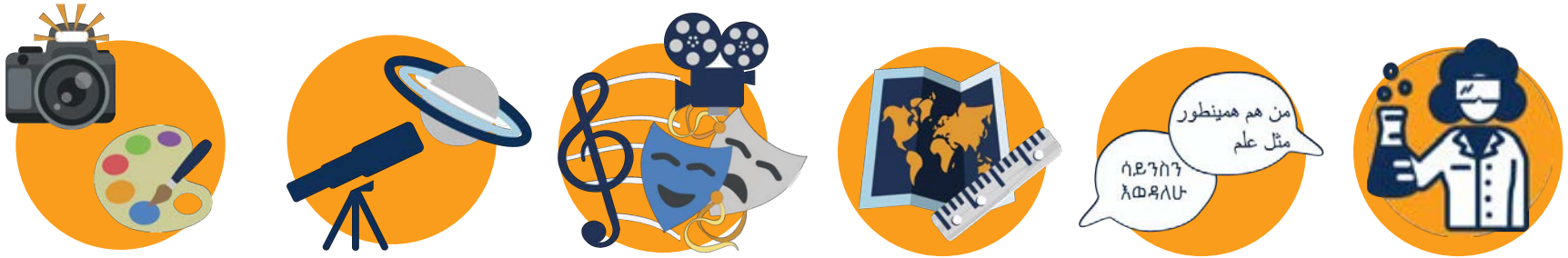
World Data System • International Technology Office



Dataset Discovery and Metadata

What's in it for researchers?

Alicia Urquidi Díaz, MLIS, PhD (ella/elle/she/sie) • aurquidi@oceannetworks.ca
WDS-ITO • Portage Webinar Series • May 19, 2021

Vancouver is on the traditional, ancestral and unceded territories of the **xʷməθkʷəy̓əm** (Musqueam), **Sḵwx̱wú7mesh** (Squamish), and **Selílwitulh** (Tsleil-Waututh) Nations.

James Bay (Victoria, BC) is on the traditional territory of the **Lekwungen**-speaking People who are known today as the **Songhees**.

To this day, the **Songhees**, **Esquimalt** and **W̱SÁNEĆ** nations maintain relationships with the land occupied by the city of Victoria and the University of Victoria, just as the Musqueam, Squamish and Tsleil-Waututh do with the city of Vancouver.

For information on these and
other Indigenous territories, please visit:
Native Land Digital Map

Today's Webinar

1. ## Metadata & Discovery
   Researchers, metadata, and discovery infrastructures.

2. ## Bringing Data Together
   An insider look into FRDR's Discovery Service
   Kelly Stathis, Discovery & Metadata Coordinator, Portage Network

3. ## Tips & tools from WDS-ITO

Discovery & Metadata

Data discovery

The ability to derive (or discover) new information and knowledge from existing data sources.

Open Metadata for Research Data Discovery in Canada (A. Garnett, A. Leahey, D. Savard, B. Towell & L. Wilson: 2017)

Metadata

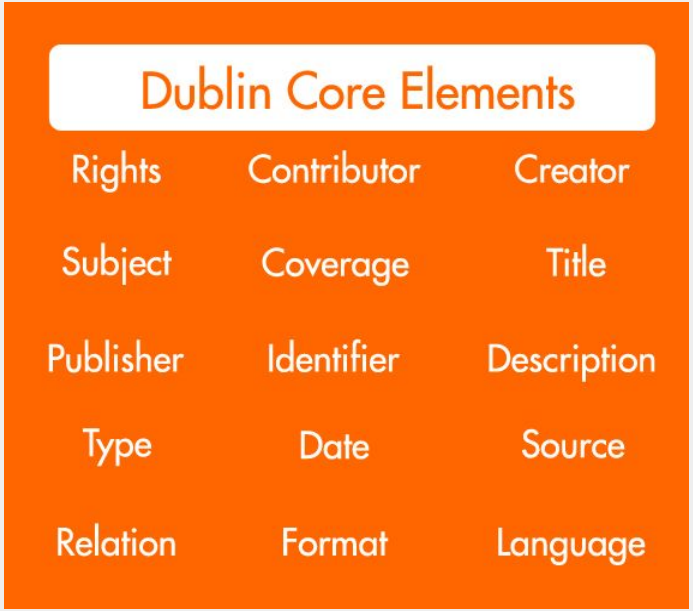
The information we create, store, and share to describe things, [which] allows us to interact with these things to obtain the knowledge we need.

Understanding Metadata (Riley & Niso: 2017)

Descriptive metadata

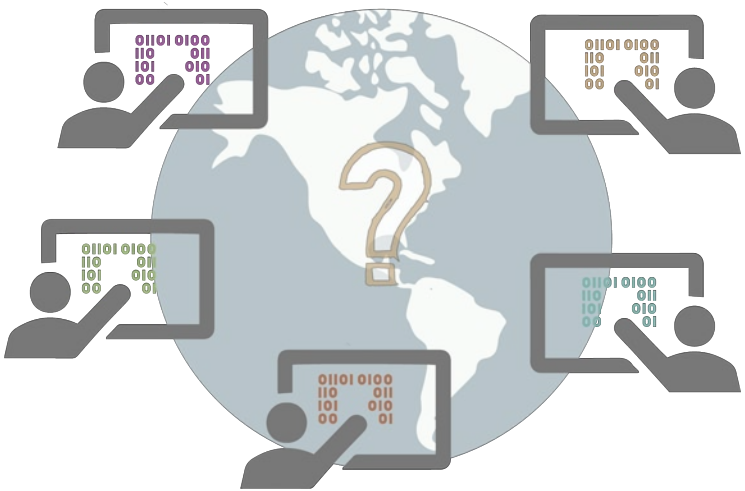
Information to facilitate the discovery (via search or browse) of resources, or provide contextual information useful in the understanding or interpretation of a resource.

Understanding Metadata (Riley & Niso: 2017)



Core elements of the Dublin Core metadata standard. (Image source)

Descriptive Metadata

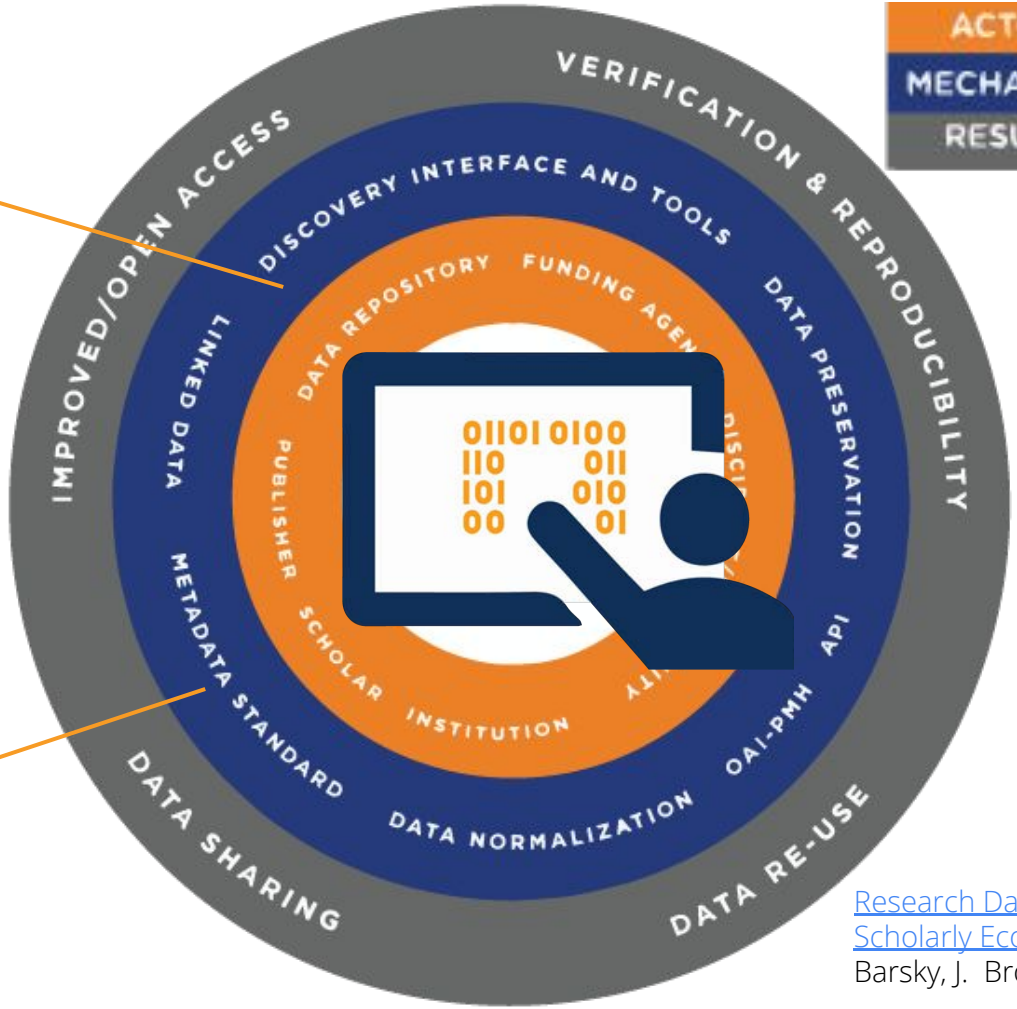


The potential for **reusing research data** is inextricably tied to how discoverable these data are to other researchers

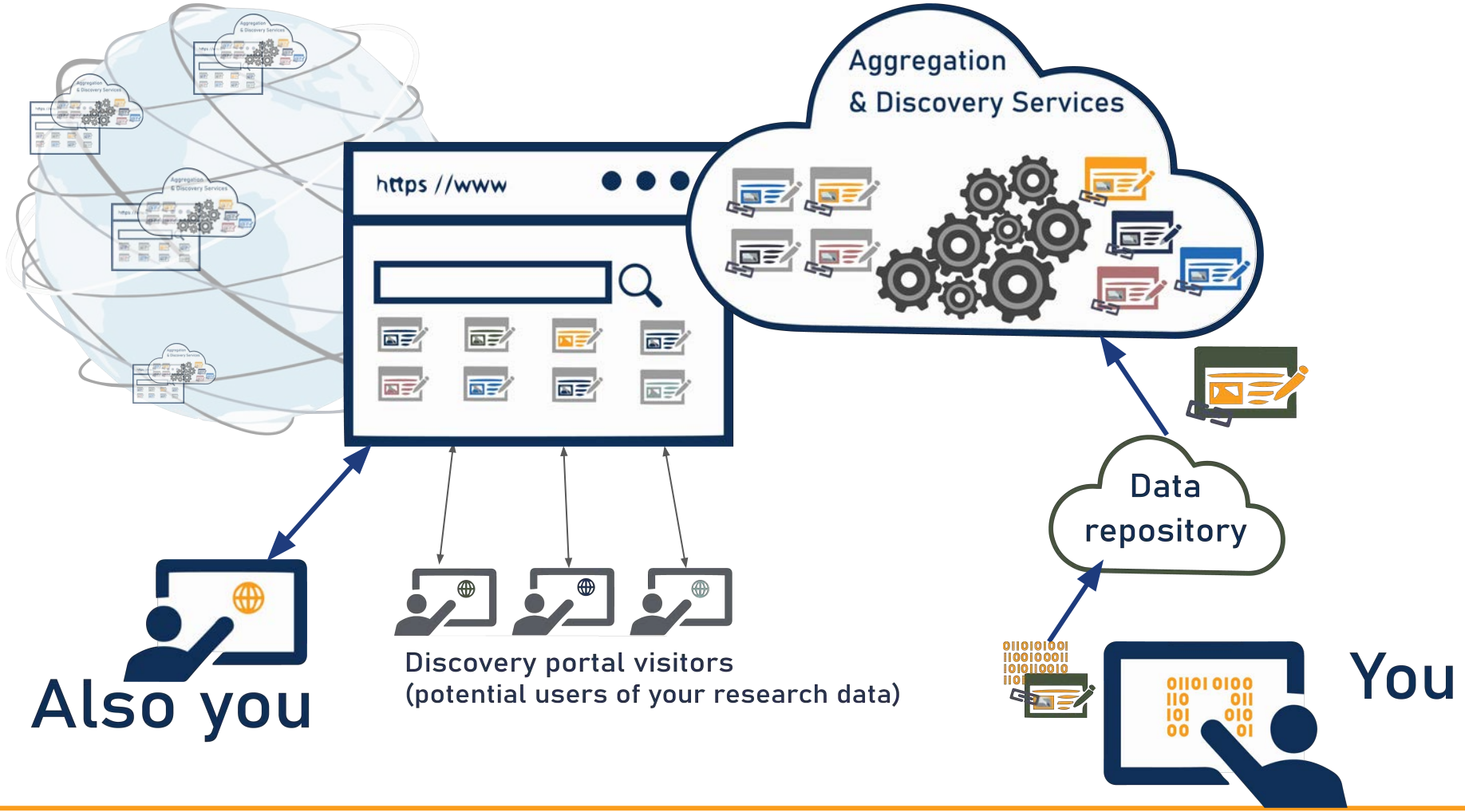
Open Metadata for Research Data Discovery in Canada (A. Garnett, A. Leahey, D. Savard, B. Towell & L. Wilson: 2017)

Discovery Infrastructures

Metadata



[Research Data Discovery and the Scholarly Ecosystem in Canada](#) (E. Barsky, J. Brosz & A. Leahey: 2016)


Aggregation
& Discovery Services
https //www
Data
repository
Also you
Discovery portal visitors
(potential users of your research data)
You

Bringing Data Together

An insider look into FRDR's Discovery Service

Kelly Stathis, Discovery & Metadata Coordinator, Portage Network

Bringing Canadian Research Data Together: The FRDR Discovery Service

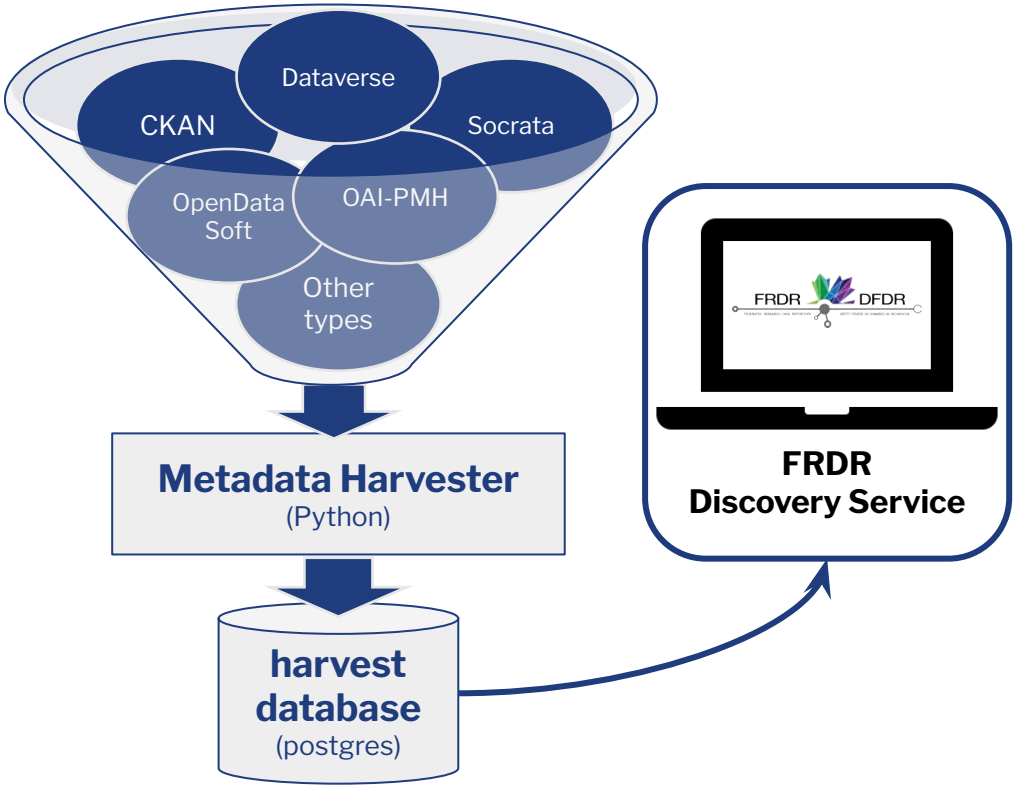
Kelly Stathis (they/elle)
Discovery & Metadata Coordinator, Portage Network
kelly.stathis@engagedri.ca
May 19, 2021

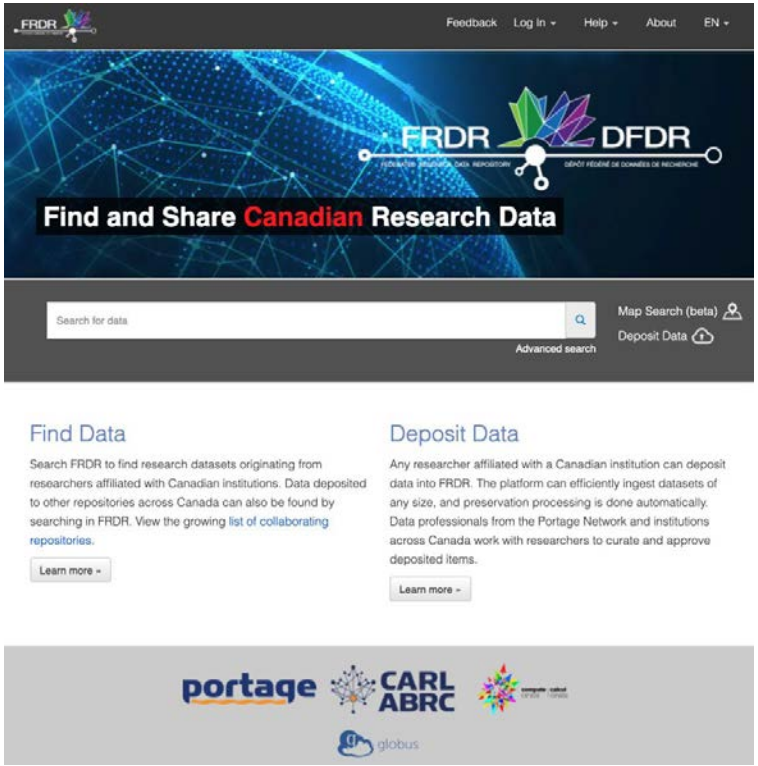
What is the FRDR Discovery Service?

- Federated Research Data Repository (FRDR) / Dépôt fédéré de données de recherche (DFDR)
- Collaboration between the New Digital Research Infrastructure Organization's Portage Network (NDRIO Portage) and Compute Canada
- FRDR Discovery Service:
  - One of FRDR's three components (discovery, deposit, and preservation)
  - A national discovery layer indexing Canadian research data repositories

https://www.frdr-dfdr.ca/

What is the FRDR Discovery Service?

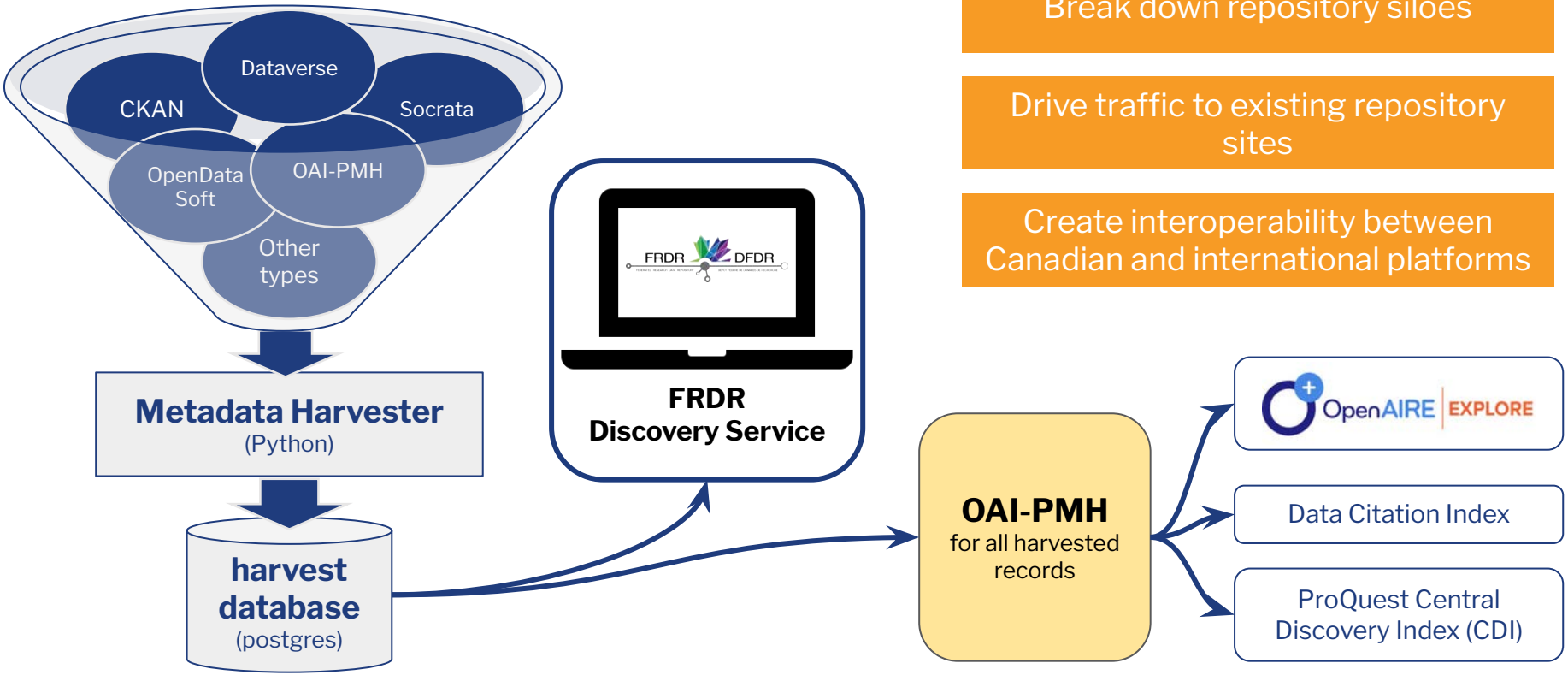




https://www.frdr-dfdr.ca/

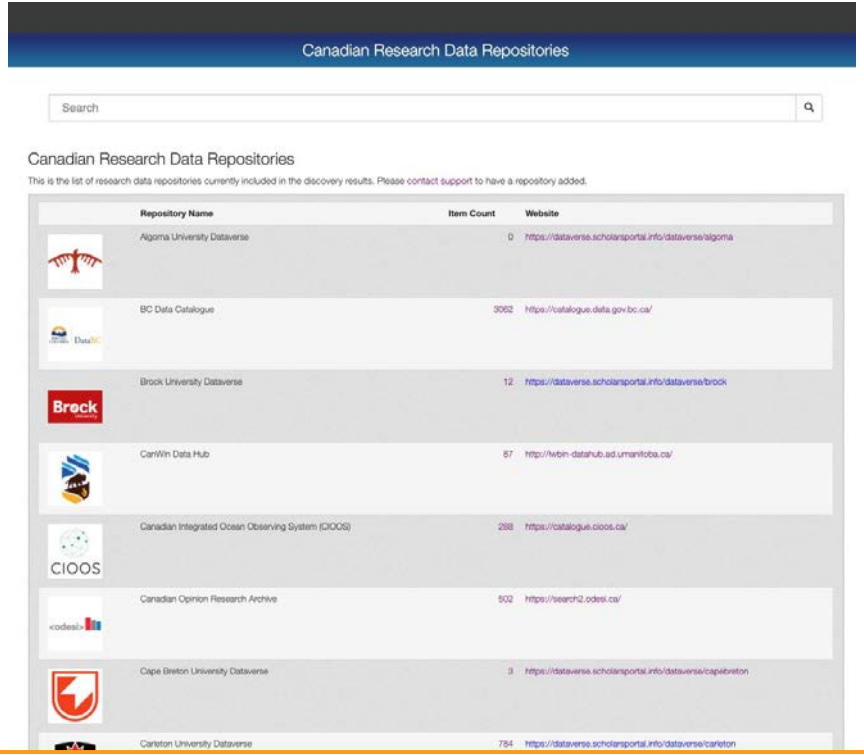
What are the benefits?

- Improve discovery of Canadian research (meta)data
- Break down repository siloes
- Drive traffic to existing repository sites
- Create interoperability between Canadian and international platforms

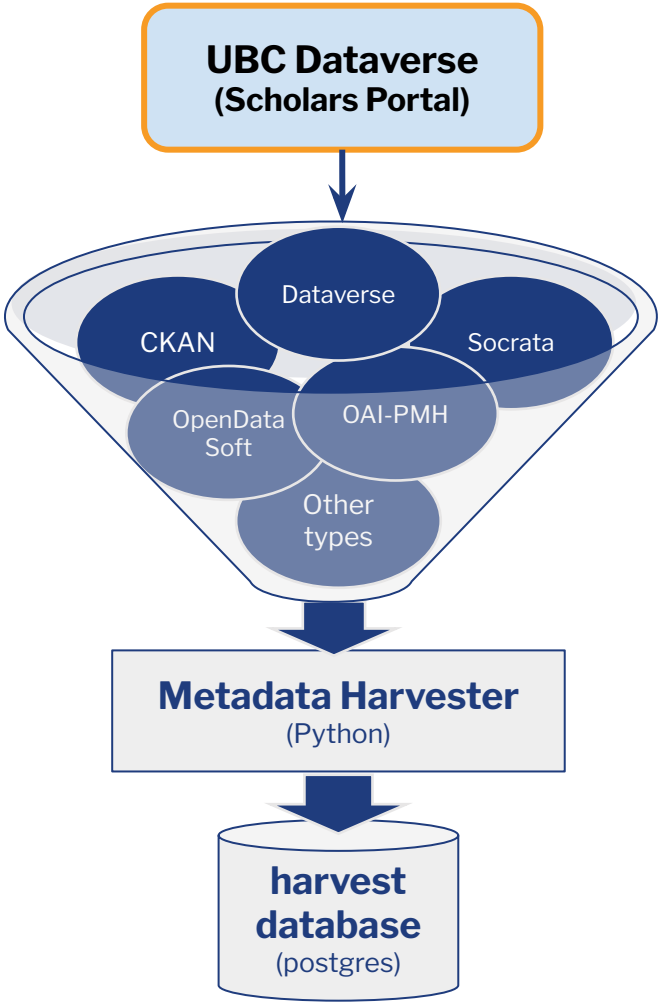


Collaborating repositories

- Over 80 discoverable research data repositories including:
  - **University** repositories, including Scholars Portal Dataverses
  - **Government** repositories at the federal, provincial, and local levels
  - **Domain-specific** repositories
  - Datasets deposited in **FRDR**

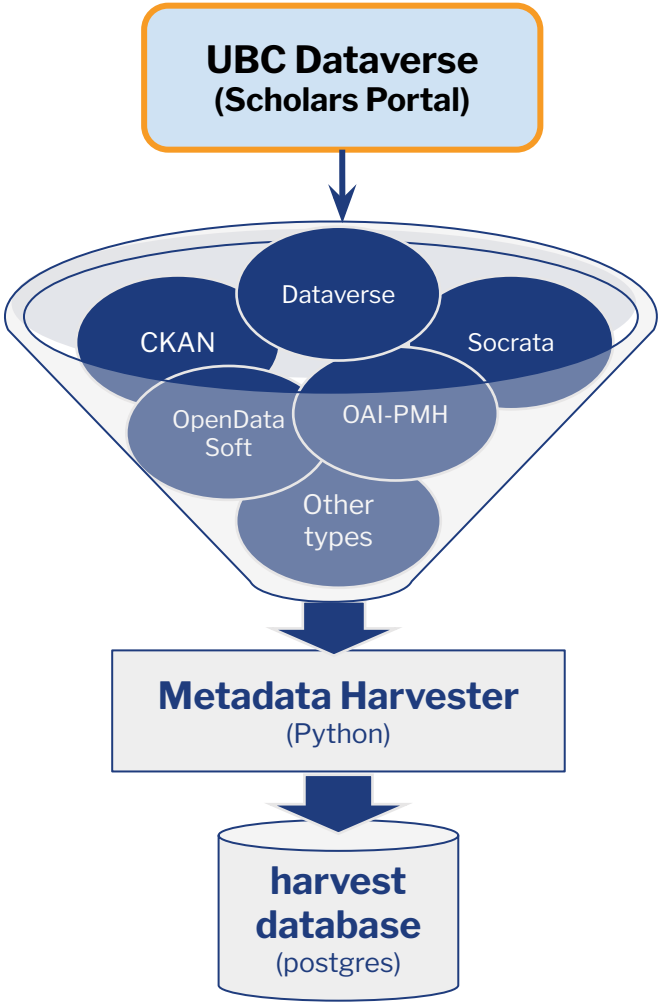


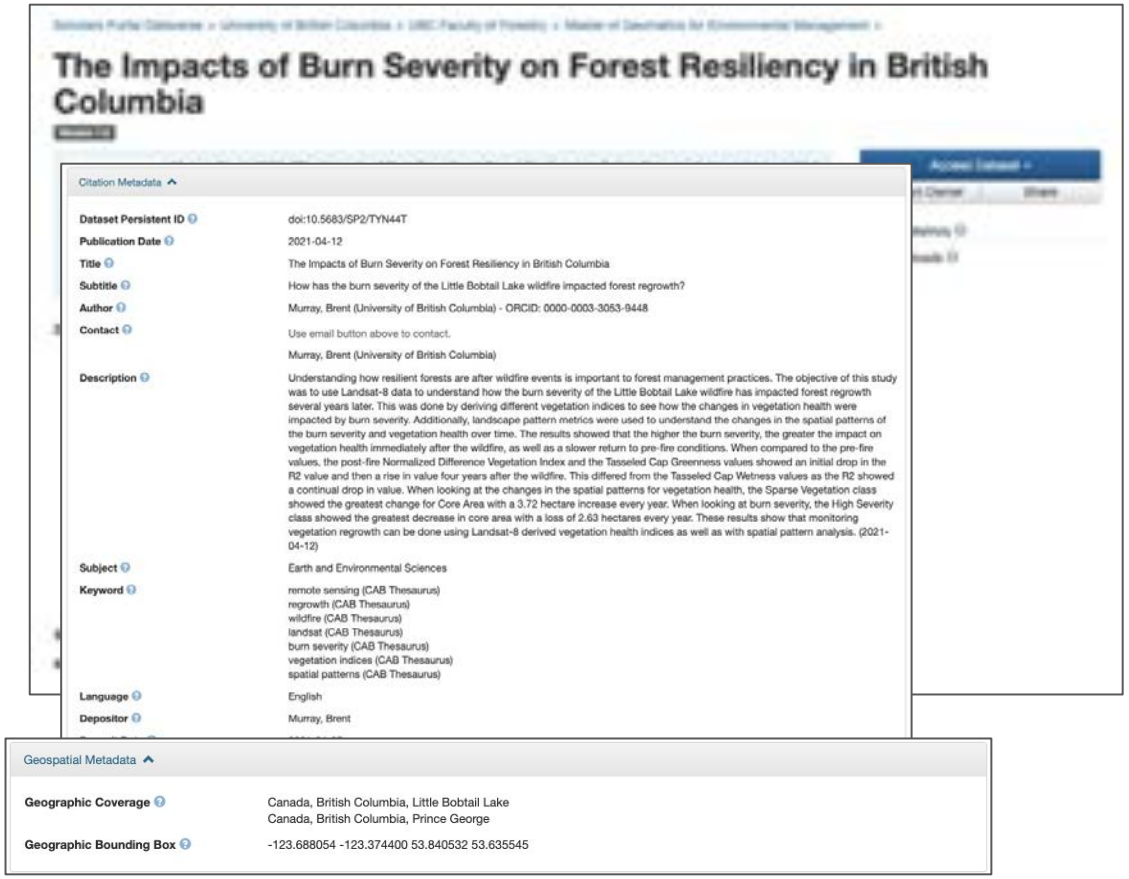
https://www.frdr-dfdr.ca/discover/html/repository-list.html



The Impacts of Burn Severity on Forest Resiliency in British Columbia

https://doi.org/10.5683/SP2/TYN44T





The Impacts of Burn Severity on Forest Resiliency in British Columbia

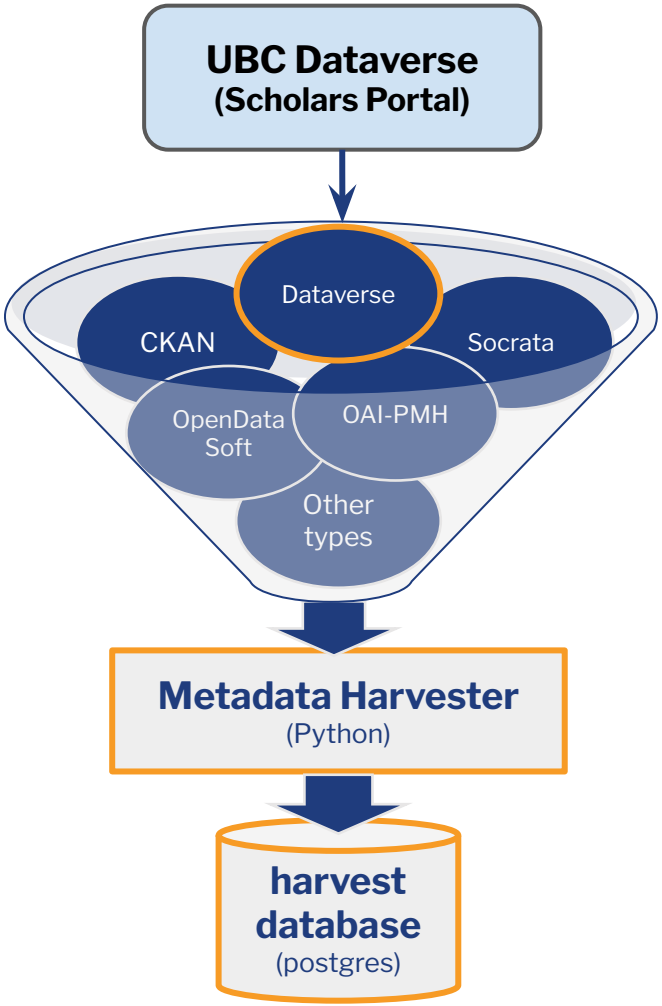
Citation Metadata

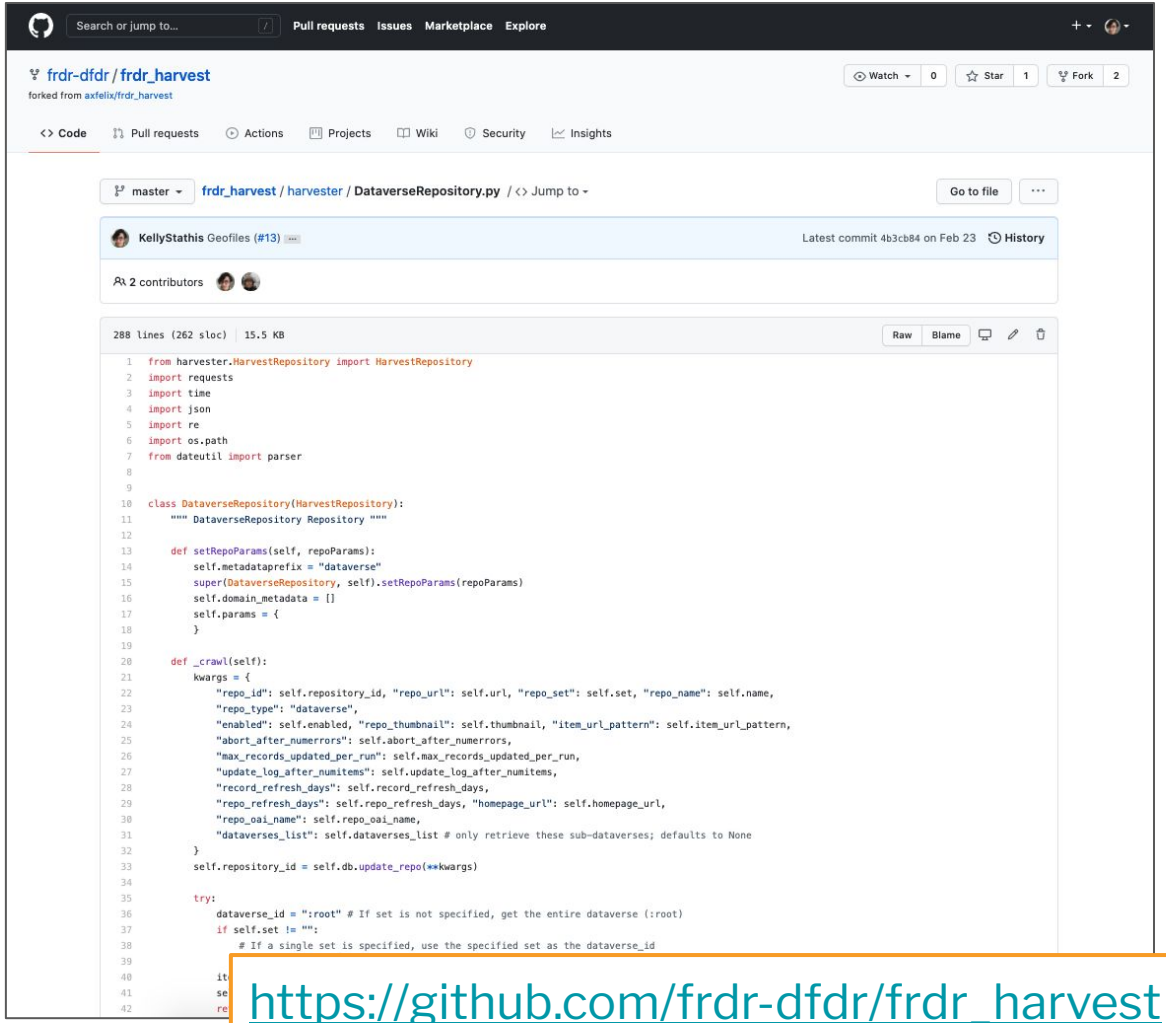
Dataset Persistent ID	doi:10.5683/SP2/TYN44T
Publication Date	2021-04-12
Title	The Impacts of Burn Severity on Forest Resiliency in British Columbia
Subtitle	How has the burn severity of the Little Bobtail Lake wildfire impacted forest regrowth?
Author	Murray, Brent (University of British Columbia) - ORCID: 0000-0003-3053-9448
Contact	Use email button above to contact. Murray, Brent (University of British Columbia)
Description	Understanding how resilient forests are after wildfire events is important to forest management practices. The objective of this study was to use Landsat-8 data to understand how the burn severity of the Little Bobtail Lake wildfire has impacted forest regrowth several years later. This was done by deriving different vegetation indices to see how the changes in vegetation health were impacted by burn severity. Additionally, landscape pattern metrics were used to understand the changes in the spatial patterns of the burn severity and vegetation health over time. The results showed that the higher the burn severity, the greater the impact on vegetation health immediately after the wildfire, as well as a slower return to pre-fire conditions. When compared to the pre-fire values, the post-fire Normalized Difference Vegetation Index and the Tasseled Cap Greenness values showed an initial drop in the R2 value and then a rise in value four years after the wildfire. This differed from the Tasseled Cap Wetness values as the R2 showed a continual drop in value. When looking at the changes in the spatial patterns for vegetation health, the Sparse Vegetation class showed the greatest change for Core Area with a 3.72 hectare increase every year. When looking at burn severity, the High Severity class showed the greatest decrease in core area with a loss of 2.63 hectares every year. These results show that monitoring vegetation regrowth can be done using Landsat-8 derived vegetation health indices as well as with spatial pattern analysis. (2021-04-12)
Subject	Earth and Environmental Sciences
Keyword	remote sensing (CAB Thesaurus) regrowth (CAB Thesaurus) wildfire (CAB Thesaurus) landsat (CAB Thesaurus) burn severity (CAB Thesaurus) vegetation indices (CAB Thesaurus) spatial patterns (CAB Thesaurus)
Language	English
Depositor	Murray, Brent

Geospatial Metadata

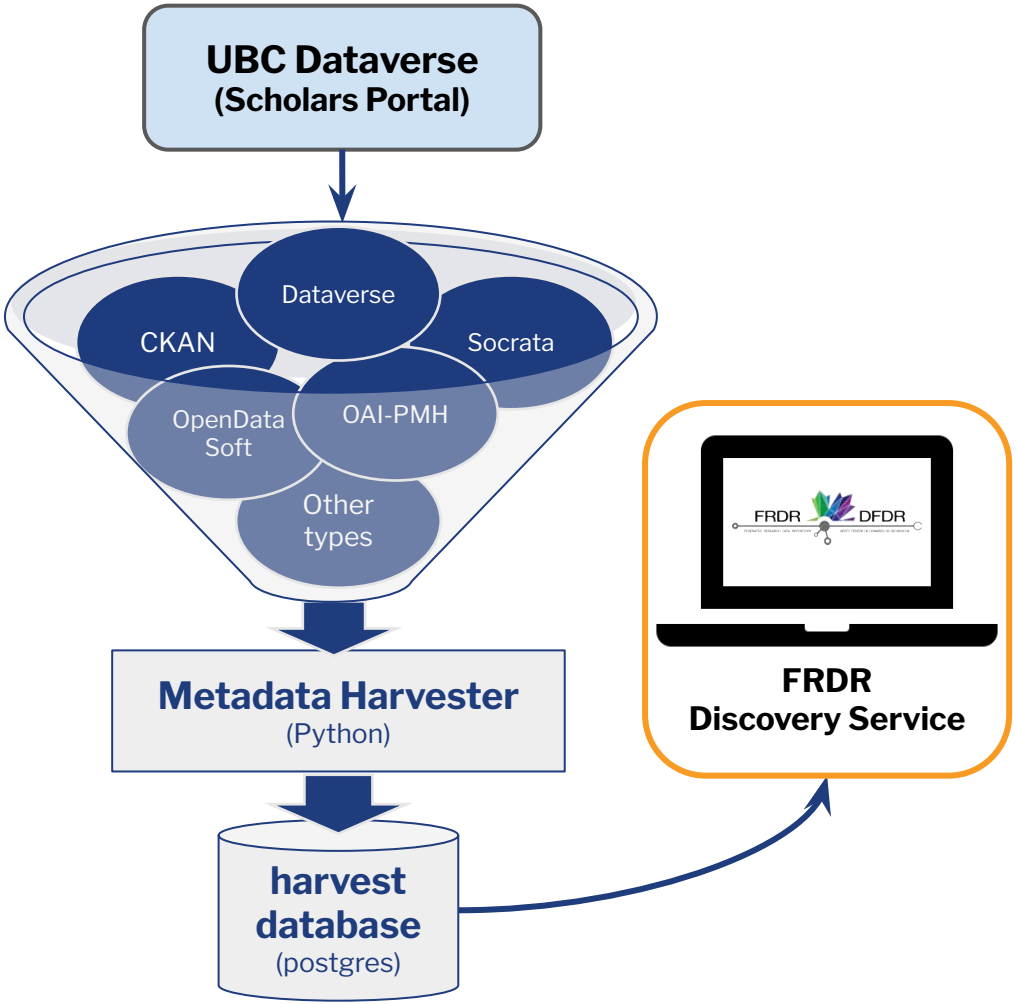
Geographic Coverage	Canada, British Columbia, Little Bobtail Lake Canada, British Columbia, Prince George
Geographic Bounding Box	-123.688054 -123.374400 53.840532 53.635545

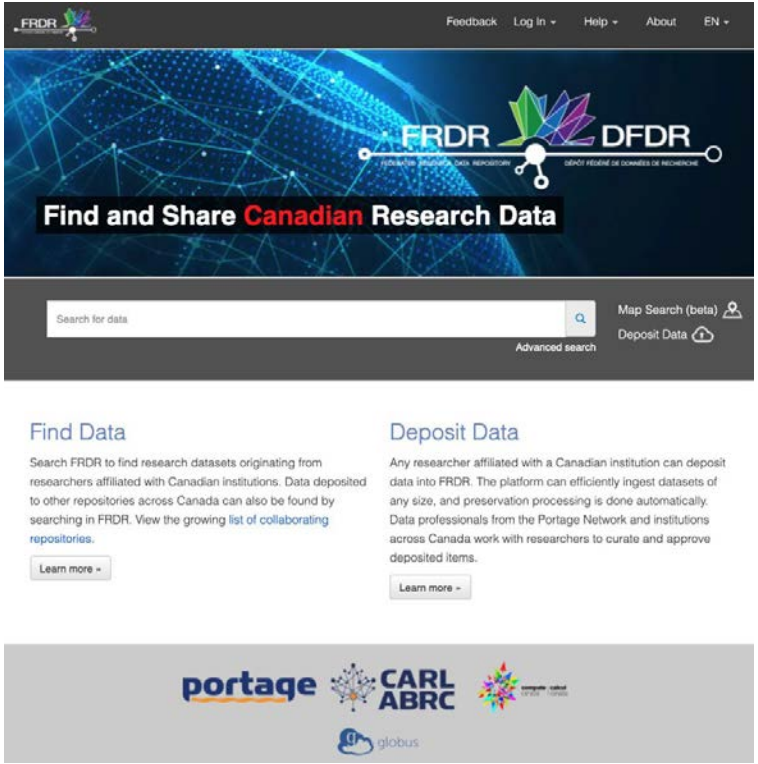
https://doi.org/10.5683/SP2/TYN44T



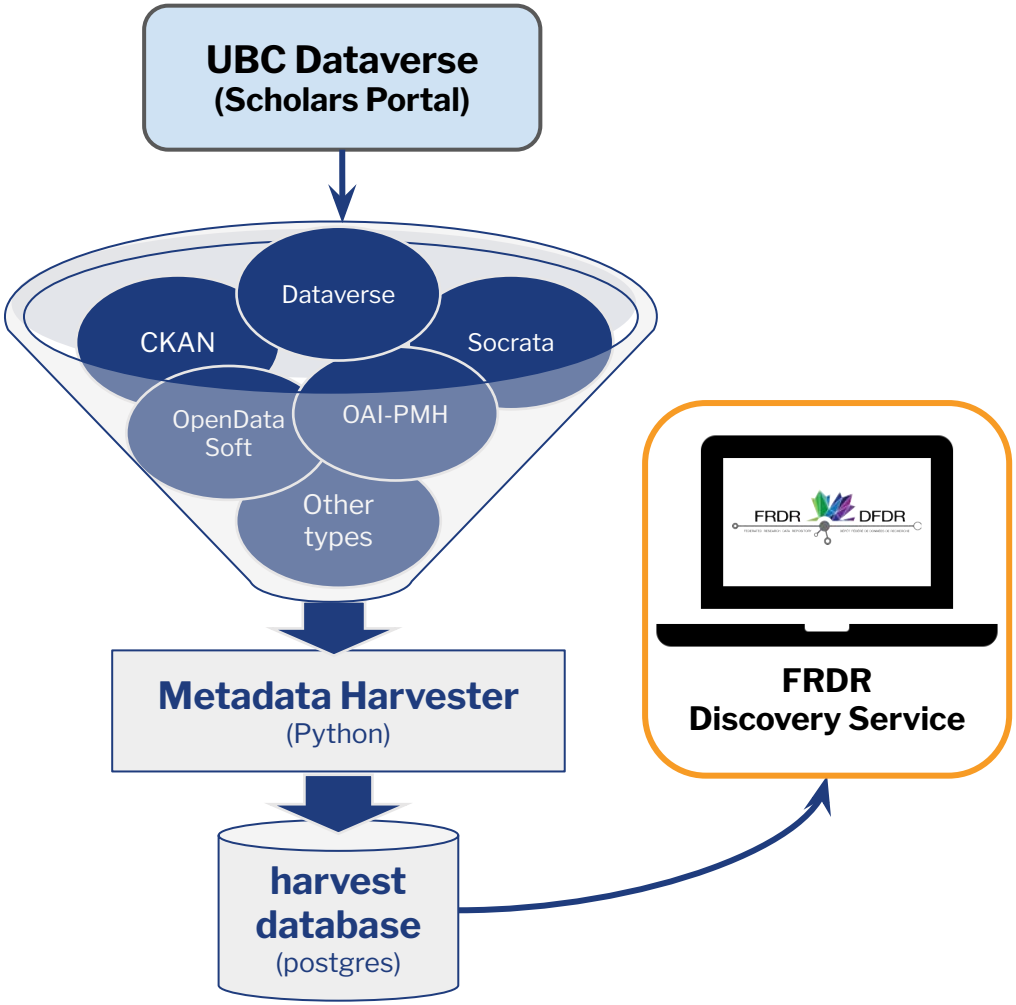


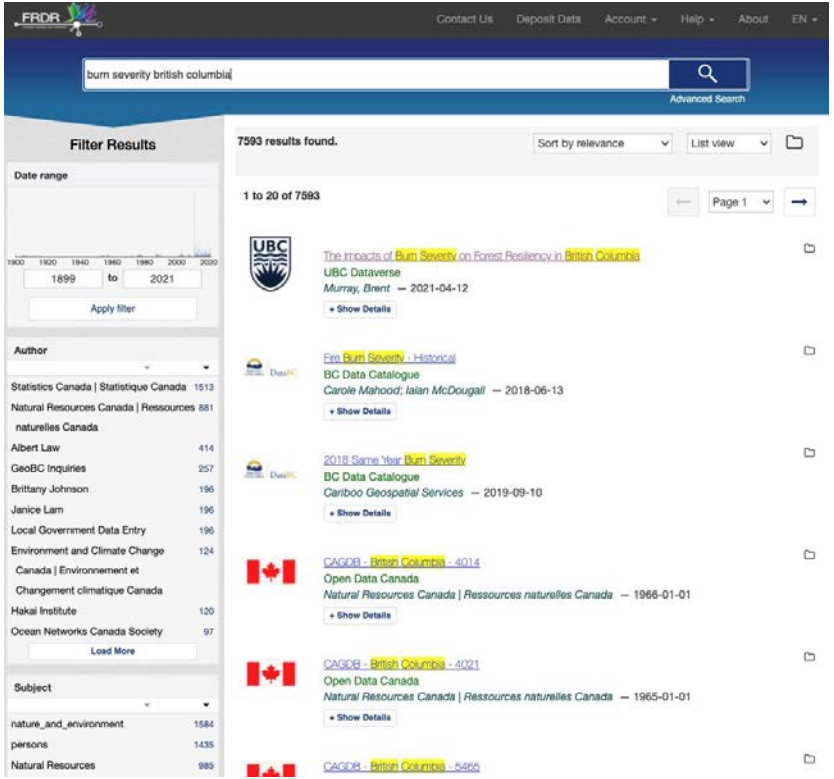
https://github.com/frdr-dfdr/frdr_harvest



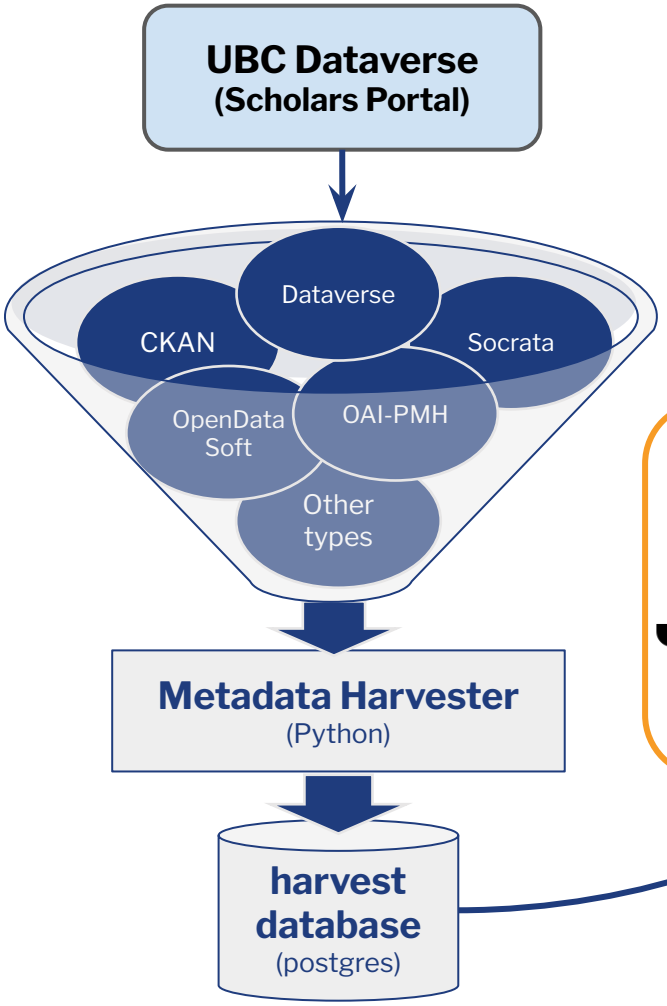


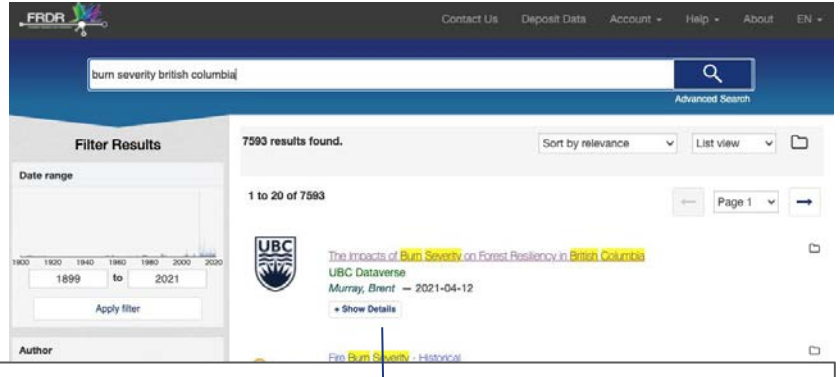
https://www.frdr-dfdr.ca/





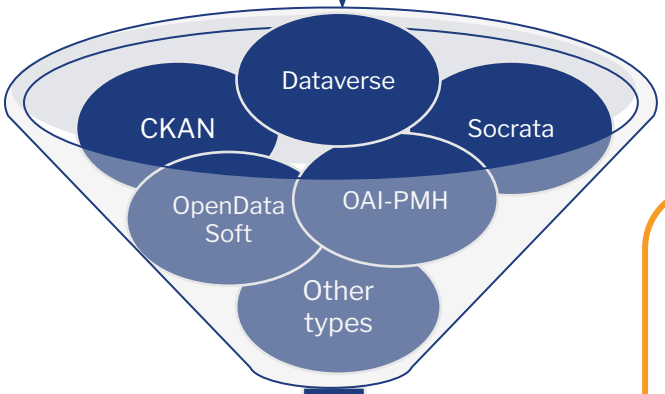
https://www.frdr-dfdr.ca/





https://www.frdr-dfdr.ca/

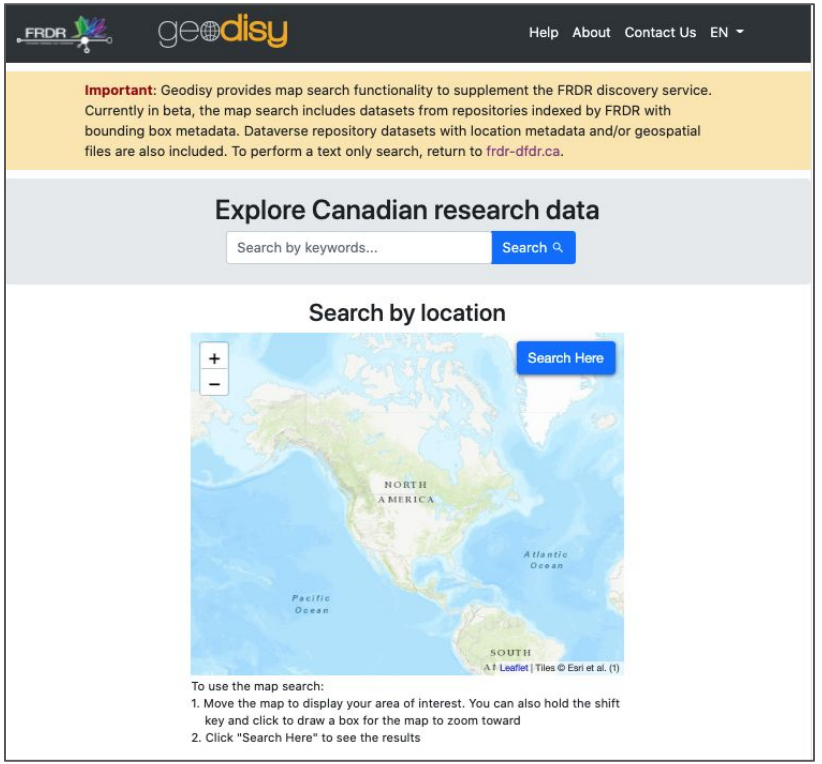
UBC Dataverse
(Scholars Portal)



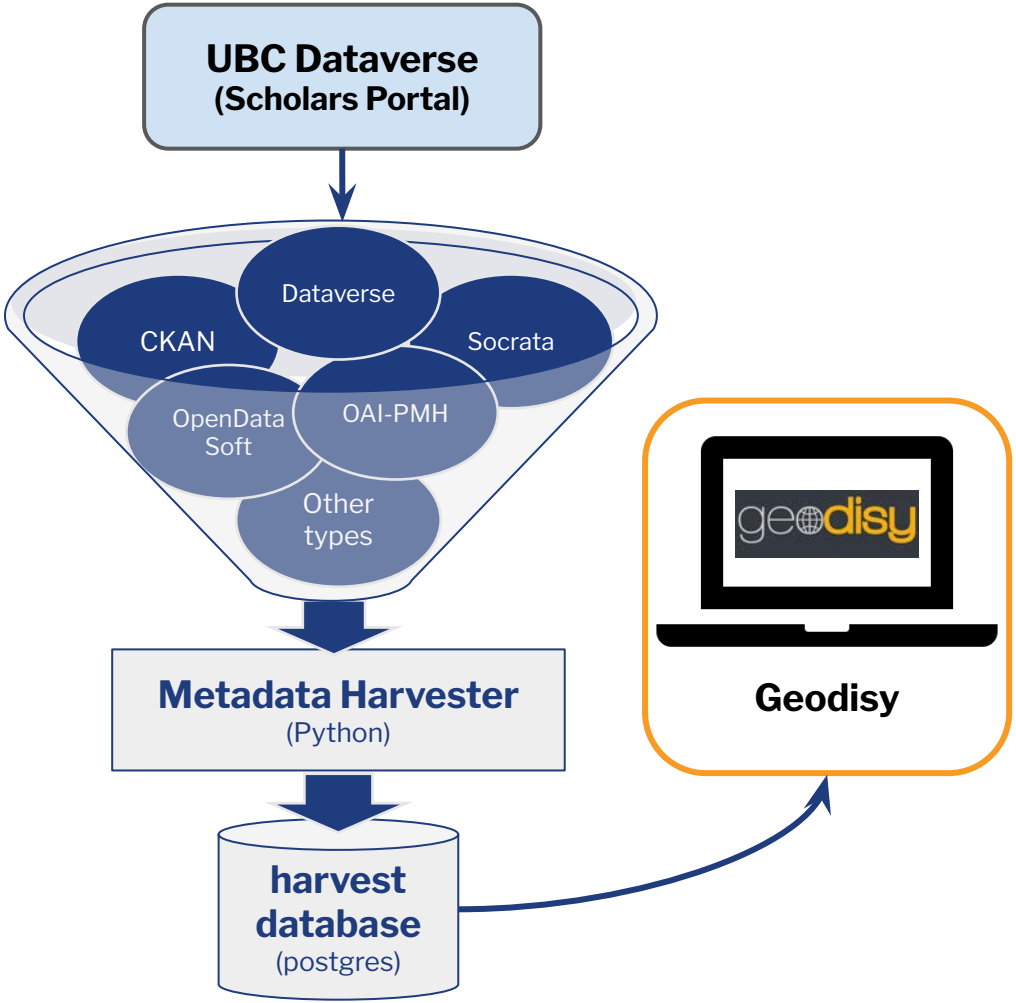
Metadata Harvester
(Python)

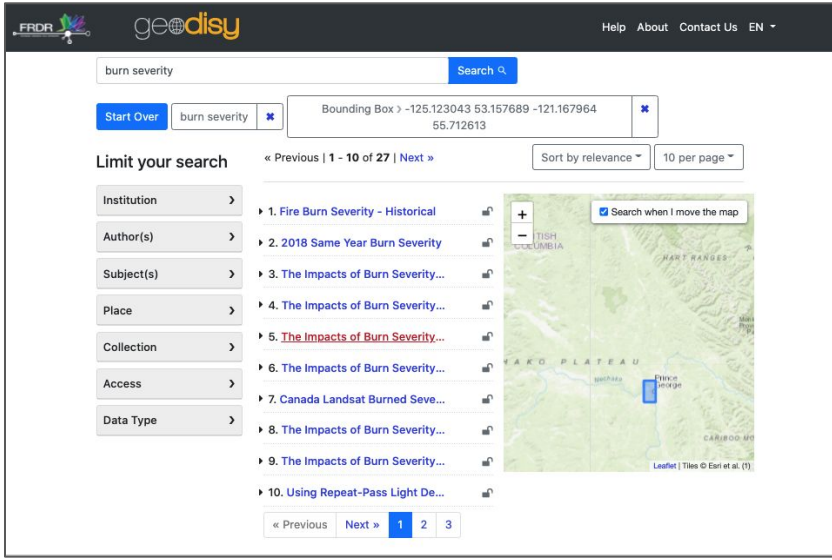
harvest database
(postgres)

Geodisy



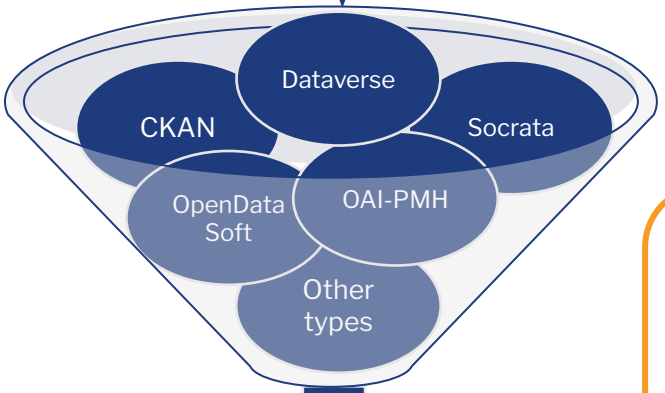
https://geo.frdr-dfdr.ca/





https://geo.frdr-dfdr.ca/

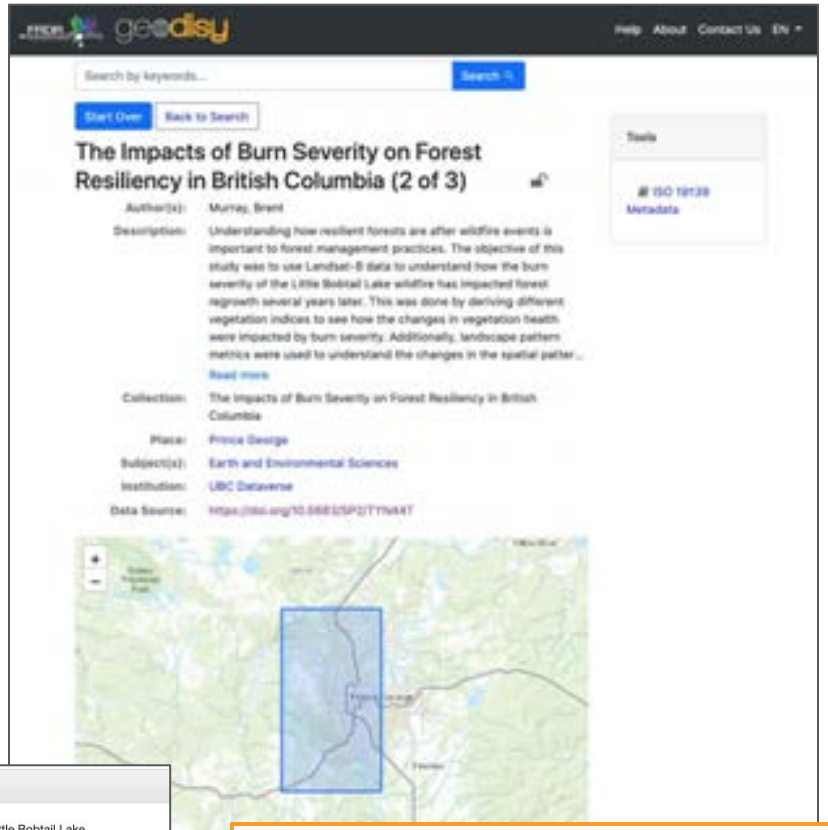
UBC Dataverse
(Scholars Portal)



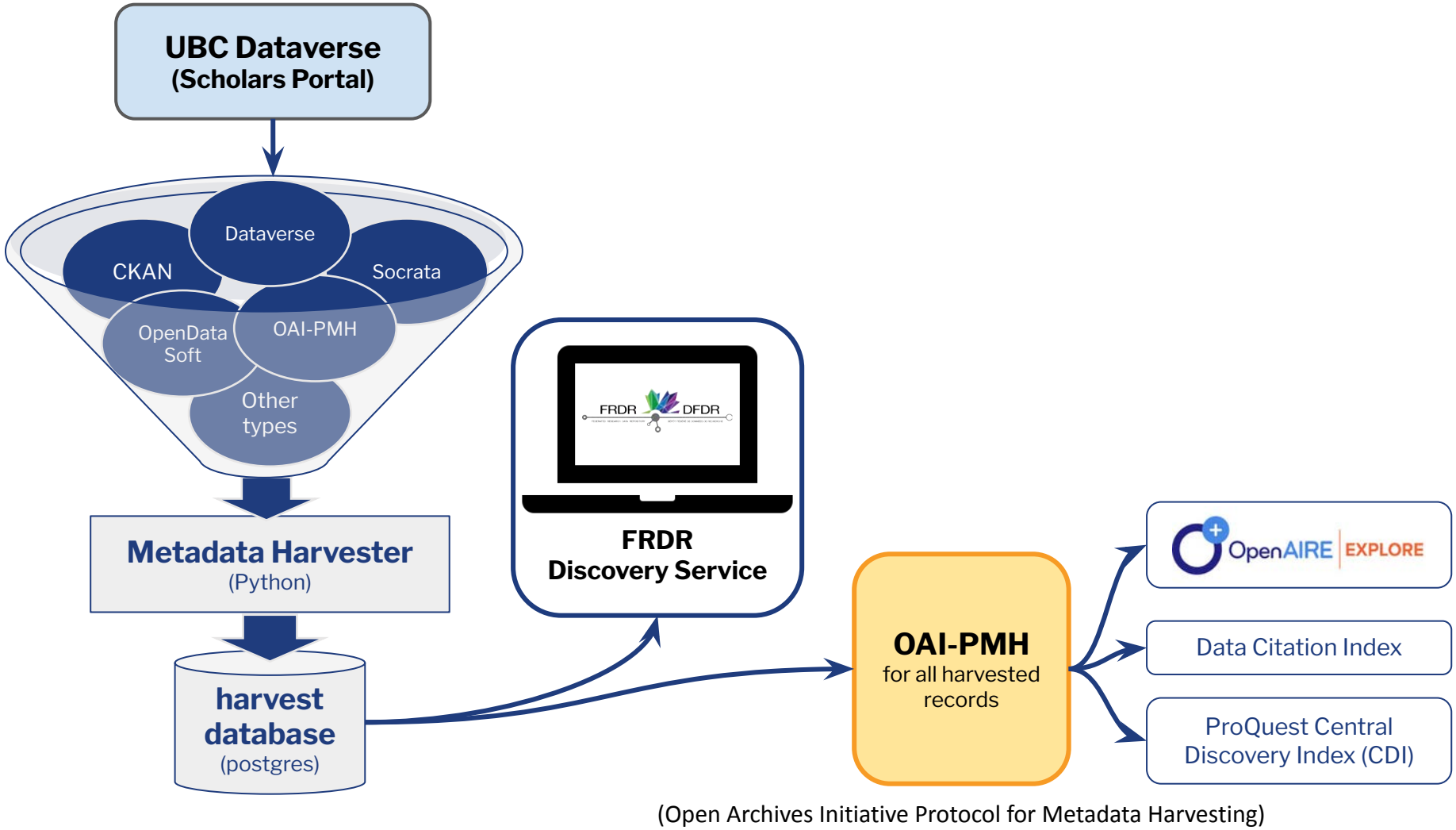
Metadata Harvester
(Python)

harvest database
(postgres)

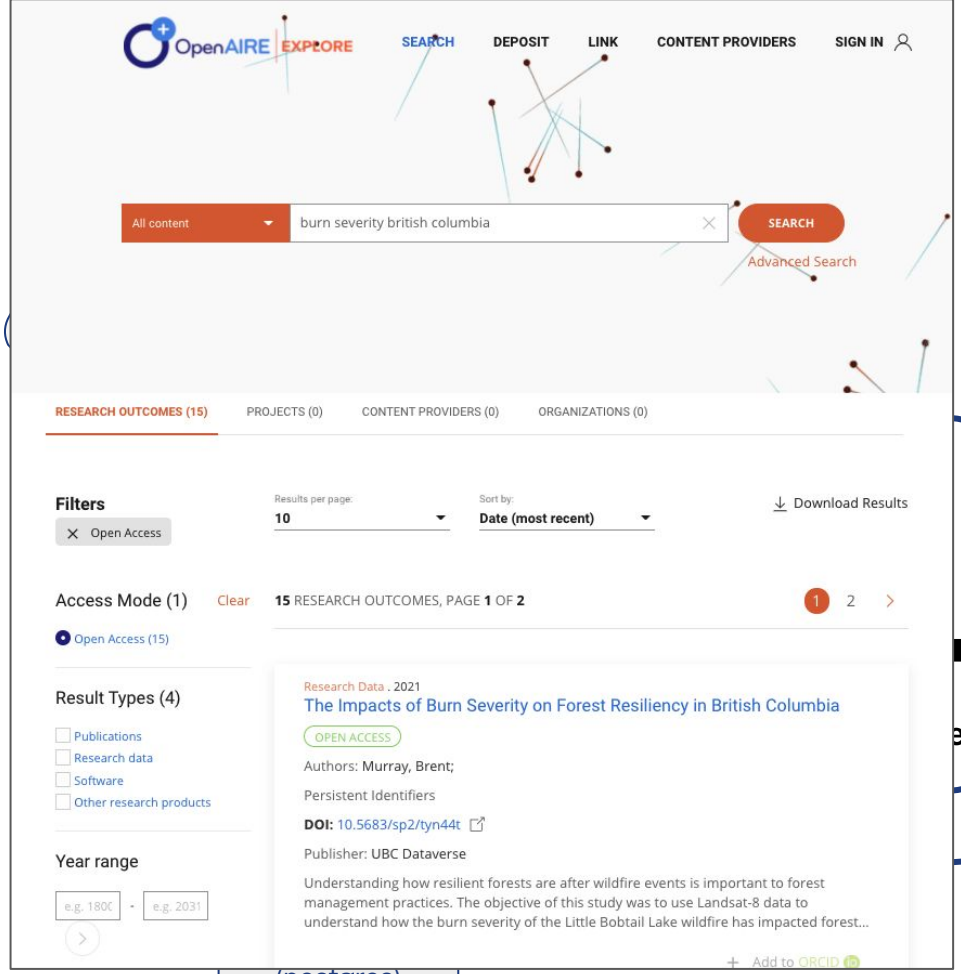
Geodisy



https://geo.frdr-dfdr.ca/



(Open Archives Initiative Protocol for Metadata Harvesting)



(postgres)

OAI-PMH
for all harvested records

OpenAIRE EXPLORE

Data Citation Index

ProQuest Central Discovery Index (CDI)

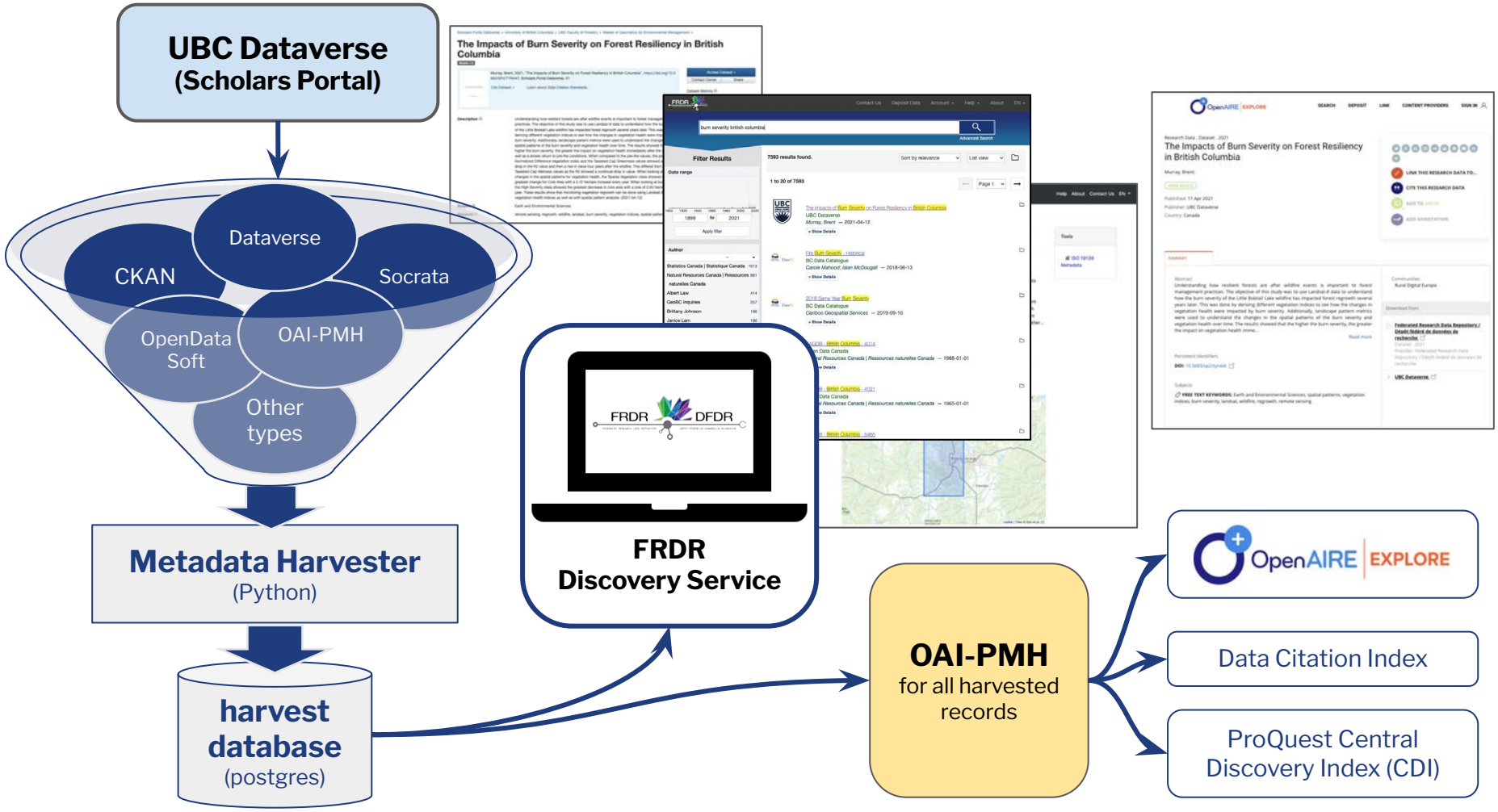
https://explore.openaire.eu/

UBC Dataverse

OAI-PMH
for all harvested records

- OpenAIRE EXPLORE
- Data Citation Index
- ProQuest Central Discovery Index (CDI)

https://explore.openaire.eu/

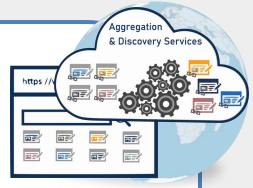

UBC Dataverse
(Scholars Portal)
Dataverse
CKAN
Socrata
OpenData Soft
OAI-PMH
Other types
Metadata Harvester
(Python)
harvest database
(postgres)
FRDR
Discovery Service
OAI-PMH
for all harvested records
OpenAIRE EXPLORE
Data Citation Index
ProQuest Central Discovery Index (CDI)

Discovery & Metadata Tips

Discovery Activity

1. Go to one of the platforms listed
2. Do a keyword search:
   earthquake monitoring vancouver island
3. Find the dataset that has 2 co-authors named Martin

Discovery Platforms



- OpenAIRE Explore
- WorldWideScience.org
- Google Dataset Search
- DataCite Commons
- DCI Web of Science*

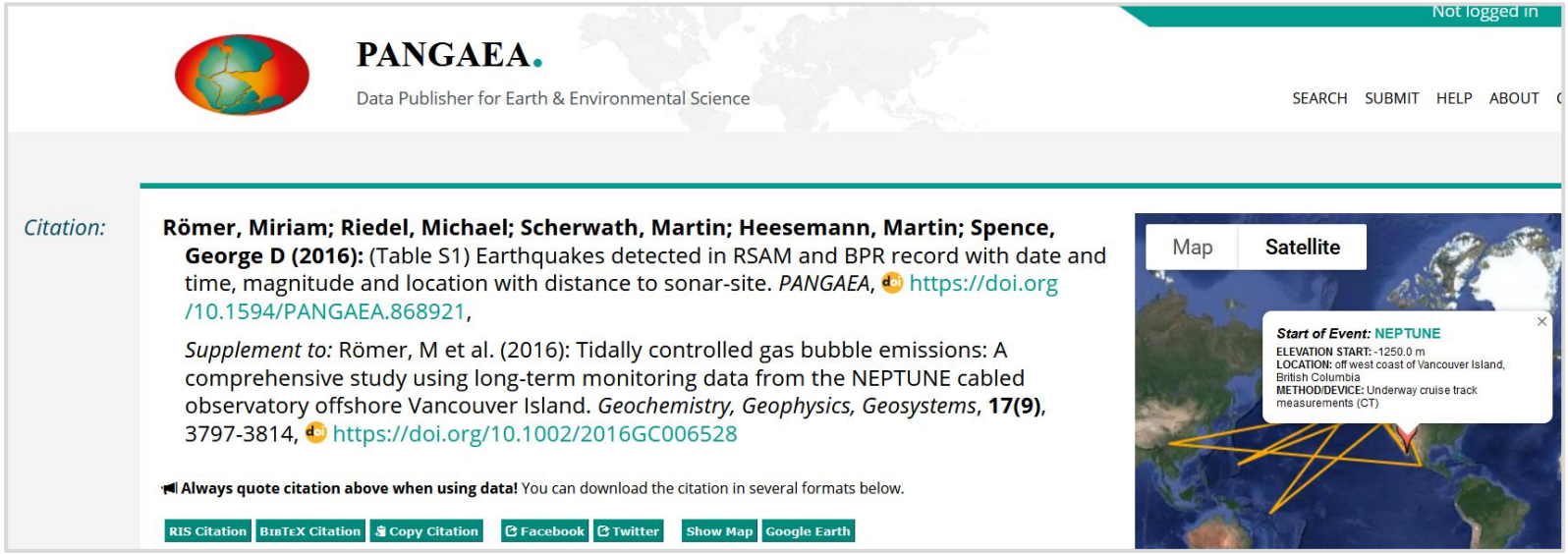
*if you have access to it

Finding and Being Found

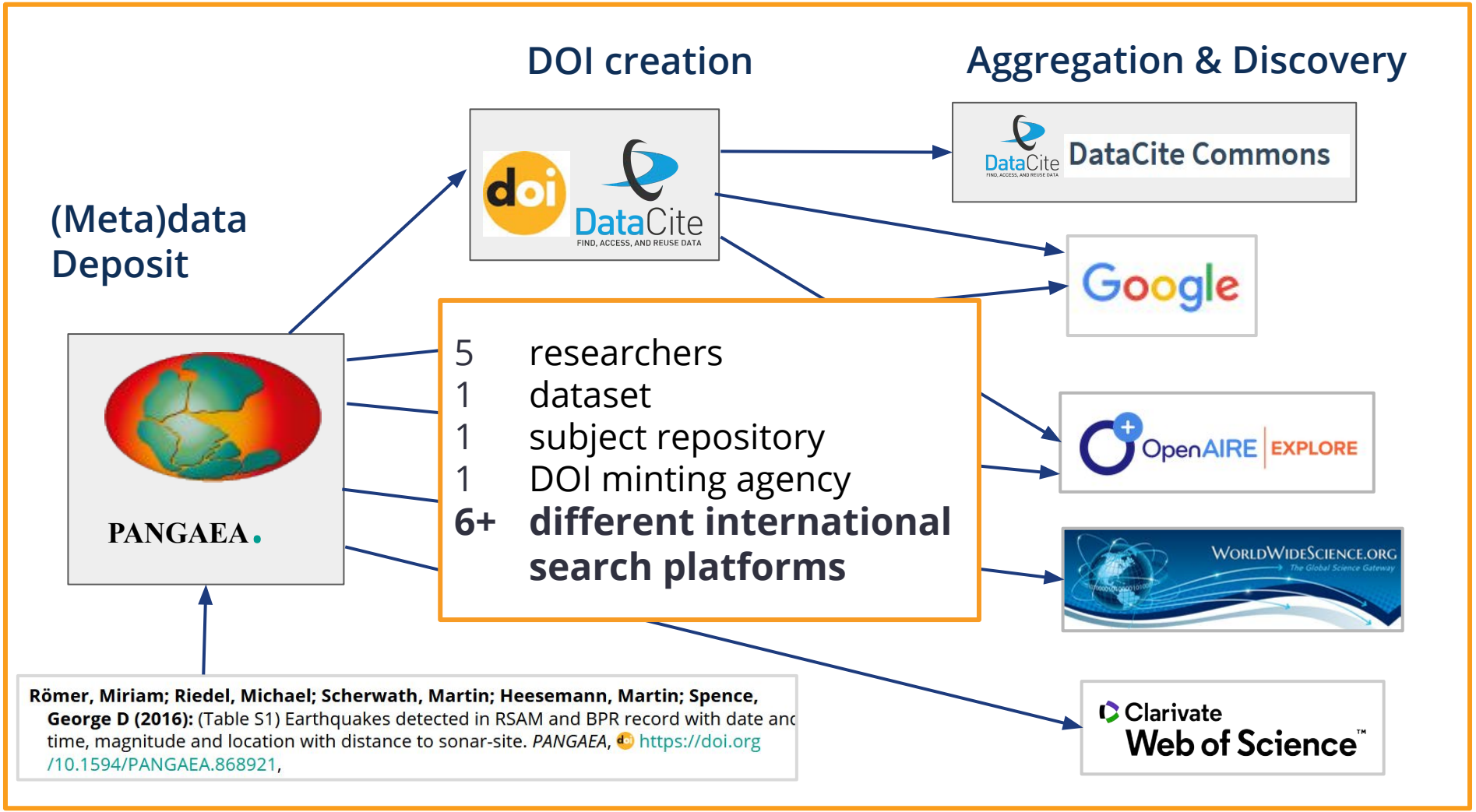
through Discovery Services

Original dataset

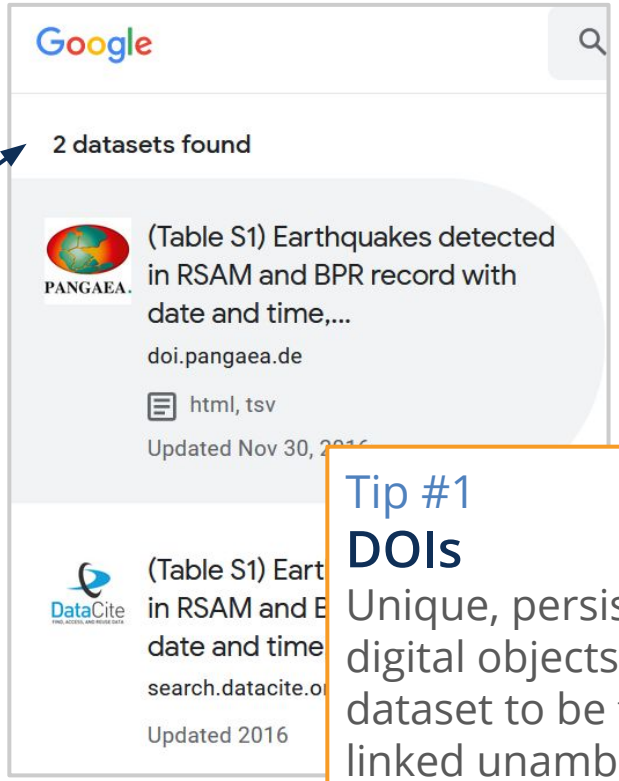
Römer, Miriam; Riedel, Michael; Scherwath, Martin; Heesemann, Martin; Spence, George D (2016): (Table S1) Earthquakes detected in RSAM and BPR record with date and time, magnitude and location with distance to sonar-site. PANGAEA, https://doi.org/10.1594/PANGAEA.868921.



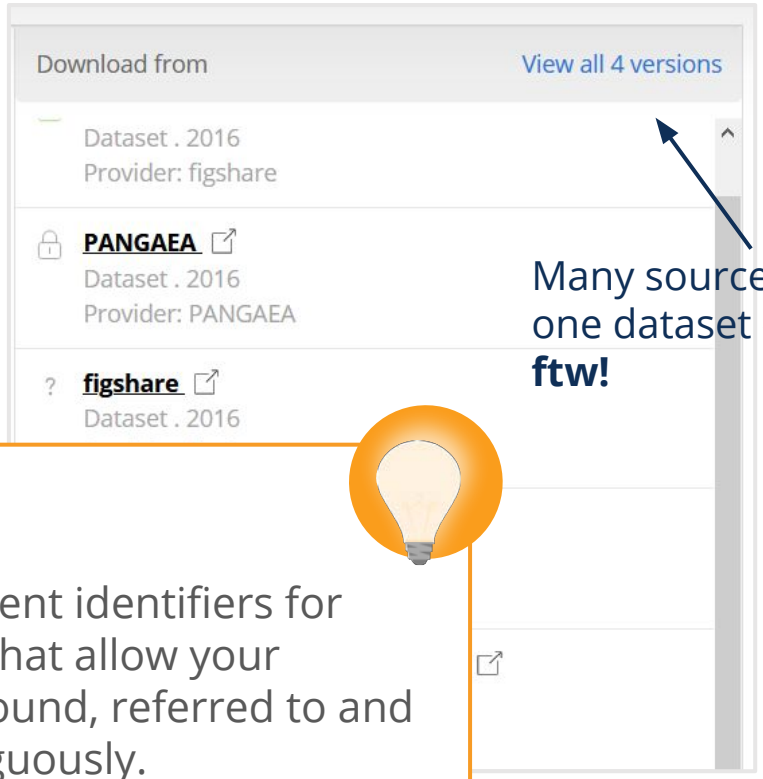
Full metadata: https://doi.pangaea.de/10.1594/PANGAEA.868921?format=metadata_jsonld


DOI creation
Aggregation & Discovery
(Meta)data Deposit
DataCite
FIND, ACCESS, AND REUSE DATA
DataCite Commons
Google
PANGAEA.
5 researchers
1 dataset
1 subject repository
1 DOI minting agency
6+ different international search platforms
OpenAIRE EXPLORE
WorldWideScience.org
The Global Science Gateway
Clarivate
Web of Science™
Römer, Miriam; Riedel, Michael; Scherwath, Martin; Heesemann, Martin; Spence, George D (2016): (Table S1) Earthquakes detected in RSAM and BPR record with date and time, magnitude and location with distance to sonar-site. PANGAEA, https://doi.org/10.1594/PANGAEA.868921,

Duplicate records?



Source: https://datasetsearch.research.google.com/



Source: https://explore.openaire.eu/search/find

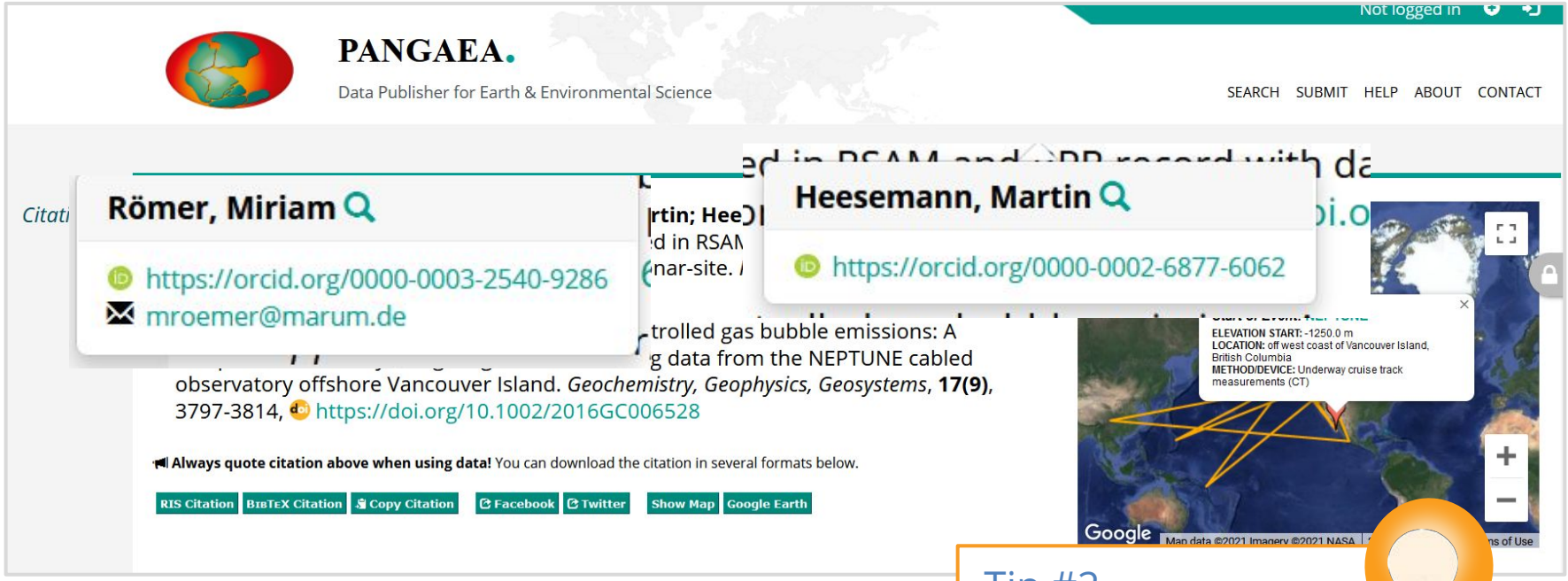
Many sources, but only one dataset → **DOIs ftw!**

Tip #1

DOIs

Unique, persistent identifiers for digital objects that allow your dataset to be found, referred to and linked unambiguously.

OpenAIRE and Google Dataset Search



ORCID
stands for
Open Researcher and Contributor ID

Tip #2

ORCiDs

Free, unique identifiers that researchers create and manage themselves.

Tip #3

Include machine-readable rights metadata

[I]nformation a human or machine needs to provide appropriate access to a resource, provide appropriate notification and compensation to rights holders, and to inform end users of any use restrictions that may exist.

Riley (2011): Seeing Standards

Many discovery platforms index datasets by access mode and kind of license, which powers the filters you can apply in your search (Source: https://explore.openaire.eu/search/find, https://commons.datacite.org/)

Access & Rights Metadata

Tip #3b

Indigenous Data Sovereignity

“The right of Indigenous peoples to control data from and about their communities and lands, articulating both individual and collective rights to data access and to privacy.

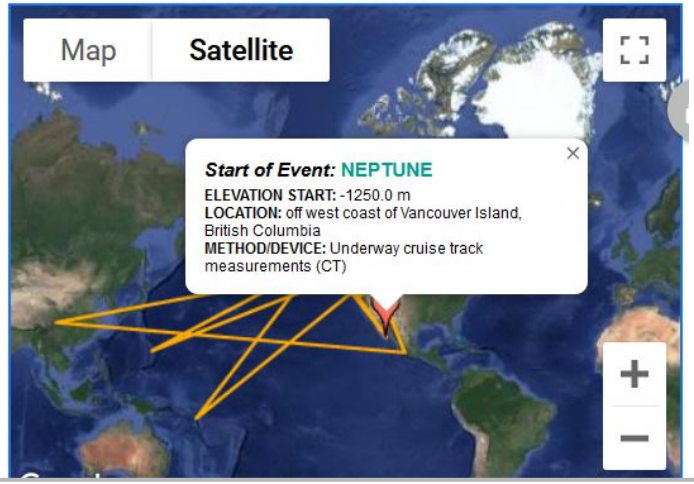
IDS also raises overarching, “fundamental questions about assumptions of ownership, representation, and control in open data communities”

Source:
First Nations Information Governance Centre - OCAP Principles

Tip #4

Add spatial metadata and view your dataset on a map

Google Dataset Search, FRDR's Geodisy, and other discovery platforms have map search functionalities, which rely on geolocation fields the metadata record.



Coverage:	*Median Latitude:* 49.192688 * *Median Longitude:* -131.117628 * *South-bound Latitude:* -10.800000 * *West-bound Longitude:* 102.890000 * *North-bound Latitude:* 56.298000 * *East-bound Longitude:* -102.180000 *Date/Time Start:* 2012-06-19T13:40:00 * *Date/Time End:* 2013-06-21T05:18:00 *Minimum Elevation:* -1250.0 m * *Maximum Elevation:* -1250.0 m
Event(s):	NEPTUNE * *Latitude Start:* 24.670000 * *Longitude Start:* -110.170000 * *Latitude End:* 40.180000 * *Longitude End:* -121.030000 * *Elevation Start:* -1250.0 m * *Elevation End:* -1250.0 m * *Location:* off west coast of Vancouver Island, British Columbia * *Method/Device:* Underway cruise track measurements (CT) * *Comment:* Ocean Networks Canada's North-East Pacific Time Series Underwater Networked Experiments (NEPTUNE) cabled ocean observatory.

Source: https://doi.pangaea.de/10.1594/PANGAEA.868921

Spatial metadata

#	Name	Short Name	Unit
1	DATE/TIME	Date/Time	
2	Magnitude	Magnitude	
3	LATITUDE	Latitude	
4	LONGITUDE	Longitude	
5	Distance	Distance	km
6	Azimuth	Azim	deg
7	Time delay	Time delay	
8	T		

Comment:

Earthquakes detected in RSAM and BPR record with date and time, magnitude and location (Latitude and Longitude) with distance to sonar-site. The azimuth is the angle between sonar and earthquake location. Time delays for P- and S-wave were calculated using standard Earth reference velocity model.

Tip #5

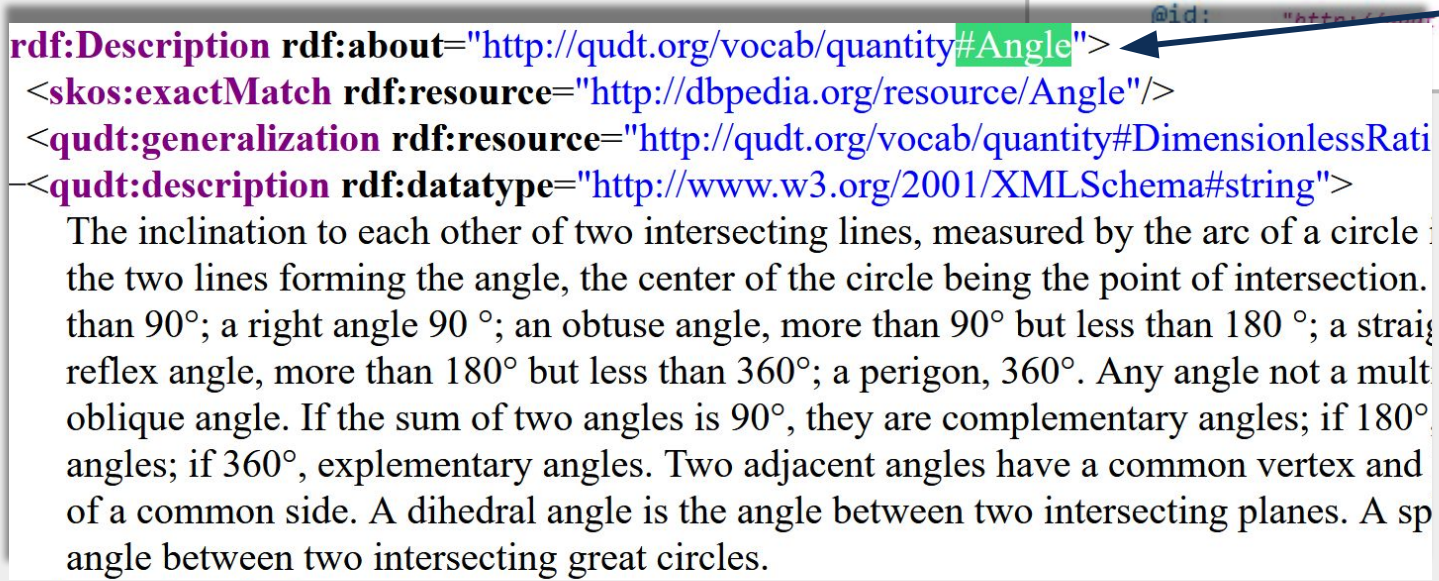
Standard terminology

Use standardized terms and (wherever possible) controlled vocabularies, and link to and use specific terminologies in your research field.

And → Metadata is for humans too! Use descriptive documents and free-text fields to include human-readable summaries.

Describing your dataset

Tip #5
Standard terminology (cont)



Tip #6

CoreTrustSeal

An international, community based, non-governmental, and non-profit organization that promotes sustainable and trustworthy data infrastructures.

- Find CTS certified repositories
- Canadian 2021 CTS certification cohort

R13. Data discovery and identification

- ➢ Inclusion in disciplinary or generic resource registries
- ➢ Presence of search facilities with
  - A searchable metadata catalogue using international standards
  - Persistent identifier systems
  - Enabled machine harvesting of the metadata
  - Recommended data citations

R14. Data reuse

- ➢ Metadata provided with dataset download
  - In formats used by target community
  - Ensure continued data understandability
  - Account for the possible evolution of formats

Source: DOI:10.5281/zenodo.3638211

Finding recommended repositories

Tip #7

Metadiscovery Tools

You can find more repositories, as well as other discovery & aggregation platforms in these service registries:

- ➔ re3data
- ➔ EOSC Marketplace
- ➔ Scientific Data's recommended repository list
- ➔ WDS-ITO's Searchable Index of Metadata Aggregators (will be available soon under doi.org/10.5281/zenodo.4589050)

Finding research data services

the "meta-meta catalogues"

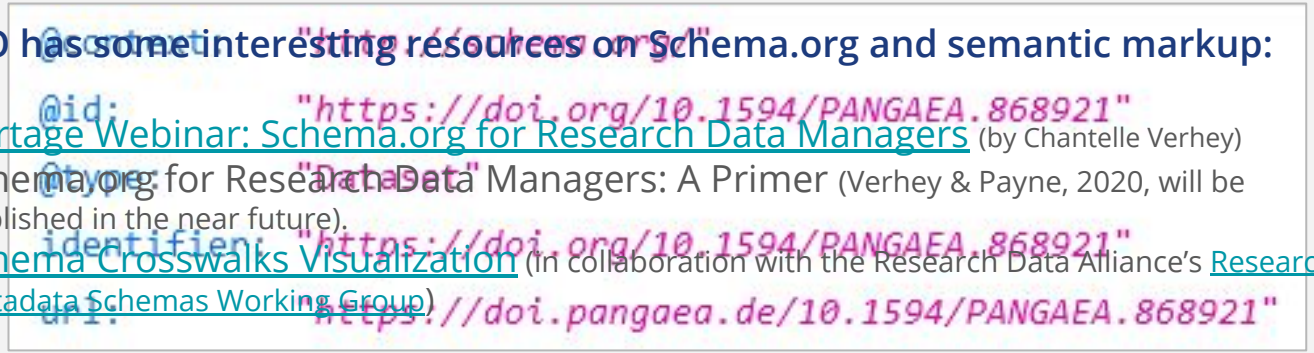
Tip #7

Schema.org

A few repositories and discovery services are now offering automatic semantic enrichment to make your metadata understandable to search engines such as Google, and other semantic digital tools.

WDS-ITO has some interesting resources on Schema.org and semantic markup:

- Portage Webinar: Schema.org for Research Data Managers (by Chantelle Verhey)
- Schema.org for Research Data Managers: A Primer (Verhey & Payne, 2020, will be published in the near future).
- Schema Crosswalks Visualization (in collaboration with the Research Data Alliance's Research Metadata Schemas Working Group)



Semantic Metadata

APRIL 15, 2021

Call for Applications: NDRIO-Portage COVID-19 Data Curation Funding

NDRIO and Portage are pleased to announce the availability of funding to support the curation of Canadian research data related to the COVID-19 pandemic. The purpose of this funding is to bring COVID-19 related data into timely compliance with the FAIR Guiding Principles (Findable, Accessible, Interoperable, and Reusable) for the management of research data, algorithms, tools and workflows.[1] Ultimately, increasing the FAIRness of datasets makes them discoverable and re-usable by both humans and machines, enabling downstream research. As a secondary objective, this initiative aims to build capacity for research data management (RDM) in Canada by supporting the training of highly-qualified personnel (HQP) in data stewardship and curation.

21 May: Deadline for applications to NDRIO-Portage COVID-19 Data Curation Funding (for Canadian researchers and research programs)

WORLD DATA SYSTEM

Thank you!