Curating Data in Repositories

Lee Wilson, Reyna Jenkyns, Meghan Goodchild,
Brad Covey, Kaitlin Newson

Canadian Data Curation Forum
October 17, 2019

https://bit.ly/2oLYgAM

Learning Objectives

Participants will learn about:

- The FAIR principles and how they relate to data repositories
- Major FAIR enablers and associated resources
- The data repository landscape in Canada, including the different types of repositories and common platforms
- Data curation practices within repositories, including curation workflows and supporting tools
- Common questions or issues curators will encounter during the data ingest process and strategies for addressing them

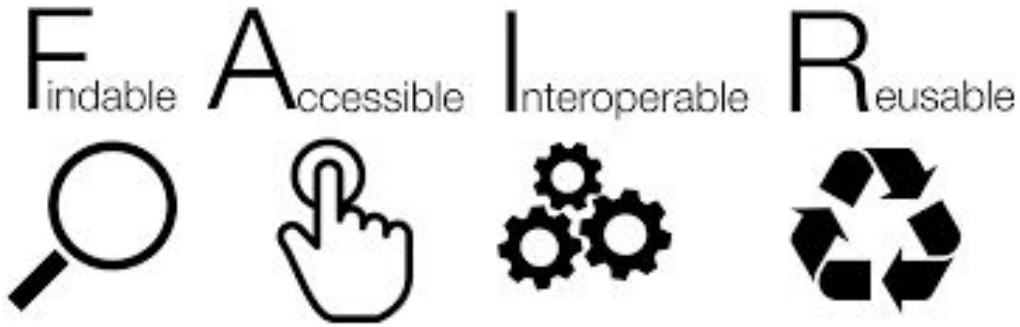
Outline

- Part 1: Making data FAIR (45 mins)
  - FAIR Principles
  - Exercise: FAIR enablers brainstorm
  - FAIR Enablers
- Part 2: Curation workflows in repositories (60 mins)
  - Overview: Canadian Repository Landscape
  - St. Lawrence Global Observatory (SLGO)
  - Federated Research Data Repository (FRDR)
  - Ocean Networks Canada (ONC)
  - Scholars Portal Dataverse
- Break (10 mins)
- Part 3: Hands on exercise -- Research Scenarios (45 mins)
- Wrap up (5 mins)

Part 1:
Making Data FAIR

What is FAIR?

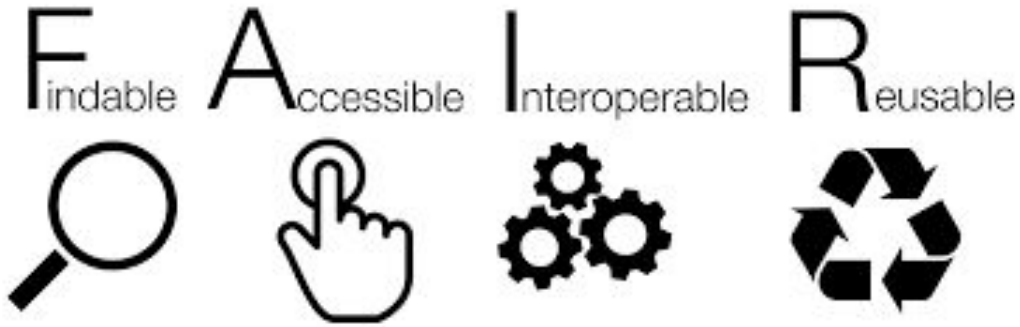
A set of principles to ensure that data are shared in a way that enables and enhances reuse by humans and machines



Wilkinson et al., 2016

What is FAIR?

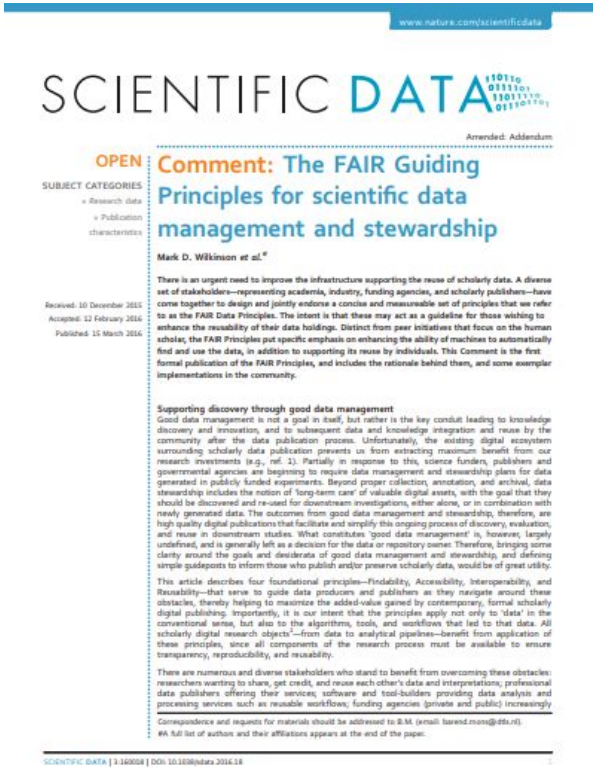
A set of principles to ensure that data are shared in a way that enables and enhances reuse by humans <u>and machines</u>



Wilkinson et al., 2016

FAIR Data Management Principles

- Findable
- Accessible
- Interoperable
- Reusable



www.nature.com/scientificdata

SCIENTIFIC DATA

Amended: Addendum

OPEN

SUBJECT CATEGORIES

» Research data

» Publication characteristics

Received: 10 December 2015
Accepted: 12 February 2016
Published: 15 March 2016

Comment: The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson et al.#

There is an urgent need to improve the infrastructure supporting the reuse of scholarly data. A diverse set of stakeholders—representing academia, industry, funding agencies, and scholarly publishers—have come together to design and jointly endorse a concise and measureable set of principles that we refer to as the FAIR Data Principles. The intent is that these may act as a guideline for those wishing to enhance the reusability of their data holdings. Distinct from peer initiatives that focus on the human scholar, the FAIR Principles put specific emphasis on enhancing the ability of machines to automatically find and use the data, in addition to supporting its reuse by individuals. This Comment is the first formal publication of the FAIR Principles, and includes the rationale behind them, and some exemplar implementations in the community.

Supporting discovery through good data management

Good data management is not a goal in itself, but rather is the key conduit leading to knowledge discovery and innovation, and to subsequent data and knowledge integration and reuse by the community after the data publication process. Unfortunately, the existing digital ecosystem surrounding scholarly data publication prevents us from extracting maximum benefit from our research investments (e.g., ref. 1). Partially in response to this, science funders, publishers and governmental agencies are beginning to require data management and stewardship plans for data generated in publicly funded experiments. Beyond proper collection, annotation, and archival, data stewardship includes the notion of 'long-term care' of valuable digital assets, with the goal that they should be discovered and re-used for downstream investigations, either alone, or in combination with newly generated data. The outcomes from good data management and stewardship, therefore, are high quality digital publications that facilitate and simplify this ongoing process of discovery, evaluation, and reuse in downstream studies. What constitutes 'good data management' is, however, largely undefined, and is generally left as a decision for the data or repository owner. Therefore, bringing some clarity around the goals and desiderata of good data management and stewardship, and defining simple guideposts to inform those who publish and/or preserve scholarly data, would be of great utility.

This article describes four foundational principles—Findability, Accessibility, Interoperability, and Reusability—that serve to guide data producers and publishers as they navigate around these obstacles, thereby helping to maximize the added-value gained by contemporary, formal scholarly digital publishing. Importantly, it is our intent that the principles apply not only to 'data' in the conventional sense, but also to the algorithms, tools, and workflows that led to that data. All scholarly digital research objects[2]—from data to analytical pipelines—benefit from application of these principles, since all components of the research process must be available to ensure transparency, reproducibility, and reusability.

There are numerous and diverse stakeholders who stand to benefit from overcoming these obstacles: researchers wanting to share, get credit, and reuse each other's data and interpretations; professional data publishers offering their services; software and tool-builders providing data analysis and processing services such as reusable workflows; funding agencies (private and public) increasingly

Correspondence and requests for materials should be addressed to B.M. (email: barend.mons@dtls.nl).
#A full list of authors and their affiliations appears at the end of the paper.

Wilkinson et al., 2016

FAIR Data Management Principles (cont.)

- **Findable**
  - (meta)data are assigned a globally unique and eternally persistent identifier.
  - data are described with rich metadata.
  - (meta)data are registered or indexed in a searchable resource.
  - metadata specify the data identifier.
- Accessible
- Interoperable
- Reusable

Wilkinson et al., 2016

FAIR Data Management Principles (cont.)

- Findable
- **Accessible**
  - (meta)data are retrievable by their identifier using a standardized communications protocol.
  - metadata are accessible, even when the data are no longer available.
- Interoperable
- Reusable

“As open as possible, as closed as necessary” -- Horizon 2020

Wilkinson et al., 2016

FAIR Data Management Principles (cont.)

- Findable
- Accessible
- **Interoperable**
  - (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
  - (meta)data use vocabularies that follow FAIR principles.
  - (meta)data include qualified references to other (meta)data.
- Reusable

Wilkinson et al., 2016

FAIR Data Management Principles (cont.)

- Findable
- Accessible
- Interoperable
- **Reusable**
  - (meta)data have a plurality of accurate and relevant attributes.
  - (meta)data are released with a clear and accessible data usage license.
  - (meta)data are associated with their provenance.
  - (meta)data meet domain-relevant community standards.

Wilkinson et al., 2016

Exercise -- How do we make data FAIR?

- Timing: 10 minutes
- Description: In pairs, brainstorm FAIR enablers
- Enter enablers into Mentimeter
- Discussion

FAIR Enablers

Overview of key FAIR enablers used in data repositories

FAIR Enablers: Persistent Identifiers

- Data Citations via DataCite, CrossRef, etc
  - Static vs dynamic data (ongoing data streams, versioning)
  - Data Citation Guidelines for Earth Science Data. https://doi.org/10.6084/m9.figshare.8441816
- ORCID for individuals
- Research Organization Registry (ROR) - Newly integrated into DataCite metadata kernel
- International Geo Sample Number for environmental samples
- Other Entities - FREYA *Survey of Current PID Services Landscape*, https://doi.org/10.5281/zenodo.1324296
- Metrics - COUNTER, impact factor?, recognition via culture change
- Working Groups (RDA, ESIP), Events (FORCE 11, Pidpalooza), FREYA (improving linkages, forum)

DataCite
FIND, ACCESS, AND REUSE DATA

ORCID

ROR

IGSN

FAIR Enablers: Certification

- 16 Requirements self-assessed online, 2 peer-reviews (1000 euro fee)
- Applications become publicly available
- Integration into re3data.org
- 68 CoreTrustSeal certified repositories
- Compliance levels (must be at 3 or 4)
  - 0 – Not applicable
  - 1 – The repository has not considered this yet
  - 2 – The repository has a theoretical concept
  - 3 – The repository is in the implementation phase
  - 4 – The guideline has been fully implemented
- Canadian WDS Members: Canadian Astronomy Data Centre, Ocean Networks Canada, Polar Data Catalogue

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

FAIR Enablers: Certification

The TRUST Principles for Digital Repositories - https://bitly.com/trustprinciples

- **T** - Transparency (R1 Missions, R2. Licenses, R4. Confidentiality/Ethics,
- **R** - Responsibility (R3. Continuity of Access, R7. Integrity and Authenticity, R6. Expert Guidance)
- **U** - User (R8 Relevance and Understandability R11 Data Quality R13 Data Discovery, R14. Data Reuse)
- **S** - Sustainability (R5. Funding Plan, R10. Preservation Plan)
- **T** - Technology (R9. Storage, R12, Archive, R15 Infrastructure, R16 Security)

ISO 16363 - more demanding standard involving a full external audit and certification

FAIR Enablers: Metadata

- Data is more than just data; metadata is data about data
- Provides a standardized way to describe contextual information, variable names, and values
- 3 main types: descriptive (e.g., title, author, etc.), structural (e.g., METS), and administrative (e.g., licensing and copyright)
- Helps with machine-to-machine interoperability
- Helps your friendly neighborhood data librarian and archivist make informed decisions around sharing and ultimately preserving your data

Example Metadata Standards:

- Dublin Core (general purpose)
- DataCite (research data)
- ISO 19115 (geospatial data)
- Darwin Core (biology)
- Ecological Metadata Language (ecology)
- Data Documentation Initiative (research data)

Metadata Standards Directory:
http://rd-alliance.github.io/metadata-directory/

Metadata Standards Catalogue:
https://rdamsc.bath.ac.uk/

FAIR Enablers: Data licensing

- Licensing defines how your data can be reused
- Common data licensing schemes:
  - https://creativecommons.org/
  - https://opendatacommons.org/
- Encourage least restrictive licenses where possible:
  - CC-0 or CC-BY vs. CC BY-ND (No Derivatives restricts data reuse)
- Custom licensing also an accepted practice
- Data Licensing Guide from the DCC: https://bit.ly/1RDEqfw
- ARDC Guide: https://bit.ly/2VutKr9 -- includes flowcharts!

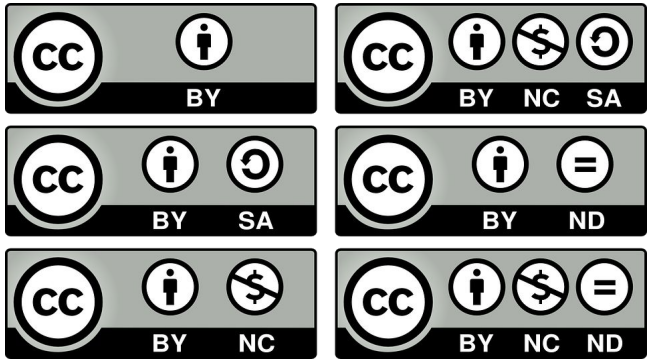


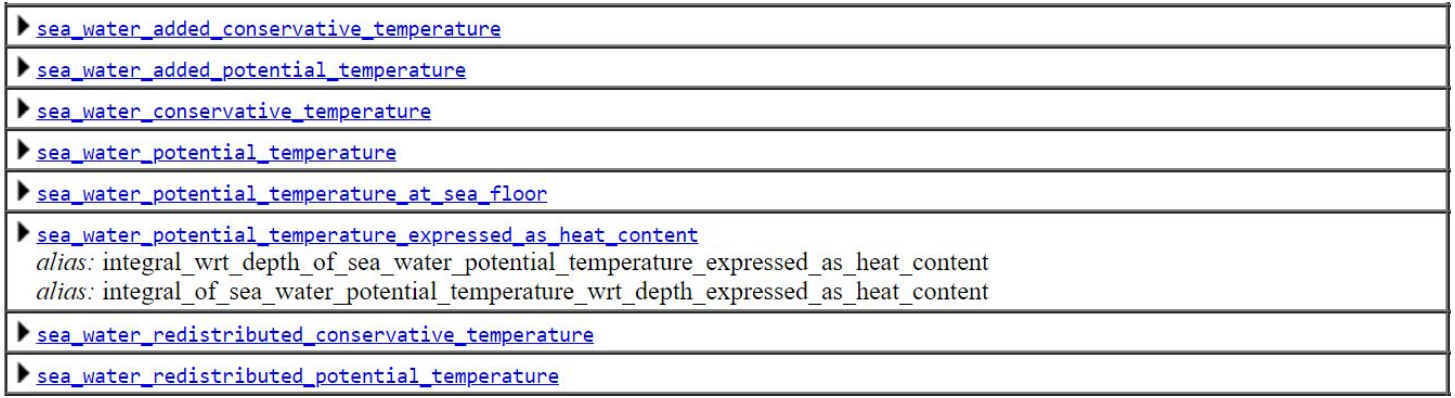
Image credit: https://pixabay.com/vectors/creative-commons-licenses-icons-by-783531/

FAIR Enablers: Controlled Vocabularies

- Controlled vocabularies allow for a common dialogue
- Increased interoperability, interoperability between standards
- Gives confidence that the same things are being referred to in the same ways
- Less important from the data provider perspective, VERY important from the data repository perspective

CF Standard

132 results for temperature



FAIR Enablers: Formats

Innovation vs Interoperability

FAIR Enablers: Formats

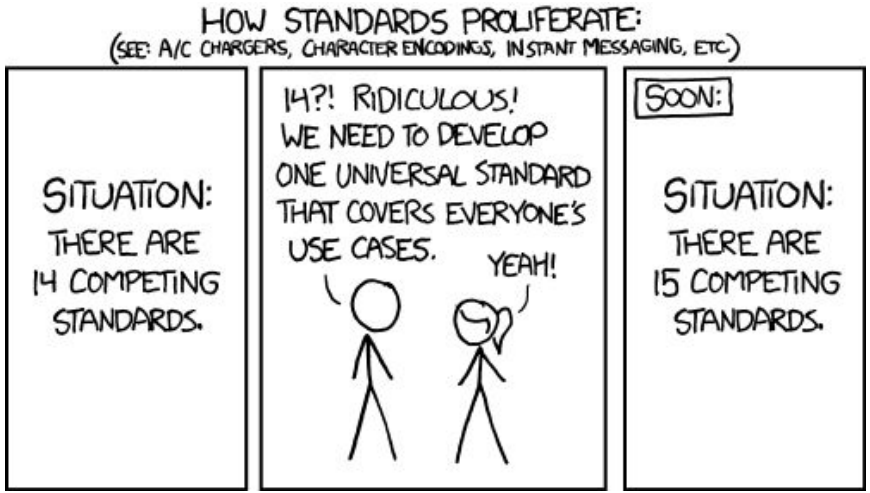
- Proprietary formats pave the way for open standards
- Open standards allow for increased interoperability & sharing
- Open standards are slow to be created, slow to be adopted, and slow to change
    - This creates stability
- Standard formats allow for easy collaboration without forcing common tools

FAIR Enablers: Provenance

- “Information concerning the creation, attribution, or version history of managed data” (Casrai)
- Documentation around origin and history of data and who generated/transformed it
- Recorded as metadata (e.g., date created, creator, instrument or software used, data processing methods):
  - README text file description
  - W3C standard, such as PROV-DM provenance data model
- Repositories:
  - Record details of data authors
  - Data documentation and metadata
  - Track versioning of metadata, files, and datasets

FAIR Enablers: Standards & APIs

- **A**pplication **P**rogramming **I**nterfaces
- Allow for the use of standard commands and language by users and systems to interact with repositories
- Harvesting & interoperability between systems
  - e.g. Dataverse & SLGO metadata to FRDR
- Examples include OAI-PMH (metadata), SWORD API (data deposit), CSW (geospatial)

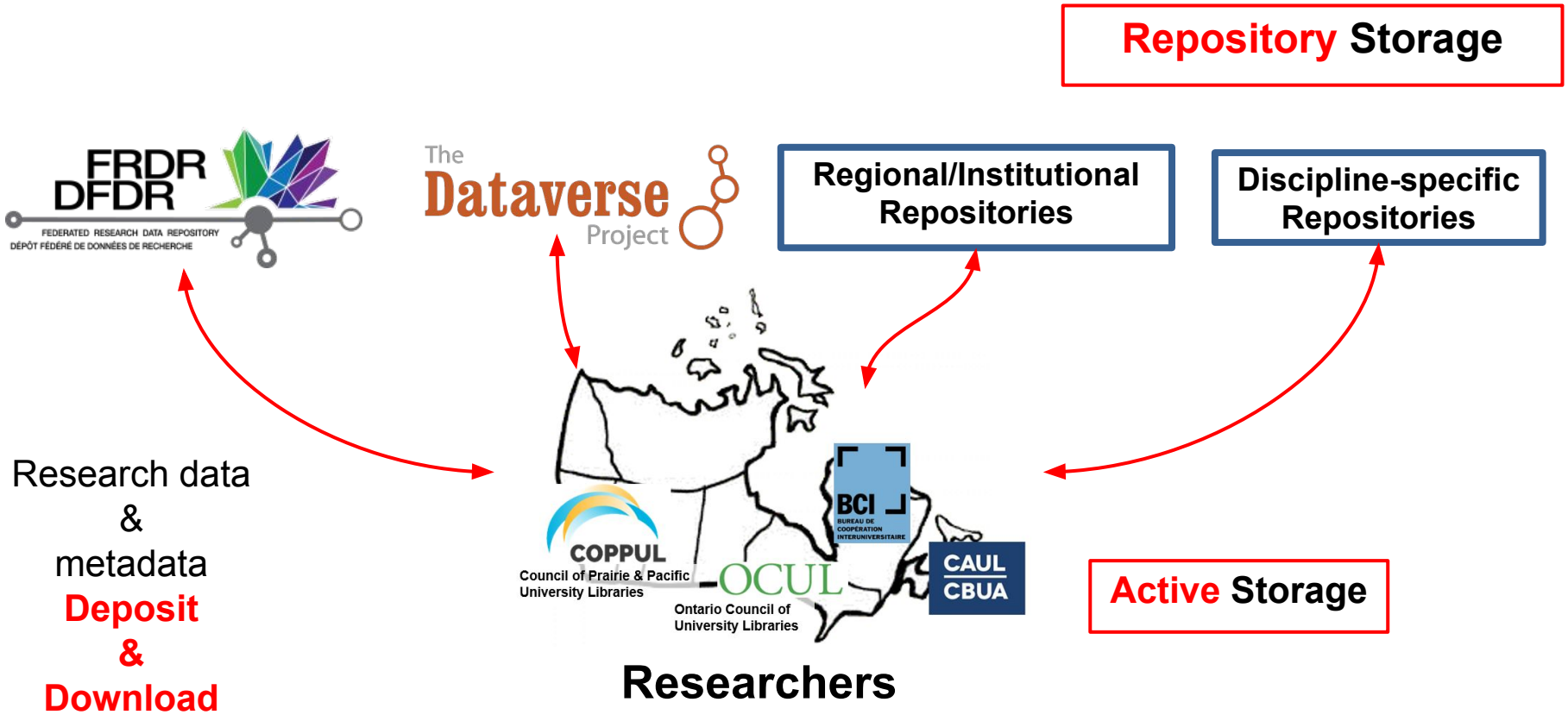


https://xkcd.com/927/

Part 2:
Curation Workflows in Repositories

The FAIR principles in action

Canadian Repository Landscape



re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

Repository Storage

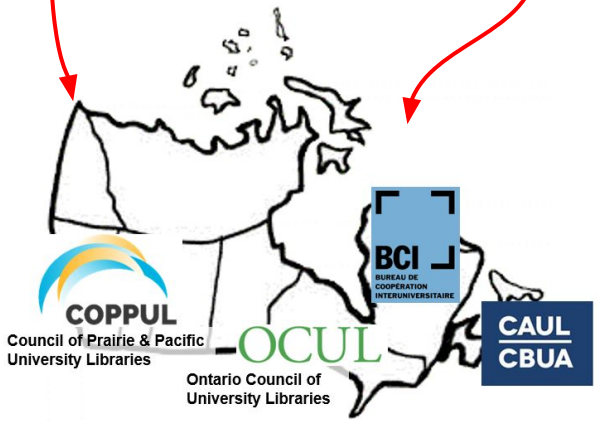
FRDR DFDR
FEDERATED RESEARCH DATA REPOSITORY
DÉPÔT FÉDÉRÉ DE DONNÉES DE RECHERCHE

The Dataverse Project

Regional/Institutional Repositories

Discipline-specific Repositories

Research data
&
metadata
Deposit
&
Download



Researchers

Active Storage

Canadian Repository Landscape

Repository Storage

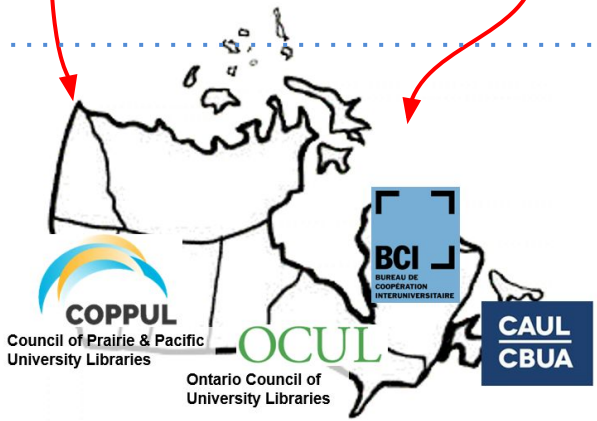
FRDR DFDR
FEDERATED RESEARCH DATA REPOSITORY
DÉPÔT FÉDÉRÉ DE DONNÉES DE RECHERCHE

The Dataverse Project

Regional/Institutional Repositories

Discipline-specific Repositories

Metadata harvested from **45+** repositories indexing over **100,000** datasets



Researchers

Active Storage

National Discovery Service

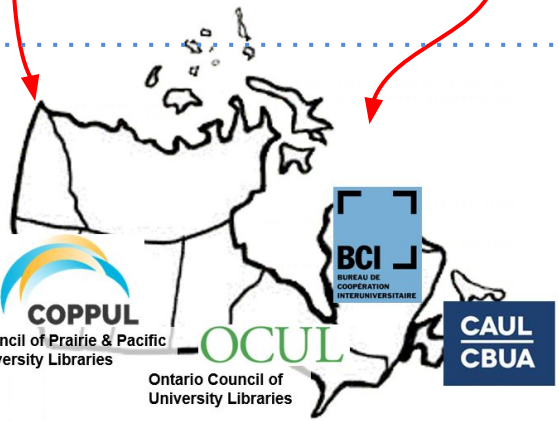
Repository Storage

FRDR DFDR
FEDERATED RESEARCH DATA REPOSITORY
DÉPÔT FÉDÉRÉ DE DONNÉES DE RECHERCHE

The Dataverse Project

Regional/Institutional Repositories

Discipline-specific Repositories



Researchers

Metadata harvested from **45+** repositories indexing over **100,000** datasets

Active Storage

National Discovery Service

- Improve discovery of research (meta)data across Canada
- Break down repository siloes
- Drive traffic to existing repository sites

https://www.frdr.ca

National Discovery Service

- Improve the discoverability of Canadian research data internationally
- Create interoperability between national and international platforms

https://www.frdr.ca

FAIRly Canadian

Exploring the FAIR principles and curation workflows in Canadian repositories

SLGOGSL

St. Lawrence Global Observatory | Observatoire global du Saint-Laurent

Brad Covey, Senior Analyst

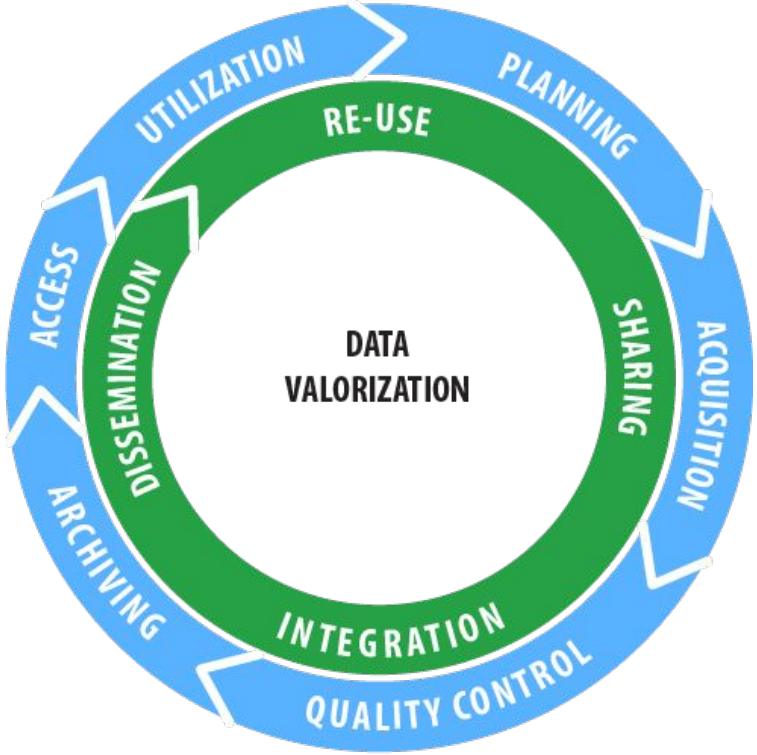
coveyb@ogsl.ca

SLGO - THE FIRST INTEGRATED OCEAN OBSERVING SYSTEM IN CANADA

Integrate, share and disseminate the data from our partners to promote the access, use and re-use of scientific data

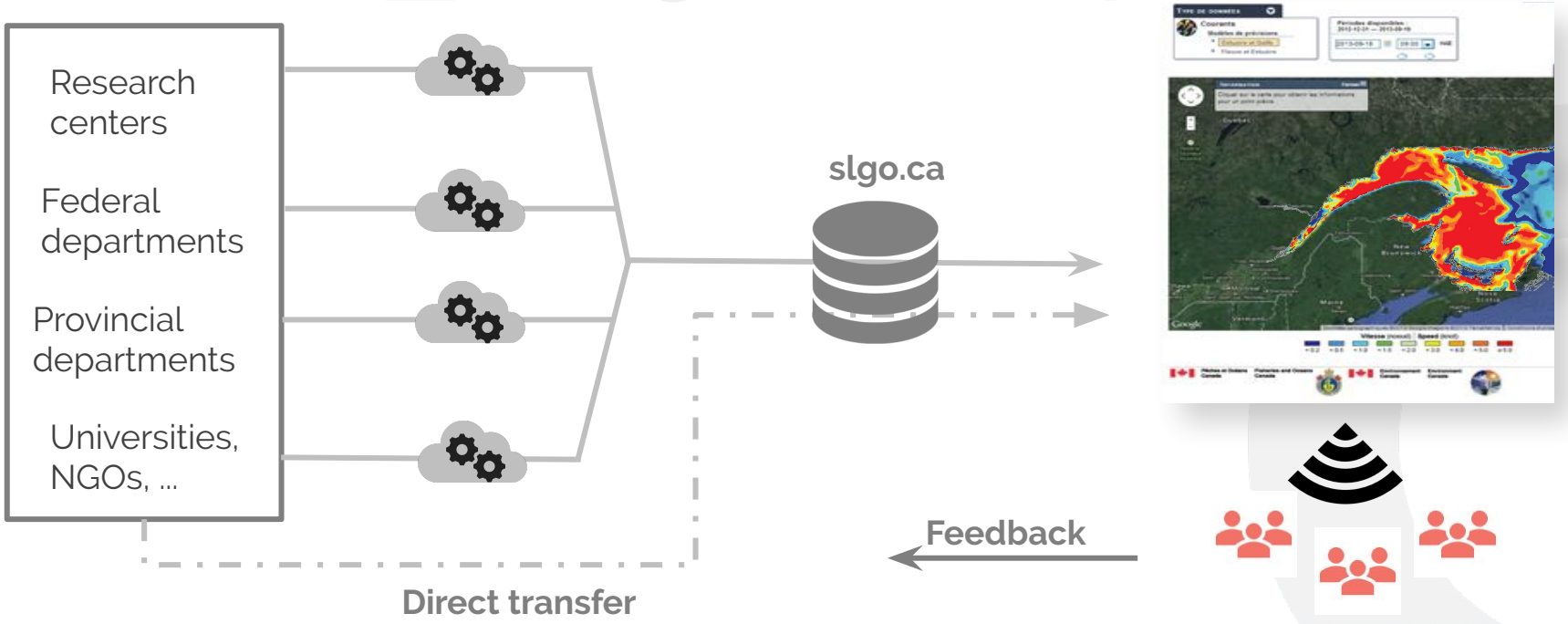

Respond to the needs of our members (producers and/or users) of data
Facilitate clear decision making in many sectors
USAGES CÔTIERS
PARTICIPATION CITOYENNE
NAVIGATION
OCÉANOGRAPHIE

Data Life Cycle



Integrated ocean observing systems

» Members database » Web services » Integration and hosting » Open access and visualisation tools



ACTIVE MEMBERS

Fisheries and Oceans Canada
Pêches et Océans Canada

Environment and Climate Change Canada
Environnement et Changement climatique Canada

Forêts, Faune et Parcs Québec

UQAR SMER
Université du Québec à Rimouski
Institut des sciences de la mer de Rimouski

AMIK
Agence Mamu Innu Kaikusseht

Agriculture, Pêcheries et Alimentation Québec

merinov
CENTRE D'INNOVATION DE L'AQUACULTURE ET DES PÊCHES DU QUÉBEC

UNIVERSITÉ LAVAL

Québec OCÉAN

ROMM
RÉSEAU D'OBSERVATION DE MAMMIFÈRES MARINS

CANADIAN WHALE INSTITUTE

AGHAMM ★ MMAFMA

RÉSEAU QUÉBÉCOIS D'URGENCES POUR LES MAMMIFÈRES MARINS

Sépaq
Aquarium du Québec

ZIP LES DEUX RIVES

Attention Fragîles

COMITÉ ZIP JACQUES-CARTIER

Comité ZIP de la rive nord de l'estuaire

Comité ZIP du Sud-de-l'Estuaire

UQTR
Université du Québec à Trois-Rivières

COMITÉ ZIPGASPÉSIE

ZIP SAGUENAY - CHARLEVOIX

CONSEIL DES ABÉNAKIS WÔLINAK

CONSEIL DES ABÉNAKIS ODANAK

COMITÉ ZIP CÔTE-NORD DU GOLFE

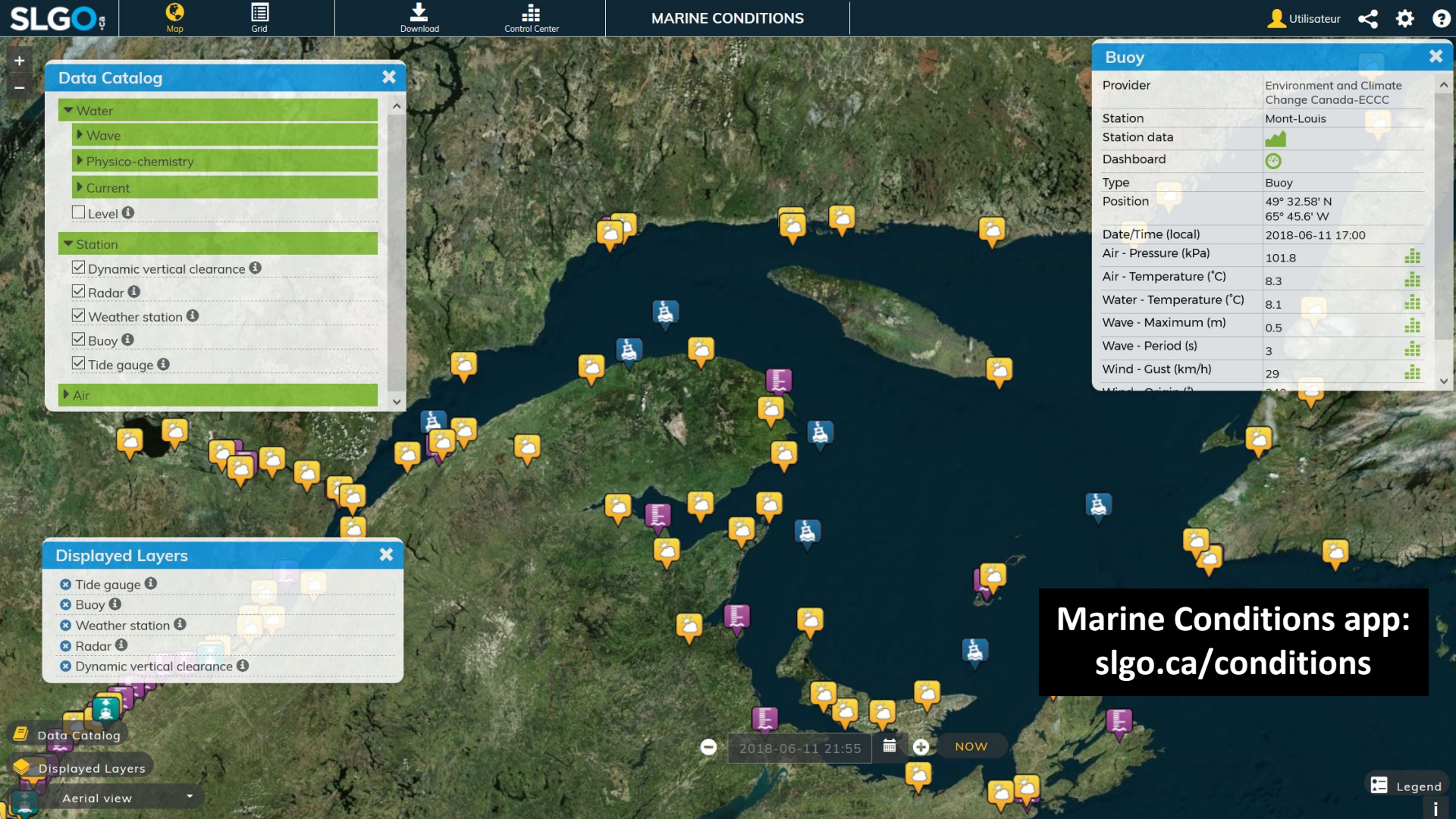
Associate and observing members

PORT MONTRÉAL

Sécurité publique Québec

Canadian Space Agency | Agence spatiale canadienne

Canada Economic Development | Développement économique Canada


SLGO
Map
Grid
Download
Control Center
MARINE CONDITIONS
Utilisateur
Data Catalog
Water
Wave
Physico-chemistry
Current
Level
Station
Dynamic vertical clearance
Radar
Weather station
Buoy
Tide gauge
Air
Buoy
Provider
Environment and Climate Change Canada-ECCC
Station
Mont-Louis
Station data
Dashboard
Type
Buoy
Position
49° 32.58' N
65° 45.6' W
Date/Time (local)
2018-06-11 17:00
Air - Pressure (kPa)
101.8
Air - Temperature (°C)
8.3
Water - Temperature (°C)
8.1
Wave - Maximum (m)
0.5
Wave - Period (s)
3
Wind - Gust (km/h)
29
Displayed Layers
Tide gauge
Buoy
Weather station
Radar
Dynamic vertical clearance
Marine Conditions app:
slgo.ca/conditions
Data Catalog
Displayed Layers
Aerial view
2018-06-11 21:55
NOW
Legend

PMZA-RIKI | 2018-06-11 21:00

Station photo

Station location

Water - Temperature
10.1 °C

Water - Salinity
23.4 PSU

Air - Pressure
101.67 kPa

Water -Density
1,017.9 kg/m³

Water - Chlorophyll
4.01 µg/l

Wave - Period
4.5 s

Wave - Mean
0.1 m

Wave - Maximum
0.2 m

Current 6m - Speed
E 0.09 m/s

Wind - Speed
O 4 km/h

Wind - Gust
O 5 km/h

Air - Temperature
10.3 °C

Air - Relative humidity
74 %

Air - Precipitation
- mm

Air - PAR radiation
3.54e+15 quanta/cm²sec

Water - CDOM fluorescence
7.45 ppb

Current surface - Speed
E 0.6 m/s

Water - pH
8.3 unit

Water - Backscattering
0 m-1

SLGO

Map | Grid | Chart | Extract | Reset | BIODIVERSITY | Utilisateur

Search
Period
Species group
Species
Sampling method
Institution
Collection

Légende

Map | Satellite

17 of 27 data

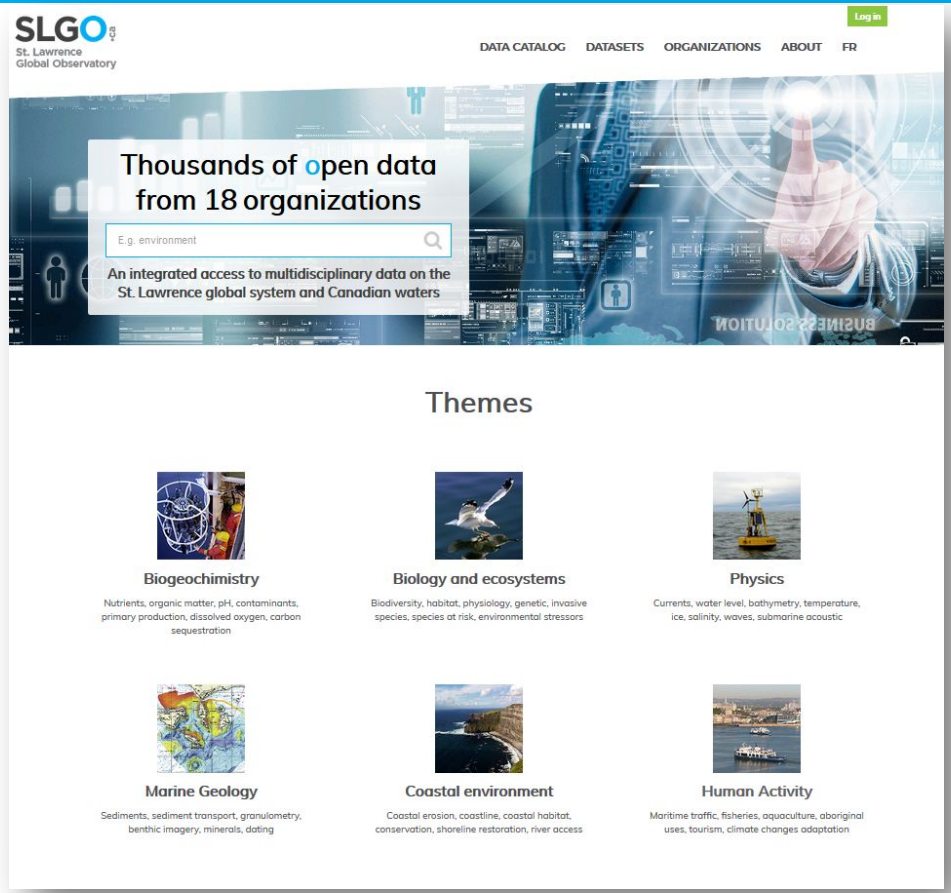
Date	2003-07-30
Group	Marine mammals
Species	Humpback whale
Scientific name	Megaptera novaeangliae
Sampling method	Visual
Collection	Marine Mammal Observat Network - ROMM Observe
Owner	Réseau d'observation de mammifères marins
Individual count	2
Presence/Absence	Presence

Biodiversity app: ogsl.ca/bio

Hudson Bay
Labrador Sea
Narsarsuaq
NEWFOUNDLAND AND LABRADOR
ONTARIO
QUEBEC
Québec City
Montreal
Ottawa
NEW BRUNSWICK
PRINCE EDWARD ISLAND
NOVA SCOTIA
St. John's
St Pierre and Miquelon
MAINE
VERMONT
NEW HAMPSHIRE
NEW YORK
MASSACHUSETTS
CT
RI
PENNSYLVANIA
New York
Philadelphia
NJ
Toronto
Detroit
Chicago
Minneapolis
WISCONSIN
MICHIGAN
IOWA
ILLINOIS
INDIANA
OHIO
Google
Map data ©2018 Google
200 km

Open Data **CATALOGUE**

catalogue.slgo.ca



Organization

Fisheries and Oceans Canada Pêches et Océans Canada

Fisheries and Oceans Canada

Fisheries and Oceans Canada (DFO)'s main activities are the management and protection of marine fisheries, marine science research, waterway management and shipping safety as... read more

License

Other (Open) OPEN DATA

Tags

- invasive species
- tunicates
- bryozoan
- crustacean
- annual survey

Dataset | Activity Stream

Dataset extent

Aquatic invasive species (AIS) program of the eastern Atlantic coast from 2006 to 2017 - Data from Québec's region.

Fisheries and Oceans Canada established a program to monitor aquatic invasive species (AIS) on the eastern Atlantic coast from 2006 to 2017 (activities into the five Atlantic provinces). The data in this database are those of the Quebec region, whose territory is divided into three maritime sectors: Îles-de-la-Madeleine, Gaspésie and Côte-Nord of the Gulf of St. Lawrence. The objectives of this program are to detect new AIS as soon as possible, to monitor their spread, and minimize the risk of their introduction and dispersal through raising awareness of the issue and rapid response. The AIS monitoring program covers nine species, including 7 tunicates, one bryozoan and one crustacean. For more information, please see the **data context**.

Parameters: presence / absence, percentage of coverage (settlement collectors), water temperature, salinity, dissolved oxygen.

View data in the **BIODIVERSITY** application.

Source identification:

Federated Research Data Repository

Lee Wilson

Service Manager | Portage/ACENET

613.482.9344 ext. 108 | lee.wilson@ace-net.ca | @lee_wilson001

FRDR = Federated Research Data Repository (DFDR en français)

https://www.frdr.ca/

A scalable, federated platform for digital research data management and the discovery of Canadian research data

Discovery **Deposit** **Preservation**

Repository Overview

FRDR is a full-featured repository for research data publication:

- A place for Canadian researchers to deposit large datasets
- Designed for scalability (TB+ scale), leveraging Globus File Transfer
- Storage and compute geographically distributed at multiple hosting sites

TO BE FINDABLE:

F1. (meta)data are assigned a globally unique and eternally persistent identifier.

FRDR assigns DOIs to all accepted submissions through DataCite Canada.

F2. data are described with rich metadata.

All content deposited directly into FRDR is required to conform to a template from which complete Dublin Core and DataCite XML can be generated; custom metadata fields are supported when applicable with the goal of ensuring these map to existing standards.

F3. (meta)data are registered or indexed in a searchable resource.

FRDR's own OAI endpoint is used to harvest metadata for discovery in FRDR's index of Canadian data repositories, and this OAI endpoint is open to the public. Conform to schema.org (Google Data Search) & datasets discoverable through DataCite.

F4. metadata specify the data identifier.

DOIs and other identifiers are visible on objects' landing pages.

TO BE ACCESSIBLE:

A1. (meta)data are retrievable by their identifier using a standardized communications protocol.

DOIs are indexed by aggregators such as DataCite, and DOIs can be used as an entry point to FRDR's own index.

A1.1 the protocol is open, free, and universally implementable.

DOIs and OAI are open standards. Upload/download via HTTPS or GridFTP (Globus File Transfer). Additionally, the code for FRDR's harvester is openly available: https://github.com/usask-rc/frdr_harvest

A1.2 the protocol allows for an authentication and authorization procedure, where necessary.

Globus authorization is implemented for any datasets which are not publicly available (i.e., unpublished or embargoed), and ORCID can be used to register for authorization. Authentication required to deposit data.

A2 metadata are accessible, even when the data are no longer available.

FRDR has been designed to maintain tombstone pages for datasets which have been deaccessioned.

TO BE INTEROPERABLE:

I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.

FRDR uses Dublin Core and DataCite as its default schema. FRDR has been working with Portage's Discovery Expert Group to facilitate the use of OpenRefine to map FAST vocabulary terms onto harvested keyword metadata.

I2. (meta)data use vocabularies that follow FAIR principles.

Dublin Core, DataCite and the FAST vocabulary is consistent with FAIR principles.

I3. (meta)data include qualified references to other (meta)data.

Wherever possible, FRDR's curators make use of "related" elements in submitted datasets in order to improve data linking.

TO BE RE-USABLE:

R1. meta(data) have a plurality of accurate and relevant attributes.

FRDR's default metadata template based on the Dublin Core and DataCite schemas is robust, with many accurate and relevant attributes from which users may choose.

R1.1. (meta)data are released with a clear and accessible data usage license.

FRDR defaults to, and encourages the use of, Creative Commons licenses wherever possible for submitted data. CC-0 (licence waived) or CC-BY (attribution) are prefered.

R1.2. (meta)data are associated with their provenance.

Provenance is automatically recorded from curator logs during our approval process and permanently associated with deposited data.

R1.3. (meta)data meet domain-relevant community standards.

In addition to implementing Dublin Core, DataCite, OAI, and DOIs as above, FRDR also allows domain-specific metadata elements when appropriate.

FRDR Curation Support

- All datasets in FRDR undergo curator review
- **Lightweight curation:** data files checked for validation and obvious errors (i.e., NULL values for tabular data), metadata reviewed and augmented where necessary, supporting documentation required and reviewed (e.g., a README file)
- **What we don’t do (yet):** Results validation or data verification, file transformations, run code or models
- **Current curation support model:** FRDR team, Curation Coordinator, Curation Expert Group (advisory), librarians/data managers working with the researcher

FRDR Curation Demo

- Curation Dashboard (current)
- Proposed Curation Workflow (mock up)

Ocean Networks Canada

Data Curation Practices

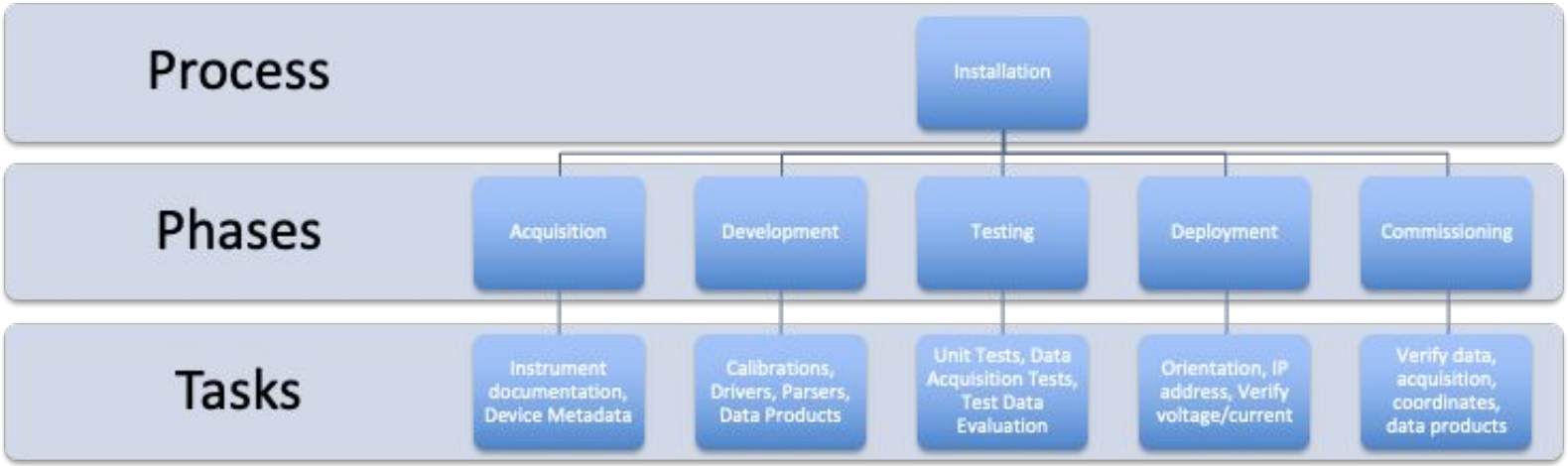
Reyna Jenkyns | reyna@uvic.ca |
ORCID: 0000-0001-6975-6816

A UNIVERSITY OF VICTORIA INITIATIVE


SOUTHERN BRITISH COLUMBIA INFRASTRUCTURE
OCEAN NETWORKS CANADA
BRITISH COLUMBIA
VANCOUVER ISLAND
VENUS Observatory
Strait of Georgia
Central
East
Explorer Plate
Middle Valley
NEPTUNE Observatory
Barkley Canyon
Endeavour
Pacific Plate
PACIFIC OCEAN
Cascadia Basin
Juan de Fuca Plate
North American Plate
WASHINGTON - USA
See Inset Map
60 km

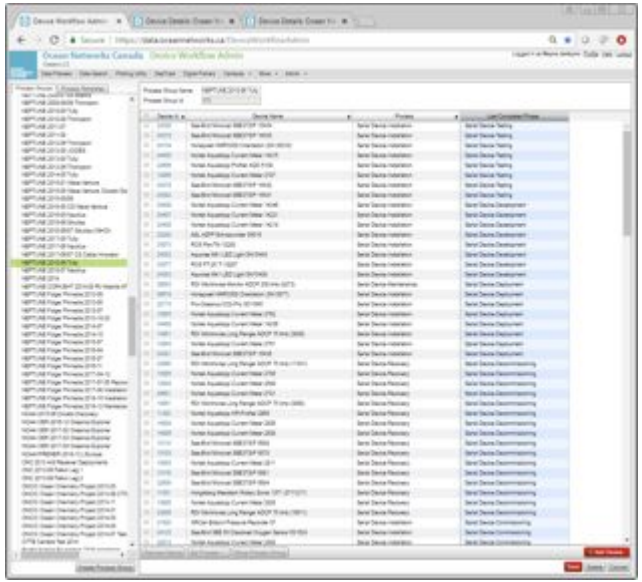
Device Workflow

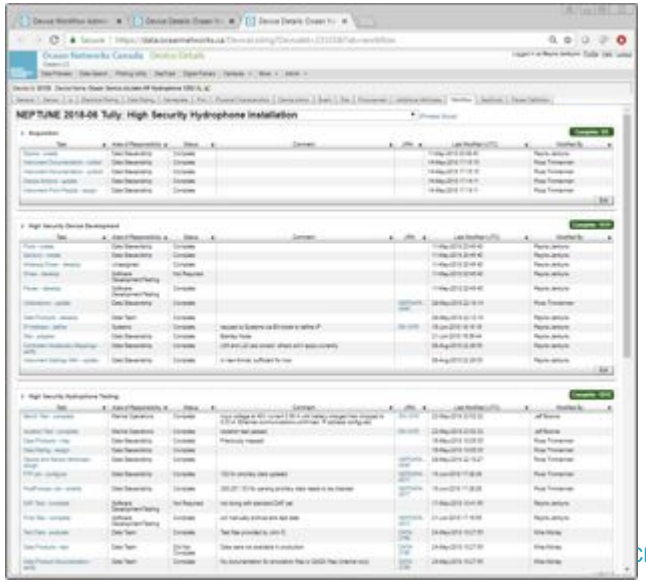
Administrators manage workflows for expedition and device operations (e.g., installations, maintenance, recoveries) as processes with ordered phases, where each phase has a list of tasks).



Device Workflow

- Customizable and versioned workflows
- Web-based Administration Tool and Worksheets with permission-controls
- Task status updates associated with user, links to further details of execution in JIRA tickets, and more detailed definitions in wiki pages
- Records events in an instrument or system's life cycle





Workflow Best Practice Enablement - Examples

Instrument Documentation - collect; Device - create	CTS R8 The repository accepts data and metadata based on defined criteria to ensure relevance and understandability for data users
Controlled Vocabulary Mappings - verify	FAIR I12: (meta)data use vocabularies that follow FAIR principles
Data agreement - complete	CTS R10 The repository assumes responsibility for long-term preservation and manages this function in a planned and documented way
Data Access - enable; Data Products - map	FAIR F4. (meta)data are registered or indexed in a searchable resource; CTS R13 The repository enables users to discover the data…
Data - verify; Automated QAQC - prepare; Manual QAQC - complete	CTS R11. The repository has appropriate expertise to address technical data and metadata quality and ensures that sufficient information is available for end users to make quality-related evaluations.
Data agreement metadata - update	CTS R2 The repository maintains all applicable licenses covering data access and use and monitors compliance
ERDDAP Datasets - update	FAIR I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation; FAIR R1.3. (Meta)data meet domain-relevant community standards

Instrument Documented Procedures

- Make and model information
- Networking details – ports, IP addresses, baud rates, time synchronization, cables, …
- Data acquisition and control description
- Hardware details – maximum depth rating, mounting requirements,…
- Calibrations – formulas, sheets, in-situ benchmarks, etc.
- Controlled Vocabulary mappings
- Site characterization assessment/needs
- Orientations – how to define and record
- Data distribution set up – data products, post-process routines, ERDDAP support, linked data relationships
- Attributes – configuration parameters for data acquisition, data products, etc.
- Manufacturer software note
- And more

Best Practices and Standards

- **Metadata:** ISO 19115; Cruise Summary Reports (nearly ready)
- **Web Services:** OPeNDAP (ERDDAP server), limited WMS and WFS, custom services leverage controlled vocabularies for discovery
- **Data Licensing/Policy:** conducted training on Creative Commons and First Nations Principles of Ownership, Control, Access and Possession as a basis for revising ONC's data policy.
- **Data formats:** NetCDF, ODV, GPX,…
- **Timestamps/Positions:** ISO 8601 timestamps, WGS84 datum,..
- **Derived Quantities:** TEOS-10, World Magnetic Model
- **Data Quality Standard:** Quality Assurance of Real Time Oceanographic Data (QARTOD), planning usage of ISO 19157
- **Sharing Ocean Best Practices:** https://www.oceanbestpractices.net/

Controlled Vocabularies

- SeaVoX Device Catalogue (dataset metadata)
- SeaDataNet Device Categories British Oceanographic Data Centre Data Storage Units (dataset metadata)
- ICES Platform Codes (expedition metadata
- SeaVoX Platform Categories (expedition metadata)
- Climate & Forecasting Standard Names (ERDDAP datasets.xml, NetCDF data products)
- Global Change Master Directory Keywords (ERDDAP datasets.xml)
- IOOS categories (ERDDAP datasets.xml)
- WORMS taxonomy (dive logging observations/biological samples)
- Coastal and Marine Ecological Classification Standard (dive logging observations of geoforms, substrate)
- ISO 3166 Country Codes (registered users)

vocab.nerc.ac.uk/collection/L22/current/TOOL0949/

-- Vemco VR2C Acoustic Monitoring Receiver --

URI	http://vocab.nerc.ac.uk/collection/L22/current/TOOL0949/
Identifier ()	SDN:L22::TOOL0949
Preferred label (en)	**Vemco VR2C Acoustic Monitoring Receiver**
Alternative label (en)	Vemco VR2C
Definition (en)	A fully autonomous acoustic receiver for monitoring the presence of tagged fauna. It is cabled and has an embedded temperature sensor in its head (typical accuracy +/- 0.5 deg C). It can have a receiver frequency of 69 kHz or 180 kHz and is depth rated to 500 m.
Version Info ()	1
Deprecated()	false
Broader	http://vocab.nerc.ac.uk/collection/L05/current/369/
Date ()	2015-11-27 09:42:45.0

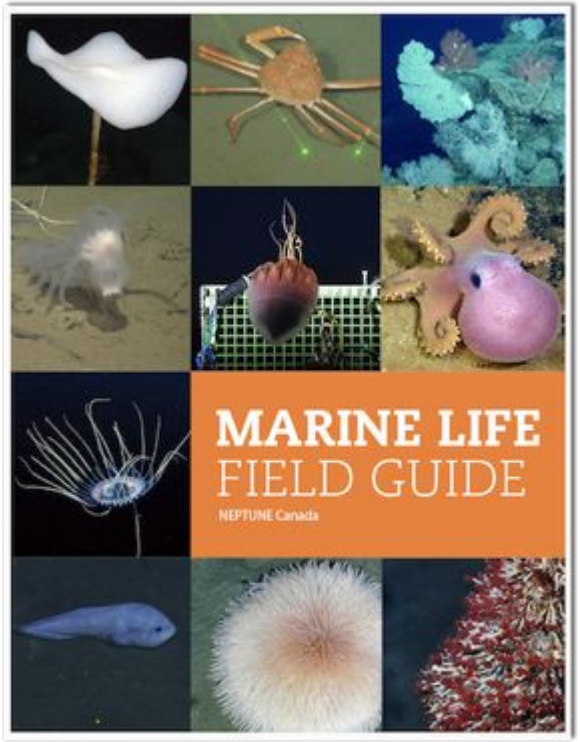
vocab.nerc.ac.uk/collection/L05/current/369/

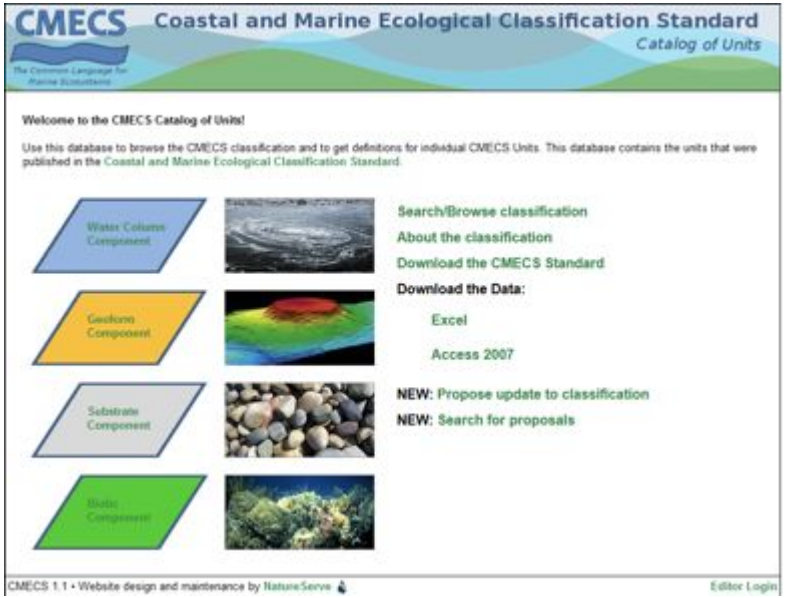
-- hydrophones --

URI	http://vocab.nerc.ac.uk/collection/L05/current/369/
Identifier ()	SDN:L05::369
Preferred label (en)	**hydrophones**
Alternative label ()	
Definition (en)	Devices containing transducers that convert underwater sound waves into electrical signals.
Version Info ()	1
Deprecated()	false
Same as	http://vocab.nerc.ac.uk/collection/L22/current/TOOL0354/
Broader	http://vocab.nerc.ac.uk/collection/L21/current/ICAT04/
Broader	http://vocab.nerc.ac.uk/collection/L19/current/SDNKG01/
Broader	http://vocab.nerc.ac.uk/collection/W06/current/CLSS0002/
Narrower	http://vocab.nerc.ac.uk/collection/L22/current/TOOL0360/
Narrower	http://vocab.nerc.ac.uk/collection/L22/current/TOOL0663/
Narrower	http://vocab.nerc.ac.uk/collection/W01/current/003/
Narrower	http://vocab.nerc.ac.uk/collection/L22/current/TOOL0942/
Narrower	http://vocab.nerc.ac.uk/collection/L22/current/TOOL0949/
Narrower	http://vocab.nerc.ac.uk/collection/L22/current/TOOL0952/
Narrower	http://vocab.nerc.ac.uk/collection/L22/current/TOOL0953/
Narrower	http://vocab.nerc.ac.uk/collection/L22/current/TOOL1159/
Narrower	http://vocab.nerc.ac.uk/collection/L22/current/TOOL1160/
Narrower	http://vocab.nerc.ac.uk/collection/L22/current/TOOL1161/
Date ()	2010-02-19 11:52:56.0

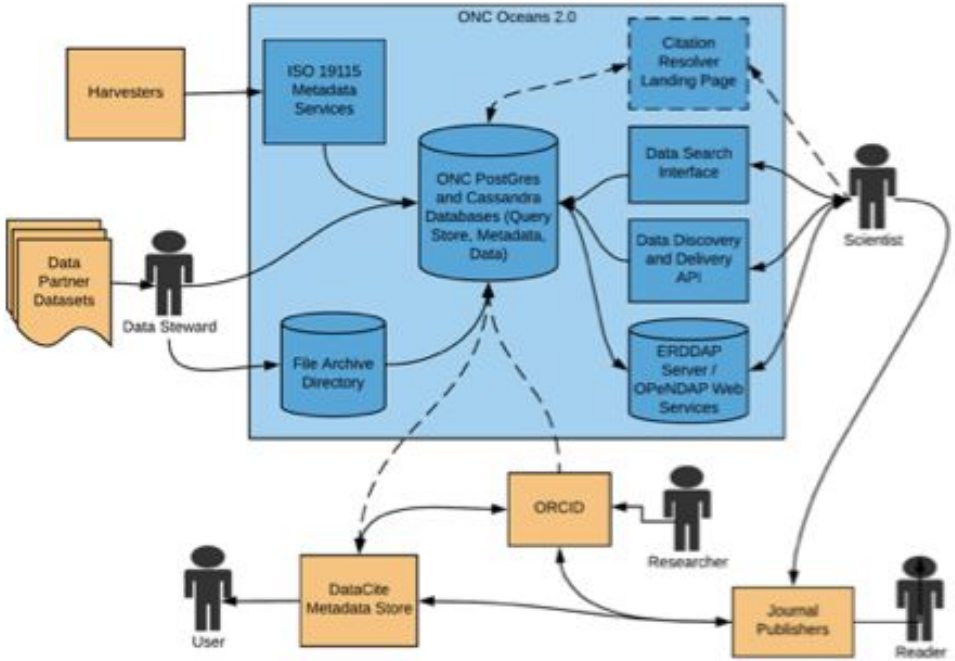
Controlled Vocabularies

ROV Dive Annotation Standards





Dynamic Datasets



- CANARIE Research Data Management funded project 2018-10 to 2020-03 to enhance FAIR compliance in Canada
- Apply 14 recommendations from RDA Data Citations WG output which covers versioning, query store, resolver landing page, technology migration resilience
- **Design challenge** given the permanent nature of DOIs – important decisions for **dataset granularity, versioning** and **sustainable architecture**
- Landing page with integration of query pid, provenance and schema.org metadata

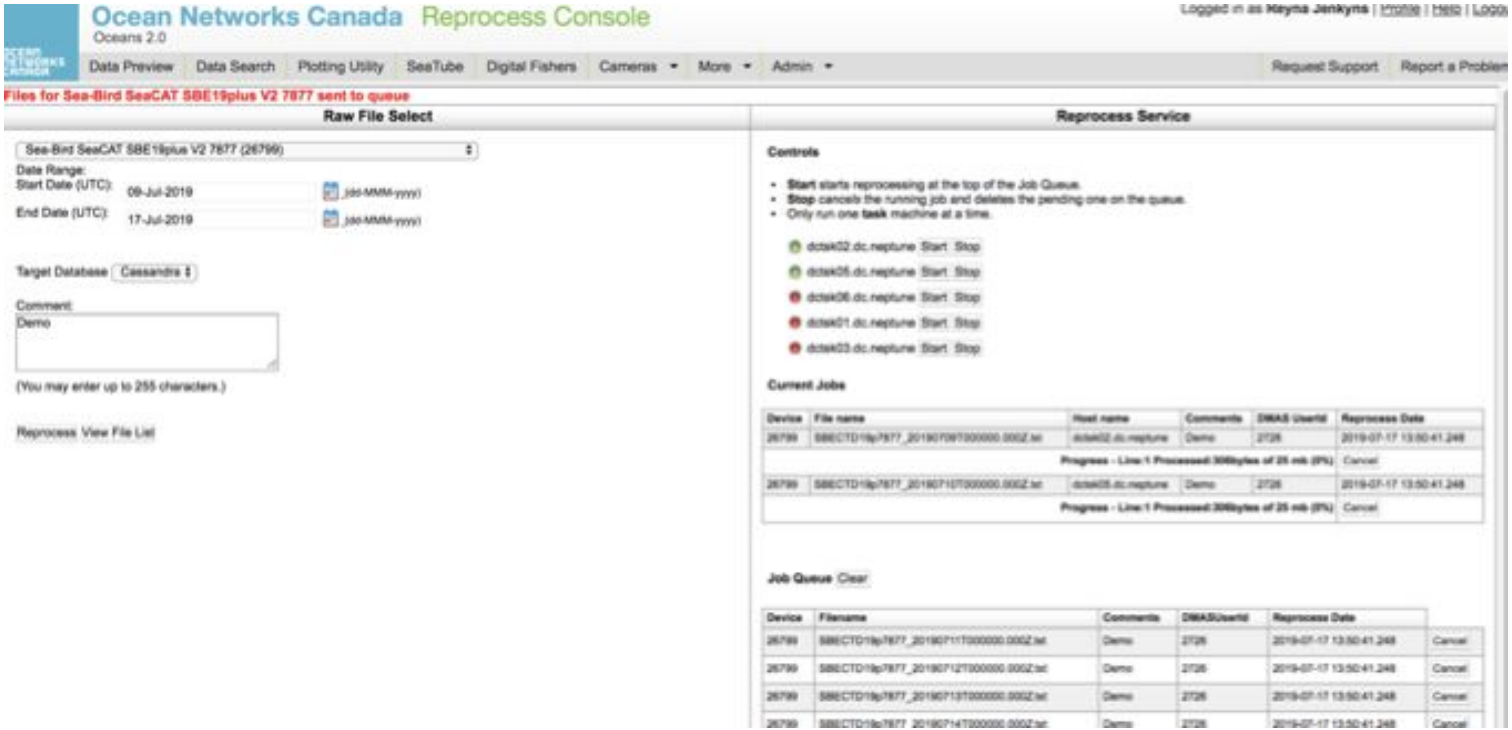
Dynamic Datasets - Dataset Granularity

- Challenges in dataset granularity decisions– by time? By geography? By instrument type? By platform? By data product level?
- Constraints to consider from DataCite metadata kernel, RDA guidelines, suitability to ONC data architecture, data partner attributions, end-users
- Decision to proceed with using data from a ***per instrument deployment on a fixed or mobile platform for all data products available*** (subset query PID allows for further granularity in citation)
- **RDA Dataset Granularity WG** – compatible community conventions important for networks of repositories (e.g., CIOOS), but many diverse existing approaches and terminology interpretations
- Intention to use **dataset collection** concept to aggregate deployments over time of same device category on same platform *(supporting metadata and data discovery already exists in Oceans 2.0)*

Dynamic Datasets - Versioning

- Enhance data versioning for compliance with RDA Guidelines
  - R1 Data Versioning & R2 Timestamping
- Many elements of data framework contribute to versioning
- Exclude changes that do not mpact actual dataset content
- Requires an **aggregated versioning solution** based on triggers with better provenance to describe reason for changes
  - software versioning (improve algorithm or visualization)
  - metadata history (position/orientation, geoextent expands, data product attributes, calibration formulas)
  - reprocessing records (correct calibration errors, add derived variables),
  - archived file modifications (manual fixes or re-generation of derived data products)
  - parser modification history (bugs, instrument firmware changed output)
- Time lag concern between versioning trigger and execution to a new version

Dynamic Datasets - Versioning



reprocess jobs enter a queue and a new dataset version is established at completion of jobs,.... jobs may only partially complete due to job failure/cancellations

Curated Datasets

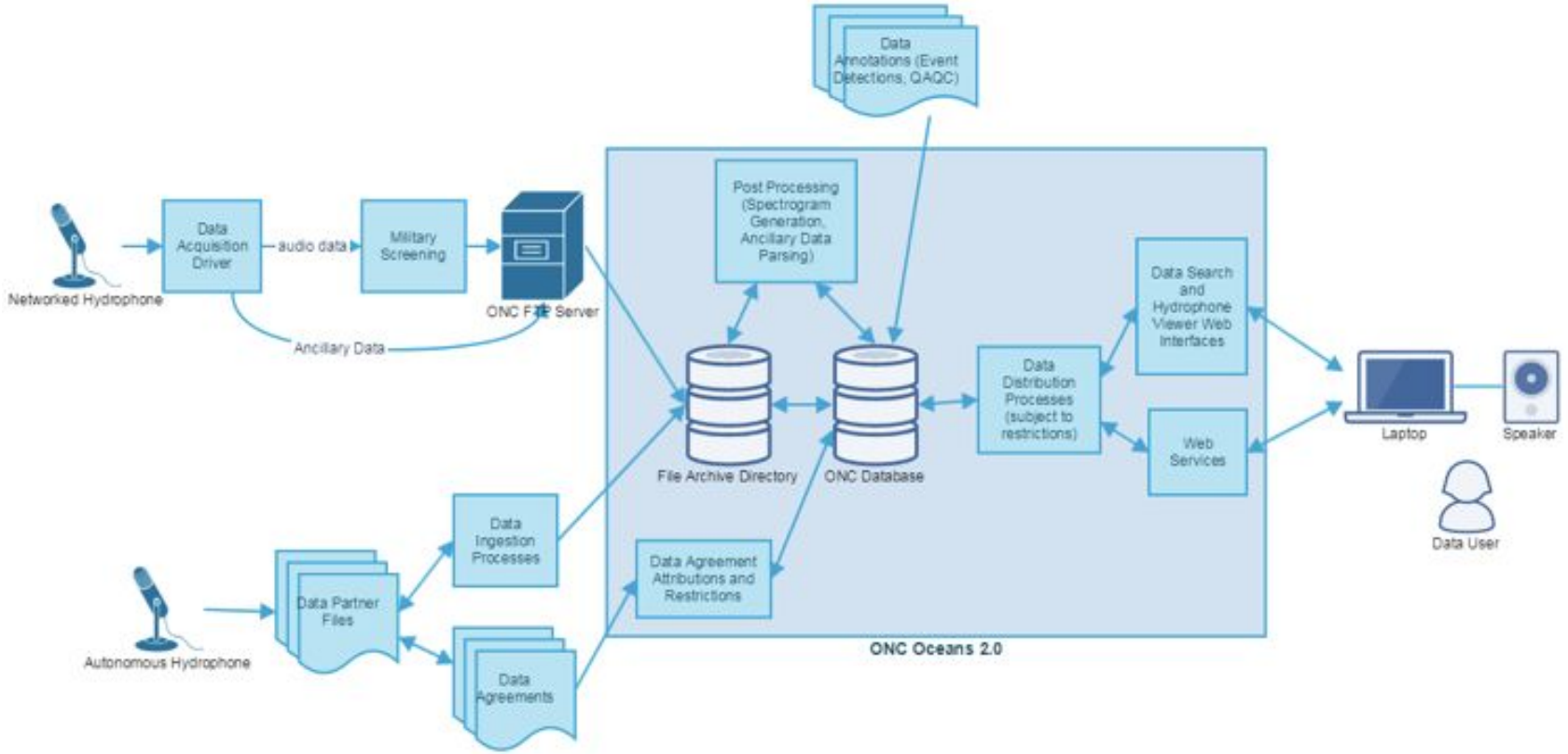
- static data sets that result from processing data that was downloaded from Oceans 2.0, either in support of a publication, or because the time series is of scientific interest
- Curation includes a manual review of QAQC flags, deriving additional variables, averaging data to a preferred temporal resolution and providing basic statistics
- multiple formats (csv, NetCDF and MatLab for scalar data) to improve reuse and interoperability. Supplemental materials, such as processing code may also be included to support data reuse.
- accompanied by metadata that thoroughly describes the provenance of the data from instrument deployment, to internal processing done by Oceans 2.0, to the final curation work done offline
- assigned a Creative Commons license.

Data Hosting and Services

Services provided: data acquisition, hosting, distribution, contributor attributions, video annotation system, granular data restriction options

Active clients/partners: Pacific Salmon Foundation, NOAA OER, DFO, FORCE, Prince Rupert Port Authority, Memorial University Marine Institute, Ocean Tracking Network, Department of Research Defense Canada, Lawrence Berkeley National Laboratory, Tsleil-Waututh First Nation, Kitsumkalum First Nation, Coastguard AIS, Bird Studies Canada, Dalhousie University, Nunavut Tunngavik Inc.,...

Data Agreement Infrastructure

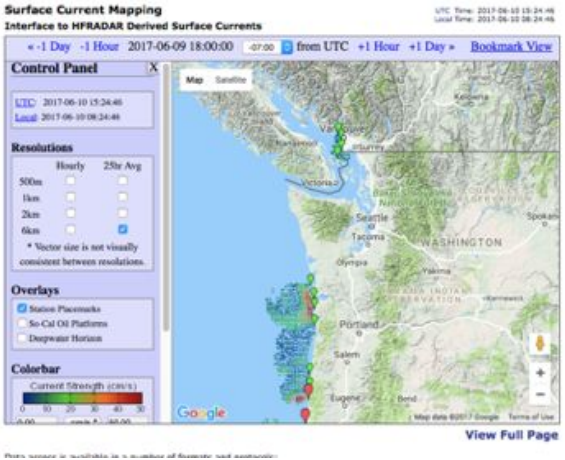


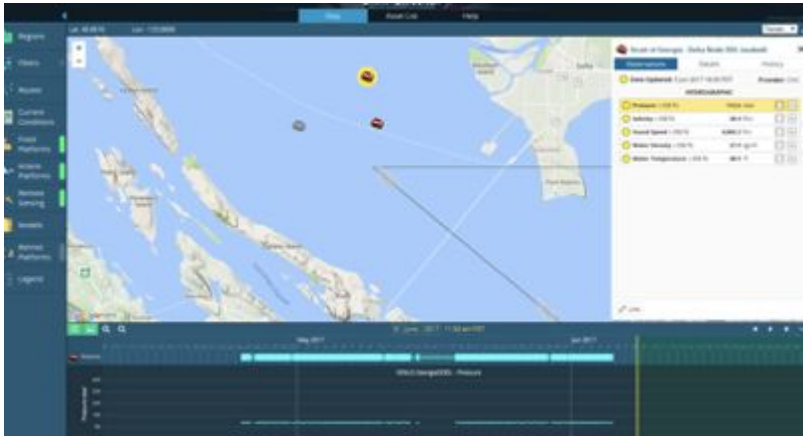
Linked Data

Existing: HFRNet, IRIS, Listening to the Deep, NOAA Tsunami,Polar Data Catalogue, NaNOOS, Pacific Northwest Seismic Network

In Progress: Canadian Integrated Ocean Observing System (CIOOS), Global Ocean Acidification Observation Network

Planned: OBIS, IGSN, World Meteorological Organization Global Tracking System





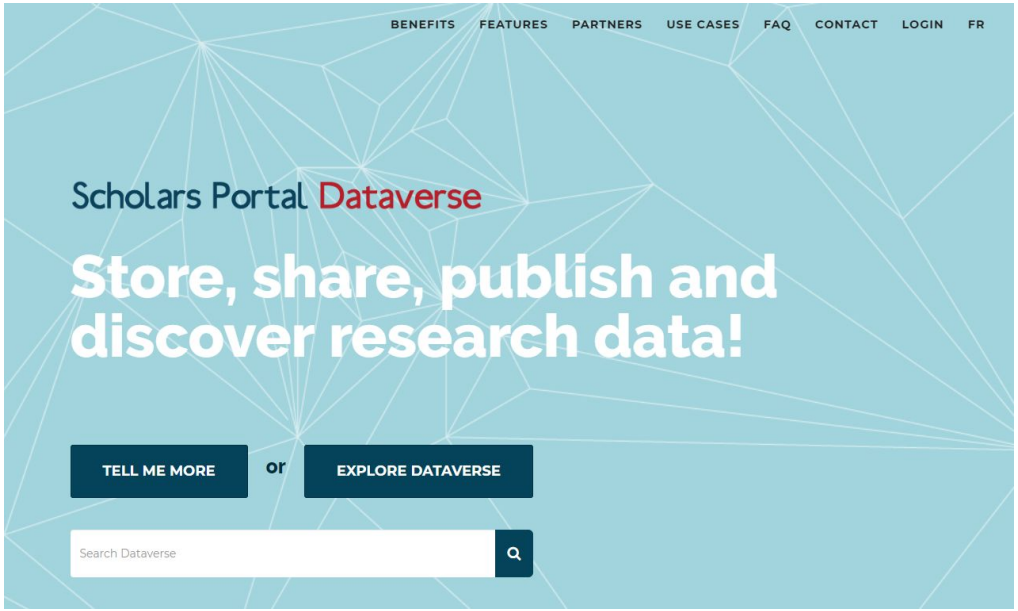
Scholars Portal Dataverse

Meghan Goodchild
Research Data Management Systems Librarian
Scholars Portal/Queen's University

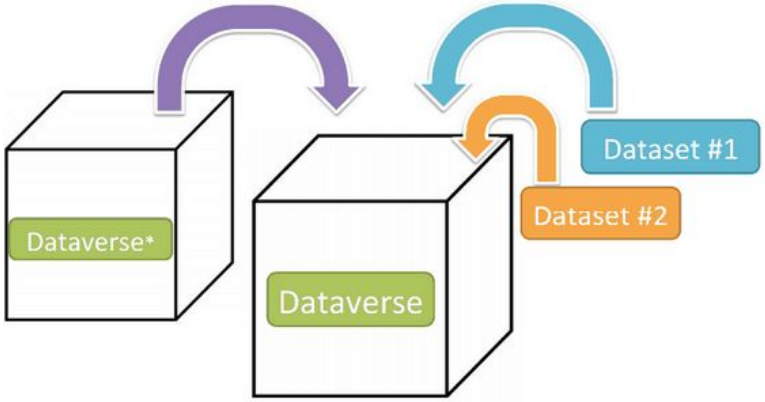
Kaitlin Newson
Digital Projects Librarian
Scholars Portal

Scholars Portal Dataverse

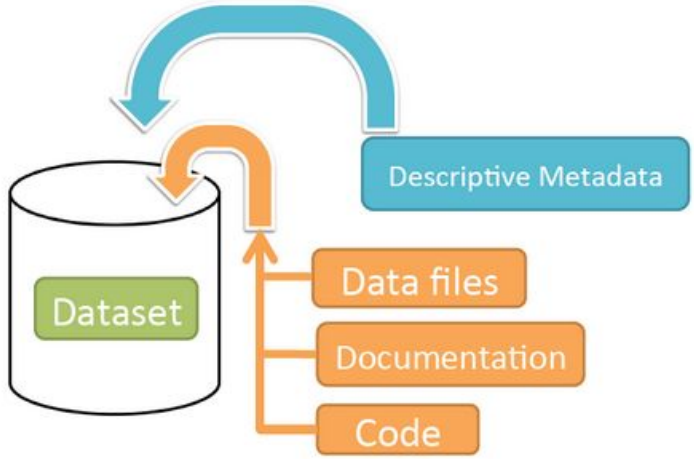
- Open-source repository platform developed by the Institute of Quantitative Social Science (IQSS) at Harvard University
- Developed with quantitative social science data in mind, but not limited to any domain
- Hosted by Scholars Portal on behalf of 47 institutions



What is a Dataverse? What is a dataset?



Dataverse = Container for datasets and/or dataverses

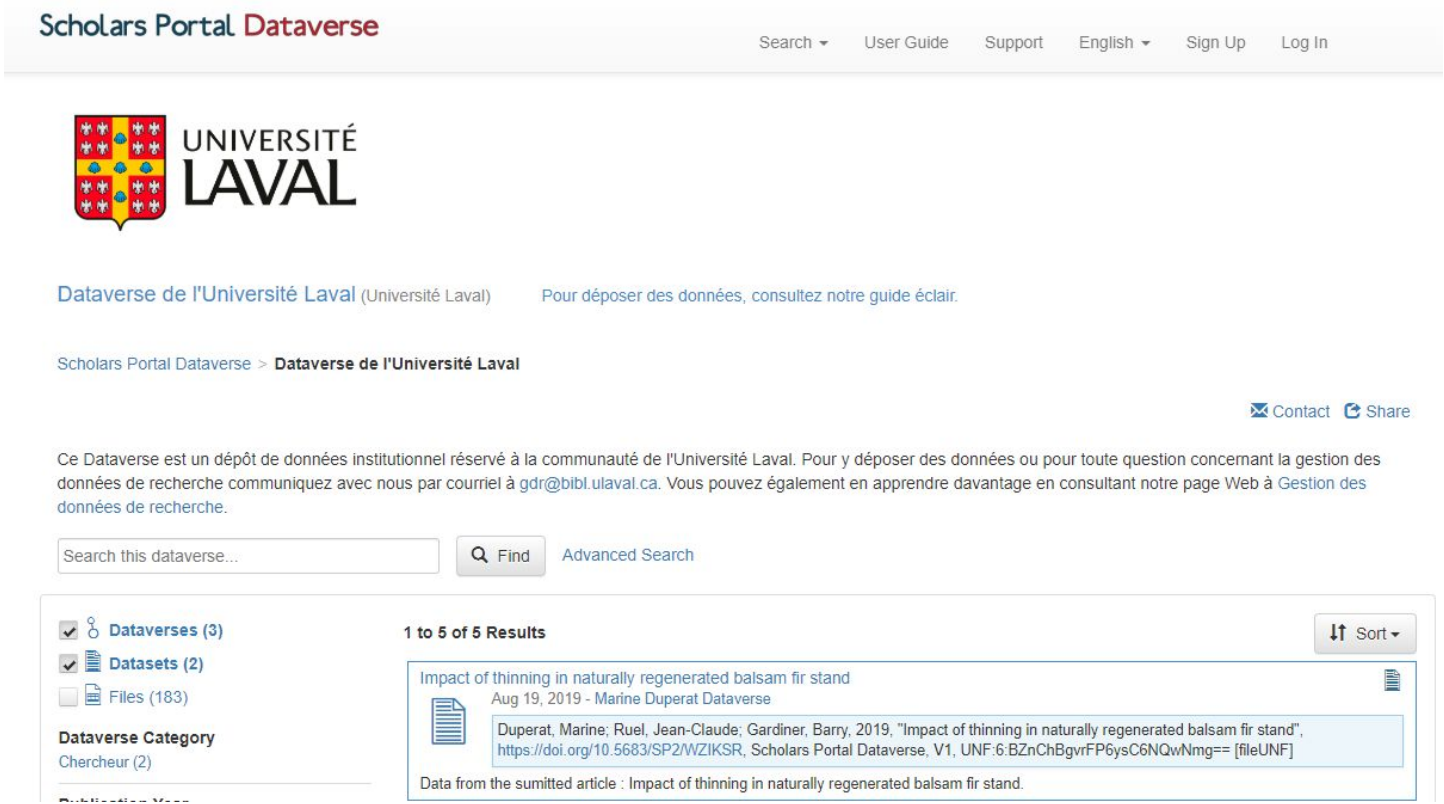


Dataset = Container for your data, documentation, and code

Institutional Dataverses

- All data organized by institution
- Library administers Institutional Dataverse space
- Researchers deposit in Institutional Dataverses defined by user affiliation
- Customizable features (branding, featured Dataverses, facets, etc.)

Institutional branding

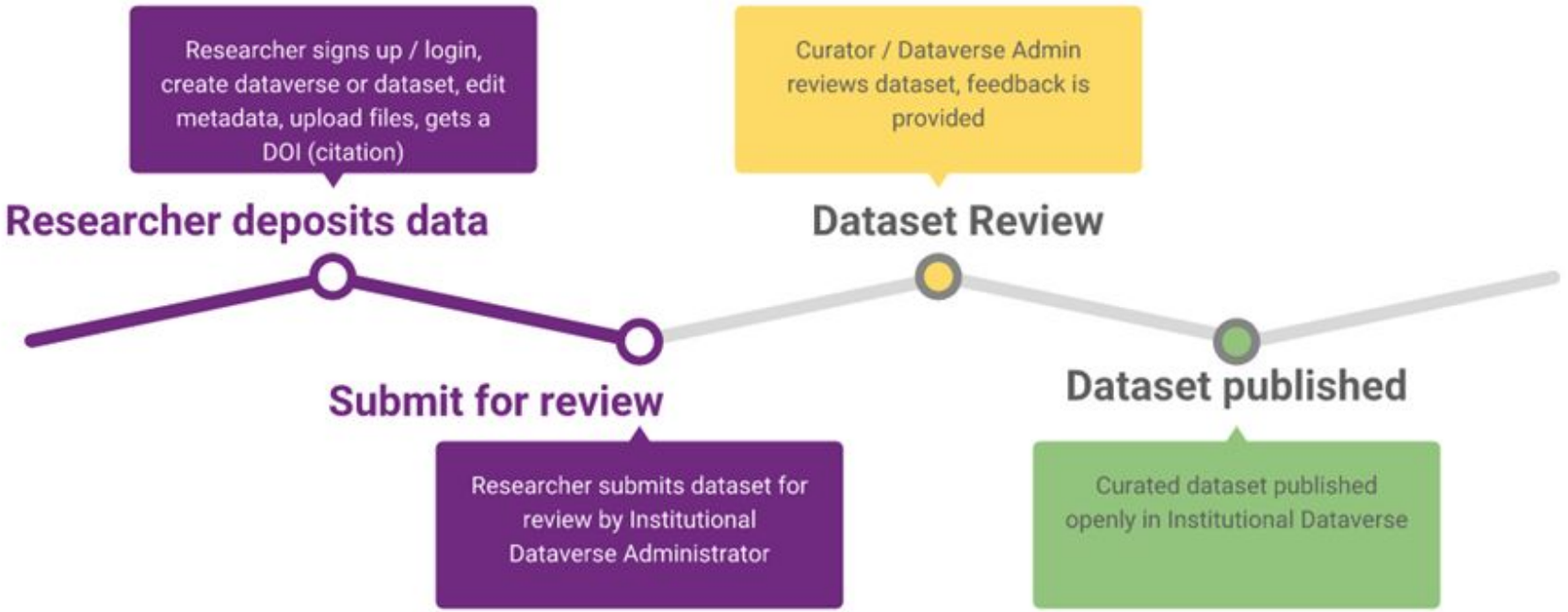


Types of deposit

Platform provides flexible deposit options for various levels of library intervention:

- Closed (Library manages deposit)
- Mediated
  - User submits dataset for review by admin
- Open (not recommended)
  - Trusted users
  - Monitoring

Mediated deposit



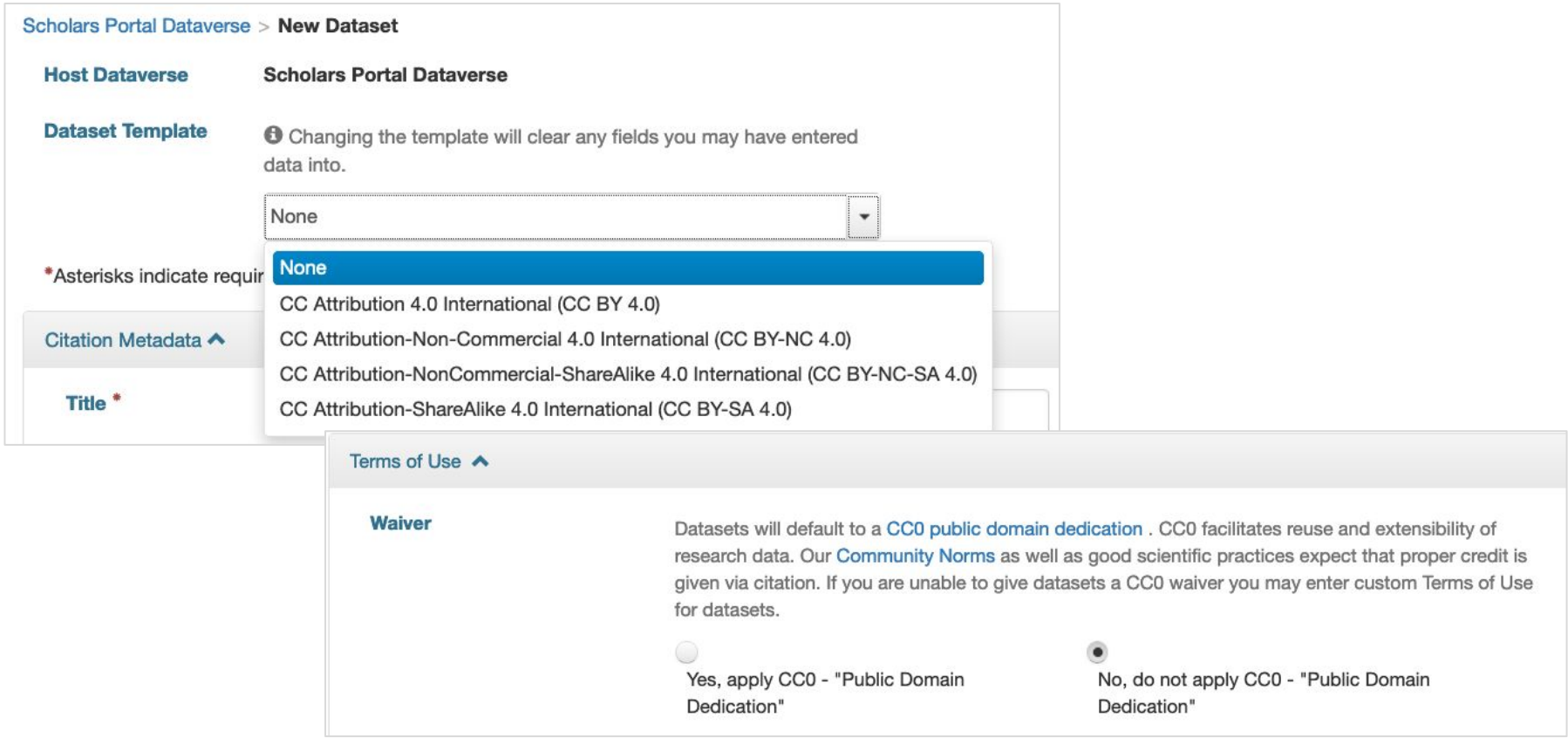
Data curation considerations

- Lowering barriers to access for researchers
- Building local capacity for reviewing datasets and mediation
- Strengthening Canadian Dataverse community

Dataverse features that enable FAIR

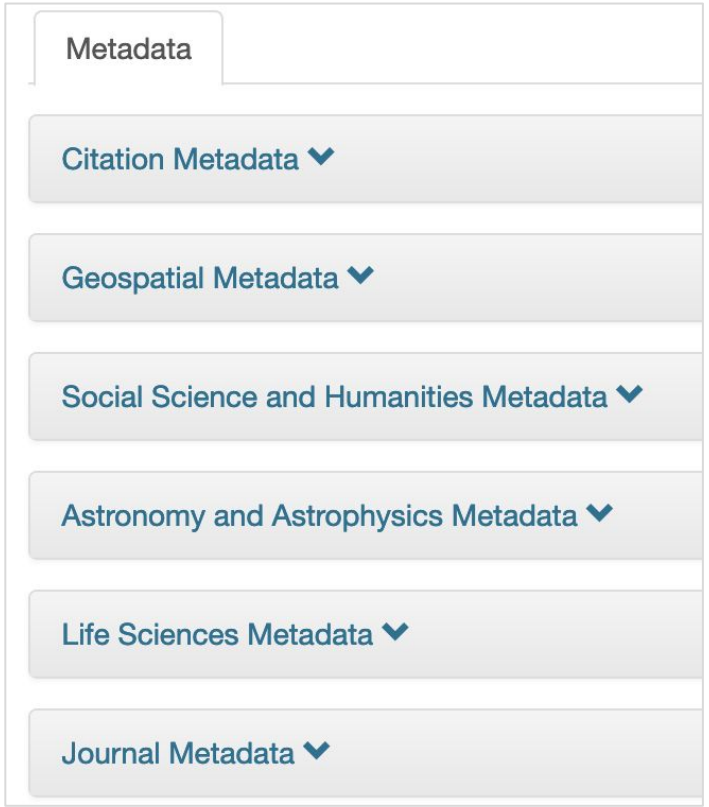
- Apply data reuse licence (CC0 default) + templates
- Rich metadata description possible:
  - Descriptive and citation metadata based on DDI standard (compliant with Dublin Core)
  - Optional discipline-specific metadata blocks (e.g., geospatial, social science & humanities, astronomy and astrophysics, life sciences)
  - Create links to other scholarly outputs
- User-friendly upload of data files, code, documentation
- Automatic data citation with DOI
- Automatic file format identification, file verification (checksums), tabular data transformation
- Dataset and file versioning
- OAI-PMH harvesting protocol; API access

Licences


Scholars Portal Dataverse > New Dataset
Host Dataverse
Scholars Portal Dataverse
Dataset Template
Changing the template will clear any fields you may have entered data into.
None
*Asterisks indicate requir
None
CC Attribution 4.0 International (CC BY 4.0)
CC Attribution-Non-Commercial 4.0 International (CC BY-NC 4.0)
CC Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0)
CC Attribution-ShareAlike 4.0 International (CC BY-SA 4.0)
Citation Metadata
Title *
Terms of Use
Waiver
Datasets will default to a CC0 public domain dedication . CC0 facilitates reuse and extensibility of research data. Our Community Norms as well as good scientific practices expect that proper credit is given via citation. If you are unable to give datasets a CC0 waiver you may enter custom Terms of Use for datasets.
Yes, apply CC0 - "Public Domain Dedication"
No, do not apply CC0 - "Public Domain Dedication"

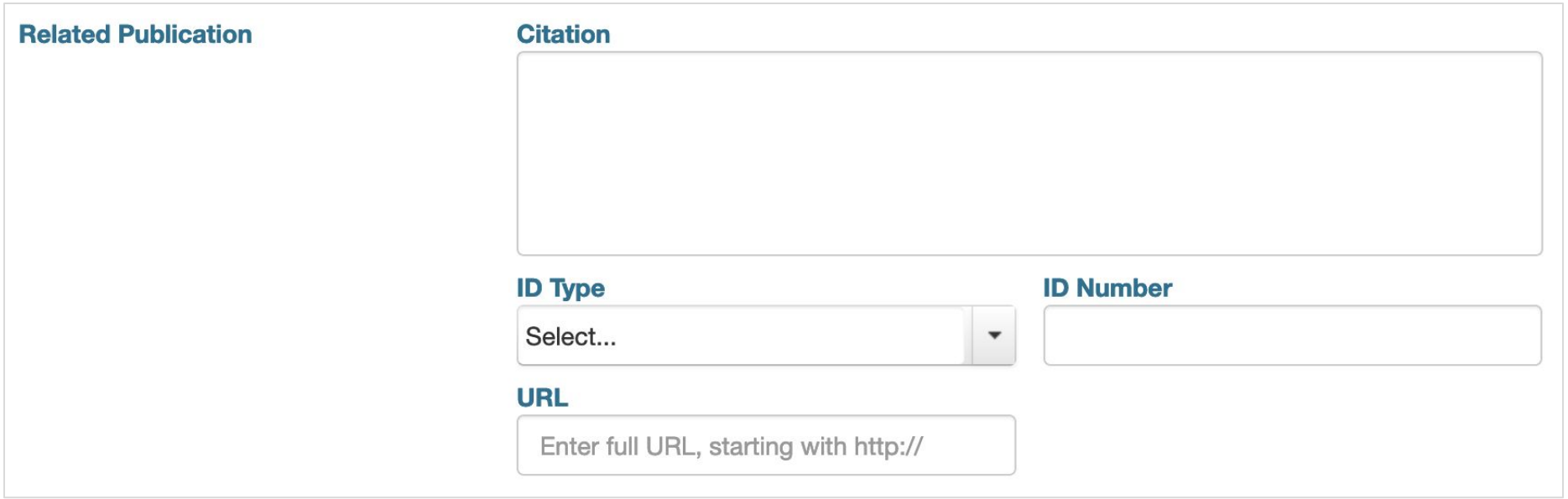
Metadata

- DDI metadata
- Maps to Dublin core and Datacite schema terms
- Metadata blocks for domain-specific metadata
- Dataset-level metadata is always publicly accessible



Metadata

Link to other scholarly outputs



File upload

Files

Upload with HTTP via your browser

File upload limit is 2.5 GB per file. Select files or drag and drop into the upload widget.

+ Select Files to Add

Drag and drop files here.

Select files from Dropbox.

Upload from Dropbox

File upload

2 Files

Edit

File Name logo.png

Preview

PNG Image
MD5: 38647c56b807b522684c89eeaa6f53ad;

Description Add file description...

Edit

File Name test_file.xlsx

MS Excel (XLSX)
MD5: 146972b470b46a6796dd6262f1d2623c;

Description Add file description...

Edit

Save Changes Cancel

File metadata

File Metadata

File UNF	UNF:6:JLfajwCp57jOpp/GXXZA3w==
Original File MD5	5768f01ce4981098138b355fddc30b68
Publication Date	2018-12-06
Size	415 B
Type	Tab-Delimited
Variables	4
Observations	6
Description	Summary Tables
Deposit Date	2018-12-06

File Metadata

MD5	4bccb2489fd7f32a799db12658e26f28
Publication Date	2018-12-06
Size	345.0 KB
Type	Adobe PDF
Description	Public Report
Deposit Date	2018-12-06

Publication

https://doi.org/10.5683/SP2/9MCJJH

Versioning

Files | Metadata | Terms | Versions

View Differences

	Dataset	Summary	Contributors	Published
☐	2.0	**Files (Added: 1; Changed File Metadata: 1);** View Details	Kaitlin Newson	Oct 12, 2019
☐	1.1	**Citation Metadata:** Subject (1 Added, 1 Changed); Description (1 Changed); **Additional Citation Metadata:** (3 Added); View Details	Kaitlin Newson	Oct 12, 2019
☐	1.0	This is the first published version.	Kaitlin Newson	Oct 12, 2019

Deaccessioning

Health regions: boundaries and correspondence with census geography, 2005 [Canada] [B2020, digital mapping files] Deaccessioned

Statistics Canada, 2016, "Health regions: boundaries and correspondence with census geography, 2005 [Canada] [B2020, digital mapping files]", https://hdl.handle.net/10864/12155, Scholars Portal Dataverse, V1, DEACCESSIONED VERSION

Deaccession Reason

The dataset has been transferred to another repository. Available at Scholars Portal Geoportal

The dataset can now be accessed at: http://geo.scholarsportal.info

Versions

Dataset	Summary	Contributors	Published
1.0	Deaccessioned Reason: The dataset has been transferred to another repository. Available at Scholars Portal Geoportal The dataset can now be accessed at: http://geo.scholarsportal.info	Alli McKinnon, Jiawei Chen	Dec 1, 2016

Interoperability

- Open Archives Initiative harvesting protocol (OAI-PMH)
  - Metadata from published, unrestricted datasets can be harvested
- Dataverse APIs
  - Search API
  - Data Access API
  - Native API
  - SWORD API (upload)

Data Curation Tool

- Funded by CANARIE RDM grant “Dataverse for the Canadian Research Community”
- Aim to improve support for data curation workflows
- External web application that connects to Dataverse to create and edit metadata at the variable level
- Works with files uploaded through tabular ingest process
- Blog post: Introducing the Data Curation Tool

	v1	v2	v3	v4	v5	v6	v7	v8	v9	v10	v11	v12	v13	v14	v15	v16
1	1.040e+14	4	26	104012	99	1	1973	2	1642	1040	1040	1	6	1	5	5
2	1.040e+14	1	19	104008	99	2	1943	1	99	99	99	1	1	6	4	5
3	1.040e+14	4	98	104003	99	1	1990	1	99	99	99	1	6	3	99	11
4	1.040e+14	2	98	104010	99	1	1983	2	1756	1040	1040	1	6	3	99	11
5	1.040e+14	1	18	104007	99	2	1927	1	99	99	99	1	1	6	4	1
6	1.040e+14	2	19	104007	99	1	1983	1	99	99	99	1	6	2	5	3
7	1.040e+14	4	15	104005	99	2	1970	1	99	99	99	1	1	2	4	6
8	1.040e+14	4	19	104006	99	1	1942	1	99	99	99	2	1	6	5	88
9	1.040e+14	4	15	104005	99	1	1965	2	1040	1276	1040	7	4	1	5	5
10	1.040e+14	7	15	104004	99	2	1955	1	99	99	99	1	1	2	5	5
11	1.040e+14	7	15	104003	99	2	7777	1	99	99	99	1	1	6	77	77
12	1.040e+14	4	18	104005	99	2	1938	1	99	99	99	1	3	6	5	5
13	1.040e+14	7	17	104005	99	1	1945	1	99	99	99	1	1	6	5	4
14	1.040e+14	4	18	104005	99	2	1949	2	1040	1380	1040	1	4	6	5	5
15	1.040e+14	2	15	104003	99	2	7777	1	99	99	99	1	1	2	4	5
16	1.040e+14	4	32	104007	99	1	1974	1	99	99	99	1	6	2	5	5
17	1.040e+14	4	98	104006	99	1	7777	1	99	99	99	1	6	3	99	11
18	1.040e+14	4	25	104010	99	2	1968	1	99	99	99	1	1	2	5	1
19	1.040e+14	4	40	104007	99	1	1967	1	99	99	99	1	1	2	6	2
20	1.040e+14	1	23	104004	99	1	1932	1	99	99	99	1	1	6	4	3
21	1.040e+14	2	18	104003	99	2	1965	1	99	99	99	1	6	2	5	5
22	1.040e+14	1	27	104011	99	1	1956	1	99	99	99	2	1	1	5	2
23	1.040e+14	4	18	104004	99	2	1923	1	99	99	99	1	3	6	3	3
24	1.040e+14	1	19	104006	99	2	1952	1	99	99	99	1	4	2	5	3
25	1.040e+14	3	16	104004	99	2	1947	2	1276	1040	1040	1	1	6	3	3

Data Curation Tool - Motivations

I-Rep Canada Poll 2007

Français

Ipsos Canada, 2017, "I-Rep Canada Poll 2007", https://doi.org/10.5683/SP/P2VFJR, Scholars Portal Dataverse, V1, UNF:6:mxzDqEjfg4jSF6aWUcpXQQ==

849 Results | Download

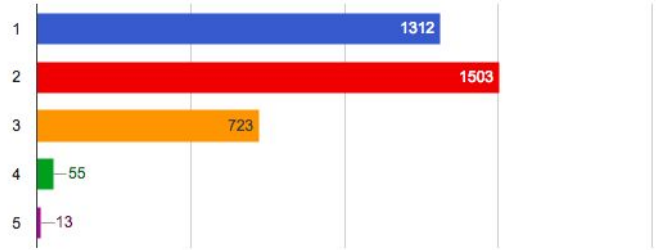
162158	q2_9	2.9) Please indicate how well you feel you know: MasterCard ?	5
162561	q2_10	2.10) Please indicate how well you feel you know: Petro-Canada ?	5
162335	q2_11	2.11) Please indicate how well you feel you know: Purolator Courier ?	5
162389	q2_12	2.12) Please indicate how well you feel you know: Shell ?	5
162625	**q2_13**	**2.13) Please indicate how well you feel you know: Tim Hortons ?**	**5**
162208	q2_14	2.14) Please indicate how well you feel you know: UPS (United Parcel Service) ?	5
161966	q2_15	2.15) Please indicate how well you feel you know: Visa ?	5

First « 3 4 5 6 7 » Last

Records Per Page 10

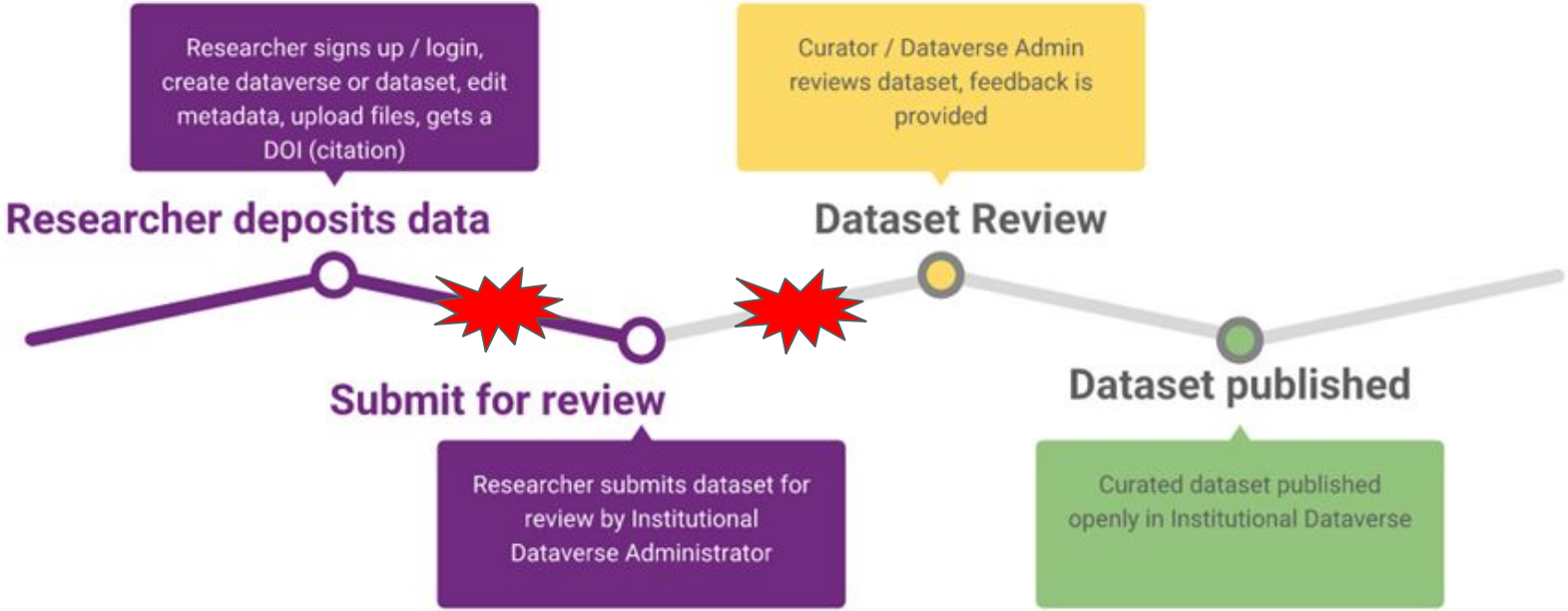
Chart View | Table View

Variable q2_13: 2.13) Please indicate how well you feel you know: Tim Hortons ?



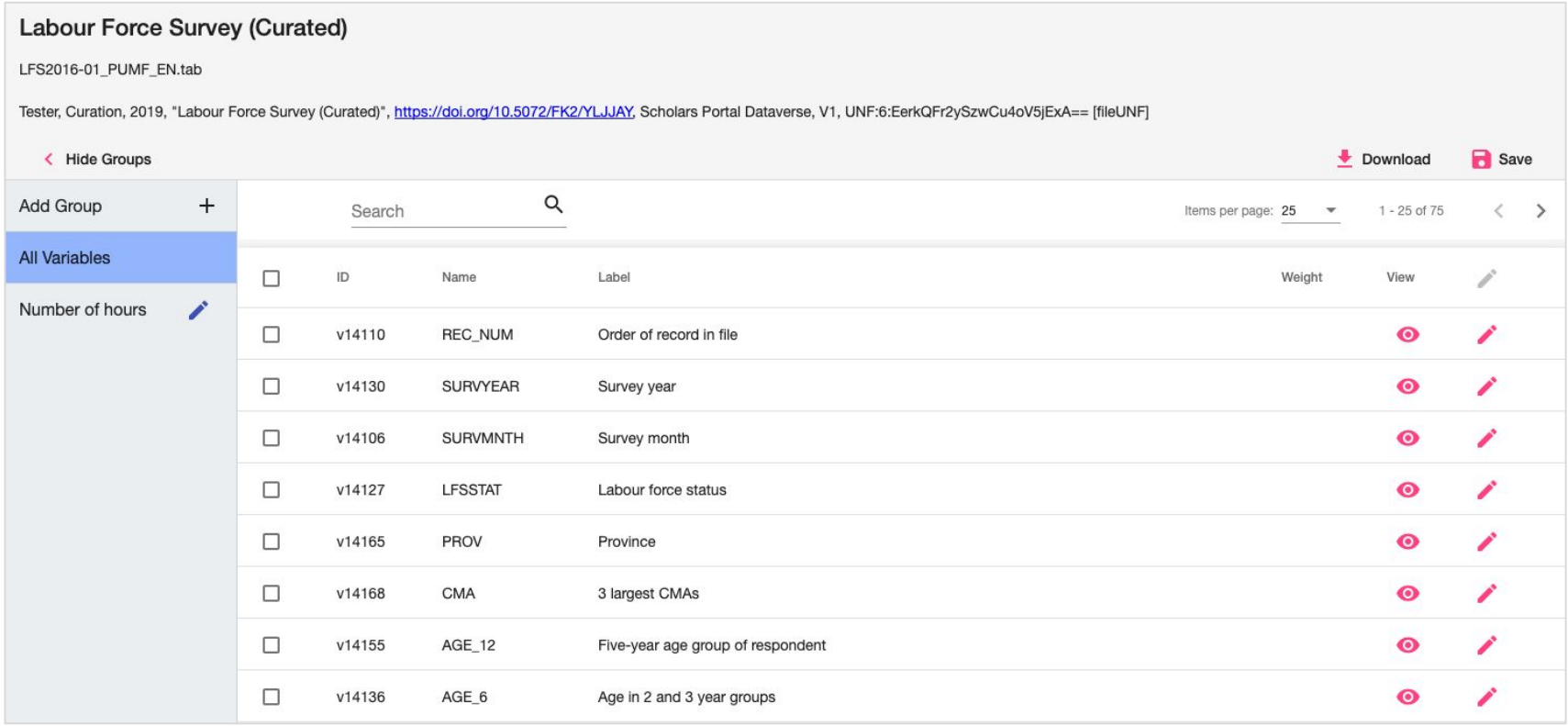
Values	Categories	N
1	1	1312
2	2	1503
3	3	723

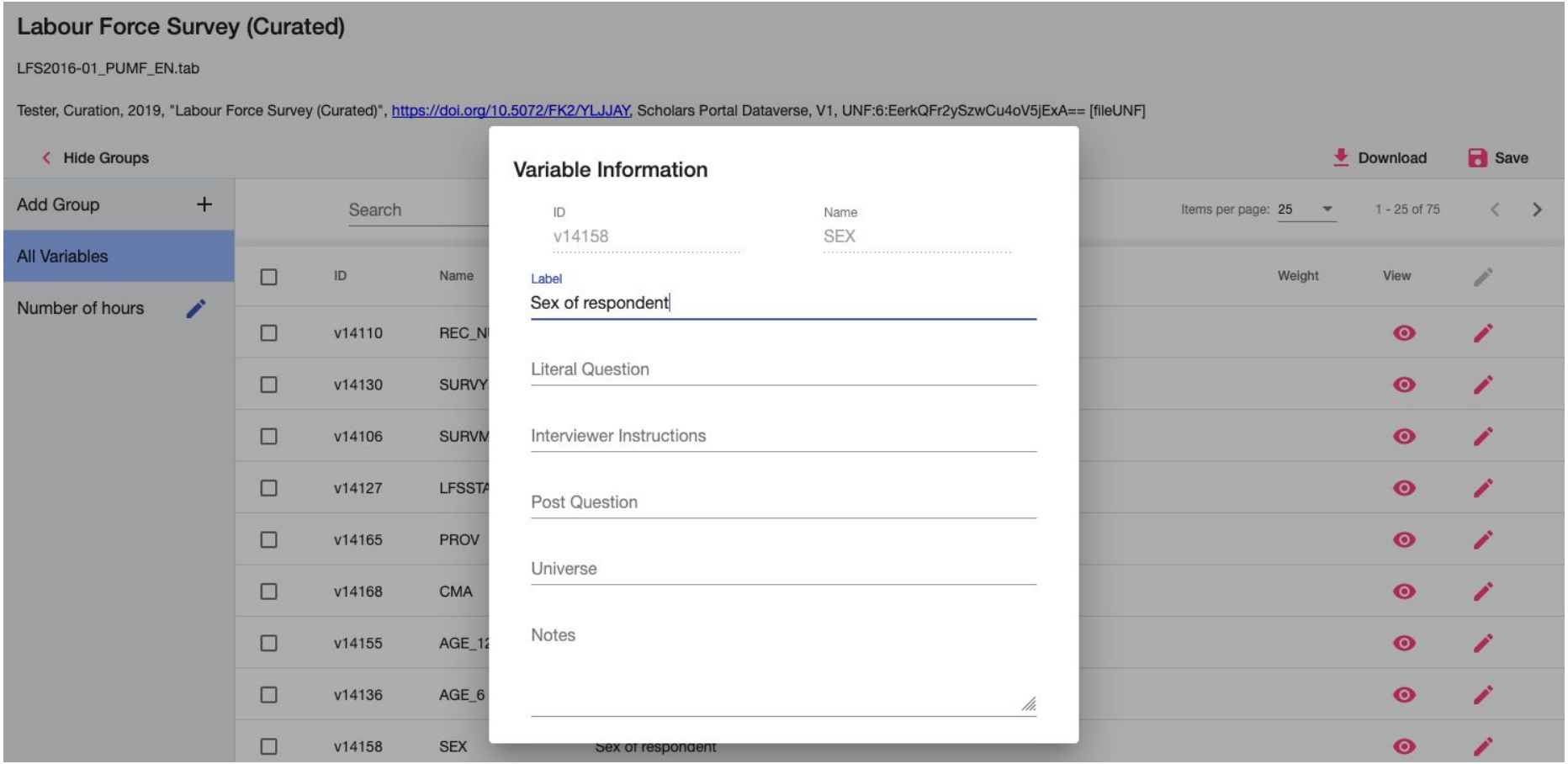
Curation Workflows



Data Curation Tool

Data Curation Tool




Labour Force Survey (Curated)
LFS2016-01_PUMF_EN.tab
Tester, Curation, 2019, "Labour Force Survey (Curated)", https://doi.org/10.5072/FK2/YLJJAY, Scholars Portal Dataverse, V1, UNF:6:EerkQFr2ySzwCu4oV5jExA== [fileUNF]
Hide Groups
Download
Save
Add Group
All Variables
Number of hours
Search
Items per page: 25
1 - 25 of 75
ID
Name
Weight
View
v14110
REC_N
v14130
SURVY
v14106
SURVM
v14127
LFSSTA
v14165
PROV
v14168
CMA
v14155
AGE_12
v14136
AGE_6
v14158
SEX
Sex of respondent
Variable Information
ID
v14158
Name
SEX
Label
Sex of respondent
Literal Question
Interviewer Instructions
Post Question
Universe
Notes

Labour Force Survey (Curated)

LFS2016-01_PUMF_EN.tab

Tester, Curation, 2019, "Labour Force Survey (Curated)", https://doi.org/10.5072/FK2/YLJJAY, Scholars Portal Dataverse, V1, UNF:6:EerkQFr2ySzwCu4oV5jExA== [fileUNF]

Hide Groups | Download | Save

Add Group

All Variables

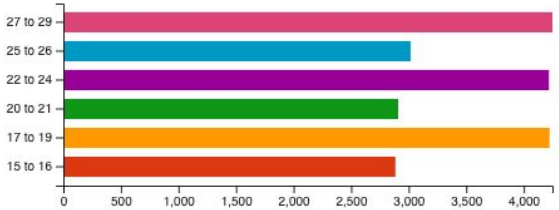
Number of hours

Search

Items per page: 25 | 1 - 25 of 75

	ID	Name		Weight	View	
☐	v14110	REC_N				
☐	v14130	SURVY				
☐	v14106	SURVM				
☐	v14127	LFSSTA				
☐	v14165	PROV				
☐	v14168	CMA				
☐	v14155	AGE_12				
☐	v14136	AGE_6	Age in 2 and 3 year groups			
☐	v14158	SEX	Sex of respondent			

AGE_6: Age in 2 and 3 year groups



Values	Categories	Count	Count Percentage(%)	Weighted Count
1	15 to 16	2,885	13.417	
2	17 to 19	4,224	19.644	
3	20 to 21	2,909	13.528	
4	22 to 24	4,219	19.621	
5	25 to 26	3,017	14.031	
6	27 to 29	4,249	19.76	

Data Curation Tool Codebook

Data Curation Tool Testing Dataset (ICPSR doi:10.5072/FK2/0TYIHL)
(DCT Testing Dataset)

View: Part 1: Document Description
Part 2: Study Description
Part 3: Data Files Description
Part 4: Variable Description
Part 5: Other Study-Related Materials
Entire Codebook

Document Description

Citation

Title:	Data Curation Tool Testing Dataset
Identification Number:	doi:10.5072/FK2/0TYIHL
Distributor:	Scholars Portal Dataverse
Date of Distribution:	2019-06-14
Version:	2
Bibliographic Citation:	Lubitch, Victoria; Leahey, Amber, 2019, "Data Curation Tool Testing Dataset", https://

Study Description

Citation

Title:	Data Curation Tool Testing Dataset
Subtitle:	DDI Test
Alternative Title:	DCT Testing Dataset
Identification Number:	doi:10.5072/FK2/0TYIHL
Authoring Entity:	Lubitch, Victoria (University of Toronto)
	Leahey, Amber (Scholars Portal)
Producer:	Leahey, Amber
Date of Production:	2019-05-22
Grant Number:	4445555
Distributor:	Scholars Portal Dataverse
Date of Distribution:	2019-06-14

Study Scope

Keywords:	Astronomy and Astrophysics, test, smoking
Topic Classification:	Metadata

Try it out!

bit.ly/2OD4zkO

Code on GitHub:
https://github.com/scholarsportal/Dataverse-Data-Curation-Tool

Available in Scholars Portal Dataverse October 31st.

Exercise -- Research Scenarios

In groups of 6, you will be assigned one of the following research scenarios, work through each of the guiding questions, and report back to the larger group (15 minutes):

- Scenario 1: The Super Radar Network
- Scenario 2: The International Alcohol Consumption Study
- Scenario 3: 3D Models of Roman Artefacts
- Scenario 4: Coastal Community Underwater Observatory
- Scenario 5: Indigenous Ocean Data

Guiding Questions

1. What are the main FAIR considerations in this scenario?
2. Where should the data be deposited? Explain rationale and possible issues.
3. What licensing should they apply? Justify decision.
4. Which metadata schema is most suitable for the dataset and why?
5. Are there any major metadata issues in the scenario (e.g., conflicting or missing information) that require clarity?
6. What information is needed to support reuse? Provenance?
7. Are there any potential file format issues that could affect reuse?
8. Are there potential sensitivities or legal issues that could limit sharing the data openly?

Discussion

Wrap up

Feedback form: http://bit.ly/CDCF-training-feedback

Get in touch:

- Dataverse team - dataverse@scholarsportal.info
- Lee Wilson - lee.wilson@ace-net.ca
- Reyna Jenkyns - reyna@uvic.ca / reyna@oceannetworks.ca
- Brad Covey - coveyb@ogsl.ca

Further Reading

- ARDC. (2019). Research Data Rights Management Guide. https://ardc.edu.au/wp-content/uploads/2019/09/Research_Data_Rights_Management_Guide.pdf
- Ball, A. How to License Research Data. Digital Curation Centre. http://www.dcc.ac.uk/resources/how-guides/license-research-data
- Barbosa, S. (2019). Dataverse and Curation Training. *Dataverse Community Meeting, 2019.* https://mfr.osf.io/render?url=https://osf.io/8q59p/?action=download%26mode=render
- Creative Commons. (2019). Creative Commons Certificate. https://certificates.creativecommons.org/
- ESIP Data Preservation and Stewardship Committee. 2019. Data Citation Guidelines for Earth Science Data. Ver. 2. Earth Science Information Partners. https://doi.org/10.6084/m9.figshare.8441816
- Lin.D, et al (2019). The TRUST Principles for Digital Repositories (v2) - https://bitly.com/trustprinciples
- Metadata Standards Catalog. https://rdamsc.bath.ac.uk/
- Portage. (2019). Repository Options in Canada: A Portage Guide. https://portagenetwork.ca/wp-content/uploads/2019/06/Repository-Options-Guide_EN.pdf

Further Reading

- Research Data Alliance. (2019). Metadata Standards Directory. http://rd-alliance.github.io/metadata-directory/
- Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data. https://doi.org/10.1038/sdata.2016.18