

# Data Ecology

LinkedData  
DigitalData

ResearchData

WebOfData

OpenData  
BigData

# Importance of Research Data Sharing

## Rationales:

- “to reproduce or verify research
- to make results of publicly funded research available to the public
- to enable others to ask new questions of extant data, and
- to advance the state of research and innovation”. (Borgman, 2007)

# Context: Research Data Services

- Rise of open data and open science data movements
- Implementation of research data management and sharing policies
- Shifting view that research data is a valid research output that should be recognized

# Why Metadata for Research Data?

- Data creation
- Data curation, archiving and preservation
- Data sharing
- Data privacy
- Data management
- Data interoperability
- Data access & discovery

*Metadata supports all of the above activities.*

# Context: Metadata in Data Repositories

- Enables subject and semantic interoperability
- Component of best practices and policies
- Need for increased awareness and understanding within the research community

# Research Data Services Examined



DataCite

Helping you to find,  
access, and reuse data



DRYAD



figshare  
credit for all your research

The  
**Dataverse  
Network**<sup>TM</sup>

Project



# Research Data Services Examined

Service	Subject area	Main services	Location
Datacite	General	Metadata, DOI	UK
Dataverse Network	General	Cite, analyze, preserve, DOI	US
Dryad	General	data underlying scholarly publications discoverable, accessible, understandable, freely reusable, and citable, DOI	US
FigShare	General	figures, datasets, media, papers, posters, presentations and filesets, altmetrics, DOI	UK

# Research Questions

- How many metadata elements are available?
- What research data specific elements are provided?
- Are standards adhered to?
- Are controlled vocabularies used?
- Are unique identifiers (e.g., DOIs) supported?
- What metadata assistance is provided?
- What are the common and unique metadata elements across the services?



	Datacite	Dataverse Network	Dryad	Figshare
<b>Number of metadata elements</b>	41	100	52	12
<b>Research specific metadata elements</b>	No	Yes	Yes	No
<b>Compliance with standards</b>	Datacite Metadata Schema, which is an application profile of Dublin Core (DC), OAI	Data Documentation Initiative (DDI) Codebook, compliant with Dublin Core (DC) and Content Standard for Digital Geospatial Metadata (CSDGM), MARC LOCKSS, OAI	Dublin Core, Darwin Core, Bibliographic Ontology, METS/ MODS OAI/DC OAI/ORE (Object Reuse and Exchange) RDF/DC CLOCKSS For now, OAI/DC is the recommended format.	CLOCKSS
<b>Use of controlled vocabularies</b>	Includes controlled vocabularies for some elements, supports use of controlled vocabularies for other elements; MESH, OBI, NCBI	Supports use of controlled vocabularies	Supports use of ontologies and controlled vocabularies such as Open Biomedical Ontologies & Gene Ontology. A trial version of HIVE is provided to support subject description. LCSH, TGN, MESH, Integrated Taxonomic Information Systems (ITIS), National Biological Information Infrastructure Biocomplexity Thesaurus, LC Name Authorities file	No formal controlled vocabularies; only 14 high level categories
<b>Support for DOI</b>	Yes	Yes	Yes	Yes
<b>Metadata assistance</b>	full documentation of metadata schema, user guidelines, full api documentation	metadata documentation available via user guide, contextual help available for each element in metadata entry form	Dryad Wiki pages provide detailed documentation including Cataloguing guidelines	Partner with DataCite

# Number and Type of Metadata Elements

	<b>Datacite</b>	<b>Dataverse Network</b>	<b>Dryad</b>	<b>Figshare</b>
<b>Number of metadata elements</b>	41	100	52	12
<b>Research specific metadata elements</b>	No	Yes	Yes	No

# Standards and Controlled Vocabularies

	Datcite	Dataverse Network	Dryad	Figshare
<b>Compliance with standards</b>	Datcite Metadata Schema, which is an application profile of Dublin Core (DC), OAI	Data Documentation Initiative (DDI) Codebook, compliant with Dublin Core (DC) and Content Standard for Digital Geospatial Metadata (CSDGM), MARC LOCKSS, OAI	Dublin Core, Darwin Core, Bibliographic Ontology, METS/MODS OAI/DC OAI/ORE (Object Reuse and Exchange) RDF/DC CLOCKSS For now, OAI/DC is the recommended format.	CLOCKSS
<b>Use of controlled vocabularies</b>	Includes controlled vocabularies for some elements, supports use of controlled vocabularies for other elements; MESH, OBI, NCBI	Supports use of controlled vocabularies	Supports use of ontologies and controlled vocabularies such as Open Biomedical Ontologies & Gene Ontology. A trial version of HIVE is provided to support subject description. LCSH, TGN, MESH, Integrated Taxonomic Information Systems (ITIS), National Biological Information Infrastructure Biocomplexity Thesaurus, LC Name Authorities file	No formal controlled vocabularies; only 14 high level categories

# DOIs, Metadata Documentation

	<b>Datacite</b>	<b>Dataverse Network</b>	<b>Dryad</b>	<b>Figshare</b>
<b>Support for DOI</b>	Yes	Yes	Yes	Yes
<b>Metadata assistance</b>	full documentation of metadata schema, user guidelines, full api documentation	metadata documentation available via user guide, contextual help available for each element in metadata entry form	Dryad Wiki pages provide detailed documentation including Cataloguing guidelines	Partner with DataCite

# Common and Unique Metadata Elements

	<b>Datacite</b>	<b>Dataverse Network</b>	<b>Dryad</b>	<b>Figshare</b>
<b>Titles</b>	title	<ul style="list-style-type: none"> <li>- title</li> <li>- subtitle</li> <li>- document title</li> </ul>	<ul style="list-style-type: none"> <li>- article title</li> <li>- journal title</li> <li>- data package title</li> </ul>	title
<b>Creators, Contributors</b>	<ul style="list-style-type: none"> <li>- creator</li> <li>- contributor</li> <li>- publisher</li> </ul>	<ul style="list-style-type: none"> <li>- author</li> <li>- producer</li> <li>- funding agency</li> <li>- distributor</li> <li>- depositor</li> <li>- contact</li> <li>- data collector</li> </ul>	<ul style="list-style-type: none"> <li>- author</li> <li>- creator</li> </ul>	<ul style="list-style-type: none"> <li>- author</li> <li>- collaborators</li> </ul>
<b>Topical subject(s)</b>	subject	<ul style="list-style-type: none"> <li>- keyword</li> <li>- topic classification</li> </ul>	<ul style="list-style-type: none"> <li>- keyword</li> <li>- scientific name</li> </ul>	<ul style="list-style-type: none"> <li>- categories</li> <li>- tags</li> </ul>
<b>General description</b>	description	abstract	<ul style="list-style-type: none"> <li>- article abstract</li> <li>- description</li> </ul>	description
<b>Object type(s)</b>	resource type	kind of data	type	type
<b>Date(s)</b>	<ul style="list-style-type: none"> <li>- date</li> <li>- publication year</li> </ul>	<ul style="list-style-type: none"> <li>- production date</li> <li>- distribution date</li> <li>- deposit date</li> <li>- version date</li> <li>- date of collection-start</li> <li>- date of collection-end</li> </ul>	<ul style="list-style-type: none"> <li>- date of issuance</li> <li>- deposit date</li> <li>- date available</li> <li>- embargo date</li> </ul>	<ul style="list-style-type: none"> <li>- date created</li> <li>- date published</li> </ul>



# Common and Unique Metadata Elements

	Datacite	Dataverse Network	Dryad	Figshare
<b>Rights, Access, Use</b>	rights	<ul style="list-style-type: none"> <li>- data access place</li> <li>- original archive</li> <li>- availability status</li> <li>- confidentiality declaration</li> <li>- special permissions</li> <li>- restrictions</li> <li>- conditions</li> <li>- provenance</li> <li>- document holdings</li> <li>- disclaimer</li> </ul>	<ul style="list-style-type: none"> <li>- rights statement</li> <li>- location of related content outside of Dryad</li> </ul>	license
<b>Object technical characteristics</b>	<ul style="list-style-type: none"> <li>- size</li> <li>- format</li> </ul>	<ul style="list-style-type: none"> <li>- software</li> <li>- software version</li> <li>- size of collection</li> <li>- study completion</li> </ul>	<ul style="list-style-type: none"> <li>- file format</li> <li>- file size</li> <li>- provenance</li> </ul>	file size
<b>Spatial subject(s)</b>	- geo location	<ul style="list-style-type: none"> <li>- country/nation</li> <li>- geographic coverage</li> <li>- geographic unit</li> <li>- geographic bounding box</li> </ul>	- spatial coverage	
<b>Identifiers</b>	<ul style="list-style-type: none"> <li>- identifier</li> <li>- alternate identifier</li> <li>- related identifier</li> </ul>	<ul style="list-style-type: none"> <li>- study global ID</li> <li>- other ID</li> </ul>	<ul style="list-style-type: none"> <li>- article identifier</li> <li>- associated Dryad data package identifier</li> <li>- data package identifier</li> <li>- identifier for related data in Dryad partner repository</li> <li>- associated Dryad publication record identifier</li> <li>- associated Dryad data file record identifier</li> <li>- data file identifier</li> <li>- issn</li> <li>- electronic issn</li> </ul>	
<b>Temporal subject(s)</b>		<ul style="list-style-type: none"> <li>- time period covered-start</li> <li>- time period covered-end</li> </ul>	- temporal coverage	



# Common and Unique Metadata Elements

	Datcite	Dataverse Network	Dryad	Figshare
<b>Citation</b>		<ul style="list-style-type: none"> <li>- citation requirements</li> <li>- depositor requirements</li> </ul>	<ul style="list-style-type: none"> <li>- journal volume number</li> <li>- journal issue</li> <li>- article start page</li> <li>- article end page</li> <li>- article pages</li> </ul>	
<b>Versioning</b>	version	version		
<b>Methodology</b>		<ul style="list-style-type: none"> <li>- unit of analysis</li> <li>- universe</li> <li>- time method</li> <li>- frequency</li> <li>- sampling procedure</li> <li>- major deviations for sample design</li> <li>- collection mode</li> <li>- type of research instrument</li> <li>- data sources</li> <li>- origin of sources</li> <li>- characteristics of sources noted</li> <li>- documentation and access to sources</li> <li>- characteristics of data collection situation</li> <li>- actions to minimize losses</li> <li>- control operations</li> <li>- weighting</li> <li>- cleaning operations</li> <li>- study level error nores</li> <li>- response rate</li> <li>- estimates of sampling errors</li> <li>- other forms of data appraisal</li> </ul>		
<b>Related resources</b>		<ul style="list-style-type: none"> <li>- series</li> <li>- series information</li> <li>- replication for</li> <li>- related publications</li> <li>- related material</li> <li>- related studies</li> <li>- other references</li> </ul>		

# Common and Unique Metadata Elements

	<b>Datacite</b>	<b>Dataverse Network</b>	<b>Dryad</b>	<b>Figshare</b>
<b>Language(s)</b>	language			
<b>Status</b>			<ul style="list-style-type: none"> <li>- status</li> <li>- article publication status</li> </ul>	
<b>Production</b>		<ul style="list-style-type: none"> <li>- production place</li> </ul>		
<b>Additional grant information</b>		<ul style="list-style-type: none"> <li>- grant number</li> <li>- grant number agency</li> </ul>		
<b>Note(s)</b>		notes		

# Conclusions

- Number and nature of metadata elements across the examined services vary
- There is a set of elements across them that allows for interoperability
- Research data specific elements are in use though not across all services
- Support for controlled vocabularies and identifiers is encouraging

# Contributions

- Improve understanding of the application of metadata in research data services
- Demonstrate the complexity and multifaceted nature of metadata for research data
- Presents a new perspective on the importance of strategic planning and policy development for managing research data
- Development of a research framework to provide a basis for interoperability and metadata sharing

# Next Steps

- Expand analysis to additional research data services (general and domain-specific)
- Detailed comparison of common and unique metadata elements
- Analysis of support for researcher identifiers such as ORCID

# Thank you!

## Questions or comments?

[sharon.farnel@ualberta.ca](mailto:sharon.farnel@ualberta.ca)

[ali.shiri@ualberta.ca](mailto:ali.shiri@ualberta.ca)