

Purdue University  
**Purdue e-Pubs**

---

Libraries Faculty and Staff Scholarship and  
Research

Purdue Libraries and School of Information  
Studies

---

5-3-2017

## Participatory Prototype Design: Developing a Sustainable Metadata Curation Workflow for Maternal Child Health Research

Amanda Harrigan  
*University of Alberta*

Saurabh Vashishtha  
*University of Alberta*

Sharon Farnel  
*University of Alberta*

Kendall Roark  
*Purdue University, roark6@purdue.edu*

Follow this and additional works at: [https://docs.lib.purdue.edu/lib\\_fsdocs](https://docs.lib.purdue.edu/lib_fsdocs)



Part of the [Cataloging and Metadata Commons](#), [Medicine and Health Sciences Commons](#), and the [Social and Cultural Anthropology Commons](#)

---

### Recommended Citation

Harrigan, Amanda, Saurabh Vashishtha, Sharon Farnel, and Kendall Roark. (2018), "Participatory Prototype Design: Developing a Sustainable Metadata Curation Workflow for Maternal Child Health Research." *International Journal of Data Curation* 2018, Vol. 13, Iss. 1, 248–270.

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact [epubs@purdue.edu](mailto:epubs@purdue.edu) for additional information.

# Participatory Prototype Design: Developing a Sustainable \* Metadata Curation Workflow for Maternal Child Health \* Research \*

Amanda Harrigan  
University of Alberta

Saurabh Vashishtha  
University of Alberta

Sharon Farnel  
University of Alberta

Kendall Roark  
Purdue University

## Abstract

This paper describes the findings from a participatory prototype design project, where the authors worked with maternal and child health (MCH) researchers and stakeholders to develop a MCH metadata profile and sustainable curation workflow. This work led to the development of three prototypes: 1) a study catalogue hosted in Dataverse, 2) a metadata and research records repository hosted in REDCap and 3) a metadata harvesting tool/dashboard hosted within the Shiny RStudio environment. We present a brief overview of the methods used to develop the metadata profile, curation workflow and prototypes. Researchers and other stakeholders were participant-collaborators throughout the project. The participatory process involved a number of steps, including but not limited to: initial project design and grant writing; scoping and mapping existing practices, workflows and relevant metadata standards; creating the metadata profile; developing semi-automated and manual techniques to harvest and transform metadata; and end project sustainability/future planning. In this paper, we discuss the design process and project outcomes, limitations and benefits of the approach, and implications for researcher-oriented metadata and data curation initiatives.

*Received 21 September 2017 ~ Revision received 2 May 2018 ~ Accepted 3 May 2018*

Correspondence should be addressed to Kendall Roark, Purdue University Libraries, 504 W. State Street, West Lafayette, IN 47907 Email: roark6@purdue.edu

The *International Journal of Digital Curation* is an international journal committed to scholarly excellence and dedicated to the advancement of digital curation across a wide range of sectors. The IJDC is published by the University of Edinburgh on behalf of the Digital Curation Centre. ISSN: 1746-8256. URL: <http://www.ijdc.net/>

Copyright rests with the authors. This work is released under a Creative Commons Attribution Licence, version 4.0. For details please see <https://creativecommons.org/licenses/by/4.0/>



## Introduction

In recent years, there has been a call for more participatory and community driven approaches to cataloguing and metadata schema creation within the context of digital archives and libraries (Bowler et al., 2011; Farnel et al., 2017) and data repositories (Yarney and Baker, 2013; Michener et al., 2012). The CLIR funded ‘Bridging the Research Data Divide’ (BRDD) project is informed by this call as well as by a desire to bridge the gap between archival and data repository approaches to scientific data curation. As part of the project, the University of Alberta Libraries (UAL) and Harvard’s Center for the History of Medicine (CHoM) described and made accessible maternal and child health (MCH) research data and the contextual records that enable their long-term access, security, and reuse. While CHoM processed and described archival and manuscript MCH research data collections, UAL processed and described contemporary, born-digital MCH research studies and records.<sup>1</sup> At the end of the project, findings from CHoM and UAL were compared to come up with a metadata element set suitable to describe MCH research data throughout its lifecycle, from active research to preservation in the archives.

This paper focuses on the processes that UAL developed to describe a selection of MCH research studies conducted by members of the Women and Children’s Health Research Institute (WCHRI)<sup>2</sup>. WCHRI, housed at the University of Alberta, is a partnership between the University of Alberta and Alberta Health Services, with core funding from the Stollery Children’s Hospital Foundation (SCHF) and the Royal Alexandra Hospital Foundation (RAHF). It supports research dedicated to improving the health and lives of women and children.

MCH is a health research field that focuses on the specific population of mothers and children, as well as adolescents, families and pregnant women, rather than a particular research method. WCHRI has a wide MCH research mandate, including healthy development and children’s health and well-being. Researchers work on both discovery research in a laboratory-setting and clinical and translational research in integrated hospitals and communities. Research is often conducted by teams with investigators based at multiple sites across Canada or internationally.

The pilot explored how to make it easier for WCHRI members to describe and share their data and documentation so that potential secondary users can easily discover them. The aim was to help researchers in the WCHRI network reduce duplication of research, make their research more visible, and promote collaborations among different groups with similar research interests.

It cannot be overlooked that sharing data takes significant time, labour and money. One researcher estimated that preparing and depositing his scientific data for a single publication took upwards of ten hours (Bruna, 2014). Data objects, alone, do not have enough information to contextualize them. Data becomes much more useful if metadata is applied, and metadata become far more valuable when maintained in standard and machine-readable ways. Creating the kind of robust, consistent and standardized metadata necessary to enable the discovery, access, reuse, linkage and preservation of data is particularly time-intensive and can be seen as an added burden by researchers (Borgman, 2008; Crystal and Greenberg, 2005; Frey, 2008). Researchers can

<sup>1</sup> For more see BRDD project page: <http://scalar.usc.edu/works/bridging-the-research-data-divide/index>

<sup>2</sup> Women and Children’s Health Research Institute (WCHRI): <https://www.wchri.org>

underestimate the value of adding metadata and its rewards, and may not know where to start or may be worried about making mistakes when applying metadata (Willoughby et al., 2014). In Federer, Lu, and Joubert's (2016) study of the data literacy training needs of biomedical researchers, median results from 190 researchers show that metadata skills, defined as the ability to "[c]apture and create metadata (descriptive information about your data, how it was collected, and other contextualizing information," were ranked as of high relevance (4), but expertise was self-ranked as only medium (3). Similarly, recent results of the Data Curation Network project's research engagement sessions indicate several gaps in support in data curation services (Johnston et al., 2017). Among this is creating and or applying metadata. Although 62.5% responded "Yes this happens" to the question of whether metadata is created, only 29% were satisfied. The report recommends that "better tools and or best practices might be welcome" to encourage more, and more satisfying, metadata creation (Johnston et al., 2017).

In addition, many of these activities necessitate fairly compensated staff even in cases when metadata creation may be partially or more fully automated. To this aim grant funding was used almost entirely to support two-year project hires: one full-time Metadata Curation Specialist (Amanda Harrigan) and a half-time Data Curation Specialist (Saurabh Vashistha). They handled most of the day-to-day intellectual and technical work associated with the project, including working directly with researchers to develop the metadata schema and data publishing workflow, and establishing the semi-automated workflows and processes.

Developing metadata guidelines and tools around specific research data is best done in concert with the researchers who will be expected to use such guidance. We sought researchers' participation in defining the appropriate elements, boundaries and level of granularity required of metadata for their research studies and data. The practical and theoretical underpinnings of this approach are informed by the principles of participatory design. The concept of participatory design rests on the user's involvement in developing effective and reasonable process change within their existing work practices and environment (Spinuzzi, 2005). This approach can be ideal for creating a bridge between researchers, metadata specialists and data curators so they can come to a consensus on what information will be necessary to describe research data.

The end goal was to create usable and sustainable metadata recommendations based on the feedback and needs of this specific MCH research community. To this end, we consulted with MCH researchers to develop and refine a platform-agnostic metadata schema, describe the studies, and display the metadata through three different prototypes. A REDCap repository was created and used to both develop and manage the metadata and research records that were collected for the project. A publicly accessible Dataverse-based prototype study catalogue<sup>3</sup> and potential data-sharing platform was also produced. As well, we created a Metadata Harvester Dashboard application to demonstrate the methodology and processes that went into harvesting the metadata. It is currently hosted on the Shiny RStudio site (RStudio Team, 2015).

While this paper will highlight the participatory approach involved in the process and prototype development, a range of stakeholders were also engaged throughout the project planning stage. For example, over 40 University of Alberta health science researchers, librarians and data managers helped map University of Alberta data flows, and to identify potential high-priority data sets and stakeholders seeking to develop sustainable data curation services and solutions (Roark, 2014; 2016). University of

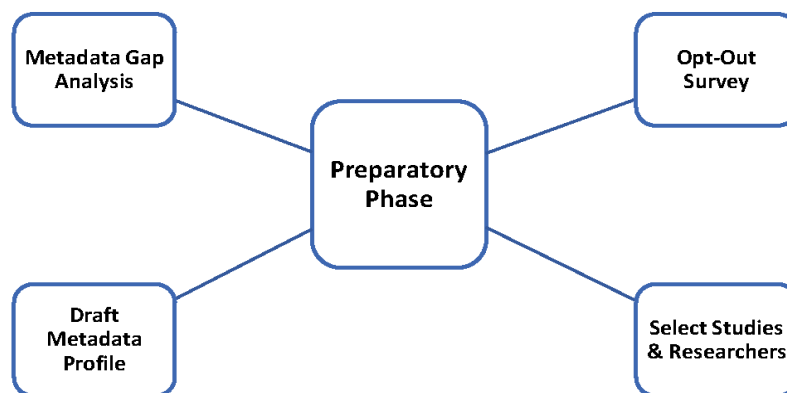
<sup>3</sup> Women and Children's Health Research Institute (WCHRI) Dataverse:  
<https://dataverse.library.ualberta.ca/dataverse/UAL-WCHRI>

Alberta affiliated faculty and administrative staff gave feedback on grant proposals and advocated for the project with administrative staff and faculty members.

The specific work reported in this paper focuses on the development of a pilot process and study data catalogue developed in collaboration with one of these stakeholder organizations, WCHRI. This was a multistage process, requiring a considerable amount of front-end work before meeting with researchers, including identifying potential studies and study types; assessing, mapping and repurposing existing metadata schemas; identifying data models and potential sources of metadata to harvest. This multi-pronged approach, involving both automating techniques and human-centred work, helped us build a usable and sustainable metadata profile.

## Workflow

The University of Alberta project team work was made up of several overlapping phases, which we have here broken into two workflows including a preparatory phase (Figure 1) and the metadata and document harvesting and creation phase (Figure 2).



**Figure 1.** Preparatory Phase.

### Survey and Map Relevant Metadata Standards

The project used the UAL instance of Dataverse (version 4.5.1) as the public platform to expose the studies, and Dataverse DDI/DC-based schema as a jumping off point for description. Dataverse is “an open source web application to share, preserve, cite, explore, and analyze research data.”<sup>4</sup> Dataverse metadata is based on the DDI schema, which is designed for and most suited to social science data. To assess whether these metadata elements were sufficient and what other fields might be required for the discovery, use and preservation of MCH research, several other schemas, standards, repositories and vocabularies, including those specific to data and clinical and health research, were consulted.

<sup>4</sup> Dataverse: <https://dataverse.org/about>

The Metadata Curation Specialist developed crosswalks between MCH relevant metadata standards and DDI/DC elements available in Dataverse. This work began with gathering a list of descriptive metadata standards and repositories, including looking at repositories listed in the registry of global repositories Re3data.org<sup>5</sup> and standards in the life sciences, broadly covering biological, natural and biomedical sciences listed in BioSharing.org, to identify metadata standards in use to describe health research and research data. This led to a list of over twelve disciplinary metadata standards for further, detailed evaluation (Table 1).<sup>6</sup>

**Table 1.** List of mapped metadata standards.

| Name of Metadata Schema, Standard or Model                                  | Website              |
|---|----------------------|
| General Standards   |                      |
| Dublin Core Metadata Terms  | <a href="#">Link</a> |
| Research Data Standards   |                      |
| Data Cite   | <a href="#">Link</a> |
| DDI-Codebook 2.5  | <a href="#">Link</a> |
| Health Research Standards   |                      |
| CONSORT Transparent Reporting of Trials 2010 Checklist                      | <a href="#">Link</a> |
| Clinicaltrials.gov element set  | <a href="#">Link</a> |
| CDISC Protocol Representation Model (PRM)                                   | <a href="#">Link</a> |
| DATS (DataTag Suite)  | <a href="#">Link</a> |
| Health Canada's Clinical Trials Database (HCCTD)                            | <a href="#">Link</a> |
| ISRCTN registry   | <a href="#">Link</a> |
| MICYRN <sup>7</sup> Birth Cohort Inventory                                  | <a href="#">Link</a> |
| SPIRIT (Standard Protocol Items: Recommendations for Interventional Trials) | <a href="#">Link</a> |
| WHO ICTRP Data Set  | <a href="#">Link</a> |

It was important to take the time to review and learn about each of the metadata standards before any mapping or cross walking began. Evaluating metadata elements involved gaining an understanding of the inherent differences between the standards, such as the granularity and discipline-specific language used to define elements. In some cases, it was straightforward to gather full field-level metadata for the different standards, while in other cases it required additional investigation including retrieval of sample metadata from known repositories and standards organization's websites.

<sup>5</sup> \* Re3data Registries of Research Data Repositories: <http://www.re3data.org/>

<sup>6</sup> \* Additional schema, standards, and vocabularies that were consulted include: W3C HCLS Dataset Description: <https://www.w3.org/TR/hcls-dataset/>; Ontology for Biomedical Investigations (OBI): <http://obi-ontology.org/>; Global Alliance for Genomics and Health: Data Use Categories and Requirements (Consent Codes) (Dyke et al., 2016); ODRL Vocabulary & Expression <http://schema.theodi.org/odrs/>; and CDISC Questionnaires, Ratings and Scales (QRS) CV: <https://www.cdisc.org/foundational/qrs>.

<sup>7</sup> \* Maternal, Infant, Child and Youth Research Network: <http://micyrn.ca/>

After the initial mapping was completed for each of the standards, they were combined into a document to provide a high-level mapping of all the disciplinary and general metadata standards to the Dataverse DDI/DC metadata schema.<sup>8</sup> Common metadata elements from the twelve schemas were mapped to the appropriate Dataverse fields, when possible. Gathering similar metadata elements in this mapping provided initial guidance about what information is most important to capture. It was helpful for determining what elements are core across all standards and identifying gaps in the Dataverse metadata schema.

Often, exact study and data collection dates that were inconsistent between different forms of documentation, such as between *initial* project documentation (grant applications and protocols) and *final* reporting to trial registries and within subsequent publications. This highlights the importance of connecting with the researcher to correct outdated or incorrect information, fill in gaps and provide context.

### Metadata Gap Analysis

Surveying a variety of metadata standards and schemas revealed a number of elements that appeared important to understand MCH research data but that are not currently captured in Dataverse metadata. The process of identifying gaps in Dataverse metadata was further informed by meetings with members of the project's Community Advisory Committee, who shared their expertise in a number of areas, including health research privacy, clinical research, and data sharing; discussions with community stakeholders, such as PolicyWise<sup>9</sup>; and through participation in the UAL Research Data Management Working Group. Through this work, a number of areas were identified that were deemed important to supplement the Dataverse-based metadata in the initial draft. Indeed, some work has been undertaken to tailor DDI to the needs of clinical researchers (Johnson and Radler, 2018; Radler and Johnson, 2014).

For example, more robust description of the conditions and restrictions around accessing and using data was deemed necessary. In particular, it was clear that more standardized ways to best describe access and use procedures for health research datasets, including associated biological materials, and communicating the availability of de-identified versions of datasets or the procedures necessary to obtain access, would likely be useful to researchers. Several further standards and practices, such as the Data Tags project, HIPAA, the Canadian Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans, and the Global Alliance for Genomics and Health, were consulted to help come up with standard definitions and language. This was then incorporated into the draft metadata form to help researchers fill out the terms and access fields in Dataverse.

Elements specific to clinical trials, such as description of outcomes measures and intervention information, were considered important to include for some researchers.<sup>10</sup> See the UAL/CHoM Joint Metadata Profile Appendix for a list of these elements.<sup>11</sup>

As well, more familiar terms were suggested to replace at times unclear DDI metadata elements. For example, several researchers considered inclusion/exclusion

---

<sup>8</sup> High-level mapping metadata standards to the Dataverse DDI/DC metadata schema: <https://bit.ly/2HFItvI>

<sup>9</sup> PolicyWise: <https://policywise.com/>

<sup>10</sup> ClinicalTrials – Protocol Registration Data Element Definitions for Interventional and Observational Studies: <https://prsinfo.clinicaltrials.gov/definitions.html>

<sup>11</sup> UAL/CHoM Joint Metadata Profile Appendix: <https://bit.ly/2rc08Ah>



criteria to be more meaningful information to include than the term ‘Universe’ or ‘Population’ to describe the group that was studied.

### **Develop Draft Metadata Profile/Collection Form**

This initial research, community consultation and metadata assessment work informed the development of a draft metadata profile/metadata collection instrument. Forms were created and data were collected and managed using REDCap electronic data capture tools hosted by WCHRI at the University of Alberta. REDCap (Research Electronic Data Capture) is a secure, web-based application designed to support data capture for research studies (Harris et al., 2009). We used REDCap for several reasons. First, we wanted to fit in with already established research data management workflows and systems. REDCap is licensed and maintained by WCHRI (on behalf of the University of Alberta), and many WCHRI researchers already use it to manage their research projects and collect data. Secondly, REDCap also gave us a chance to validate the metadata with researchers before creating a public Dataverse record, which comes with an automatically generated DOI (digital object identifier). Once a DOI is created, it is not deleted, but rather deaccessioned leaving a public record. Third, REDCap also already has survey functionality, which helped when communicating with researchers. Fourth, the data dictionary function and API allowed us to import data from external sources, and easily manipulate, log changes, move and interact with the metadata we collected in helpful ways. Lastly, the use of REDCap allowed the iterative development of the metadata schema, while still maintaining versions of previous data. Although REDCap was a useful tool to develop the metadata schema and describe and manage the research studies selected for this pilot, there were also a number of limitations. The main limitation of using REDCap for metadata collection is that the forms have to include a maximum number of potential fields for repeatable elements, which creates cumbersome .csv files with potentially many blank fields.

The first draft of the metadata application profile was an attempt to meet the initial objective of a set of elements suitable to describe studies by our MCH research community. It included the full Dataverse 4.0 metadata as well as several other elements identified during the gap analysis. In keeping with participant design, this was not intended as a perfect or final solution. The idea was that the first draft would be fed back to the researchers, and then modified with feedback in order to come up with the final schema.

### **Select Studies and Researchers**

The project attempted to describe a representative selection of the types of research conducted by WCHRI-affiliated researchers, with the majority of the studies described being randomized controlled trials and observational, prospective cohort studies. Other types of research studies, such as a systematic review and knowledge translation study, were also included. As well as touching on the diverse sorts of research that WCHRI members are doing, these studies also represent a variety of statuses and conditions: some of the studies have been completed, some are still ongoing, and some are still recruiting participants. As well, since terminated or withdrawn studies can be important for other researchers to be aware of in terms of collaboration and reducing duplication of work, a terminated trial was also described and included in the catalogue.

In order to come up with a selection of researchers and studies to work with, we first contacted all 413 WCHRI members directly through a REDCap survey. The survey



simply described the project and gave researchers the opportunity to pre-emptively “opt-out” of being contacted further about the project. Only two researchers indicated that they would like to opt out. Several researchers also explicitly expressed interest in participating. We set about coming up with the collection of studies for the catalogue from the researchers who did not opt out of the project. The majority of researchers and studies were identified for inclusion by comparing ClinicalTrials.gov against a list of WCHRI researchers.<sup>12</sup> We purposively over selected clinical trials for drugs and devices, as these types of studies often entail ongoing institutional reporting and archival responsibilities. This process was repeated until we had the desired number of studies to describe for this pilot. A total of 27 participants completed this study, and 38 studies were described.

### **Pre-populate Record with Publicly Available Information**

In order to maintain fruitful relationships with researchers it is important to understand their busy schedules and respect their time (Crystal and Greenberg, 2005; Federer et al., 2015; Federer et al., 2016; Johnston et al., 2017; Read et al., 2015). Clinical researchers are very busy people, so to be most efficient with their time we pre-populated the metadata record as much as possible before contacting them. To that end, we developed semi-automated processes for harvesting, transforming and repurposing already available sources of metadata from identified sources to streamline metadata production. Semi-automatic metadata creation involves a combination of software/programming and manual human processes (Park and Lu, 2009).

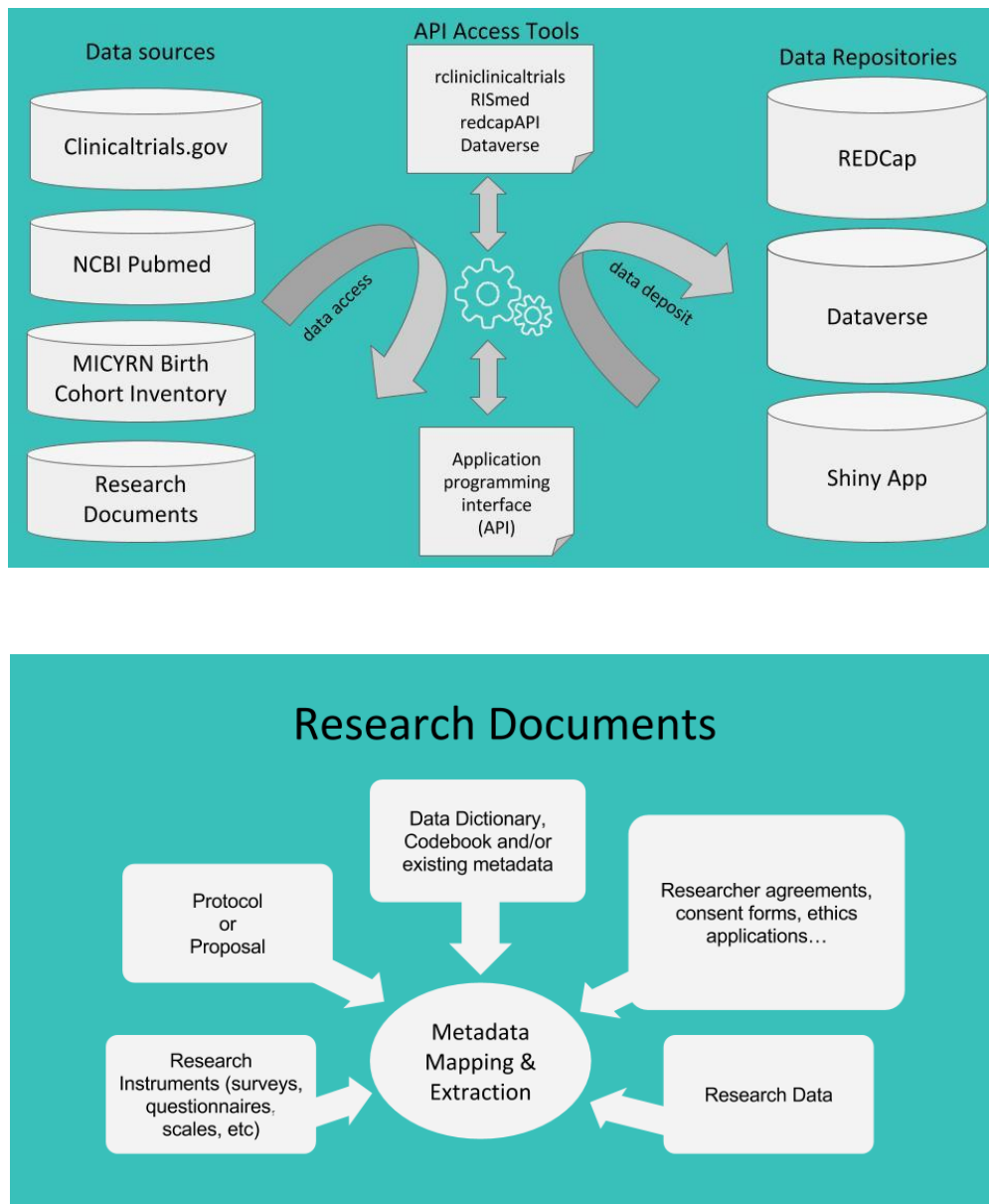
Along with the Dataverse-based metadata collection form, REDCap forms that mimic the metadata structure and elements drawn from ClinicalTrials.gov and the MICYRN Birth Cohort Inventory, which provides detailed information about Canadian birth cohort studies, were also created and used to capture already publicly available information about the studies. R-based scripts were used to pull relevant metadata via publicly available APIs into the REDCap project to populate the ClinicalTrials.gov metadata form. Specifically, an R-based script was written to extract complete metadata stored in ClinicalTrials.gov via an API in the form of multiple .csv files. An R-based package, ‘reclinicaltrials’(Sachs, 2017), was used to access the ClinicalTrials.gov API. We were also able to use the conceptual and semantic mappings done earlier in the project to import common elements from the ClinicalTrials.gov registry to our Dataverse-based schema in REDCap using APIs. Similarly, we used a package called ‘RISmed’ to access the Pubmed API for relevant publication information. This returned a refined list of publications related to each clinical trial included in the study. Unfortunately, only the primary author was reported by this tool, and so we used the corresponding information provided by the ‘RISmed’ (Kovalchik, 2017) tool and pulled additional information from NCBI Pubmed. The collected metadata was manually formatted into a single .csv file consisting of all fields mentioned in the REDCap form. And finally, a R-based script using an R-based package, ‘redcapAPI’ (Nutter and Lane, 2015), was written to submit the metadata stored in .csv files into the pre-created REDCap form.

Publications and other available information sources were manually searched for relevant metadata. Publications are good sources for identifying rich descriptive metadata about research data methodology and analysis, and can be sufficiently mapped to existing metadata standards, like DDI (Chao, 2015). The metadata drawn from these

---

<sup>12</sup> ClinicalTrials.gov: <https://clinicaltrials.gov/ct2/home>

publicly available sources was quite extensive, and provided a great deal of critical contextual information, such as titles, names and contact information of principal investigators, and related publications.



**Figure 2.** Metadata and document harvesting and creation phase. \*

### Connect with Researchers

After we populated the metadata forms as much as possible from publicly available sources, selected participants were sent another REDCap online survey. The survey included a sample of some of the metadata we had already captured to illustrate the sort of information a metadata record of their study would contain. We also hoped that sharing metadata we had already collected would encourage the researcher to participate

because he or she would see that much work had already been done. Along with this, the researchers were asked whether they were willing to share research documents, such as study protocols, grant and ethics applications, and data collection forms, so that we could use them to further fill out the metadata form. Researchers were also asked if they were willing and able to meet with us to go over the completed metadata record.

If they responded that they did not want to participate further, we sent them the full basic study record we had created from the publicly available sources as a REDCap survey. The researcher could then look it over, make additions or changes, and approve it for inclusion in the WCHRI study catalogue Dataverse. Six researchers checked and validated the metadata record in REDCap but were not able or willing to meet. Although this was not ideal, it gave an indication of what level of completeness could be expected if there was not a mediator to walk researchers through revising the initial draft of the metadata form. Generally, those who checked over the metadata form without meeting with the Metadata Curation Specialist did not add much information that had not already been captured. Two attempts were made to reach out to researchers, after which a lack of response resulted in the researcher's removal from the list of potential participants. This process was repeated until the target number of studies (minimum 36) outlined in our grant was reached.

### **Further Populate Records with Information from Study Documentation**

If the researcher was willing to share study documentation, they sent them as attached files in the REDCap survey, through email, or, in one case, by sharing a physical binder. Shared study documentation included protocols, ethics submissions, consent forms, data collection forms, case report forms, code sheets, information sheets, data dictionaries, grant applications, publications, and one de-identified dataset. The researcher was asked if we could add study documents and de-identified datasets, if available and approved, to the metadata record.

Information from the study documentation was also used to complete a more robust metadata record. This is part of trying to fit in with already existing processes and documentation in the data management cycle, and utilizing existing metadata already collected in research documents. The documents were first organized into types and searched to find common information to map to our metadata fields. For example, some of the protocols have the same headings, some of which can be semantically mapped to a metadata element in the REDCap metadata form. Looking at the documentation systematically also helped us see any important information captured in the documentation that was not already represented in the metadata profile.

Although the study documentation provided a wealth of information to describe the context of studies and data, including detailed information about data collection methods, study contributors, and sampling procedures, manually searching individual documents was a time-consuming and labour-intensive process. An attempt was made to simplify this process and make it more efficient through text mining techniques. However, we did not have a large enough sample to make this possible and, more significantly, documentation formats were too varied to create usable mappings. The trend towards standardizing documents, such as CDISC Protocol Representation Model (PRM) for organizing study protocols, could potentially facilitate this process in the future.

## Conduct Metadata Consultations

Engaging and collaborating with researchers helped us to test the metadata schema in the field, and further our aim to simplify the creation of sound metadata. The 26 researchers who participated fully in metadata consultations were either study principal investigators, research coordinators or others members of a research team to whom the PI delegated the task. 15 consultations were one-on-one interviews with individuals and five consultations were with research teams of two to three people.

In keeping with the principles of the participatory design process, the Metadata Curation Specialist encouraged pilot participants to guide the discussion, probing for more information when necessary. A discussion guide<sup>13</sup> was created and consulted beforehand, although the consultations were more open and conversational in nature, allowing the researcher to lead the conversation as much as possible. Consultations with a conversational tone encourage researchers to elaborate on their answers and provide more in-depth information (Read et al., 2015). This approach enabled the discussion to go in unanticipated directions, informed and led by the researcher perspective rather than steered solely by the Metadata Curation Specialist. As Read et al. (2015) suggest, when conducting data interviews with researchers it is also important not to require that researchers adopt the language of the library. Instead librarians and data specialists should try to speak to the researchers in their own language (Read et al., 2015). As such an effort was made to avoid using too many library-centric words. In addition, an attempt was made to explain and offer education on important concepts like “metadata” and “controlled vocabulary,” rather than assume a shared understanding.

If a researcher agreed to meet, the Metadata Curation Specialist arranged a one-hour consultation to go through the pre-populated metadata record. The aim of consultation was to validate, add to and amend information in the pre-populated metadata form and to get feedback on the metadata schema/form itself, including feedback on specific elements, language, controlled vocabularies, and the order and number of elements. A broader discussion of metadata and research data was also encouraged, although the one-hour time limit of the meetings somewhat hindered this from fully developing. We wanted to discover what information would help them find and understand data, or what information would be helpful in order to search for collaborators. We also wanted to learn what elements were not necessary. We started out with very extensive study-level metadata, resulting in a very long form, and wanted researchers to help us whittle it down into something meaningful, that could realistically be sustained. These conversations were not recorded. Throughout the conversation, the Metadata Curation Specialist took notes on a printed out copy of the pre-populated REDCap metadata form. A printout was used rather than directly entering data into REDCap to preserve changes in metadata. Data was later changed in REDCap to reflect the amended data.

## Process Feedback and Revise Metadata Profile

Information gleaned from the consultations was analysed and incorporated into recommendations for refining the metadata profile. The full metadata data from REDCap was exported as a .csv file and analysed for similarities in responses to attempt to create lists of useful elements, and to come up with language for standardized lists for elements. Elements with majority blank/non-responses were also noted, as in discussion many of the elements were deemed non-essential or unsuited to the type of research conducted by the researchers. Discussion with researchers was also taken into account.

<sup>13</sup> Pilot Data Catalogue Discussion Guide: <https://bit.ly/2I2I99L>

The feedback from researchers was then compared with the schema gap analysis results and synthesized to come up with a metadata profile suited to researcher needs. A new REDCap form was created to reflect these changes.

A number of researchers were unsure what was expected for the ‘Contributor’ and ‘Contributor Type’ Dataverse elements and expressed that more discipline-specific elements reflecting the work roles of clinical research, such as biostatistician, study coordinator, advisor, etc, would be much more intuitive and lead to more consistent data collection. Many of the Dataverse/DDI elements are grounded in statistical social science so many elements such as Time Method, Type of Research Instrument, Major Deviations for Sample Design, Cleaning Operations, and Estimates of Sampling Error, were not seen as relevant and thus left blank.<sup>14</sup>

At the end of the project, CHoM and UAL came together to discuss and compare their separate metadata recommendations and requirements. These have been synthesized into a set of metadata elements suited to describing both active and long-term archival research data.<sup>15</sup> A detailed guidelines document around applying these elements is currently being finalized and will be published at a later date.

### **Transfer Metadata to Publicly Accessible Platforms**

The approved metadata and research documentation was later transferred from REDCap to the UAL publicly accessible and searchable Dataverse instance to enable discovery of the studies by other researchers. In addition, an R-based web application, ‘shiny’, was used to develop a prototype dashboard to provide alternative access to the full metadata which is currently only available in REDCap, which is open to only a certain community. This dashboard application (Vashishtha, 2017) is also capable of providing access to metadata from ClinicalTrials.gov. The shiny-based application is freely accessible to users on the web (RStudio Team, 2015).

## **Usability Testing and Future % Workshop/Sustainability Planning %**

During the last month of our project, the team conducted a future workshop for participant collaborators, as well as usability testing sessions for the pilot WCHRI data catalogue. Our research protocol for both was approved by the University of Alberta Research Ethics Board.<sup>16</sup>

### **Future Workshop**

We hosted a future workshop in order to begin collectively identifying potential next steps and sustainability issues for the pilot process and data catalogue. We invited the researchers who participated in the pilot project and representatives from WCHRI and PolicyWise. In the end, four invitees were able to attend: one researcher/pilot participant and three administrative/research support staff from WCHRI and PolicyWise.

---

<sup>14</sup> Refer to data analysis sheet for full information: <https://bit.ly/2Fw50VX>

<sup>15</sup> At-a-Glance Element Tables: <https://bit.ly/2HKSLGK>

<sup>16</sup> The REB (Research Ethics Board) in Canada is the equivalent of an IRB (Institutional Review Board).



### Workshop design and findings

The two co-PIs acted as facilitators of the session and the project Metadata Specialist and Data Curator (both with more direct involvement in the day to day operations of the project) acted as participant observers taking detailed notes and contributing to the discussion when they felt it was appropriate. Workshop participants were asked to read an information sheet and sign a consent/audio-video release before participating in the sessions. The design of our future workshop was based on a modified four phase structure (Jungk, 1987), with the addition of a fifth phase focused on discussion and a Futures Wheel exercise (Lauttamäki, 2014; Glenn, 1994). In Phase I (Introduction), we reviewed the history of the project and current outcomes and explained the rationale for the future workshop. In Phase II (Critique), participants were asked to identify challenges to data curation and sharing and how the pilot project may or may not have addressed these problems. Tables 2 and 3 identify key themes that were identified during the critique phase and verified with project participants throughout the workshop. Co-PI Roark performed in vivo coding of the workshop transcripts post-workshop, which was later verified by the full project team. The findings reported in this paper were also reviewed and validated by workshop participants.

**Table 2.** Benefits/Positive Outcomes – Future Sustainability Workshop.

| Challenges   | Code | Meaning                                |
|--|------|--|
| “Definitely for us the real benefit is education and cultural change within the research community to start thinking about metadata and data management in general. And the promotion of standards within this community which helps us because [...] making sure they think about metadata and data in the future.”   | 1    | Culture change                         |
| “I think from our perspective we found it to be a very useful catalogue at the end of the day. It was a really interesting product that showcased work that for WCHRI maybe goes under the radar. And maybe some of this never does get published or isn’t maybe WCHRI or WCHRI’s funders aren’t acknowledged within the publication. This is one place where we can kind of pull together and showcase work that’s been done. So for us, I think that was kind of illuminating in terms of a positive, and really got us thinking about how we want to use this going forward.”   | 2    | Orphan Data/Studies and Research Waste |
| “I had a lengthy conversation with somebody who felt, after participating, and you know creating this record and actually getting to the point now of sharing the actual data, that we spent so much time describing, and they were completely on board, suddenly changed their mind because they might want to keep that data and use it for their own purposes. And they [collaborator/PI] didn’t want it out there. [...] I wanted to see the value of something that I contributed to the creation of and then didn’t see it realized in that last moment because of concerns that this is proprietary, this is mine.” | 3    | Understanding and Value of Metadata    |

Continued over /-



| Challenges   | Code | Meaning                                 |
|--|------|---|
| “I think it has potential to increase awareness of the issues and the solutions around data sharing, which is becoming more of an issue as time. Well, it’s becoming more widely discussed as time goes on.”   | 4    | Understanding and Value of Data Sharing |
| “I can agree definitely the benefit of creating connections and networking and in a way discovering other like-minded people who also care about metadata, care about maternal child health metadata in particular. So that was a benefit of the project.” | 5    | Understanding and Value of Data Sharing |

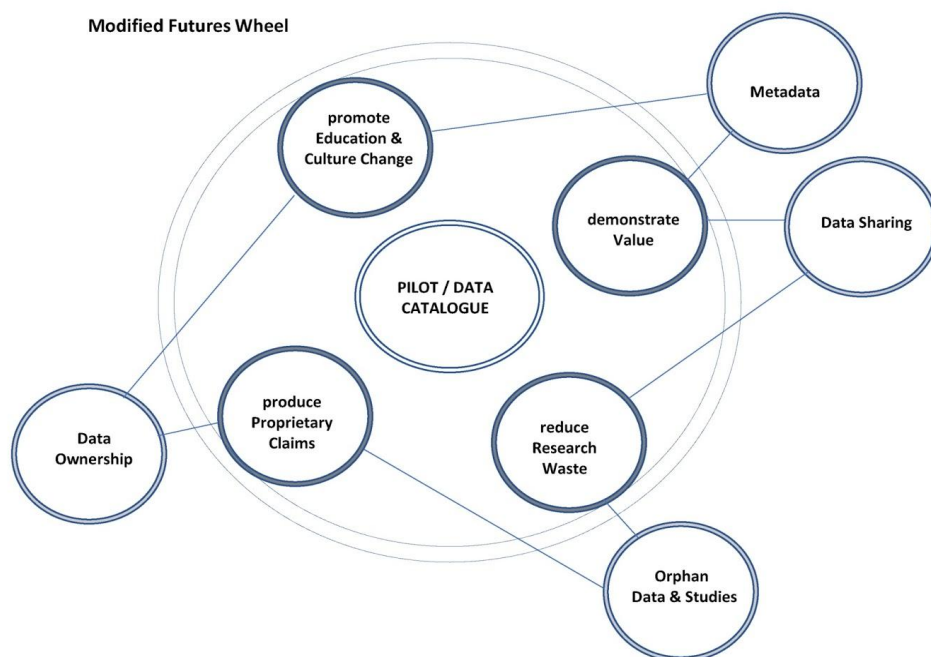
**Table 2.** Benefits/Positive Outcomes (continued)

**Table 3.** Challenges - Future Sustainability Workshop.

| Challenges   | Code | Meaning                                 |
|--|------|---|
| “So one was communicating the value of the product and metadata in general and making that case for trying to reveal research. And yeah, had a lot of misunderstanding.”   | 3    | Understanding and Value of Metadata;    |
|  | 4    | Understanding and Value of Data Sharing |
| “I don’t know if it’s relevant to this particular discussion, but when we invited people to participate I was quite pleased by the fact that people actually wanted to participate, but at the same time I was a little disappointed that it was relatively few. I was disappointed in that I thought we could have perhaps gotten more people who were interested in participating.”  | 6    | Participation                           |
| “I had a lengthy conversation with somebody who felt, after participating, and you know creating this record and actually getting to the point now of sharing the actual data, that we spent so much time describing, and they were completely on board, suddenly changed their mind because they might want to keep that data and use it for their own purposes. And they [collaborator/PI] didn’t want it out there. [...] I wanted to see the value of something that I contributed to the creation of and then didn’t see it realized in that last moment because of concerns that this is proprietary, this is mine.” | 7    | Proprietary / Ownership of Data         |
| “That was a big concern too amongst our management team. Just how people would feel about that. And was it allowed? It was the whole range of feelings around sharing data. Even though everyone kind of talks about it in this ‘Oh, yeah. Sharing data is really good and we all should be sharing data.’ But when it comes right down to it, people are very reticent to do that. They feel a lot of ownership around that. I don’t know how you overcome that challenge.”   | 7    | Proprietary / Ownership of Data         |
|  | 8    | Range of feelings around Data Sharing   |

In Phase III (Visionary), participants were asked to shift their Phase II insights into a group visualization exercise based on the Futures Wheel (Lauttamäki, 2014; Glenn, 1994). This exercise provided the opportunity to collectively imagine changes and effects that the pilot process/data catalogue might intensify or bring about. This allowed us to move from a discussion of challenges and benefits that stakeholders experienced while trying to gain support for and/or while participating as pilot researcher/depositor, toward near future scenarios and sustainability issues that could potentially arise. In Phase IV (Establishing), workshop participants further elaborated, consolidated and evaluated the scenarios. The facilitators pushed the group to discuss how these ideas could inform issues of project sustainability. In Phase V (Discussion), participants provided further insights into their overall experience of participating in various ways with the project.

Figure 3 is a synthesis of the key potential impacts of the pilot process/data catalogue discussed during the group visualization and scenario exercises. We were not able to complete the full exercise within the one-and-a-half-hour session. After the session, co-PI Roark condensed multiple Futures Wheels into one focused around the pilot/data catalogue. The inner ring represents the central issue/artefact, outer rings represent first order effects and their link to broader intersecting themes respectively. Not all themes highlighted in Tables 2 and 3 were discussed in depth during this exercise.



**Figure 3.** Modified Futures Wheel. \*

In the Fall of 2017 co-PI Farnel continued to raise awareness about the data catalogue through a presentation at the WCHRI Lunch and Learn series. Next steps include working with WCHRI administration to elicit further participation in design and sustainability planning. The team may also explore further use of techniques informed by future-oriented and values-based participatory design practices (Lauttamäki, 2014; Shilton, 2012) to explore data ownership and other issues around data sharing that sparked strong emotional responses from participants.

### **Usability Testing: WCHRI Pilot Study Catalogue**

In the last month of the project, we hosted a series of 30-45 minute usability testing sessions with potential WCHRI catalogue users. Four individual sessions were conducted with participants recruited through the WCHRI newsletter, and university associated postdoctoral fellows and medical humanities listservs. At least two team members were present at each session and acted as either facilitator or note-taker/videographer. All participants were given an information sheet and consent form with audio-visual release. In addition to audio-visual and detailed note-taking, the team also captured moving images of the computer screen using Camtasia software, while the user was performing the tasks. Usability testing participants were asked to perform a series of tasks related to information discovery and retrieval using a series of different platform interfaces related to the pilot MCH data catalogue (WCHRI Pilot Study Catalogue). The research team also asked a series of questions and probes that encouraged the user tester to “think aloud” and describe the reasoning behind their decision-making process.

There were two findings from the usability testing sessions which may need further attention. The first occurred when participants were asked to search for project pilot studies across both Dataverse and Datacite<sup>17</sup> (in that order). Users expressed some frustration when search strategies and options from other platforms (PubMed Central<sup>18</sup>, Dataverse) were not available within the new context. Users also tended to interpret mediated data access terms (e.g. contact the researcher for data access) as meaning that study data were either unavailable to anyone or only available to members of a research team. Both findings can be considered in subsequent iterations of testing, design and development activity.

## **Discussion**

Research data is complex and the researchers involved in its creation are needed to ensure that metadata to describe it is accurate and sufficient (Willoughby et al., 2014; Willoughby et al., 2015; Frey, 2008). At the same time, researchers often lack the time and experience to create the sort of rich and effective metadata needed to describe and support the use of research data. Researchers need support to help them create effective metadata. This pilot study revealed a number of practices that could help researchers to create metadata to describe their research data.

<sup>17</sup> Datacite: <https://search.datacite.org/> \*

<sup>18</sup> PubMed Central: <https://www.ncbi.nlm.nih.gov/pmc/> \*

## Communicate the Importance of Metadata

Although MCH researchers understand the context of their studies more than anyone, the majority are unlikely to have specialized metadata knowledge or experience. Metadata specialists, librarians and other data curation professionals possess this metadata expertise, but it can be a challenge communicating the value of creating metadata to researchers. For example, during metadata consultations several researchers asked for a better understanding of why we were creating the metadata record and its overall purpose. Researchers will be more likely to put time and effort into metadata creation if they understand the value of metadata in discovery, citation, collaboration, and their own professional development. The pilot data catalogue seems to have provided at least some stakeholders with a concrete example of the importance of metadata.

## Metadata for Restricted Data

None of the researchers consulted during this project were comfortable with making data available for public, unmediated download. A number of reasons for a reluctance to openly share data were cited, including concerns about participant privacy, the belief that their data are too small or specialized to be of value to others, and the work that would be involved to organize their data before it could be shared. Researchers who do not want to or cannot share their research data can create a metadata record about the data to let others know that the data exist and to provide them with information on access procedures for restricted data. Guidelines exist for enabling access to collections containing confidential or personal health information (PHI) in archival and data repository collections (e.g. Novak Guistainis and Evans Letocha, 2015; NIH, 2004; ICPSR<sup>19</sup>). However, more insight is needed into how potential secondary data users perceive metadata about restriction, and the meaning assigned to mediated or restricted access terms in general.

## Short and Simple

Creating metadata to describe research studies and data should be as user-friendly and intuitive as possible for researchers, while at the same time retaining the potential for rich description and critical engagement with the creation metadata records. This involves balancing what is ideal to capture and what is realistic to expect from researchers. Usable metadata creation forms are needed to improve the extent and quality of researcher-generated metadata. Several researchers noted the need to keep metadata forms as short and straightforward as possible to simplify and encourage metadata creation. Iterative, user-centered metadata design can help improve the usability of metadata creation forms, and responsible automated metadata creation.

## Clear and Relevant Language

During metadata consultations with researchers, confusion and uncertainty over the meaning of metadata elements was common. For many researchers, it was unclear how relevant some elements were, and there were often different interpretations of the same element amongst researchers. Researchers frequently vacillated over the meaning of

---

<sup>19</sup> \*ICPSR – Sharing Sensitive Data: <https://library.stanford.edu/research/data-management-services/share-and-preserve-research-data/sharing-sensitive-data>

elements, indicating that they were trying to figure out the metadata as they created it. This ambiguity not only causes frustration for researchers trying to enter metadata, it also creates inconsistent metadata across records.

As was observed by Crystal and Greenberg (2005), researchers may “struggle to apply their detailed local knowledge to global, generic schemas.” Targeted assistance by way of clear descriptions, relevant examples, and standardized lists targeted at specific researcher communities will help researchers determine appropriate inputs for standardized metadata fields. For example, researchers were often unsure what input was expected for the ‘Contributor’ and ‘Contributor Type’ elements. Providing a list of more specific types of contributors, based on typical MCH studies, like Study Coordinator, Statistician, or Data Manager, would help dispel this frustration and promote completion of the ‘Contributor’ metadata element.

### **Augmenting Automation**

Creating tools, scripts and applications for automating or semi-automating the routine work of metadata generation would ease the time, labour and money needed for manual metadata entry. This would allow researchers, librarians, and others involved in data curation to focus on more intellectual tasks (Crystal and Greenberg, 2005) For example, reusing publicly available metadata through APIs and other harvesting tools could greatly simplify researcher’s work. However, it is also important to check the quality of harvested metadata. The metadata may need to be further refined manually or through other semi-automatic processes. In this pilot we found researcher input and oversight invaluable.

### **Secondary Use of Metadata**

When possible, metadata should be streamlined and reused across existing documents, tools and processes in the research data lifecycle. Local processes can be developed that fit into how researchers already think about and manage their data. Tools and guidelines that fit into already existing data collection and management activities can help ease metadata creation, ensuring interoperability across the various documents created and systems used during the research data lifecycle, such as the data capture and survey development tool REDCap, data deposit and description involving Dataverse, and the research documents that are produced before, during, and after the study, such as grant applications, study protocols, and questionnaires. The trend towards more standardization in study documentation, such as consent forms, protocols, ethics applications, and grant proposals will allow this practice to grow and greatly simplify metadata production in the future.

## **Conclusion**

Creating metadata for research data is a complex and time-consuming task that can greatly benefit from well-designed metadata creation tools and targeted support. This project contributed to understanding the metadata needs of a small sample of MCH research studies and data. By taking a researcher-centered perspective on metadata creation, it is hoped that this pilot can provide inspiration for future studies in different research contexts. Moving forward it will be important to consider how values and

expectations around data ownership might be built into the design process. In addition, while restricted access metadata in part addressed concerns around the confidentiality of human-subjects data, research participants are another important stakeholder group to include in future discussions (Manhas et al., 2016; Geary et al., 2013; Hardy et al., 2016). Data varies across disciplines and study types, and it follows that this may result in different metadata needs. Regardless of differences, an approach that emphasizes engagement with researchers and which seeks to identify and build appropriate tools may be more likely to be incorporated into research workflows. Participatory design of schemas and tools with close attention to local needs can help create useful processes to simplify and embed metadata creation within existing workflows and research data practices.

## Acknowledgements

The work described in this paper and the ‘Bridging the Research Data Divide: Rethinking long-term value and access for historical and contemporary maternal, infant, and child research’ project are supported by a Cataloging Hidden Special Collections and Archives grant from the Council on Library and Information Resources (CLIR)<sup>20</sup>, funded by The Andrew W. Mellon Foundation<sup>21</sup> and awarded to the Center for the History of Medicine (Harvard University) and the University of Alberta Libraries, and a Purdue University International Travel Grant Award (PI Roark). We would like to acknowledge the contributions of our collaborators Emily Novak Gustainis and Amber LaFountain at the Center for the History of Medicine (Harvard). We offer a special thanks to all the individual researchers and teams that helped develop and contributed to the WCHRI Data Catalogue and our Community Advisory Committee Members and collaborators: Lawrence Richer, Rick Watts, and Tanya Voth (WCHRI, University of Alberta), Robin Featherstone (Alberta Research Centre for Health Evidence (ARCHE), University of Alberta), Xinjie Cui and Jason Lau (PolicyWise), Elodie Portales-Casamar and Anne Junker (MICYRN) and Christine Wagoner (Faculty of Medicine and Dentistry, University of Alberta) for volunteering their time and expertise.

## References

- Borgman, C.L. (2008). Data, disciplines, and scholarly publishing. *Learned Publishing*, 21(1), 29–38. doi:10.1087/095315108X254476
- Bowler, L., Koshman, S., Oh, J.S., He, D., Callery, B.G., Bowker, G., & Cox, R. (2011). Issues in user-centered design in LIS. *Library Trends*, 59(4): 721-752.
- Bruna, E.M. (2014, September 4). The opportunity cost of my #OpenScience was 36 hours + \$690 (UPDATED) [Blog post]. Retrieved from <http://brunolab.org/blog/2014/09/04/the-opportunity-cost-of-my-openscience-was-35-hours-690/>

<sup>20</sup> CLIR: <https://www.clir.org>

<sup>21</sup> The Andrew W. Mellon Foundation: <https://mellon.org>



- Chao, T. (2015). Mapping methods metadata for research data. *International Journal of Digital Curation*, 10(1), 82-94. doi:10.2218/ijdc.v10i1.347
- Crystal, A., & Greenberg, J. (2005). Usability of a metadata creation application for resource authors. *Library and Information Science Research*, 27(2), 177–189. doi:10.1016/j.lisr.2005.01.012
- Dyke, S.O.M., Philippakis, A.A., Rambla De Argila, J., Paltoo, D.N., Luetkemeier, E.S., Knoppers, B.M., ... Sherry, S.T. (2016). Consent codes: Upholding standard data use conditions. *PLoS Genetics*, 12(1), e1005772. doi:10.1371/journal.pgen.1005772
- Farnel, S., Shiri, A., Campbell, S., Cockney, C., Rathi, D., Stobbs, R. (2017). A community-driven metadata framework for describing cultural resources: The Digital Library North project. *Cataloging & Classification Quarterly*, 55(5), 289-306.
- Federer, L.M., Lu, Y.L., & Joubert, D.J. (2016). Data literacy training needs of biomedical researchers. *Journal of the Medical Library Association (JMLA)* 104(1).
- Federer, L.M., Lu, Y.L., Joubert, D.J., Welsh, J., & Brandys, B. (2015). Biomedical data sharing and reuse: Attitudes and practices of clinical and scientific research staff. *PLoS ONE*, 10(6), e0129506. doi:10.1371/journal.pone.0129506
- Frey, J.G. (2008). Curation of laboratory experimental data as part of the overall data lifecycle. *International Journal of Digital Curation*, 3(1), 44–62. doi:10.2218/ijdc.v3i1.41
- Geary, J., Jardine, C.G., Guebert, J., Bubela, T. (2013). Access and benefits sharing of genetic resources and associated traditional knowledge in northern Canada: Understanding the legal environment and creating effective research agreements. *International Journal of Circumpolar Health*, 72(1). doi:10.3402/ijch.v72i0.21351
- Glenn J.C. (1994). The Futures Wheel. In *Futures Research Methodology*, AC/UNU Millennium Project.
- Greenberg, J., Crystal, A., Robertson, W., & Leadem, E. (2003). Iterative design of metadata creation tools for resource authors. Paper presented at the International Conference on Dublin Core and Metadata Applications. North America. Retrieved from <http://dcpapers.dublincore.org/pubs/article/view/733>
- Hardy, L., Hughes, A., Hulen, E., Schwartz, A.L. (2016). Implementing qualitative data management plans to ensure ethical standards in multi-partner centers. *Journal of Empirical Research on Human Research Ethics*, 11(2), 191–198. doi:10.1177/1556264616636233
- Harris, P.A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., Conde, J.G. (2009). Research electronic data capture (REDCap) – A metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics* 42(2), 377-81.

- Heidorn, P.B. (2011) The emerging role of libraries in data curation and eScience. *Journal of Library Administration*, 51(7-8), 662-672. doi:10.1080/01930826.2011.601269
- Johnson, D.K. & Radler, B. (2018). Longitudinal research center in a box: Using DDI to enhance the mission of the UC Davis Alzheimer's Disease Center and the Midlife in the U.S. (MIDUS) study. Workshop presented at the North American Data Documentation Initiative (NADDI) Conference. doi:10.5281/zenodo.1213320
- Johnston, L.R. Carlson, J., Hudson-Vitale, C., Imker, H., Kozlowski, W., Olendorf, R., & Stewart, C. (2017). Data Curation Network special report: Results of the fall 2016 researcher engagement sessions. Retrieved from <https://docs.google.com/document/d/1ANiIdIChe6VPNloykMJFQEfAPJRGGA8-zO-g8z7-XH7k/edit#>
- Jungk, R. (1987). Future workshops: How to create desirable future. London: Institute for Social Inventions.
- Kovalchik, S. (2017). RISmed: Download content from NCBI databases. R package version 2.1.7. Retrieved from <https://CRAN.R-project.org/package=RISmed>
- Lauttamäki, V. (2014) Practical guide for facilitating a futures workshop. Finland Futures Research Centre. Retrieved from <https://www.utu.fi/FI/YKSIKOT/FFRC/KEHITTAMISPALVELUT/FUTURESFOCUS/DOCUMENTS/FUTURES-WORKSHOPS.PDF>
- Manhas, K.P., Page, S., Dodd, S.X., Letourneau, N., Ambrose, A., Cui, X., & Tough, S. C. (2016). Parental perspectives on consent for participation in large-scale, non-biological data repositories. *Life Sciences, Society and Policy*, 12(1). doi:10.1186/s40504-016-0034-6
- Michener, W.K. Allard, S. Budden, A. Cook, R.B. Douglass, K. Frame, M. Kelling, S. Koskela, R. Tenopir, C. Vieglaiss, D.A. (2012). Participatory design of DataONE—Enabling cyberinfrastructure for the biological and environmental sciences. *Ecological Informatics*, 11(5-15). doi:10.1016/j.ecoinf.2011.08.007
- NIH. (2004). Research repositories, databases, and the HIPAA privacy rule. NIH Publication Number 04-5489. Retrieved from [https://privacyruleandresearch.nih.gov/research\\_repositories.asp](https://privacyruleandresearch.nih.gov/research_repositories.asp)
- Novak Gustainis, E.R. & Evans Letocha, P. (2015). The practice of privacy. In *Innovation, Collaboration and Models: Proceedings of the CLIR Cataloging Hidden Special Collections and Archives Symposium, March 2015 (CLIR pub 169)*, 163-176. Washington D.C.: Council on Library and Information Resource. Retrieved from <http://nrs.harvard.edu/urn-3:HUL.InstRepos:24830902>
- Nutter, B. & Lane, S. (2015). redcapAPI: Accessing data from REDCap projects using the API. doi:10.5281/zenodo.11826

- Park, J.R., & Lu, C. (2009). Application of semi-automatic metadata generation in libraries: Types, tools, and techniques. *Library & Information Science Research, 31*, 225–231. doi:10.1016/j.lisr.2009.05.002
- Park, J.R. (2002) Hindrances in semantic mapping involving thesauri and metadata. *Journal of Internet Cataloging, 5*(3), 59-77. Retrieved from [http://www.tandfonline.com/doi/abs/10.1300/J141v05n03\\_04](http://www.tandfonline.com/doi/abs/10.1300/J141v05n03_04)
- Radler, B. & Johnson, D. (2014). On health research uses of DDI. Workshop presented at the North American Data Documentation Initiative (NADDI) Conference. Retrieved from <http://summit.sfu.ca/item/13945>
- Read, K.B., Surkis, A., Larson, C., McCrillis, A., Graff, A., Nicholson, J., & Xu, J. (2015). Starting the data conversation: Informing data services at an academic health sciences library. *Journal of the Medical Library Association, 103*(3), 131–135. doi:10.3163/1536-5050.103.3.005
- Roark, K. (2014). Curating risk, mediating access: Mapping the legal, policy and organizational context of data sharing in the health sciences. Health Research Data Symposium 2014, University of Alberta, Educational Research Archive (ERA). doi:10.7939/R3SX64C62
- Roark, K. (2016). Health research data landscape map for curating risk, mediating access: Mapping the legal, policy and organizational context of data sharing in the health sciences. *UAL Dataverse, V1*. doi:10.7939/DVN/10953
- RStudio Team. (2015). RStudio: Integrated development for R. RStudio, Inc. Retrieved from <http://www.rstudio.com/>
- Sachs, M.S. (2017). Rclinicaltrials version 1.4.7: R interface to clinicaltrials.gov. Retrieved from <https://github.com/sachsmc/rclinicaltrials>
- Saldana, J. (2009). The coding manual for qualitative researchers. London; Thousand Oaks, California: Sage.
- Shilton, K. (2012) Values levers: Building ethics into design. *Science, Technology & Human Values, 38*(3): 374-397.
- Spinuzzi, C. (2005) The methodology of participatory design. *Technical Communication, 52*(2), 163-174.
- Vashishtha, S. (2017). WCHRI Shiny App. Retrieved from [https://vashishtha.shinyapps.io/wchri\\_shiny\\_app/](https://vashishtha.shinyapps.io/wchri_shiny_app/)
- Willoughby, C., Bird, C.L., Coles, S.J., Frey, J. (2014). Creating context for the experiment record - User-defined metadata: Investigations into metadata usage in the LabTrove ELN. *Journal of Chemical Information and Modeling, 54*(12):3268-83. doi:10.1021/ci500469f

- Willoughby, C., Bird, C., & Frey, J. (2015). User-defined metadata: Using cues and changing perspectives. *International Journal of Digital Curation*, 10(1), 18–47. doi:10.2218/ijdc.v10i1.343
- Wright, A. (2016). REDCap: A tool for the electronic capture of research data. *Journal of Electronic Resources in Medical Libraries*, 13(4), 197-201. doi:10.1080/15424065.2016.1259026
- Yarney, L. Baker, K.S. (2013). Towards standardization: A participatory framework for scientific standard-making. *The International Journal of Data Curation*, 8(1): 157-172. doi:10.2218/ijdc.v8i1.252