





Identifying metadata commonalities across restricted health data sources: A mixed methods study exploring how to improve the discovery of and access to restricted datasets

Kevin B. Read, University of Saskatchewan, Saskatoon, Saskatchewan, Canada, kevin.read@usask.ca 

Grant Gibson, Canadian Research Data Centre Network and McMaster University, Hamilton, Ontario, Canada 


Amber Leahey, Scholars Portal, Toronto, Ontario, Canada 

Lynn Peterson, National Research Council of Canada, Ottawa, Ontario, Canada 

Sarah Rutley, University of Saskatchewan, Saskatoon, Saskatchewan, Canada 

Julie Shi, University of Toronto, Toronto, Ontario, Canada 

Victoria Smith, Digital Research Alliance of Canada, Ottawa, Ontario, Canada

Kelly Stathis, DataCite, Hannover, Germany 

Abstract

Background: While open datasets are adopting FAIR principles to improve their discovery and use, restricted data—those only accessible via request or application—have fallen behind. Metadata is not an inherent characteristic of restricted data, which limits its ability to be found and used. To better understand discoverability and accessibility of restricted data, this study reviewed restricted health data sources to determine how they describe their datasets and access procedures, what descriptive commonalities exist across data sources, and to what extent the commonalities we found can be accommodated within existing metadata schemas.

Received: January 19, 2024 **Accepted:** July 7, 2024 **Published:** August 16, 2024

Keywords: restricted data, data discovery, data access, metadata, data sharing, data reuse

Citation: Read, Kevin B., Grant Gibson, Amber Leahey, Lynn Peterson, Sarah Rutley, Julie Shi, Victoria Smith, and Kelly Stathis. 2024. "Identifying metadata commonalities across restricted health data sources: A mixed methods study exploring how to improve the discovery of and access to restricted datasets." *Journal of eScience Librarianship* 13 (2): e907. <https://doi.org/10.7191/jeslib.907>.

Data Availability: Datasets that include the dataset and access information extraction process, metadata commonality categorization, metadata definitions, and the metadata mapping are available in the Open Science Framework at: <https://osf.io/txrve>.

The *Journal of eScience Librarianship* is a peer-reviewed open access journal. © 2024 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0), which permits unrestricted use, distribution, and reproduction in any medium non-commercially, provided the original author and source are credited. See <https://creativecommons.org/licenses/by-nc/4.0>.

OPEN ACCESS

Abstract Continued

Methods: This study extracted dataset and access information provided by a sample of 48 restricted data sources, identified commonalities across these data sources to develop possible metadata elements for restricted data, and mapped these metadata elements to existing metadata schemas (e.g., DataCite) to evaluate how well they accommodate information supplied by restricted data sources.

Results: Restricted data sources describe their datasets (35 commonalities) and access procedures (27 commonalities) in similar ways. Dataset descriptions aligned with existing metadata schemas, with the DDI-Lifecycle and -Codebook schemas receiving 91.4% and 85.7% exact matches respectively with the dataset elements we identified. Access procedures did not align with metadata available in existing schemas.

Discussion: While descriptive dataset metadata for restricted data sources will make their data more findable, the *accessibility* of these datasets could be significantly improved by structured metadata capturing data access information. Presently, metadata schemas do not accommodate the level of detail restricted data sources provide about access procedures and requirements.

Background

With the advent of the FAIR principles for data in 2016 (Wilkinson et al. 2016), researchers and data stewards were given a clear set of aspirations for enhancing the value of their data by making them as Findable, Accessible, Interoperable, and Reusable as possible. Guidelines and other resources for the “fairification” of datasets quickly followed (“FAIR Cookbook,” n.d.). While many creators and curators of open datasets in academic, government, and not-for-profit contexts are adopting these principles, restricted data have fallen behind in being made FAIR. Restricted data, in the context of this study, are defined as data with potential value for research that are only accessible via request or application.

Restricted data for research

The category of restricted data is particularly relevant to researchers in the health sciences, who regularly require access to data measuring individual and group characteristics and outcomes that are considered sensitive in nature. A recent review of restricted health data sources in Canada by the present authors (Read et al. 2024a) found that none of the sources investigated employed standardized and machine-readable descriptive metadata of any kind, and few provided information that would adequately describe the data for even a human reader. This circumstance is decidedly un-FAIR and presents significant barriers to use for anyone not already familiar with a dataset’s existence, availability, and content. In many cases, the information provided by data sources was also not sufficient to allow a researcher to assess whether pursuing access to a dataset would be appropriate in the context of their research or their role (e.g., student vs. faculty); these

issues are well documented in the literature on this topic, particularly with respect to navigating the access process for acquiring restricted health data (Bekemeier et al. 2019; Clayton et al. 2021; Hanna et al. 2021; Ho, Gorges, and Portales-Casamar 2018; Mpango and Nabukenya 2019; Pongiglione et al. 2021; Prince et al. 2018; Saulnier et al. 2019; Siu et al. 2016; Sydes et al. 2015).

Insufficient or nonexistent metadata is not an inherent characteristic of restricted access data. While datasets themselves may not be available openly, in most cases there is little need for data sources to conceal basic information that would ease the discovery of their content, and help potential users understand procedures and eligibility for access. As one study points out: “FAIR is not equal to open” (Mons et al. 2017). Rather, implementation guidelines and interpretations of the FAIR principles are agnostic as to dataset openness (Jacobsen, de Miranda Azevedo, et al. 2020; Jacobsen, Kaliyaperumal, et al. 2020), and have been applied to even very sensitive data (van der Velde et al. 2022; Ghardallou et al. 2022). A more FAIR landscape of restricted data—in health spaces and beyond—would incorporate adequate descriptive and access information in the form of structured metadata. This would allow researchers to find and evaluate restricted datasets more easily and allow systems such as data indexers and aggregators to harvest and expose these valuable resources for discovery.

To learn more about how to improve the discovery of and access to restricted data, this study explored 48 online sources of restricted health data in Canada to:

- identify what kinds of information these sources provide about datasets themselves and the procedures for accessing them;
- identify commonalities across sources to inform the creation of possible metadata elements for restricted data; and
- assess whether and to what extent these elements can be accommodated within existing metadata schemas.

This research aimed to determine gaps and opportunities in the ways that restricted data sources make their data discoverable and accessible, and in the extent to which available metadata schemas facilitate discovery and access. This manuscript was originally made available as a pre-print publication (Read et al. 2024b).

Methods

This study used an iterative, three step process to:

- 1) Extract dataset and access information provided by a sample of 48 restricted data sources that did not utilize structured metadata (Read et al. 2024a);
- 2) Identify commonalities across these data sources to develop possible metadata elements for restricted data; and

- 3) Map these metadata elements to existing metadata schemas to evaluate how well the latter currently accommodate information supplied by restricted data sources.

This study began in December 2021 and was completed in March 2023.

Step 1: Extract dataset and access information from restricted data sources

We examined the 48 restricted data sources from our previous study to identify the “dataset information” they provided to describe their datasets, and the “access information” they used to describe their access request processes and requirements. We reviewed each data source website as well as any documents we could download that provided information about the dataset or the actions required to access the dataset. Some data sources contained many datasets with varying levels of access, meaning the dataset and access information we extracted could appear more than once in a single data source. Because the information we found pertained to different datasets, we considered this information unique and included them as separate even though they originated from the same source. All dataset and access information from each data source was compiled into two spreadsheets, which are available for download in the supplemental files.

Dataset information

We defined “dataset information” as a discrete piece of information the data source provided about their datasets themselves (e.g., a dataset description, the date the data was collected, the dataset population). We extracted this information from 14 of the 48 data sources because they were the only sources that described their data to some extent.

Access information

We defined “access information” as a discrete piece of information the data source provided related to the access request process (e.g., who is eligible to access the data, the cost of the dataset, what documents are required when submitting an application). We extracted this information from all 48 data sources.

Step 2: Identify common dataset and access elements in restricted data sources

To identify commonalities across data sources, we manually grouped the dataset and access information together into meaningfully similar categories and recorded the frequency in which they appeared. Dataset and access information was considered “common” if it appeared more than once. We then ascribed definitions to the common categories that captured the nature of the information provided or requested; these categories became the “dataset elements” and “access elements” that we mapped to existing metadata schemas in step 3.

Step 3: Map common dataset and access elements to existing metadata schemas

Using our newfound common “dataset elements” and “access elements”, we mapped each element to the following metadata schemas to determine to what extent they aligned: DataCite (2021), DDI-Lifecycle

(n.d.), DDI Codebook (n.d.), DCAT Vocabulary (2023), and DATS (Sansone et al. 2017). These metadata schemas were selected because of their focus on describing data.

We performed our mapping using an “exact”, “partial”, or “none” matching system. An “exact” match was assigned if an element we identified matched exactly to an element in a metadata schema (e.g., Dataset title = Title in the DataCite Metadata Schema). A “partial” match was assigned if an element we identified had some relation to an element in a metadata schema but did not match exactly (e.g., Data Custodian = contributorType “DataManager” in the DataCite Metadata Schema). “No match” was assigned if we could find no corresponding element in a metadata schema.

Once the mapping process was complete, we recorded which of our dataset elements and access elements had the strongest alignment with elements across multiple existing metadata schemas. We designated strong alignment when we found four or more exact matches for that element across the five schemas we examined.

Results

Dataset and access information

We extracted 275 examples of dataset information in 14 data sources, and 2,059 examples of access information in 48 data sources.

Common elements

Dataset elements

From the 275 examples of dataset information, we identified 35 common “dataset elements” which were subsequently grouped into broader categories (Table 1). Within these broader categories, Data Characteristics (n=91), Access and Availability (n=34), Description (n=33), and Provenance (n=27) represented the most common elements identified across data sources. These commonalities indicate that the restricted data sources in our sample provided similar information about their datasets.

Access elements

From the 2,059 examples of access information we extracted, we identified 27 common “access elements” which are categorized and listed in Table 2. This categorization identified strong commonalities between access elements for Request Requirements, in particular Research Team Information (n=668), Research Plan (n=374), Data Management (n=215), Description of Request (n=184), and Ethics Approval (n=151). These findings indicate that the restricted data sources in our sample provided similar information with respect to their data access request processes and requirements.

Table 1: Frequency of restricted dataset element commonalities across data sources.

Metadata Category	Dataset Element	Definition	Frequency
Data characteristics			91
	Population	Population details within dataset	26
	Data quality	Data quality measures related to dataset (e.g., accuracy, completeness)	19
	Geographic	Geographic location where data was collected	12
	Variables	Variable level information about dataset	11
	Temporal, Date range	The available date range for dataset	9
	Temporal, Start date	The start date for data collection	4
	Temporal, End date	The end date for data collection	4
	Temporal, Reference period	The calendar year when data was made available for access	3
	Size of dataset	Size of dataset	3
Access and availability			34
	Data access	Access parameters for acquiring dataset	17
	Release schedule	When (and how often) data is released for access	14
	Legal, licensing, security	Legal, license, and data security parameters for dataset	3
Data description - Narrative			33
	Data description	Narrative free-text description of dataset	17
	Keyword	Keywords supplied pertaining to the subject of dataset	5
	Purpose of data collection	Background free-text information about nature of data collection	5
	Industry/sector	Industry or sector where data was collected from	4
	Study description	Narrative free-text description of study	2
Provenance			27
	Data source	The association/organization responsible for dataset	14
	Contact information	Contact information pertaining to dataset support	9

Table 1 Continued: Frequency of restricted dataset element commonalities across data sources.

Metadata Category	Dataset Element	Definition	Frequency
Provenance			27
	Data custodian	Person(s)/Organization(s) responsible for the dataset	2
	Data storage	How and where data is stored	2
Methods			24
	Study design	Narrative free-text description of the study design	16
	Data collection	Narrative free-text description of the data collection process	8
Administration			21
	Record history	Administrative details on version history of dataset	7
	Status	Current status of dataset (e.g., complete, in process)	6
	File type and format	File types and formats of dataset	5
	Granularity of data	Granularity provided within dataset (e.g., variable level, anonymization)	3
Additional information			19
	Additional information	Free-text information about the dataset not included elsewhere (e.g., related products)	13
	Classifications	Classifications/standards applied to the data (e.g., ICD-10)	4
	Publications	Publications associated with dataset	2
Title information			16
	Study title	Title of study where dataset was collected	9
	Dataset title	Title of dataset	7
Unique identifiers			10
	Study identifier	Unique identifier assigned to the study within a data source	5
	URLs	Hyperlinks associated with dataset (e.g., administrative, publications)	3
	Record identifier	Unique identifier assigned to dataset record within a data source	2

Table 2: Frequency of restricted access elements commonalities across data sources.

Metadata Category	Metadata Subcategory	Access Element	Definition	Frequency
Request requirements				1765
Research team information				668
		Requestor Name	Name of individual making a data request	212
		PI	Principal investigator overseeing project for which data is being requested	185
		Team members	Names of all individuals working on study where data would be used	96
		Primary contact	Name and contact information of the primary person responsible for managing/ stewarding the data	82
		Student info	Name and contact information of students engaging with data	50
		Educational / professional background	The educational and professional qualifications of each person who will be interacting with the data	25
		Conflict of interest	Conflicts of interest present with of the data requestors	18
Research plan				374
		Study purpose	Requestor(s) describe study purpose	113
		Study design	Requestor(s) describe design of study	82
		External linkages	Requestor(s) indicate additional data linkages that may occur with requested data	50
		Timeline	Requestor(s) provide expected length of study where data will be used	49
		Project title	Requestor(s) provide title of project where data will be used	47
		Support needed	Requestor(s) indicate data management, storage, security, or analysis support required to use data	20
		Scientific review	Requestor(s) indicate the level of scientific review their study has undergone	13
Request data description			Requestor(s) provides free-text description of what data they intend to use	184
Ethics approval				151
		Ethics review	Requestor(s) must indicate that ethics approval has been obtained	100
		Risks/benefits	Requestor(s) must list the risks and benefits of using data	34
		Participant recruitment	Requestor(s) must indicate how they will recruit participants for their study	17

Table 2 Continued: Frequency of restricted access elements commonalities across data sources.

Metadata Category	Metadata Subcategory	Access Element	Definition	Frequency
	Data management			215
		Storage and security	Data source describes how requested data will be securely stored and managed	134
		Processing	Data source indicates how data will be processed	42
		Access restrictions	Data source indicates what restrictions are placed on access of the data	39
	Funding		Data source states that prior funding must be received to access the data	96
	Intended use			48
		Request rationale	Requestor(s) must provide rationale why access to the data is necessary to carry out their study	48
	Dissemination plan		Requestor(s) must provide rationale why access to the data is necessary to carry out their study	29
		Project title	Requestor(s) provide title of project where data will be used	47
		Support needed	Requestor(s) indicate data management, storage, security, or analysis support required to use data	20
		Scientific review	Requestor(s) indicate the level of scientific review their study has undergone	13
Terms of use				134
	Sign off		Data source provides information about ethical and legal sign off for transferring access to data	96
	Legal matters		Data source outlines legal terms of access, use, and management	38
Pricing				27
	Cost to acquire data		Data source provides the costs required to access and use data	27

Mapping elements to metadata schemas

Mapping dataset elements

When mapping our common dataset elements to existing metadata schemas, we found that there was relatively strong alignment between the information our 48 restricted data sources provided and the information that metadata schemas already capture (Figure 1). Notably, the DDI Lifecycle and DDI Codebook metadata schemas had very strong alignment with our common dataset elements, for which we assigned 91.4% and 85.7% “exact” matches, respectively. The DATS (40%) and DataCite (37.1%) metadata schemas received the fewest “exact” matches. Metadata schemas that received the most “no match” equivalents included DataCite (34.3%), DCAT (20%), and DATS (14.3%).

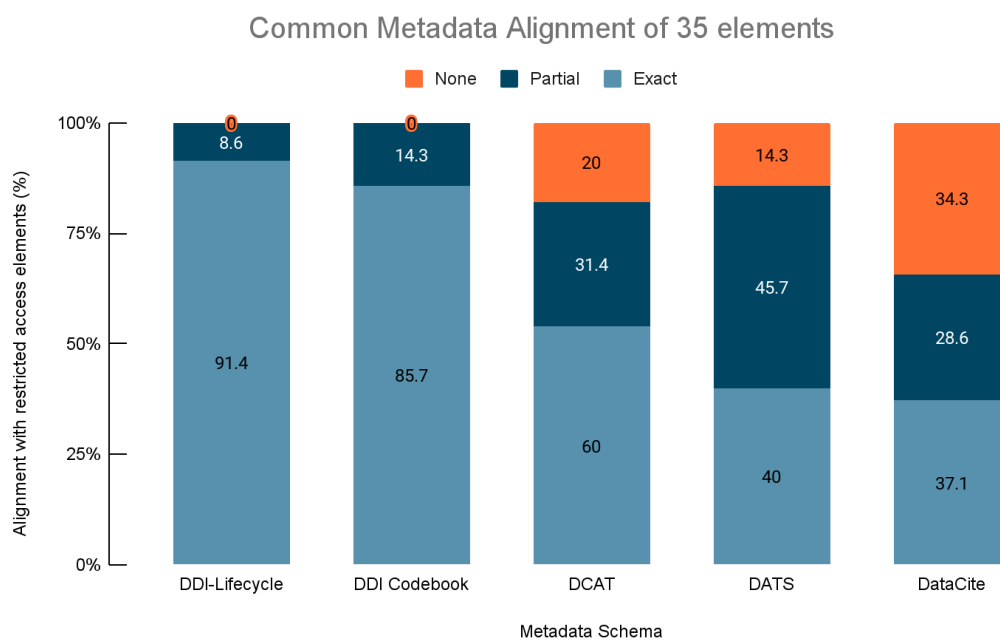


Figure 1: Dataset elements alignment with existing metadata schemas.

The dataset elements that demonstrated the strongest alignment across metadata schemas (≥ 4 schemas) were:

- Dataset title
- File type and format
- Keyword
- Additional information
- Classifications
- Record identifier

- Size of dataset
- Publications
- Legal, licensing, security considerations
- Study title
- Study identifier

These results indicate that the restricted data sources in our sample tended to capture information about their datasets that align with existing metadata schemas.

Mapping access elements

Although the 48 restricted data sources in our sample described their access procedures in similar ways, our attempt to map the common access elements we identified to existing metadata schemas was unsuccessful because no schemas accommodated the level of detail used across sources. While existing schemas include elements such as `accessRights`, `accessRestrictions`, `accessConditions`, and `licenceInformation`, our restricted data sources employ much more granular concepts when describing access processes and requirements (Table 2). Our inability to map the common access elements to any of the metadata schemas we selected suggests that these schemas are not currently suited to capturing information about access and request processes for restricted data.

Discussion

Our results suggest that the descriptive dataset information provided by restricted data sources, even where minimal, could be accommodated by an existing metadata schema, while information about access requirements and procedures could not. In practice, this means that it would be possible for many restricted health data sources to adopt an existing metadata schema, while metadata standards bodies could take action to better accommodate the access information that restricted data sources provide. If undertaken, these efforts could significantly improve the FAIR-ness of restricted access datasets by making them more findable and accessible.

Improving restricted data discovery

However, because none of these data sources utilize structured metadata, at present their discoverability is limited. Researchers locate data through prior awareness (either personally or via others), a search engine or data aggregator (e.g., Google Dataset Search, HealthData.gov, the European Data Portal, and Canada's Lunaris), or a data support professional such as a librarian (who would similarly rely on existing knowledge, searching, and data discovery infrastructure) (Krämer et al. 2021; Koesten et al. 2017). Because the data sources in this study did not make use of structured metadata, at present their "visibility" to search engines and aggregators is, respectively, limited or non-existent. Many of the data sources in our study shared

commonalities in the ways of how they described their datasets, and these commonalities align well with existing metadata schemas. If the data sources in this study—and restricted data sources generally—were to take advantage of their “metadata readiness” by adopting an existing metadata schema, it would significantly enhance researchers’ ability to find them using modern discovery infrastructure.

For restricted data sources to adopt metadata standards, data stewards will need an understanding of metadata and its value, and the resources necessary to adopt and implement it. This may include an understanding of how difficult it is to find their datasets currently, why metadata improves discoverability, how to apply a metadata standard to their data, and where their datasets will be findable once that is achieved. It may also include financial or human resources. Many of the data sources we identified in this study represented small, one-time projects that presumably do not have large teams of personnel to take on this task. If national funding bodies or data initiatives are interested in improving the discovery of restricted data sources, we recommend that they incentivize, support, and help implement the adoption of metadata among restricted data sources. For example, because of this work, the Digital Research Alliance of Canada (Alliance) is developing a strategy to work with data sources included in this study to adopt a metadata schema, with the goal of making them findable in Canada’s data discovery index, Lunaris leveraging its Network of Experts (Digital Research Alliance of Canada 2022b). Organizations and initiatives in other geographical or disciplinary contexts could offer similar support to restricted data sources in their areas.

While the adoption of metadata standards by sources will be an important step toward improving the discovery of restricted data, additional challenges remain. To date, restricted data is generally not registered with a persistent identifier such as a DOI or handle. However, because any prospective restricted metadata record would not be a direct access point to the data itself, the potential utility of such a practice for improving data discovery is not always clear to data stewards. Greater awareness is needed of the utility of persistent identifiers—and, their accompanying open metadata, which is widely used by data aggregators—for improving the discovery of restricted data. Future directions for this research should consider both the discovery of restricted data and the preservation of that data for the long term. Questions we posit include: what are the recommended practices for assigning DOIs to open metadata records for restricted datasets?; and is it feasible to create national infrastructure that can reliably make restricted datasets discoverable in one location but maintain its security and preservation in another? One emerging initiative attempting to address these issues is the Alliance’s Controlled Access Management project, which aims to enable researchers, institutions and repositories to manage controlled access to research data through new software, tools and workflows (Digital Research Alliance of Canada 2022a).

Improving restricted data access

While adding descriptive dataset metadata to restricted data sources will make data more findable, the *accessibility* of these datasets could be significantly improved by structured metadata capturing information and requirements related to data access (“access metadata”). The widespread adoption of access metadata

for restricted data would help researchers understand, at a glance, whether a particular dataset is worth pursuing (e.g., timeline to access, eligibility requirements, access requirements, cost). Within our sample, access information was presented inconsistently and in an unstructured way (if at all) or was only presented during the request process itself (e.g., while populating and clicking through a multi-step application) (Read et al. 2024a). Ideally, when a researcher examines a dataset record in a repository, catalog, or index, they would be able to understand not only its contents and characteristics, but also the restrictions, procedures, and conditions for use (Figure 2). Access metadata could also be used from a search perspective to help researchers filter for datasets for which access will be achievable in their particular context. For example, a graduate student seeking health data that was free of charge and which they could access within two months could quickly identify datasets that satisfy these criteria.

Dataset Metadata	Access Metadata
<p>Patient Centered Measurement: The Acute Patient Survey 2022</p> <p>Creator British Columbia Office for Patient Centered Measurement (BCOPCM)</p> <p>Data sources 78 Acute Care Hospitals 2 freestanding rehabilitation hospitals 6 health authorities</p> <p>Date range 2022-09-01 - 2023-03-01</p> <p>Description The Acute Inpatient Survey 2022 asked patients about their health-related quality of life and their experiences with the quality of care and services received as an inpatient in one of 78 acute care hospitals and two freestanding hospitals in British Columbia.</p> <p>Size of dataset 500 MB</p> <p>File type .csv</p>	<p>Access eligibility criteria Must have current Tri-agency grant funding to accommodate the access and management of the data.</p> <p>Access requirements</p> <ul style="list-style-type: none"> • Ethics approval • Project proposal • Dissemination plan <p>Conditions of use Data can only be accessed on virtual environments made available by the BCOPCM.</p> <p>Timeline to access Expect an application to take 4-6 weeks to be processed from the time of submission.</p> <p>Cost of data</p> <ul style="list-style-type: none"> • Access: \$259 • Additional support needed from data provider: \$40/hour

Figure 2: Mock restricted data record with access metadata

At present, metadata schemas do not accommodate the level of detail restricted data sources commonly provide about access procedures and requirements, instead typically providing one free-text element to capture all related information (e.g., accessRights, accessRestrictions). With only one metadata element devoted to access, restricted data sources lack a template for providing the types of information important to researchers when evaluating a dataset for use. Additionally, this single-element approach precludes the ability of access characteristics to be filtered in discovery infrastructure. To improve the availability, findability, and clarity of access-level details for researchers, we strongly recommend that metadata standards bodies expand their schemas to accommodate more detail about dataset access, in alignment with the elements identified in this study. While metadata schemas do not currently accommodate restricted data sources, we acknowledge that libraries have developed data catalogues with custom access metadata as a stopgap to help researchers locate restricted data (Yee et al. 2023).

Future directions

This study examined dataset and access information supplied by restricted health data sources that are stewarded by academic, government, and non-profit organizations in the Canadian context. We believe that our findings can inform efforts to develop access-specific extensions to existing metadata schemas, where the commonalities identified in this study could serve as the basis for new metadata elements. The addition of access-specific metadata to existing schemas could then support existing data repository infrastructure to ensure that restricted data can be discoverable and that the access process is transparent. Data catalogues and indices could then also aggregate restricted datasets alongside those that are made public. Finally, adding access metadata to existing standards can support individual researchers who are collecting restricted data and are required to comply with emerging funder and publisher data sharing policies. These researchers would benefit from clear and consistent restricted access metadata standards, which could allow them to make their research data more discoverable and accessible either via a repository, a publication data availability statement, or institutional/organizational data source.

Data Availability

Datasets that include the dataset and access information extraction process, metadata commonality categorization, metadata definitions, and the metadata mapping are available in the Open Science Framework at: <https://osf.io/txrve>.

Competing Interests

The authors declare that they have no competing interests.

References

- Bekemeier, Betty, Seungeun Park, Uba Backonja, India Ornelas, and Anne M. Turner. 2019. "Data, Capacity-Building, and Training Needs to Address Rural Health Inequities in the Northwest United States: A Qualitative Study." *Journal of the American Medical Informatics Association: JAMIA* 26 (8–9): 825–834. <https://doi.org/10.1093/jamia/ocz037>.
- Clayton, Gemma L., Daisy Elliott, Julian P. T. Higgins, and Hayley E. Jones. 2021. "Use of External Evidence for Design and Bayesian Analysis of Clinical Trials: A Qualitative Study of Trialists' Views." *Trials* 22 (1): 789. <https://doi.org/10.1186/s13063-021-05759-8>.
- "Data Catalog Vocabulary (DCAT)." 2023. Metadata Schema. <https://www.w3.org/TR/vocab-dcat-3>.
- DataCite. 2021. "DataCite Metadata Schema." Website. DataCite Schema. March 30, 2021. <https://schema.datacite.org>.
- "DDI-Codebook | Data Documentation Initiative." n.d. Accessed December 15, 2023. <https://ddialliance.org/Specification/DDI-Codebook>.
-

“DDI-Lifecycle | Data Documentation Initiative.” n.d. Accessed December 15, 2023.
<https://ddialliance.org/Specification/DDI-Lifecycle>.

Digital Research Alliance of Canada. 2022a. “Canadian Digital Research Infrastructure Investment Overview 2023-25.”
<https://alliancecan.ca/sites/default/files/2023-03/2023-25%20DRI%20Investment%20Overview.pdf>.

———. 2022b. “Network of Experts.” 2022.
<https://alliancecan.ca/en/services/research-data-management/network-experts>.

“FAIR Cookbook.” n.d. Accessed July 17, 2023.
<https://fairplus.github.io/the-fair-cookbook/content/home.html>.

Ghardallou, Mariem, Morgane Wirtz, Sakinat Folorunso, Zohra Touati, Ezekiel Ogundepo, Klara Smits, Ali Mtiraoui, and Mirjam van Reisen. 2022. “Expanding Non-Patient COVID-19 Data: Towards the FAIRification of Migrants’ Data in Tunisia, Libya and Niger.” *Data Intelligence* 4 (4): 955–970.
https://doi.org/10.1162/dint_a_00181.

Hanna, Catherine R., Elizabeth Lemmon, Holly Ennis, Robert J. Jones, Joy Hay, Roger Halliday, Steve Clark, Eva Morris, and Peter Hall. 2021. “Creation of the First National Linked Colorectal Cancer Dataset in Scotland: Prospects for Future Research and a Reflection on Lessons Learned.” *International Journal of Population Data Science* 6 (1): 1654. <https://doi.org/10.23889/ijpds.v6i1.1654>.

Ho, Hoi Ki Kiki, Matthias Gorges, and Elodie Portales-Casamar. 2018. “Data Access and Usage Practices Across a Cohort of Researchers at a Large Tertiary Pediatric Hospital: Qualitative Survey Study.” *JMIR Medical Informatics* 6 (2): e32. <https://doi.org/10.2196/medinform.8724>.

Jacobsen, Annika, Rajaram Kaliyaperumal, Luiz Olavo Bonino da Silva Santos, Barend Mons, Erik Schultes, Marco Roos, and Mark Thompson. 2020. “A Generic Workflow for the Data FAIRification Process.” *Data Intelligence* 2 (1–2): 56–65. https://doi.org/10.1162/dint_a_00028.

Jacobsen, Annika, Ricardo de Miranda Azevedo, Nick Juty, Dominique Batista, Simon Coles, Ronald Cornet, Mélanie Courtot, et al. 2020. “FAIR Principles: Interpretations and Implementation Considerations.” *Data Intelligence* 2 (1–2): 10–29. https://doi.org/10.1162/dint_r_00024.

Koesten, Laura M., Emilia Kacprzak, Jenifer F. A. Tension, and Elena Simperl. 2017. “The Trials and Tribulations of Working with Structured Data: -a Study on Information Seeking Behaviour.” In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 1277–1289. CHI ’17. New York, NY, USA: Association for Computing Machinery. <https://doi.org/10.1145/3025453.3025838>.

Krämer, Thomas, Andrea Papenmeier, Zeljko Carevic, Dagmar Kern, and Brigitte Mathiak. 2021. “Data-Seeking Behaviour in the Social Sciences.” *International Journal on Digital Libraries* 22 (2): 175–195.
<https://doi.org/10.1007/s00799-021-00303-0>.

Mons, Barend, Cameron Neylon, Jan Velterop, Michel Dumontier, Luiz Olavo Bonino da Silva Santos, and Mark D. Wilkinson. 2017. “Cloudy, Increasingly FAIR; Revisiting the FAIR Data Guiding Principles for the European Open Science Cloud.” *Information Services & Use* 37 (1): 49–56.
<https://doi.org/10.3233/ISU-170824>.

Mpango, Jonathan and Josephine Nabukenya. 2019. “A Qualitative Study to Examine Approaches Used to Manage Data about Health Facilities and Their Challenges: A Case of Uganda.” *AMIA Annual Symposium Proceedings*. *AMIA Symposium 2019* (101209213): 1157–1166.

- Pongiglione, Benedetta, Aleksandra Torbica, Hedwig Blommestein, Saskia de Groot, Oriana Ciani, Sarah Walker, Florian Dams, et al. 2021. "Do Existing Real-World Data Sources Generate Suitable Evidence for the HTA of Medical Devices in Europe? Mapping and Critical Appraisal." *International Journal of Technology Assessment in Health Care* 37 (1): e62. <https://doi.org/10.1017/S0266462321000301>.
- Prince, Karl, Matthew Jones, Alan Blackwell, Alexander Simpson, Sallyanne Meakins, and Alain Vuylsteke. 2018. "Barriers to the Secondary Use of Data in Critical Care." *Journal of the Intensive Care Society* 19 (2): 127–131. <https://doi.org/10.1177/1751143717741082>.
- Read, Kevin B., Grant Gibson, Amber Leahey, Lynn Peterson, Sarah Rutley, Julie Shi, Victoria Smith, and Kelly Stathis. 2024a. "Understanding the Challenges Associated with Finding and Accessing Restricted Data in Canada: A Mixed Methods Study." *FACETS*. 9(): 1-9. <https://doi.org/10.1139/facets-2023-0102>.
- . 2024b. "Identifying Metadata Commonalities Across Restricted Health Data Sources: A Mixed Methods Study Exploring How to Improve the Discovery of and Access to Restricted Datasets." OSF Preprints [Preprint]. May 6, 2024. <https://doi.org/10.31219/osf.io/pujry>.
- Sansone, Susanna-Assunta, Alejandra Gonzalez-Beltran, Philippe Rocca-Serra, George Alter, Jeffrey S. Grethe, Hua Xu, Ian M. Fore, et al. 2017. "DATS, the Data Tag Suite to Enable Discoverability of Datasets." *Scientific Data* 4 (June): 170059. <https://doi.org/10.1038/sdata.2017.59>.
- Saulnier, Katie M., David Bujold, Stephanie O. M. Dyke, Charles Dupras, Stephan Beck, Guillaume Bourque, and Yann Joly. 2019. "Benefits and Barriers in the Design of Harmonized Access Agreements for International Data Sharing." *Scientific Data* 6 (1): 1–6. <https://doi.org/10.1038/s41597-019-0310-4>.
- Siu, Lillian L., Mark Lawler, David Haussler, Bartha Maria Knoppers, Jeremy Lewin, Daniel J. Vis, Rachel G. Liao, Fabrice Andre, Ian Banks, and J. Carl Barrett. 2016. "Facilitating a Culture of Responsible and Effective Sharing of Cancer Genome Data." *Nature Medicine* 22 (5): 464–471. <https://doi.org/10.1038/nm.4089>.
- Sydes, Matthew R., Anthony L. Johnson, Sarah K. Meredith, Mary Rauchenberger, Annabelle South, and Mahesh K. B. Parmar. 2015. "Sharing Data from Clinical Trials: The Rationale for a Controlled Access Approach." *Trials* 16 (1): 1–6. <https://doi.org/10.1186/s13063-015-0604-6>.
- Velde, K. Joeri van der, Gurnoor Singh, Rajaram Kaliyaperumal, XiaoFeng Liao, Sander de Ridder, Susanne Rebers, Hindrik H. D. Kerstens, et al. 2022. "FAIR Genomes Metadata Schema Promoting Next Generation Sequencing Data Reuse in Dutch Healthcare and Research." *Scientific Data* 9 (1): 169. <https://doi.org/10.1038/s41597-022-01265-x>.
- Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, and Philip E. Bourne. 2016. "The FAIR Guiding Principles for Scientific Data Management and Stewardship." *Scientific Data* 3. <https://doi.org/10.1038/sdata.2016.18>.
- Yee, Michelle, Alisa Surkis, Ian Lamb, and Nicole Contaxis. 2023. "The NYU Data Catalog: A Modular, Flexible Infrastructure for Data Discovery." *Journal of the American Medical Informatics Association* 30 (10): 1693–1700. <https://doi.org/10.1093/jamia/ocad125>.