaluit, Nunavut	*Franssen, Jan; Chiasson-Poirier, Gabriel; Lamoureux, Scott*
15-Dec-2015	Preliminary freshwater sampling of selected streams in the watershed of Greiner Lake, Cambridge Bay, Nunavut.	*McLennan, Donald*

CARL ABRC
CANADIAN ASSOCIATION OF RESEARCH LIBRARIES
ASSOCIATION DES BIBLIOTHÈQUES DE RECHERCHE DU CANADA

compute canada | calcul canada

Preservation Processing



Type: Standard

Transfer name

Accession no.

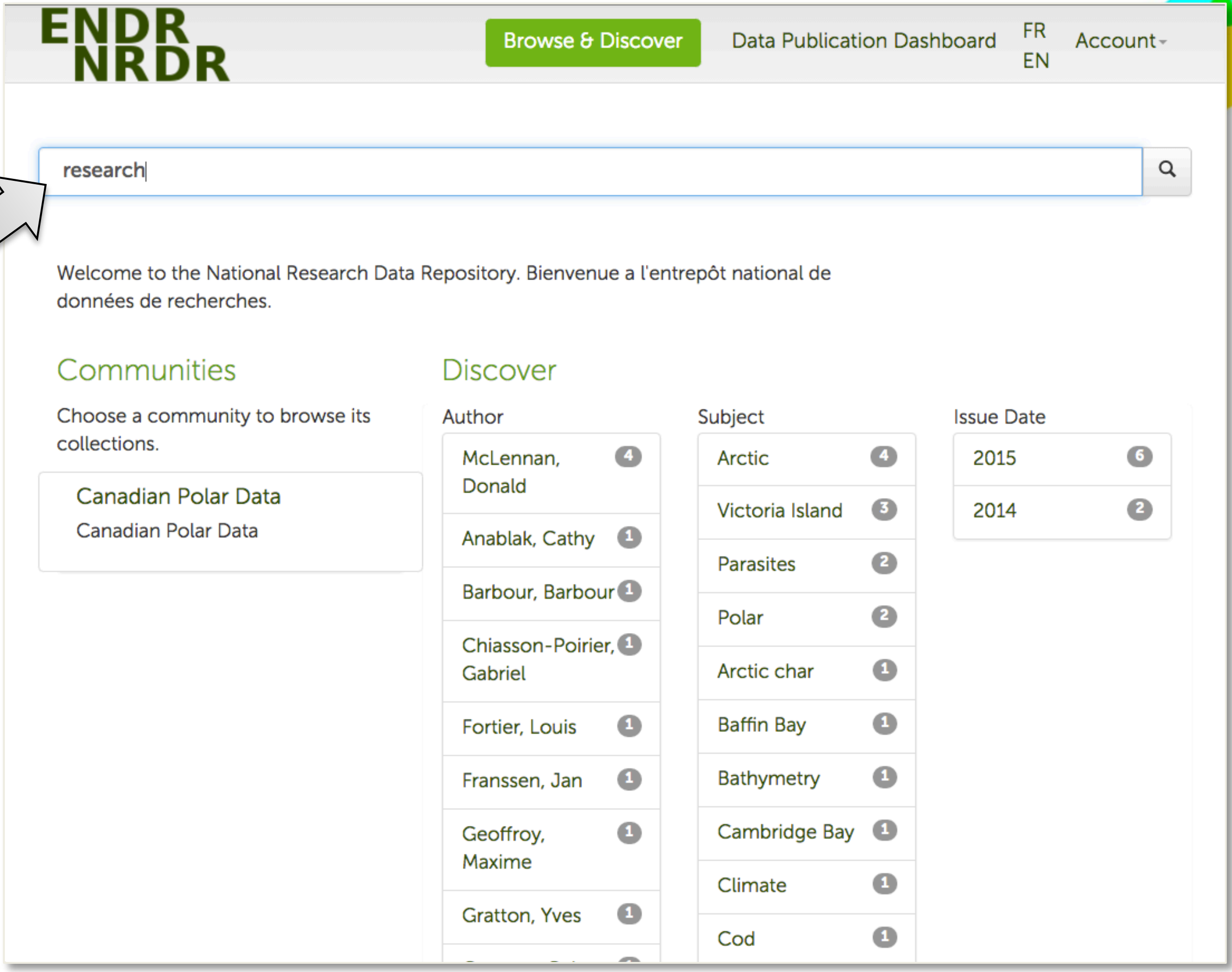
/home

Transfer	UUID	Transfer start time
✓ 39	ddb09136-4eaa-400d-85f5-4da6fc52fc9b	2016-06-09 14:20
▸ Micro-service: Create SIP from Transfer		
▸ Micro-service: Complete transfer		
▸ Micro-service: Examine contents		
▸ Micro-service: Validation		
▸ Micro-service: Characterize and extract metadata		
▸ Micro-service: Update METS.xml document		
▸ Micro-service: Extract packages		
▸ Micro-service: Identify file format		
▸ Micro-service: Clean up names		
▸ Micro-service: Generate transfer structure report		
▸ Micro-service: Scan for viruses		
▸ Micro-service: Quarantine		
▸ Micro-service: Generate METS.xml document		
▸ Micro-service: Verify transfer checksums		
▸ Micro-service: Reformat metadata files		
▸ Micro-service: Assign file UUIDs and checksums		
▸ Micro-service: Include default Transfer processingMCP.xml		
▸ Micro-service: Verify transfer compliance		
▸ Micro-service: Rename with transfer UUID		
▸ Micro-service: Approve transfer		
✓ 38	b3ad79ed-6a12-4bc6-9572-b70a21e810d4	2016-06-08 14:50
✓ 36	2a8c2021-4348-4731-8a0b-50ea5462b90b	2016-06-08 14:40

CARL ABRC
CANADIAN ASSOCIATION OF RESEARCH LIBRARIES
ASSOCIATION DES BIBLIOTHÈQUES DE RECHERCHE DU CANADA

compute canada | calcul canada

Starting A Search



Welcome to the National Research Data Repository. Bienvenue a l'entrepôt national de données de recherches.

Communities

Choose a community to browse its collections.

Canadian Polar Data
Canadian Polar Data

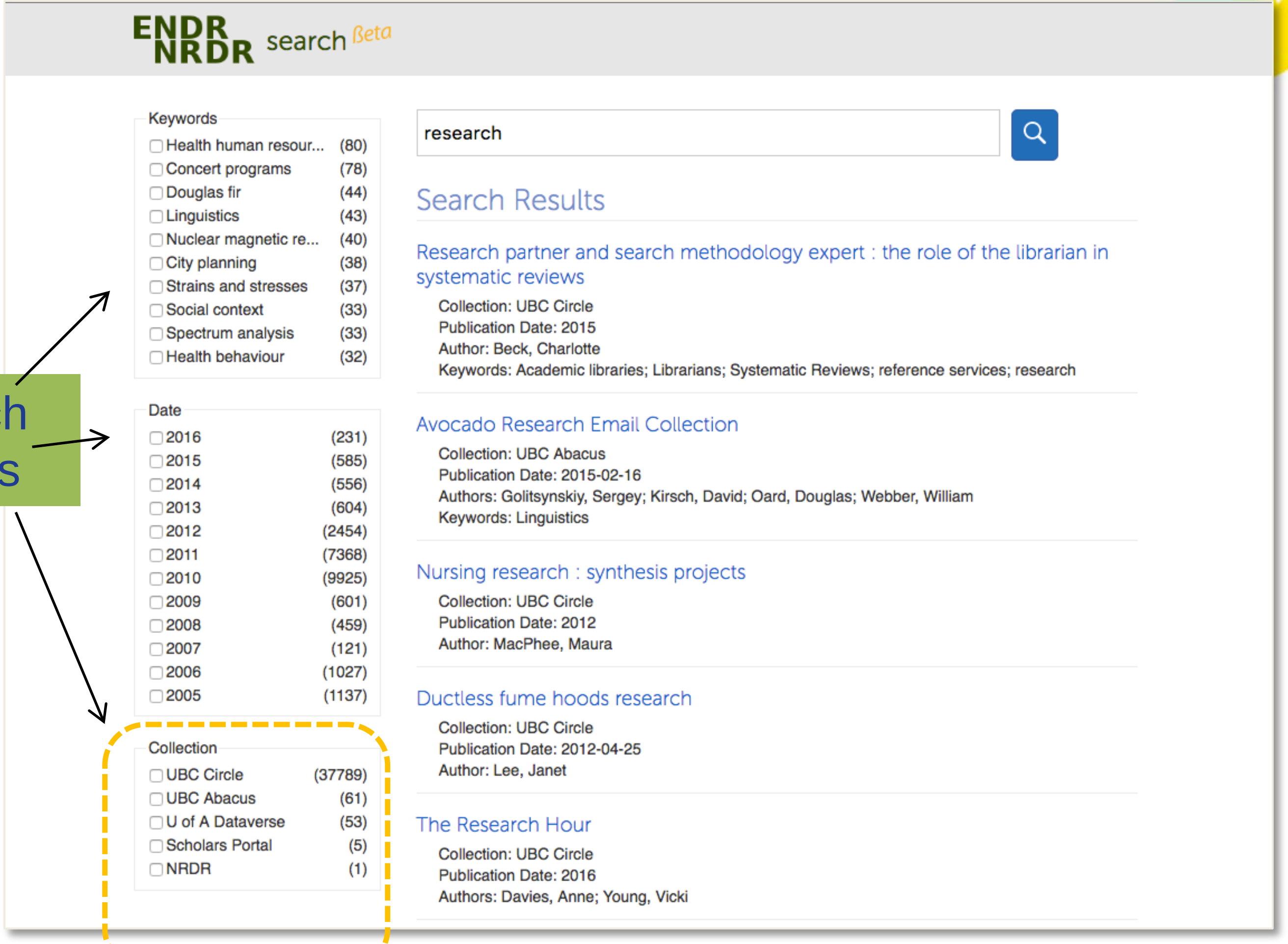
Discover

Author		Subject		Issue Date	
McLennan, Donald	4	Arctic	4	2015	6
Anablak, Cathy	1	Victoria Island	3	2014	2
Barbour, Barbour	1	Parasites	2		
Chiasson-Poirier, Gabriel	1	Polar	2		
Fortier, Louis	1	Arctic char	1		
Franssen, Jan	1	Baffin Bay	1		
Geoffroy, Maxime	1	Bathymetry	1		
Gratton, Yves	1	Cambridge Bay	1		
		Climate	1		
		Cod	1		

CARL ABRC
CANADIAN ASSOCIATION OF RESEARCH LIBRARIES
ASSOCIATION DES BIBLIOTHÈQUES DE RECHERCHE DU CANADA

compute canada | calcul canada

Search Results



ENDR NRDR search Beta

Keywords	
□ Health human resour...	(80)
□ Concert programs	(78)
□ Douglas fir	(44)
□ Linguistics	(43)
□ Nuclear magnetic re...	(40)
□ City planning	(38)
□ Strains and stresses	(37)
□ Social context	(33)
□ Spectrum analysis	(33)
□ Health behaviour	(32)

Date	
□ 2016	(231)
□ 2015	(585)
□ 2014	(556)
□ 2013	(604)
□ 2012	(2454)
□ 2011	(7368)
□ 2010	(9925)
□ 2009	(601)
□ 2008	(459)
□ 2007	(121)
□ 2006	(1027)
□ 2005	(1137)

Collection	
□ UBC Circle	(37789)
□ UBC Abacus	(61)
□ U of A Dataverse	(53)
□ Scholars Portal	(5)
□ NRDR	(1)

research

Search Results

Research partner and search methodology expert : the role of the librarian in systematic reviews

Collection: UBC Circle
Publication Date: 2015
Author: Beck, Charlotte
Keywords: Academic libraries; Librarians; Systematic Reviews; reference services; research

Avocado Research Email Collection

Collection: UBC Abacus
Publication Date: 2015-02-16
Authors: Golitsynskiy, Sergey; Kirsch, David; Oard, Douglas; Webber, William
Keywords: Linguistics

Nursing research : synthesis projects

Collection: UBC Circle
Publication Date: 2012
Author: MacPhee, Maura

Ductless fume hoods research

Collection: UBC Circle
Publication Date: 2012-04-25
Author: Lee, Janet

The Research Hour

Collection: UBC Circle
Publication Date: 2016
Authors: Davies, Anne; Young, Vicki

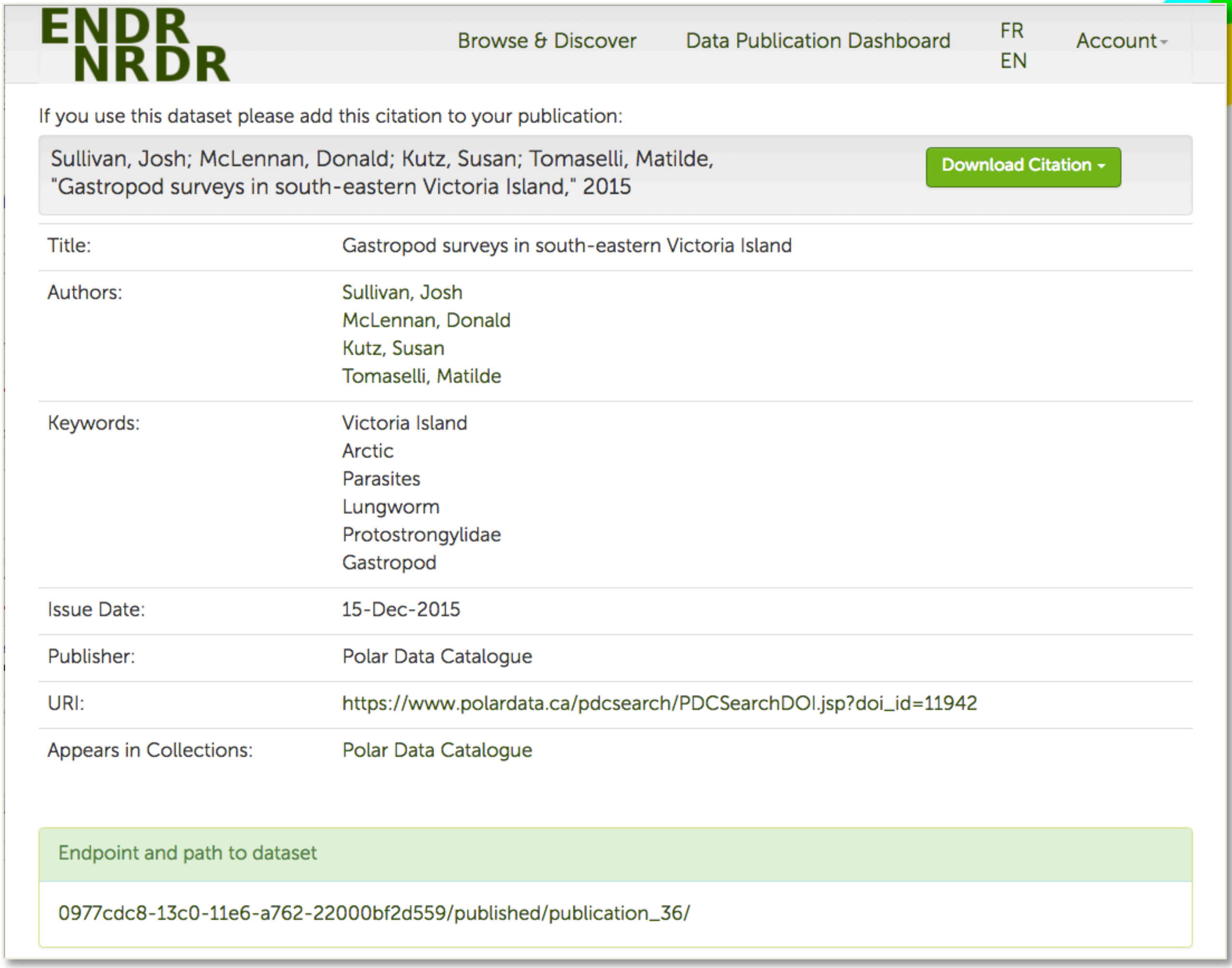
Search Facets

Source Repositories

CARL ABRC
CANADIAN ASSOCIATION OF RESEARCH LIBRARIES
ASSOCIATION DES BIBLIOTHÈQUES DE RECHERCHE DU CANADA

compute canada | calcul canada

Item Page



If you use this dataset please add this citation to your publication:

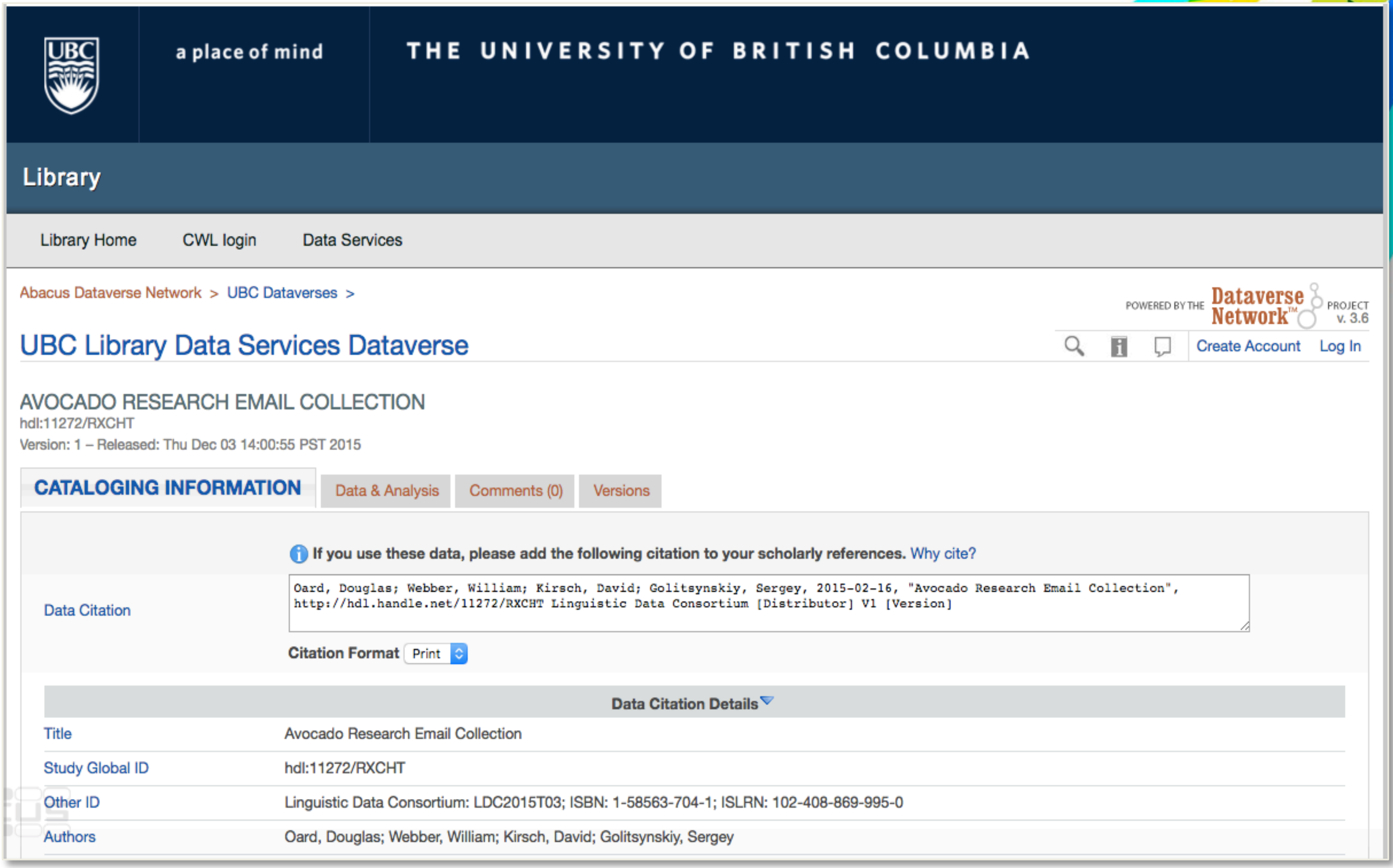
Sullivan, Josh; McLennan, Donald; Kutz, Susan; Tomaselli, Matilde, "Gastropod surveys in south-eastern Victoria Island," 2015

Title:	Gastropod surveys in south-eastern Victoria Island
Authors:	Sullivan, Josh McLennan, Donald Kutz, Susan Tomaselli, Matilde
Keywords:	Victoria Island Arctic Parasites Lungworm Protostrongylidae Gastropod
Issue Date:	15-Dec-2015
Publisher:	Polar Data Catalogue
URI:	https://www.polardata.ca/pdcsearch/PDCSearchDOI.jsp?doi_id=11942
Appears in Collections:	Polar Data Catalogue

Endpoint and path to dataset

0977cdc8-13c0-11e6-a762-22000bf2d559/published/publication_36/

Item Page



UBC
a place of mind
THE UNIVERSITY OF BRITISH COLUMBIA
Library
Library Home
CWL login
Data Services
POWERED BY THE Dataverse Network™ PROJECT v. 3.6
Create Account
Log In

Abacus Dataverse Network > UBC Dataverses >

UBC Library Data Services Dataverse

AVOCADO RESEARCH EMAIL COLLECTION

hdl:11272/RXCHT

Version: 1 – Released: Thu Dec 03 14:00:55 PST 2015

ⓘ If you use these data, please add the following citation to your scholarly references. Why cite?

Data Citation

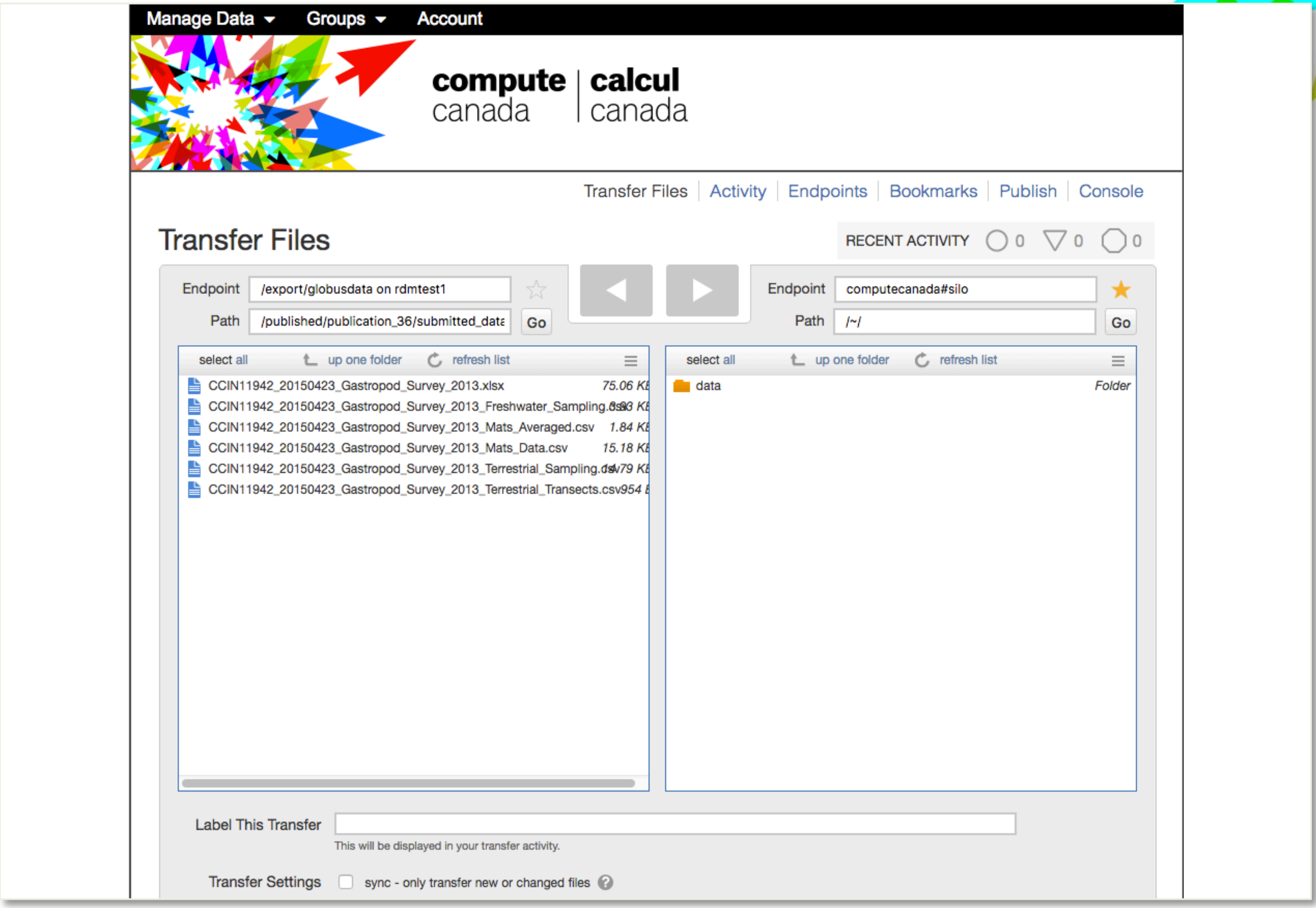
```
Oard, Douglas; Webber, William; Kirsch, David; Golitsynskiy, Sergey, 2015-02-16, "Avocado Research Email Collection",
http://hdl.handle.net/11272/RXCHT Linguistic Data Consortium [Distributor] V1 [Version]
```

Citation Format: Print

Data Citation Details

Title	Avocado Research Email Collection
Study Global ID	hdl:11272/RXCHT
Other ID	Linguistic Data Consortium: LDC2015T03; ISBN: 1-58563-704-1; ISLRN: 102-408-869-995-0
Authors	Oard, Douglas; Webber, William; Kirsch, David; Golitsynskiy, Sergey

Dataset File Transfer



Transfer Files

Endpoint /export/globusdata on rdmtest1 ☆

Path /published/publication_36/submitted_data

CCIN11942_20150423_Gastropod_Survey_2013.xlsx	75.06 KB
CCIN11942_20150423_Gastropod_Survey_2013_Freshwater_Sampling.csv	8.93 KB
CCIN11942_20150423_Gastropod_Survey_2013_Mats_Averaged.csv	1.84 KB
CCIN11942_20150423_Gastropod_Survey_2013_Mats_Data.csv	15.18 KB
CCIN11942_20150423_Gastropod_Survey_2013_Terrestrial_Sampling.csv	14.79 KB
CCIN11942_20150423_Gastropod_Survey_2013_Terrestrial_Transects.csv	954 B

Endpoint computecanada#silo ★

Path /~/

data	Folder

Label This Transfer

This will be displayed in your transfer activity.

Transfer Settings □ sync - only transfer new or changed files ?

Globus Transfer
(large file capability)

Further Information

Public web site:

www.computecanada.ca/rdm