The Current State of Meta-Repositories for Data

Tushar B. Kute,
http://tusharkute.com

Introduction

- Researchers have many options available to them in order to fulfill individual, funder, and publisher requirements to deposit and share research data.

- Thus many of their research outputs, including data, may be scattered across various institutional, domain, funder, publisher-supported, and general repositories and websites.

- Given this, a faculty member searching for data sets similar to his or her own research may find it difficult, if not impossible, to discover relevant sources.

Introduction

- Without direct connections among the various research outputs, there are few mechanisms for anyone to understand what data, article, and code are related to the same research.

- This is a significant scholarly communications issue. Recently, much work has developed around online solutions to federate and link the records across these dispersed repositories, creating large meta-repositories of data.

Introduction

- A previous study of digital libraries that are more similar to these meta-repositories of data, compiled by the Digital Repository of Ireland (DRI), focused on how digital objects were being cared for internationally.

- The authors found three different models: the metadata aggregator, the single-site repository, and the multi-site repository.

- DRI also indicated that funding agencies place a greater emphasis on access rather than preservation of the digital content, which may ultimately put the ongoing availability of content at risk.

Community Initiatives

- Founded in 2009, the Confederation of Open Access Repositories (COAR) seeks to create community and support for repositories worldwide.

- Current members include the Vienna University Library and Archive Services, the University of Antwerp, McMaster University Library, bepress, and the World Bank, to name a few.

- The COAR organization and community builds capacity, aligns policies and practices, and acts as a global voice for the repository community.

tusharkute
.com

Community Initiatives

- Organizations that support the quality and accessibility of data are not new.

- The International Council for Science: Committee on Data for Science and Technology (CODATA) is an organization established over forty years ago.

- One of CODATA's main objectives is to facilitate international cooperation among those institutions collecting, organizing, and using data.

- This work is primarily facilitated through the committees and working groups focused on projects of specific scope, such as legal interoperability.

Methods

- A unified term to describe meta-repositories of data currently does not exist, which makes conducting Web searches to identify these systems impossible.

- Conducting Web searches using the terms federated repositories and repository aggregator resulted in zero relevant systems.

- Thus, the meta-repositories of data described here were primarily identified through the author's knowledge of such systems and suggestions from colleagues.

Meta-repositories

The Meta-Repositories Chosen for Analysis in this Study

Meta-Repository	Mission	URL
1. Australian Research Data Commons (ANDS)	ANDS is a system built and maintained in Australia to • "make Australian research data collections more valuable by managing, connecting, enabling discovery and supporting the reuse of this data" • "enable richer research, more accountable research; more efficient use of research data; and improved provision of data to support policy development."[a]	http://ands.org.au
2. Beilefeld Academic Search Engine (BASE)	BASE is a portal established by Bielefeld University Library, United Kingdom that integrates Open Archives Initiative (OAI) resources as one information type among others into the local digital library environment, together with catalogs, article databases, and digitized collections.	https://www.base-search.net/

Meta-repositories

3. COnnecting REpositories (CORE)	CORE is a UK-based meta-repository that seeks "to aggregate all open access research outputs from repositories and journals worldwide and make them available to the public."[b]	https://core.ac.uk/
4. Data.gov	Data.gov is the home of US government metadata. Non-federal data sources can also be added to the data set voluntarily.	http://www.data.gov/
5. Data Archiving and Networked Services (DANS)	Developed in the Netherlands, DANS is a service institute that promotes sustained access to digital research data.	http://www.dans.knaw.nl/en
6. DataBridge	DataBridge is a cross-institutional collaboration that aims to make the "long tail" of data more discoverable.	http://databridge.web.unc.edu/
7. DataCite	DataCite is an organization that works with data centers to assign digital object identifiers to research assets.	https://www.datacite.org
8. EUDAT	EUDAT is a system that includes data access, deposit, sharing, archiving, identification, and discovery of research data produced across the European Union.	https://eudat.eu

Meta-repositories

9. ICSU/World Data System (WDS)	Launched in Japan, ICSU/WDS research data system seeks to enable universal and equitable access to scientific data.	https://www.icsu-wds.org
10. OpenAIRE	Initiated in the European Union, OpenAIRE brings together scholarly metadata to support open scholarship and improve the reuse of publications and data.	https://www.openaire.eu/
11. OpenDOAR	OpenDOAR is a directory of open-access academic repositories.	http://opendoar.org/
12. OneRepo	OneRepo is a system that seeks to bring together all open-access scholarly articles.	http://onerepo.net
13. SHARE	SHARE is a metadata data set about research and scholarly activities through the research life cycle (such as data management plans, funder information, articles, data sets, etc.)	http://share-research.org

Meta-repositories

The Meta-Repository Website Analysis Used Variables Categorized into Four Areas

Area	Variables Collected
Background	Date founded, goals/vision, mission, funding model
Content Coverage	Time span, spatial/geographic parameters, domain specificity, data types, providers, number of records, update frequency
Metadata	Standards, elements
Functionality	Faceted searching, feeds/alerts

Results

- The results of the website analysis show various points of similarities among the thirteen meta-repositories of data. Six of the meta-repositories were created to support national missions to ensure quality data and accessibility (meta-repositories 1, 4, 5, 8, 9, 10), while the remaining were created as responses to growing scholarly communication needs, to maximize research impact, and to otherwise promote science.

- For example, the mission of the ANDS is to make Australian research data collections more available "by managing, connecting, enabling discovery and supporting the reuse of this data." In contrast, SHARE's mission is "to maximize research impact by making a comprehensive inventory research widely accessible, discoverable, and reusable."

Contents

- From a content perspective, the majority of the meta-repositories were harvesting content from repositories worldwide (n=9), while two were limited to nations and one was unknown in spatial coverage.

- None of the repository aggregators were limited to a specific domain (i.e., gathering source information only from a specific scientific discipline).

- While all of the repository aggregators had metadata about data sets in their systems, many also had articles, theses and dissertations, and conference papers and presentations

Contents

- One repository aggregator, OpenDOAR, also included content such as audiovisual material and learning objects.

- Most systems were simply aggregating the metadata, but a handful of the meta-repositories had the actual digital asset stored, including CORE and Data.gov.

Functionality

- In regard to search features and faceting found in the meta-repositories for data, all working systems had some type of advanced search limiters.

- The most common types of features were facets that allowed the user to limit the results by a subject area, institution, or publication year.

- The Australian National Data Service had a unique function that allowed the user to find related people and related organizations from a search query.

Metadata

- One of the most glaring areas where many of the meta-repositories for data systems did not align was in their use of metadata standards.

- Of the thirteen systems, only two used the same standard: DataCite and OpenAIRE (DataCite metadata standard).

- The remaining eleven systems all used a local standard—RIOXX, DDI Lite, panFMP, DDI, RIF-CS— or were not using a standard for various reasons.

Conclusion

- Scholarly communication is in need of systems to pull together and link dispersed research objects.

- Just as Netflix revolutionized film discovery and rental, meta-repositories are needed to discover and highlight research from varying providers, make recommendations, show relationships between research and researchers, and make connections among the digital assets.

- The whole story of research, and the complete scholarly record, is more than just the final publication.

- It includes funder information, data sets, documentation, and code in many cases.

Thank you

@mitu_skillologies

/mITuSkillologies

@mitu_group

/company/mitu-skillologies

MITUSkillologies

Web Resources
https://mitu.co.in
http://tusharkute.com

contact@mitu.co.in

tushar@tusharkute.com