



OSPO

OPEN SCHOLARSHIP POLICY OBSERVATORY



Draft Report: Research Data Management Support in the Humanities: Challenges and Recommendations

by Caroline Winter | 23 November 2021 | Community News, English, Observations and Responses | 2 comments



The following draft report by Stefan Higgins, Lisa Goddard, and Shahira Khair outlines discussions and findings from [Research Data Management for Digitally-Curious Humanists](#), an online event sponsored by the Social Sciences and Humanities Research Council (SSHRC) and held on June 14, 2021 as a Digital Humanities Summer Institute (DHSI) 2021 — Online Edition aligned event.

The authors invite readers to share feedback about the report using the Leave a Reply feature at the bottom of this page or by emailing Lisa Goddard (lgoddard@juvic.ca).

Research Data Management Support in the Humanities: Challenges and Recommendations [DRAFT]

Stefan Higgins, Lisa Goddard, and Shahira Khair, University of Victoria

Draft v. 3 November 8, 2021

Table of Contents:

1.0 Recommendations for Research Data Management Support in the Humanities

2.0 Introduction

3.0 First Principles and Definitions

3.1 Data require labour

3.2 Data require design

3.3 Data require production

4.0 The Relationship between Humanists and Data

4.1 Pre-Event Survey and Other Trends

4.2 What Does “Data” Mean in the Humanities? Dr. Miriam Posner, UCLA

4.3 Data Considerations Across Humanities Disciplines, Research Roundtable

4.4 Data Management Planning Workshop, Post-Event Survey

4.5 The “Three P’s” of Data Work

4.6 What are the consequences for humanists if they do not accept data management?

5.0 Bibliography

1.0 Recommendations for Research Data Management Support in the Humanities

This report summarizes proceedings from *Research Data Management for Digitally-Curious Humanists*, a virtual event sponsored by Social Sciences and Humanities Research Council (SSHRC) on Research Data Management Capacity Building. This event was held as a Digital Humanities Summer Institute 2021 aligned conference, and was led by the University of Victoria Libraries and the Electronic Textual Cultures Lab on June 14th, 2021. The program, presentations, and related resources are openly available on the project site: <https://osf.io/6vepj/wiki/home/>

The following recommendations reflect conversations with Humanist researchers and students before, during, and after the Connections event, based on pre- and post-surveys of attendees, and on presentations and discussion during the event. We have included the recommendations at the beginning of this report for easy reference, but please read further sections for much more detail.

1. Many humanists are uncertain about what constitutes “data” in the context of their research projects. Better guidance on defining research data must be developed in consultation with both digital and non-digital humanists from a variety of different disciplines.
2. As RDM policies become more mature, it is imperative to spend time examining edge cases, including analogue scholarship, and fine arts research processes. Directed effort must be made to engage researchers who identify their research as “not fitting into current data management policy” rather than focussing on successful, masthead DH projects, which generally already have institutional support, funding, and technical capabilities. This recommendation can be summarized as looking at the boundaries and edge cases of policies, as well as the centre.

3. Humanist researchers are not necessarily convinced about the relevance or value of “data” in their disciplines. This fosters a reluctance to engage with data management planning, and a tendency to see RDM as a bureaucratic burden. Clear examples of the value of high quality, sustainable, reusable humanities data sets are necessary to convince humanist researchers of the importance of RDM work.
4. Humanist researchers continue to feel that they need support for data management at *all* stages of the research process: on conceptual and theoretical approaches to data; on guidance for meeting new Tri-Agencies funding requirements; on making choices about data infrastructure; on defining appropriate metadata frameworks; on capturing and recording metadata according to standards; and on how to ensure their research does not change in kind in order to meet data policy.
5. Humanist researchers would like to receive funding increases that reflect the additional cost of research data management to projects, including the need to hire and train team members who can oversee the design and creation of data and metadata to ensure that practice aligns with the data management plan.
6. Humanists require RDM support and training sessions over the course of their whole careers, and not simply when they are ready to apply for funding. Ideally, data management concepts and basic skills will be developed at the undergraduate and graduate level. Asking researchers to try to absorb and apply all of this information at the point of grant application is likely to generate frustration and shallow engagement, as material becomes outdated or forgotten over the award timespans.
7. Many senior humanities researchers and instructors do not feel that they have enough RDM knowledge to confidently teach the necessary concepts and tools. A great deal of RDM instruction is aimed at experienced researchers, but it is also necessary to develop instructional resources that are aimed at undergraduate and graduate audiences. Ideally these instructional materials will include asynchronous options, and hands-on learning exercises that can be evaluated in a for-credit context.
8. Humanist data are extremely diverse. Most data are not highly structured or machine-generated, and a significant amount of what might be considered data are not digital. For funding bodies, institutions, and humanities researchers, one central task of research data management will be developing infrastructures that achieve a measure of standardization that supports widespread access, while ensuring researchers do not lose the ability to critically engage with different theories, methods, and practices of categorization in their own work. Research software, publishing platforms, and data repositories need to be flexible enough to support humanist research objects and processes without unduly constraining them.
9. Avoid applying over-standardized solutions for diverse research across different disciplines and fields. Although some measure of standardization is necessary for any RDM work at scale, an overemphasis on standardization risks conflating and confusing different types of research and their needs.
10. Research data management and digital research infrastructure (DRI) are closely connected. Ideally the Tri-Agencies will work closely with Digital Research Alliance of Canada (the Alliance) to ensure that digital infrastructure and research software are developed in ways that incorporate RDM principles, and facilitate the production of good data and metadata that can easily be exported for ingest into repositories that will provide long-term access and preservation.
11. Platform, software, and tool choices will significantly affect the way in which project data is organized, described, and accessed. In order to produce good data, humanists will need expert guidance on the way in which their technology and tool choices will impact their ability to export data and related

metadata for deposit into repositories. This is closely related to the kinds of general data research infrastructure needs that humanists have articulated in several of the 2020 NDRIO white papers. There is a critical need for improved access to research tools and infrastructure, but technology alone cannot fully address researcher needs. Human experts who can provide support and guidance are equally important.

12. Humanist researchers continue to struggle with project sustainability, but are often loathe to divorce back-end data from its front-end context for the purposes of preservation. One way to address this is to include more contextual and interpretive information in metadata, and to design projects from inception so that data can stand alone, outside the context of the user interface. Not only will this produce more reusable data, but it will help a great deal with the problem of project preservation. Humanist researchers require much more theoretical and practical training on metadata creation, ideally beginning at the undergraduate and graduate level.
13. Data work will not always be perfect from the beginning, and so data as practice involves a willingness to *experiment*, or to be prepared for the changes that projects undergo, and the contingencies they may encounter. It is extremely unlikely that humanist researchers will be able to create accurate and detailed plans at the application stage. Data management planning tools must support the evolving nature of data management plans with document versioning, alerts to remind researchers to revisit data plans periodically, and authorization tools that can accommodate changing team membership.
14. The Tri-Agencies should clearly articulate how DMPs will be evaluated during the application review process. Given the lack of DMP expertise among many humanists, it is imperative that clear direction to reviewers is provided about how to evaluate the DMP component of an application. There is some risk that SSHRC reviewers who do not accept the importance of “data” in their disciplines will not place weight on data management as a criteria for evaluation.
15. The Tri-Agencies should clearly articulate the oversight process and reporting requirements related to Data Management Plans. Without some kind of formal follow-up, there is a strong chance that DMPs created at the point of funding application will never again be consulted, updated, or put into practice.
16. Humanist researchers strongly agree that research projects involving human subjects must prioritize consultation with communities of practice. Ethical concerns must trump data-sharing benefits in all cases. Indigenous data is out of the scope of the current policy, which is appropriate, but funding for community-designed and -owned solutions are also necessary so that Indigenous people are able to control, access, and use their data over time.

2.0 Introduction

This report summarizes proceedings from *Research Data Management for Digitally-Curious Humanists*, a virtual event sponsored by Social Sciences and Humanities Research Council (SSHRC) on Research Data Management Capacity Building. This event was held as Digital Humanities Summer Institute 2021 aligned conference, and which was led by the University of Victoria Libraries and the Electronic Textual Cultures Lab on June 14th, 2021. The program, presentations, and related resources are openly available on the project site: <https://osf.io/6vepj/wiki/home/>

The event had three goals:

1. To provide a community-building forum to better understand the unique considerations of managing humanities research data;
2. To provide hands-on training in developing data management plans (DMPs) for humanities researchers; and
3. To produce tangible outputs that can be shared with the humanities research community to support the creation of DMPs and broad adoption of RDM practices.

What is a Data Management Plan? The Canadian Tri-Agencies define a Data Management Plan (DMP) as “a living document, typically associated with an individual research project or program that consists of the practices, processes and strategies that pertain to a set of specified topics related to data management and curation.” (Tri-Agencies, FAQ 2021)

Many humanities researchers already do research and research preparation work that looks something like a DMP: a historian deciding how to classify archival documents, or an internet researcher deciding which ephemera count as notable, and which do not. In the humanities, these processes may not always be explicit, especially when it comes to tasks like choosing and storing digital research materials. In this regard, DMPs can be a valuable tool for thinking through processes for collecting, organizing, curating, and sharing data when paired with the necessary knowledge and understanding.

This report explores what it means to conduct good research data management in the humanities using a series of reflections on data, as well as recommendations for its management. It is *not* a plan for data management plans, nor is it a guide to research data management (RDM). Instead, our primary aim is to communicate to both researchers and funding agencies some of the thinking and labour required to develop good data research practices that are suited to research processes used in humanities disciplines.

The report is structured as follows. In the first section, we address some of the theoretical considerations around data and data management in the humanities. We identify and describe three principles for “what data mean” in the humanities. Our goal in this section is to provide an essential lexicon for data work in the humanities.

The second section focuses on the relations between humanists and data. Divided into 3 parts, it is a summary of the proceedings of our 2021 event. In the first subsection, we display some results of our pre-event survey, noting where researcher concerns and anxieties about RDM appear. In the second subsection, we summarize the main proceedings from our event: our keynote presentation on “Data Trouble,” and a roundtable with 5 speakers on “*Data considerations across humanities disciplines*.” These speakers are all either working on projects that use, or are thinking about, RDM in the humanities. Links to specific projects exemplify how some scholars are doing data work in the humanities. Summaries of the talks gesture to some of the concerns and problems related to data work in the humanities. Finally, our third subsection pulls together some threads on the relations between humanists and data, especially by considering the results of our post-event survey.

The third section asks what support humanists will require for their data, developing an extended and detailed series of recommendations for RDM in Canada that build on the prior two sections. In addition to pragmatic recommendations noted in a bulleted format, the section develops two major short- and long-term requirements for RDM support from funding bodies in Canada. Our primary recommendations are: that much more humanities-focussed learning support is necessary for researchers at all levels, and that instruction material should be developed in consultation with humanists researchers who do not consider themselves to be “digital” humanists; that planning and development of digital research infrastructure (DRI) and expert DRI support for humanists be closely aligned with best practices for the design and production of reusable research data; that clear guidance is needed regarding the evaluation of DMPs as

part of grant submissions, and regarding the oversight and reporting expectations for researchers.

In particular, our report concludes that open dialogue about the challenges of RDM, including the ways it could go wrong, is preferable to unrealistic expectations about onboarding that could produce new problems to solve.

3.0 First Principles and Definitions

The word “data” is, today, ubiquitous. The Canadian Tri-Agencies define research data as follows: “Research data are data that are used as primary sources to support technical or scientific enquiry, research, scholarship, or creative practice, and that are used as evidence in the research process and/or are commonly accepted in the research community as necessary to validate research findings and results.” (Tri-Agencies, FAQ 2021) This is, generally speaking, a solid definition to begin with. It is worth, however, looking a little more closely at the word “data,” which can mean many different things in different contexts; in the humanities, especially, the meanings and uses of data are contested. Data are complicated. Our event’s keynote presenter, Dr. Miriam Posner, got us started by presenting on “Data Trouble,” or the difficulty humanists have defining the concept of data in disciplines that have not traditionally considered their research outputs in this way. That presentation can be accessed via the event website [here](#), and is summarized in greater detail in Section 3.2.

As Posner noted, the *Oxford English Dictionary* defines data as “related items of (chiefly numerical) information considered collectively, typically obtained by scientific work and used for reference, analysis, and calculation,” and, in the context of computing, “quantities, characters, or symbols on which operations are performed by a computer, considered collectively” (*OED*, data., n.). This definition treats data as primarily quantitative and numerical. One datum is the smallest possible unit of analysis, and *only thereafter* are data interpreted. By contrast, the humanities are often defined primarily by their emphasis on *interpretation*, and spanning fields from literature, to jurisprudence, to philosophy, to history. If we imagine a spectrum, with quantitative, numerical data at one end, it would seem that the concerns of humanities research lie at the far end, with no close relationship to data. But readers of this report will know that to not be true. So what gives?

One of the major problems raised by Posner’s talk was that of “categorization.” What sets the humanities apart is the degree to which the meaning and interpretation of “this” and “that” are always in conversation, contestation, and revision. But the standard approach to “data” assumes that the “this” and “that” are *not* up for discussion: data are always transparently observable and classifiable, and therefore self-evidently produced for indisputable analysis and calculation. Humanists—and, in fact, many other researchers who utilize data in one way or another—know this assumption is wrong.

In an effort, then, to acknowledge much of the humanities thinking on data, and to centre it on research data management practices, we explore aspects of the creation process that challenge the idea of data as something “given”.

1. Data require *labour*.
2. Data require *design*.
3. Data require *production*.

3.1: Data require *labour*

While the *OED* provides a useful contemporary definition of “data,” the origins of the word are in Latin, suggesting “the things having been given,” implying, as media studies scholar Alexander Galloway (2011) has argued, that “something has already taken place and, via a gift or endowment, it enters into presence.” The implication that data are a gift (from the world, or from elsewhere) suggests that data do

not require *labour* on the part of the recipient or user. This section argues the opposite.

1. Data mean *something given*.

Some of the first early modern uses of the word “data” in the 17th century were in clerical texts, adopting the Latinate meaning of “something given” as unambiguously transcendental or “God-given,” and thereafter used as the basis for philosophical or moral reasoning and other inferences (*OED*). Data, then, refer to a world that is assumed to be “raw,” “unspoiled,” or “given” to its interlocutor, and that may be extracted, measured, and harnessed to produce information and knowledge (Gitelman ed. 2013). This definition of data eliminates the concept of labour: there is no measuring device used to produce the observations called data, nor are there any measurers (humans) involved in decisions about how to collect data, what to collect, where, or when. Good data management planning pays attention to the tools used, the people involved, and the methods for data production; in so doing, data are *produced* (see 3). This report, following extensive work in the humanities (see for example Gitelman ed. 2013, Halpern 2014, Chun 2016, Thylstrup et al. 2021), *recommends in the strongest terms possible that data never be called “raw.”*

2. Data mean *something taken*.

The digital humanities scholar Johanna Drucker (2011) makes a critical distinction between “the act of observation” and the “phenomena observed.” The claim that “data are given” collapses that distinction by assuming that the phenomena in question are not, in some way, resistant to their observation (e.g., microscopic organisms are “resistant” to observation with the eye alone). For Drucker, the act of observation involves two central characteristics: the observer, and the tool used for observation. Human observers bring social, cultural, historical, economic, political, and other forms of baggage; our histories structure what and how we observe and the inferences that follow. Likewise, our tools are the products of historical enquiry and design. Therefore, neither are neutral. Calling data “something taken” shifts the focus of data work from the assumption that measured objects are transparently observable to the action of collection and production. The action of the collection and production of data involves *labour*. Focussing on the *work* of data pays closer attention to just how that labour affects data themselves: for example, how human or tool error can radically alter the data produced. Moreover, acknowledging the role of “capture” or “extraction” in data work continues modes of humanistic knowledge production that acknowledge knowledge and data’s partial and situated construction.

3. Data are shaped by systems.

In a western world with presumed access to ubiquitous computing power, and where it can sometimes seem like everything can be “made digital,” it can be easy to assume that “computers do the work.” In addition to erasing the human labour involved in data work—from researchers developing plans and methods, to librarians developing protocols for storage, to name a few—this assumption can erase critical distinctions about access to computing, computing power and capabilities, and the myriad hard- and soft-ware differences that influence the diversity of research methods and results. As the **Endings Project** has argued, computational differences are project differences. Today, institutions often rely to a degree on software- and platforms-as-a-service. Significant computing power can be under corporate control (e.g. commercial cloud, proprietary software packages). Moreover, the programming languages (e.g., JavaScript), or encoding standards (e.g., TEI XML) used to support data work are not the same: while some, like TEI XML, are standards with longer lives than others, research stored digitally may still not be future-proofed, leading to “dead” projects or data that cannot be reused. *Over reliance on computational power places the control of infrastructure, data, and the labour required to sustain it in the hands of whoever controls that computing power.* One way to do data work with less hard- and soft-ware dependency is via “**minimal computing**,” which recentres human researchers and the tools they use as doing the essential labour of data work. It asks questions like “is significant computing power actually needed to do this research?” This is not to argue for an “offline” approach to research data management. On the contrary, minimal computing centres *specific projects as labour* in the data production process, requiring researchers, funding bodies, and institutions to consider platform and

computing choices as part of data management planning and project sustainability. Software choices made at the beginning of a project will greatly impact the way in which data is collected, organized, described, and accessed. These choices will impact the availability and reusability of data and must be considered as part of good data management planning.

Good data work starts before there is data. Thinking about labour requires us to approach data work as always including situated human researchers; as including the act of producing data from observable phenomena; and as dependent on computational infrastructures for their production. In the next claim, we expand on what some of that labour might look like.

3.2: Data require *design*.

Design work involves the *craft* or *practice* of creation. It involves the experimentation with, and use of (often complex) tools (Tenen 2016). It involves choices: which data to collect, how, and why; or how to classify, analyse, and represent data research. All of these choices involve work, and as any crafts person knows, the labour of design alters the object produced. Here, we argue that researchers, institutions, and funding bodies can centre the labour of data work by placing greater emphasis on project design as a situated practice.

1. Data produce *information*.

Data and information are intricately related, but not the same. Often, they are used in a tautology to define each other: data are pieces of information, while information is a collection of data. In reality, the word “information” has Latin roots that differentiate it from “data”: as Galloway (2011) notes, the root of information stems from “the act of taking form or being put into form.” In the assumption that data are “raw,” information is, then, “cooked”: it has been shaped, sculpted, or interpreted. In reality, as this report has argued, since data work involves labour at every step of the process, and since labour and design are intricately connected, data are always on the cusp of becoming information. At any point that data are interpreted, or re/presented to others, they become information. *Information is what data are in systems of relations (i.e., datasets).* Since data are always used in *data sets* (see 2.B), the labour and production of data work makes data into information. The choices researchers make about their data (e.g., developing DMPs) *form* their data and datasets, placing data into a fundamental organization of and with information. As we will suggest later (section 4.2), much of Humanities work in research data management is actually centered around the management and representation of information—not necessarily numerical datasets to be analysed, but collections of objects with (often complex) metadata that may be interacted with by researchers and other users in ways that go beyond simple reproduction of calculations.

2. Data produce *narratives*.

Data tell stories (Chun 2016) because data are used to do things. A study of the mineral composition of rocks is used to narrate a geological history. The dataset of 9/11 victim names tells a story about an event, about the people involved, and about how they related to their communities. All stories have politics (Lampland & Star 2009). As a result, *data, put to use, have politics too.* This statement does not mean that data are intrinsically a form of domination, or can only be used in that way. Instead, it pushes back against the optimistic belief that data are inherently “true” in their observations about the world in a way that belies the possibility of harm. As this report has argued, data work starts before there are data. Choices about what and how to observe structure data and the insights gleaned therein. Data and datasets may go so far as to structure empirical realities, whether in a “positive” sense by producing “facts” and “truth” to add to a system of knowledge, or in a “negative” sense by being denied or misconstrued as “disinformation” or “post-truth” (Harsin 2018). Researchers and funding bodies must be mindful of those facts, and of the concerns that follow, not in order to foreclose data work, but rather, to encourage *project and data design* that anticipates how the choices they make to extract, organize, and classify data have consequences for knowledge and politics. Geoffrey Bowker and Susan Leigh

Star (1993) call this work the act of “sorting things out.”

Having examined the labour involved in data work, the preceding section supplemented that claim by emphasizing the role of *design*. Data work is design work, and therefore always has to be mindful of the extent to which it “crafts” knowledge and reality. Design work begins before there are data. It involves recognising how data produce information, and how the choices made in data work “cook” data from the start. Design work also involves thinking about data as intrinsically systematic, going beyond technical considerations to think about how systems of data operate successfully or not. Finally, design work involves a recognition of the role that narrative and politics will always play in data work. In the next, and final, claim, we look at how some of the previous arguments get operationalized (or not) as a way to set the stage for some of the recommendations that follow in this report.

3.3: Data require *production*.

To say data are produced is to argue that data are *made* in the act of observation, measuring, and collection. Rather than refer to data as “raw” or “cooked,” this paradigm would refer to data as unprocessed or processed, respectively: a language choice that reflects the role of humans in data work at every step. There are already serious attempts to develop paradigms for the management of this production. The FAIR (Findability, Accessibility, Interoperability, Reusability) guiding principles for research data management are a major objective of the Canadian Tri-Agencies data management policy (see, for reference, Wilkinson et al. 2016). The Tri-Agencies are also committed to **Indigenous data sovereignty**, including via paradigms like **OCAP**[®] (Ownership, Control, Access, Possession), but, as the Tri-Agencies note themselves, these paradigms don’t necessarily “respond to the needs and values of distinct First Nations, Métis, and Inuit communities, collectives and organizations.” Further to these two paradigms, **CARE** (Collective Benefit, Authority to Control, Responsibility, Ethics) provides a model for Indigenous Data Governance. Finally, the TRUST (Transparency, Responsibility, User Focus, Sustainability, Technology) principles provide a paradigm for the use and management of digital repositories (Lin et al. 2020). A major concern for these paradigms moving forward is the operationalization of principles into choices about standardization or not, ethical guides, and frameworks for data management.

1. *People.*

This report has already covered how people, and their labour, are involved in the design of data work, as well as every other stage. But in the Humanities, people are also often the *subject* of data work. *It is imperative that research data management in the Humanities not reduce policy to technical matters. The risk of such a reduction is the development of data policy and management that can actively harm individuals and communities that are involved in, or the subject of, data production.*

2. *Practice.*

Practice means the actions and methods of research encompassing human actors, labour, and the craft of design. The production of data is an outcome of the design of research data management plans, but also of software and metadata choices. Moreover, data production requires extensive and intensive knowledge of relevant tools and methods, whether to avoid failure, or more simply to maximize benefit (Tenen 2016). Critically, data work will not always be perfect from the beginning, and so data as practice involves a willingness to *experiment*, or to be prepared for the changes that projects undergo, and the contingencies they may encounter.

3. *Presentation.*

By stating data mean presentation, we are emphasising the role that *audiences* play in the reception and interpretation of data. Researchers cannot assume that they have full control over the “public” life of their data and information. Moreover, we want to emphasize that while good data storage and archival practices are paramount, these are not the sole concerns of data work, especially in the humanities, where what count as data tend to be a mixture of media (see section 3.1) that precipitate a wide variety of ways of storing, displaying, and sharing data work.

This section has attempted to collect and collate many of the claims made about “data” in fields like the Humanities, Science and Technology Studies, and Media Studies. It has developed a tiered system for discussing data revolving around three (3) claims about what data mean: *labour* (something given, something taken, minimal computing); *design* (information, systems, narrative); and *production* (people, practice, presentation). A major challenge for RDM in Canada is the need to create digital infrastructure and data management tools that work for all disciplines, while balancing the need for standardization and interoperability against the need for flexibility and discipline-specific nuance. Humanists have diverse perspectives on the definition of “data”, and on the relevance of data to their work and research methods.

In the following section, “The Relations between Humanists and Data”, we summarize our event proceedings, including presentations from humanities scholars who are trying to implement data management in the context of diverse projects. We also highlight some of the pre-and post-survey responses we received from attendees. This section provides the context for many of the recommendations that are found in the first section of this paper. Finally, in the last section, we ask “What are the consequences for Humanists if they do not accept data management?” Although the question is framed around researchers, we connect researcher success to funding body and institutional success, arguing that major mistakes in RDM policy in Canada could, down the line, imperil the position of Canadian research funding bodies as a privileged and protected sector of the Canadian research landscape.

4.0 The Relationship between Humanists and Data

If you asked a person at random to describe types of data, they would likely give you one of three answers: data is either the product of scientific inquiry, a part of computer science, or “stuff” that gets collected by corporations and states about people, usually through online or offline surveillance. None of those answers involve the humanities. Many humanists would also not consider their research to involve data. There are important reasons to differentiate data from other research materials, but as the previous section has demonstrated, the actual, complex meaning of “data” means that many humanists will encounter or work with data during their research.

This section’s primary purpose is to recap some of the conversations that occurred about humanists and data during our day-long event on Research Data Management for Digitally-Curious Humanists. In the first subsection, we elaborate on several trends that emerged in our pre-event survey and during the event. In the second subsection, we provide a summary of our keynote presentation, as well as brief summaries on the presentations from our roundtable, “Data considerations across Humanities disciplines.” In the third subsection, we summarize our afternoon workshop on Data Management Plans, as well as trends that emerged in our post-event survey, with an eye to recommendations for data management policy in the following sections.

4.1 Pre-Event Survey and Other Trends

Our pre-event survey received 73 responses (n=73); in order to complete the survey, respondents were required to answer each question, so the sample size for each piece of information given here is the same. These answers can be summarized in the following three points:

- While many humanists are familiar with the concept of data and the idea that it might be related to their work, there is little consensus about how to define “data” in a humanities context, given the diversity of humanists’ objects of study;
- Many humanists are uncertain about what data work would look like for them (e.g., a data repository

for quantitative data vs. a project website that links to data materials);

- There is a lack of conceptual and knowledge-based resources, as well as a lack of support, service, and maintenance for technical infrastructure, to support research data management in the humanities.

We asked respondents (n=73) “What information are you most interested in gaining from attending this event?”:

- 66% of respondents hoped for “general knowledge on research data management”;
- 62% of respondents hoped for “practices for managing my research materials”;
- 66% of respondents were looking for direction on “how to create a data management plan.”

These survey responses demonstrate a need for both general and specific knowledge and support for research data work. They highlight the degree of inexperience amongst humanists with respect to research *as data*, and data management. Even if there are big, successful DH projects with significant funding and institutional support, it is imperative to assume that “onboarding” RDM in the humanities will require more than simple access to digital infrastructures, especially because the processes of “creating” DMPs and developing “practices” for research management require both specialized and distributed time and labour.

In a question asking respondents (n=73) about their experience with research data management policies, only 3% called themselves “very experienced,” while 58% called themselves “capable” (with some knowledge), and 40% called themselves “inexperienced.” Attendees were not merely looking for advice on how to create data management plans, but also for general knowledge and practices about research materials and data.

From these responses, we can make the following recommendation:

Humanists require frequent support sessions, both to distribute general knowledge on research data, and to provide consultation and advice on data management and planning (in other words, different *types* of support sessions).

Finally, survey respondents were asked to briefly describe their areas of research, and what types of materials they work with that could be considered as “data.” Here are some of their responses:

- Text
- Digital Audio files
- Digital Video files
- Digital Image files
- Physical archival records
- Medieval texts
- Library catalogue data
- Policy papers and records
- VHS
- 8mm film
- Camcorder tapes
- Community relations (impact, contributions, feedback)
- Religion in the 20th century
- Social movements in the 20th century
- Organizational memberships, budgets, bureaucratic documents
- Paper correspondence
- Interviews
- Maps
- Cultural Analytics (e.g. Wikipedia links)

- Programming languages
- Software
- Geodata
- Book history
- American literature
- American culture
- Translations
- Children's literature (books and magazines)
- Cultural materials (e.g. cards, toys, games)
- Radio tapes
- Ancient texts
- Prosopography
- Artefacts (and images of artefacts)
- State and private financial records (e.g. grant records)
- Text corpora
- Clippings, photos, ephemera, journals, logs
- Web usage and analytics
- CSV files

These examples represent a sample of 73 responses. Humanities work undoubtedly includes other research materials that could also be considered as “data,” particularly since respondents were asked to name individual objects or materials, rather than systems. Researchers from other disciplines and the general public might assume much of the humanist data noted here could be identified as “texts,” but, like the term data, the meaning of the word “text” in the humanities varies across disciplines and subdisciplines, across contexts, and across materials. Moreover, lumping research materials into a super-category like “texts” reduces the precision and accuracy of those materials, the data they are used to produce, and the metadata that describe them.

Humanists have different conceptions of research form, methodologies, content, storage, archiving, sharing, and more. These differences mean humanities data management is difficult to standardize. Given some of the considerations outlined in the first two sections, standardization of these research materials into neatly classifiable and interoperable systems risks shoeorning research materials into inaccurate or harmful categorizations. For example, the internet database ImageNet—which collected 14 million images, scraped from the internet, and standardized them into categories based on what they were images *of*, all for the purpose of developing training sets of objects for machine vision—has been widely criticized for producing a classification system that is replete with bias, discriminatory, and therefore actively harmful (Crawford and Paglan 2019). Similarly, attempts to shoehorn humanistic data policy into *one* policy, *one* storage infrastructure, *one* standardization, will certainly produce gaps in, and misrepresentations of, research materials, at best, and other harms at worst. Potential problems, for example, stretch from technical problems with digital objects that become corrupted or produce errors in database infrastructures, to epistemological problems with research materials being turned into “data” and therefore altered in content, form, and meaning.

It was evident from our survey (and this thread was continued during our event) that it is not possible to reduce humanities research to data work. Much of humanities research is, in terms of methodology, either *hermeneutic* (interpretive) or *heuristic* (involving discovery through mixed methods that may not be replicable in the way a scientific study is designed to be). The humanities use a diverse cast of research materials and methods. The same materials may have different uses and meanings in different projects.

From these insights, we can make the following recommendation:

While a degree of standardization may be necessary for research data management to enable

interoperability between datasets and infrastructure, flexibility is also required to support humanistic methodologies. Support sessions for humanists should ask them about their research materials and endeavour to develop or extend infrastructure options and standards that are adaptable and flexible to different humanistic research materials and methods.

4.2 What does “data” mean in the Humanities? Dr. Miriam Posner, UCLA

Our keynote speaker, Dr. Miriam Posner, a Digital Humanist and professor in Information Studies at UCLA, presented on “[Data Trouble](#),” asking whether there is a specifically “humanist” way of working with data. The answer Posner gave was, roughly, “it’s complicated”—because humanities scholars don’t generally think of their work as data, even if they are not generally opposed to the idea of “data” per se. So what gives? Citing the work of Christine L. Borgman (2015), Posner suggests that the right question to ask about data is not *what* it is but *when* it is; in other words, when do we call research materials data, since the “dataness” of any set of materials may inhere not in the qualities of the materials themselves, but in *how* the scholar intends to use them?

When do we get to the moment that marks a set of sources as data or not data? In particular, humanities work goes beyond positivism, venturing into speculation and interpretation, both methods of labour that are more difficult to classify into a “this” or “that” of a dataset. At the same time, the humanities are not “anti-evidence.” Posner instead argues the answer lies in some fundamental (but complicated) assumptions that get activated as soon as a source is designated as “data”:

- The task of *demarcation* of the boundaries of, and between, research objects, events, and people for computational processing, when humanists tend not to see these as cleanly disconnected from one another;
- The task of *parameterization*, or whether one data point can be sensibly measured against another, and the difficulty of developing standard scales of measurement for a morass of materials that may not be clearly demarcated;
- The trouble with *ontological stability*, or the way dividing the world into categories expresses assumptions about how we see the world, and the trouble this factor can induce for making generalisations at scale from data;
- The problem of *replicability* of the results of data work by other researchers;
- The problem of the *boundedness* of data versus the plenitude of the empirical world;
- The trouble with *deracination*, or data’s propensity to mean something different as they become distanced from the object or phenomenon they represent.

Each of these assumptions, Posner argued, fall under the umbrella term *categorization*, which is both the axiom of data work (i.e., turning a *ménagerie* of things into a dataset), and at the heart of humanistic reluctance to embrace data. Humanists do not inherently agree on what categorization is, let alone on its application in specific contexts. This is not to say that categorization is bad (it is, after all, the central task of humanities work too), but that the ability to critically examine categorization itself is integral to humanities work. For funding bodies, institutions, and humanities researchers, one central task of research data management will be developing infrastructures that achieve a measure of standardization that supports widespread access, while ensuring researchers do not lose the ability to critically engage with different theories, methods, and practices of categorization in their own work.

4.3 “Data considerations across Humanities disciplines” Roundtable

This roundtable had 5 speakers, who shared pre-recorded talks with attendees, and then answered questions in a roundtable format.

Dr. Jon Bath, Department of Art and Art History, University of Saskatchewan

Dr. Bath is a specialist in the visual arts, and focussed on the role of documenting *process* as data in research materials. The Canadian Tri-Agency [Research Data Management Policy](#) defines data as “primary sources,” a definition Bath troubles by asking what actually counts as, and can be validated as, a primary source in Fine Arts research? Depending on the research in question, a primary source could be any of a series of sketches used to plan out an artistic artefact, the artefact itself, or any number of copies or pieces inspired by the artefact. What portions of these processes count as the object to be stored in an institutional repository? Because, as Bath argues, the Fine Arts are often left out of conversations about research data management, or are treated as exceptional, practitioners, by failing to become involved in research management planning, and by failing to consider what parts of their work and research should be stored in repositories, risk the loss of an expanded definition of future primary sources. For many artists and researchers, this loss would encompass not only the loss of archives, but of the theories and ideas that spring from them. For Bath, the critical recommendation for research data management is the inclusion of the *process* of “data” creation and research outputs into data management policy. This means data management plans and infrastructure that are prepared and able for researchers to add a diverse cast of materials, without siloing that work into an excluded “Fine Arts” field. On a conceptual level, this recommendation involves not only an inclusion in systems of categorization, but a chance to examine what those categories include and leave out.

Dr. Constance Crompton, Department of Communication, University of Ottawa

Dr. Crompton is a co-director of the Lesbian and Gay Liberation in Canada project, archiving the 20th-century gay—and contemporary queer—liberation movement in Canada. The major challenge of the project has been balancing innovative archival data work with assurances of the stability of physical materials being digitized, analysed, and visualized for a public history website. As one example, because of the size of the archive, and because of the goal of producing a sustainable public history website for the long term, using boutique and bespoke programming languages for this task is not sustainable. A public website that “dies” when underlying code no longer functions on the platform used is not a useful website. Many of the questions the project has asked are about what types of data, and what encoding formats, are useful for project longevity, when future users may not be familiar with the standards used at the time of publication. For Crompton, the critical recommendation for research data management is thinking through how even the most standardized datasets may change with time, and how different audiences may rely on different interpretive frameworks even for those standards. In other words, successful standardization today is not a guarantee of standardization tomorrow.

Dr. Ewa Czaykowska-Higgins, Department of Indigenous Education and Humanities, University of Victoria

Dr. Czaykowska-Higgins’ presentation focussed on Indigenous language documentation and its management alongside communities. Czaykowska-Higgins emphasizes the degree to which reducing language to “data” empties it of culture, history, communication, connection to place and space, and experience. As one example of Humanities “data,” language is a type of public or collective memory, which may be lost if languages are reduced to a standardized reference like a digital dictionary. Data work, in collaboration with language revitalization efforts, relies on extensive collaboration with Indigenous communities, whose goals and needs should be the *first* priority of the work, rather than an add-on later. Much of what can be called “public memory,” or archival, institutions, have roots in the colonial history of Canada, and are liable to continue practices of data work as “extraction,” recalling the notion that data is always “captured” in uneven and power-laden ways. For this sort of data work, the processes of tool choice, planning, management, design, and collaboration must begin *before* “data work,” more literally speaking, begins. For Czaykowska-Higgins, the critical recommendation for research data management involves prioritizing consultation with communities of practice, engaging with histories of the data *and* the archive, and preventing harm by choosing people before data.

Dr. Felicity Tayler, Interim Head, Research Services, Arts and Special Collections of the University of Ottawa Library

Dr. Tayler is a co-applicant on the SSHRC-funded partnership the [SpokenWeb](#), which includes 28 institutions across Canada and the USA, and produces a significant amount of humanities data from scholarly and creative work with audio recordings (present or past). For SpokenWeb, the key to successful data curation and sharing has been the inclusion of librarian co-applicants at each partner institution, ensuring the inclusion of necessary expertise in data management. Included in this process was the development of robust, bespoke metadata and cataloguing systems (to ensure a standardized but specific method for engaging with project data across institutional partners), as well as shared ethics protocols. Where SpokenWeb partners were not able to come up with standardized conventions, they nevertheless developed centralized documentation practices that would allow other and future researchers to contend with these contingencies and idiosyncrasies. For Tayler, the critical recommendation for research data management involves taking the time to think about and organize data for the long-term, with contingencies in mind. Two practices relevant to this recommendation are the inclusion of essential support staff *as part of the project from the beginning*, and the development of methods and standards for documenting uncertainty for different researchers and audiences.

Dr. Caroline Winter, Dr. Graham Jensen, Alyssa Arbuckle, Dr. Raymond Siemens, University of Victoria

Members of the Electronic Textual Cultures Lab at the University of Victoria presented together about the [Canadian HSS Commons](#), an open online space for researchers and other stakeholders to collaborate around research. The members were keen to point out the disparities between institutional expectations for research involving data, and the lack of training and support for researchers in research data management. Additionally, in Canada, there are considerations related to differences in copyright law, privacy, and storage regulations between provinces, and compliance with data management policies—all of which are deserving of further scrutiny, policy, and practice. The members' critical recommendation for research data management involves building stronger connections with research communities to support a diversity of needs, and to prevent the buildup of siloed research. The members emphasized that simply building data infrastructure without proper consultation—at both national and institutional levels—may lead to either misuse or nonuse of those infrastructures.

4.4 Data Management Planning Workshop, Post-Event Survey, and Other Considerations

Our afternoon began with a workshop on “[Creating a Data Management Plan with the DMP](#),” facilitated by James Doiron (University of Alberta), Shahira Khair (University of Victoria), and Robyn Nicholson (NDRIO-Portage). The workshop was structured as an Introduction to Data Management Planning, with key components, a roundtable discussion on the many associated challenges, and finally an overview of Portage's DMP Assistant.

The speakers defined Research Data Management as managing data throughout all phases (active and beyond) of the research cycle. Tri-Agency RDM Policy currently comprises three (3) parts: an Institutional Strategy, the addition of Data Management Plans (DMPs) to certain funding application processes, and finally a data and metadata deposit into digital repositories during or after agency-supported research. For the Tri-Agencies, the DMP—a formal document that articulates the strategies and tools that will be used to manage research data—is integral to a successful lifecycle of research planning, strategy, implementation, and outcomes. Workshop attendees were given a detailed breakdown of DMPs, as well as some examples, available [here](#). Attendees were also introduced to the [Portage Network](#), a national RDM network dedicated to the capacity-building and coordination of RDM in Canada. Of key use to researchers, attendees were introduced to Portage's DMP Assistant, a tool designed to guide and

support researchers as they engage with the Tri-Agencies' new RDM requirements. A breakdown of the Assistant is available [here](#). In addition to supporting researchers, the Assistant is able to provide institution-specific templates, guidance, and information, with an emphasis on new templates designed for the Arts, the Humanities, and mixed methods.

As readers will discover in the following pages, while there is significant interest in new RDM practices from the Tri-Agencies, researchers attending the workshop were especially concerned about the added administrative burden of RDM without increased funding and interpersonal support. This concern will remain a major barrier to researcher buy-in to RDM, even with increased outreach and instruction from the Tri-Agencies and other supporting institutions.

Following our event, the Post-Event Survey asked two questions of attendees. We received fewer responses (n=10), but the responses we did receive were detailed and summarized some of the points discussed during our event. All questions in the Post-Event Survey were mandatory; for ease of reading, responses are summarized in point form below, in order to get a general sense of the concerns and anxieties for humanities researchers working with data. The first question is more general in nature, and survey responses follow many of the trends engaged in the pre-event survey, and in our panels.

What kind of training and support would be most helpful for you in improving your capacity for research data management?

- Support and training for Data Management Plans;
- A glossary of terms;
- A nation-wide data curation service that includes researchers without institutional affiliation;
- Ensured data security (and information about security) across different platforms and storage options;
- Guidance on how to make decisions about which platforms are best for the research in question;
- Workshops according to disciplines and their lexicons and methods;
- Assurances that researchers without institutional affiliations will not be at a disadvantage during these processes;
- Information and guidance on the adjudication of data management by funding bodies;
- Further guidance on how diverse and complex objects like culture or art may be categorized as “data”;
- Further guidance on how projects will be adjudicated; how will the Tri-Agencies ensure that grant funding does not privilege specific types of data work that uses specific infrastructures?;
- Mentorship with continuity (as opposed to a help desk format).

What concerns you most about the new Tri-Agency requirements for research data management?

- That RDM will be yet another hoop to jump through for funding, prompting researchers to do the bare minimum data management;
- That FAIR data needs to be curated and there won't be the proper resources to do this;
- A lack of sustained resources to assist and support humanists in this area;
- Discouraging smaller or non-affiliated researchers and projects from applying for funding;
- Archival processes that are standardized and simplified to meet a baseline and unable to contend with the diversity of humanities research (e.g. Indigenous languages, or typeface issues);
- The additional work now involved to secure funding, particularly for students, early-career researchers, and smaller projects;
- Sensitive data issues and public/private repository questions;
- Technical vocabulary requirements and the imposition of protocols without established sufficient support;
- How might DMPs be helpful tools rather than more bureaucratic labour?

These responses vary in their complexity and scope, but some trends emerge. In addition to calls like

those from [this](#) report (Rockwell, Huculak, and Château-Dutier 2020), researchers continue to feel like they need support at *all* stages of the research data management process: on conceptions and theories of data; on guidance for Tri-Agencies Policy; on how to meet new funding requirements; on making choices about data infrastructure use; and on how to ensure their research does not change in kind in order to meet data policy. Some of these concerns may be regarded as anxieties about new policies and methods for performing research. However, it is worth re-emphasizing the extent to which researchers have significant questions about *all* aspects of research data management.

NDRIO's 2020 call for white papers generated a decent number of reports on the ways in which Canada's Digital Research Infrastructure (DRI) can better support humanist research. (Estill 2020; Evalyn et al. 2020; O'Donnell 2020; Rockwell et al. 2020; Siemens and Arbuckle 2020). These whitepapers generally focus on infrastructure, staffing concerns, and technical fixes, with only a few recognizing the need for increased attention to data and data policy in Canada. Moreover, while there are principles for data management like FAIR, CARE, and TRUST, the question of how these principles can be operationalized, and how they may be applied to humanities data, is not settled.

This report does not dispute the need for humanist-centred infrastructure, but instead suggests that while there is a technical distinction between the back and front end, the ways in which humanist data work engages with research data management makes the line between back and front end less neat. In other words, humanists need infrastructural support and capacity in both the back and front end. As this report has argued, Humanists think intensively about *processes of categorization*, about *interpretation*, and about the *presentation of research*. The back end, or the "data access layer," is used to store and access data in categories; research materials become data through categorization, and the back end's ability to establish and maintain these categories and parameters is what allows data to be processed computationally. The front end often contains substantial contextual information that helps to situate and interpret the data held on the back end. Significant technical skill, as well as data expertise, is required to interact with a back end in a confident and comprehensive way. In the first instance, then, requiring Humanists to store their data digitally (i.e., in "back ends") without requisite technical training or support establishes a *skills gap*. Humanists can become alienated from their research materials, and data can become alienated from their contexts.

As DH scholar Johanna Drucker has noted, "one point that needs to be made clearly and strongly is that 'the machine' does not do the work. While various procedures can be automated along the way, the bulk of the labor [of data] requires skilled, trained, thoughtful, and patient professionals, who are the key factors in these projects" (2017). In this quotation, the "machine" can refer to both the literal tool (e.g. servers, hardware, software) and technical support staff as the "mindings" of the machine. There are 2 problems with this kind of approach. First, in both of these senses, merely presenting researchers with more or "better" machines in the form of storage options like national public repositories does not solve the major issue of a lack of knowledge and conceptual support to serve researcher needs. Second, much of data work consists not of the collection or taking of data, but of the preparation of the system in advance, the processing during collection, and storage as a form of display of information, or, a representation of data. All of these processes are labour-intensive.

As this report has argued, much of humanist data work involves the production of data from research materials that are already themselves a kind of information. This labour includes processing work that is interpretive or experimental, rather than merely "cleaning" numerical data so that it can be read by a machine. This report calls for further consultation with institutions and researchers to explore edge cases that are difficult to fit into current RDM practice, including "analogue" research that is based mostly in print material, and research processes in the Fine Arts. The needs of large DH projects working with highly structured data will not be the same as small independent projects working with scanned corpora or ephemera. As a result, standardization of humanist RDM, as it stands, is likely to exclude or constrain more research materials than it includes or supports (or, it is likely to alter some research materials to

such an extent that the research is no longer the same).

This report recognizes that the implementation of principles and recommendations to RDM operations is not always straightforward. We have stressed the importance of project *design* before data work starts, as well as an accounting of the *labour* involved in prospective projects: who will be needed to support the project, when, and where? The DMP questions related to Responsibilities and Resources offer an important way to articulate the *labour* involved in the research project in question. What tools will be required to complete this research? Who will need to be employed by the project to use these tools to perform tasks, at what stage, and for how long? What kinds of expertise will these workers need to have? In addition to centring the role of human labour in RDM, this record will also establish project *histories* before they begin, engaging more completely with the communities of practice involved, and presenting future researchers with the means to understand how labour is divided amongst projects, as well as to establishing networks of *people* involved with research projects. Projects that anticipate and record the labour involved (even if only in terms of roles rather than persons) before beginning are better positioned to ensure that the work is able to happen when the project moves from planning into production.

4.5 The “Three P’s” of Data Work

For many researchers, developing data management plans will be a novel, and perhaps uncomfortable exercise. There are, and will continue to be, questions to be asked about how to operationalize many of the recommendations made in this report. In a summary format, we recommend that institutions, practitioners, researchers, and other interested parties take away the following 3 objectives for data work. They can be expressed as the “Three P’s” of data work:

Data as Principles

- All data work requires preparatory *conceptual* and planning work, and should follow the principles outlined in the second section of this report.

Data as Practice

- Practice means both “doing” and “*re*-doing”; it refers to the actions and methods of research (rather than the ideas or desires, though these are related);
- Practice is *labour* (by people and by machines);
- Practice is *experimentation* (projects and data work change, and projects that are prepared for these contingencies will produce better information and knowledge);
- The combination of “labour” and “experimentation” equals *design* (the notion that projects cannot perfectly anticipate endings, and that standards need to be designed to fit projects).

Data as Presentation

- Research Data Management is not only about “raw data storage” but about “information processing and representation”;
- A humanities focus on *mixed media*,
 - The objects of inquiry in the humanities are not the same, and as such, data will not be;
 - Instead, data presentation should focus on *accessibility* to other researchers or the public, and *reusability* (less about interoperable metadata on large numerical datasets, than whether intended audiences can understand project outputs and reuse them);
- Data, or processed information, is always presented to an *audience*,
 - Researchers (and the institutions and infrastructures that support them) should regard these audiences as other political actors.

4.6 What are the consequences for humanists if they do not accept data management?

The consequences for humanists if they do not accept data management are fairly straightforward. They are primarily focussed on insufficient or improper resources; on bad or unusable project outputs; on research outputs that disappear once project funding dries up; and finally on consolidation of humanities research into exclusive mega-project categories. In point form, here are some examples:

- Loss of funding;
- Inadequate or irrelevant tools;
- Improperly planned projects nevertheless subject to the bureaucratic requirements of RDM, meaning more labour, more time, and less useful distribution of funding;
- Difficulty communicating research to intended audiences (especially beyond esoteric and exclusive academic journals);
- The political stakes of not engaging with the infrastructure of research (leading to a technologically determinist set of outcomes that can harm researchers and subjects);
- All humanist research materials treated as data, but not in ways humanists want or agree with;
- Humanists becoming seen incorrectly as the “analogue” to the other, more digital disciplines;
- The shoehorning of humanist research work into categories like “ethics” rather than the full fields it encompasses.

If researchers are not provided with adequate support, the Tri-Agencies risk a scenario where researchers prioritizing policy requirements at the expense of research requirements end up producing bad research with unusable data. The consequences of insufficient and improper support are twofold. First, SSHRC is likely to see a concomitance of increases in mega-projects from cash- and support-rich institutions, alongside decreases in smaller research projects with boutique needs or without institutional support. Second, SSHRC is likely to see the arrival of a lot of unusable datasets: an unfortunate increase in the amount of projects that, due to new regulations, have data and use repositories, but that are nonfunctional, whether because they are not interoperable, cannot be read, or are insufficiently documented.

5.0 Bibliography:

Amoore, Louise. 2020. *Cloud Ethics: Algorithms and the Attributes of Ourselves and Others*. Durham, NC: Duke University Press.

Ananny, Mike. 2020. “Public Interest and Media Infrastructures: Regulating the Technology Companies That Make ‘Pictures in Our Heads.’” Research Report. Max Bell School of Public Policy at McGill University.

https://static1.squarespace.com/static/5ea874746663b45e14a384a4/t/5f086e1dda5c0354abd84aa4/1594387999164/MTD_Report_Ananny.pdf.

Antoniuk, Jeffrey, and Susan Brown. 2020. “Interface Matters.” 054. University of Alberta; The Canadian Writing Research Collaboratory. https://engagedri.ca/wp-content/uploads/2020/12/CWRC_Interface_Matters_NDRIO_White_Paper.pdf.

Apprich, Clemens, Wendy Hui Kyong Chun, Florian Cramer, and Hito Steyerl. 2018. *Pattern Discrimination. In Search of Media*. Minneapolis, MN: University of Minnesota Press.

- Austin, Claire C. 2018. "A Path to Big Data Readiness." https://www.researchgate.net/publication/329591852_A_Path_to_Big_Data_Readiness.
- Birch, Kean. 2019. "Personal Data Isn't the 'new Oil,' It's a Way to Manipulate Capitalism." *The Conversation*. 2019. <http://theconversation.com/personal-data-isnt-the-new-oil-its-a-way-to-manipulate-capitalism-126349>.
- Borgman, Christine L. 2015. *Big Data, Little Data, No Data: Scholarship in the Networked World*. Cambridge, MA: MIT Press.
- Bosker, Bianca. 2019. "Why Everything Is Getting Louder." *The Atlantic*. October 8, 2019. <https://www.theatlantic.com/magazine/archive/2019/11/the-end-of-silence/598366/>.
- Bowker, Geoffrey, and Susan Leigh Star. 1999. *Sorting Things Out: Classification and Its Consequences*. Cambridge, MA: MIT Press.
- Boyd, Jason. 2020. "DRI, University Libraries and Digital Humanities Research Centres." 091. Centre for Digital Humanities, Ryerson University3. https://engagedri.ca/wp-content/uploads/2020/12/Ryerson_NDRIOWhitePaper.pdf.
- Brin, Sergey, and Lawrence Page. 1998. "The Anatomy of a Large-Scale Hypertextual Web Search Engine." <http://ilpubs.stanford.edu:8090/361/1/1998-8.pdf>.
- Broman, Karl W., and Kara H. Woo. 2018. "Data Organization in Spreadsheets." *The American Statistician* 72 (1): 2–10. <https://doi.org/10.1080/00031305.2017.1375989>.
- Brown, Susan. 2020. "Sustaining Digital Research Infrastructure in the Humanities." 106. University of Guelph. https://engagedri.ca/wp-content/uploads/2021/01/SBrown-Sustaining_DRI_in_the_Humanities.pdf.
- Chun, Wendy Hui Kyong. 2015. "On Hypo-Real Models or Global Climate Change: A Challenge for the Humanities." *Critical Inquiry* 41 (3): 675–703. <https://doi.org/10.1086/680090>.
- . 2016. "Big Data as Drama." *ELH* 83 (2): 363–82. <https://doi.org/10.1353/elh.2016.0011>.
- Chun, Wendy Hui Kyong. 2008. "The Enduring Ephemeral, or the Future Is a Memory." *Critical Inquiry* 35 (1): 148–71. <https://doi.org/10.1086/595632>.
- Crawford, Kate, and Trevor Paglen. 2019. "Excavating AI." <https://excavating.ai>.
- Doiron, James. 2020. "Data Management Plan Exemplar #3: Mixed Methods Fictional Exemplar," September. <https://doi.org/10.5281/zenodo.4019563>.
- Drucker, J. 2011. "Humanities Approaches to Graphical Display." *Digital Humanities Quarterly* 5 (1). <http://www.digitalhumanities.org/dhq/vol/5/1/000091/000091.html>.
- Drucker, Johanna. 2017. "The Back End: Infrastructure Design for Scholarly Research." *The Journal of Modern Periodical Studies* 8 (2): 119–33.
- Eberle-Sinatra, Michael, and Emmanuel Chateau-Dutier. 2020. "Développer une infrastructure de services numériques pour les Humanités numériques canadiennes: livre blanc pour la NOIRN." 016. Centre de recherche interuniversitaire sur les humanités numériques. <https://engagedri.ca/wp-content/uploads/2020/12/crihnLivreBlancNOIRN.pdf>.
- Estill, Laura. 2020. "All Researchers Use Digital Resources: On Campus Support, Grants, Labs, and Equity." 022. StFX Digital Humanities Centre. https://engagedri.ca/wp-content/uploads/2020/12/NDRIO-White-Paper-w-CenterNet-SUBMITTED_fixed-1.pdf.

- Evalyn, Lawrence, Elizabeth Parke, Patrick Keilty, and Elspeth Brown. 2020. "Gaps in Digital Research Infrastructure for Canadian Digital Humanities Researchers." 018. University of Toronto.
https://engagedri.ca/wp-content/uploads/2020/12/NDRIO_DH_white_paper_digital_humanities_university_of_toronto.pdf.
- Galloway, Alexander. 2011. "Are Some Things Unrepresentable?" *Theory, Culture & Society* 28 (7–8): 85–102. <https://doi.org/10.1177/0263276411423038>.
- Gitelman, Lisa, ed. 2013. *Raw Data Is an Oxymoron*. Infrastructures Series. Cambridge, MA: The MIT Press.
- Government of Canada, Innovation. n.d. "Frequently Asked Questions Tri-Agency Research Data Management Policy." Accessed May 3, 2021. https://www.ic.gc.ca/eic/site/063.nsf/eng/h_97609.html.
- . n.d. "Public Consultation Summary – Science.Gc.Ca." Accessed May 3, 2021. https://www.ic.gc.ca/eic/site/063.nsf/eng/h_97905.html.
- . n.d. "Tri-Agency Research Data Management Policy – Science.Gc.Ca." Accessed May 3, 2021. https://www.ic.gc.ca/eic/site/063.nsf/eng/h_97610.html.
- Gray, Jonathan, Carolin Gerlitz, and Liliana Bounegru. 2018. "Data Infrastructure Literacy." *Big Data & Society* 5 (2). <https://doi.org/10.1177/2053951718786316>.
- Gray, Vincent, and Alexandra Cooper. 2020. "Data Management Plan Exemplar #2: Digital Humanities and Secondary Data," September. <https://doi.org/10.5281/zenodo.4019309>.
- Green, Ben. 2021. "The Contestation of Tech Ethics: A Sociotechnical Approach to Ethics and Technology in Action." *ArXiv:2106.01784 [Cs]*, June. <http://arxiv.org/abs/2106.01784>.
- Grguric, Eka. 2019. "Minimal Computing: One Approach to the Challenge of Computational Reproducibility – UBC Library Open Collections." 2019.
<https://open.library.ubc.ca/cIRcle/collections/ubclibraryandarchives/494/items/1.0387127>.
- Halpern, Orit. 2014. *Beautiful Data: A History of Vision and Reason since 1945*. Durham, NC: Duke University Press.
- Harris, Amy. 2021. "Nanna Bonde Thylstrup – Dataset Ethics: Deleting Archives, Encountering Remains | Digital Democracies Institute." <https://digitaldemocracies.org/nanna-bonde-thylstrup-dataset-ethics-deleting-archives-encountering-remains/>.
- Harrower, Natalie, Maciej Maryl, Timea Biro, and Beat Immenhauser. 2020. "Sustainable and FAIR Data Sharing in the Humanities: Recommendations of the ALLEA Working Group E-Humanities – Digital Repository of Ireland." <https://repository.dri.ie/catalog/tq582c863>.
- Harsin, Jayson. 2018. "Post-Truth and Critical Communication Studies." In *Oxford Research Encyclopedias. Communication*.
<https://oxfordre.com/communication/view/10.1093/acrefore/9780190228613.001.0001/acrefore-9780190228613-e-757?print=pdf>.
- Hong, Sun-ha. 2020. *Technologies of Speculation: The Limits of Knowledge in a Data-Driven Society*. New York, NY: NYU Press.
- Junker, Marie-Odile, and Delasie Torkornoo. 2020. "Indigenous Language Technologies and Online Resources: Algonquian Dictionaries Project and Algonquian Linguistic Atlas." 029. Carleton University.
https://engagedri.ca/wp-content/uploads/2020/12/Junker_Torkornoo_white_paper_ndrio_2020.pdf.
- Kirschenbaum, Matthew. 2014. "Software, It's a Thing." Medium. July 26.
<https://medium.com/@mkirschenbaum/software-its-a-thing-a550448d0ed3>.

- Kitchin, Rob. 2014. *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. London, UK: SAGE Publications.
- Lacroix, Denis, and Sathya Rao. 2020. "Data Management Plan for Belgians and French in the Prairies (Exemplar)," October. <https://doi.org/10.5281/zenodo.4062484>.
- Lampland, Martha, and Susan Leigh Star. 2009. *Standards and Their Stories: How Quantifying, Classifying, and Formalizing Practices Shape Everyday Life*. Ithaca, NY: Cornell University Press.
- Levy, Michelle. 2020. "Data Management Plan Exemplar #1: Digital Humanities," October. <https://doi.org/10.5281/zenodo.4064168>.
- Lincoln, Matthew D. 2020. "Scoping _The Index of Digital Humanities Conferences_ for Now and for Later." *Matthew Lincoln, PhD* (blog). October 12. <https://matthewlincoln.net/2020/10/12/index-of-digital-humanities-conferences.html>.
- Mattern, Shannon. 2021. "How to Map Nothing." *Places Journal*, March. <https://placesjournal.org/article/how-to-map-nothing/>.
- Mirzoeff, Nicholas. 2011. *The Right to Look: A Counterhistory of Visuality*. Durham, NC: Duke University Press.
- Mulvin, Dylan. 2021. *Proxies: The Cultural Work of Standing In*. Cambridge, MA: The MIT Press.
- O'Donnell, Daniel Paul. 2020. "'Good Things Come in Small Packets': How (Inter)National Digital Research Infrastructure Can Support 'Small Data' Humanities and Cultural Heritage Research." 001. "Good Things" Research Team, University of Lethbridge. <https://engagedri.ca/wp-content/uploads/2020/12/DRI-Infrastructure-White-paper-Good-things-come-in-small-packets.pdf>.
- Onuoha, Mimi. 2016. "The Library of Missing Datasets — MIMI QNŪQHA." MIMI QNŪQHA. 2016. <https://mimionuoha.com/the-library-of-missing-datasets>.
- Paquette-Bigras, Ève. 2020. "Data Management Plan: Soundscape Study (Exemplar)," September. <https://doi.org/10.5281/zenodo.4056785>.
- Pasquale, Frank. 2015. *The Black Box Society*. Cambridge, MA: Harvard University Press.
- Quinless, Jacqueline, and Shahira Khair. 2019. "The Enduring Potential of Data: An Assessment of Researcher Data Stewardship Practices at the University of Victoria." <https://dspace.library.uvic.ca/handle/1828/10509>.
- Rockwell, Geoffrey, Matt Huculak, and Emmanuel Château-Dutier. 2020. "Canada's Future DRI Ecosystem for Humanities & Social Sciences (HSS)." 020. The Canadian Society of Digital Humanities. https://engagedri.ca/wp-content/uploads/2020/12/CSDH_NDRIO_WhitePaper.pdf.
- Sadowski, Jathan. 2019. "When Data Is Capital: Datafication, Accumulation, and Extraction." *Big Data & Society* 6 (1). <https://doi.org/10.1177/2053951718820549>.
- Sandvig, Christian. 2013. "The Internet as Infrastructure." In *The Oxford Handbook of Internet Studies*. Oxford, UK: Oxford Handbooks Online. <https://doi.org/10.1093/oxfordhb/9780199589074.013.0005>.
- Sayers, Jentery. 2016. "Minimal Definitions." *Minimal Computing* (blog). 2016. <https://go-dh.github.io/mincomp/thoughts/2016/10/02/minimal-definitions/>.
- ed. 2018. "Remediation, Data, Memory." In *The Routledge Companion to Media Studies and Digital Humanities*. New York, NY: Routledge.
- Siebert, Bernhard. 2011. "The Map Is the Territory." *Radical Philosophy* 169 (Sept/Oct).

<https://www.radicalphilosophy.com/article/the-map-is-the-territory>.

Siemens, Ray, and Alyssa Arbuckle. 2020a. "HQP Pathways: Engaging the Canada's Different Disciplinary Models for HQP Training and Funding to Facilitate DRI Uptake in Canada." 055. Digital Humanities Summer Institute. https://engagedri.ca/wp-content/uploads/2020/12/DHSL_NDRIO-whitepaper_12-14-20.1.pdf.

———2020b. "Steps to Success in Ensuring DRI Engages and Mobilizes Humanities and Social Science Research." 063. Implementing New Knowledge Environments. https://engagedri.ca/wp-content/uploads/2020/12/INKE_NDRIO-whitepaper_12-14-20.1.pdf.

Star, Susan Leigh. 1992. "The Trojan Door: Organizations, Work, and the 'Open Black Box.'" *Systems Practice* 5 (4): 395–410. <https://doi.org/10.1007/BF01059831>.

Taylor, Felicity, and Maziar Jafari. 2021. "Shifting Horizons: A Literature Review of Research Data Management Train-the-Trainer Models for Library and Campus-Wide Research Support Staff in Canadian Institutions." *Evidence Based Library and Information Practice* 16 (1): 78–90. <https://doi.org/10.18438/ebliip29814>.

Taylor, Felicity, Chantal Ripp, and Maziar Jafari. 2020. "RDM Readiness Report: Shifting Horizons II: Realities of Research Data Management." Working Paper. <https://doi.org/10.20381/s1pj-1x65>.

Tenen, Dennis. 2016. "'9. Blunt Instrumentalism: On Tools and Methods' in 'Debates in the Digital Humanities 2016' on Debates in the DH Manifold." *Debates in the Digital Humanities*. <https://dhdebates.gc.cuny.edu/read/untitled/section/09605ba7-ca68-473d-b5a4-c58528f42619>.

Thorp, Jer. 2017. "You Say Data, I Say System." Medium. July 19. <https://blprnt.medium.com/you-say-data-i-say-system-54e84aa7a421>.

———2021. *Living in Data: A Citizen's Guide to a Better Information Future*. New York, NY.: Farrar, Straus and Giroux.

Thylstrup, Nanna Bonde, Daniela Agostinho, Annie Ring, Catherine D'Ignazio, and Kristin Veel, eds. 2021. *Uncertain Archives: Critical Keywords for Big Data*. Cambridge, MA: MIT Press.

Viljoen, Salome. 2021. "Data Relations." *Logic Magazine*. May 17. Accessed June 10, 2021. <https://logicmag.io/distribution/data-relations/>.

Wernimont, Jacqueline. 2018. *Numbered Lives: Life and Death in Quantum Media*. Cambridge, MA: MIT Press.

Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, et al. 2016a. "The FAIR Guiding Principles for Scientific Data Management and Stewardship." *Scientific Data* 3 (1): 160018. <https://doi.org/10.1038/sdata.2016.18>.

———2016b. "The FAIR Guiding Principles for Scientific Data Management and Stewardship." *Scientific Data* 3 (1): 160018. <https://doi.org/10.1038/sdata.2016.18>.

Wilson, Greg, D. A. Aruliah, C. Titus Brown, Neil P. Chue Hong, Matt Davis, Richard T. Guy, Steven H. D. Haddock, et al. 2014. "Best Practices for Scientific Computing." *PLOS Biology* 12 (1): e1001745. <https://doi.org/10.1371/journal.pbio.1001745>.

Zuckerman, Ethan. 2020. "What Is Digital Public Infrastructure?" Center for Journalism & Liberty. November 17. Accessed May 9, 2021. <https://www.journalismliberty.org/publications/what-is-digital-public-infrastructure>.

2 Comments

1.



Alyssa Arbuckle on November 25, 2021 at 5:03 pm

Thank you for this excellent, comprehensive, and nuanced report! I especially appreciated your consideration of “edge cases” as well as your in-depth engagement with how we talk (and think) about data.

One suggestion, if I may: in Section 4.0, you write “If you asked a person at random to describe types of data, they would likely give you one of three answers: data is either the product of scientific inquiry, a part of computer science, or “stuff” that gets collected by corporations and states about people, usually through online or offline surveillance. None of those answers involve the humanities.” I would argue that, in fact, **all** of these answers involve the humanities in some way! I do take your point — that there is a cognitive disconnect between the concept of data and the disciplinary output of the humanities, but I wonder if there is another way to express this??



•

Shahira Khair on November 26, 2021 at 11:34 am

Thanks Alyssa. I think that’s a good catch and we’ll address it in revisions. The point we’re trying to make is that preconceptions of “Data” do not immediately resonate with many humanists’ understanding of their research materials. I agree clarification needed.

Search



Archives

Categories

Community News

English

French

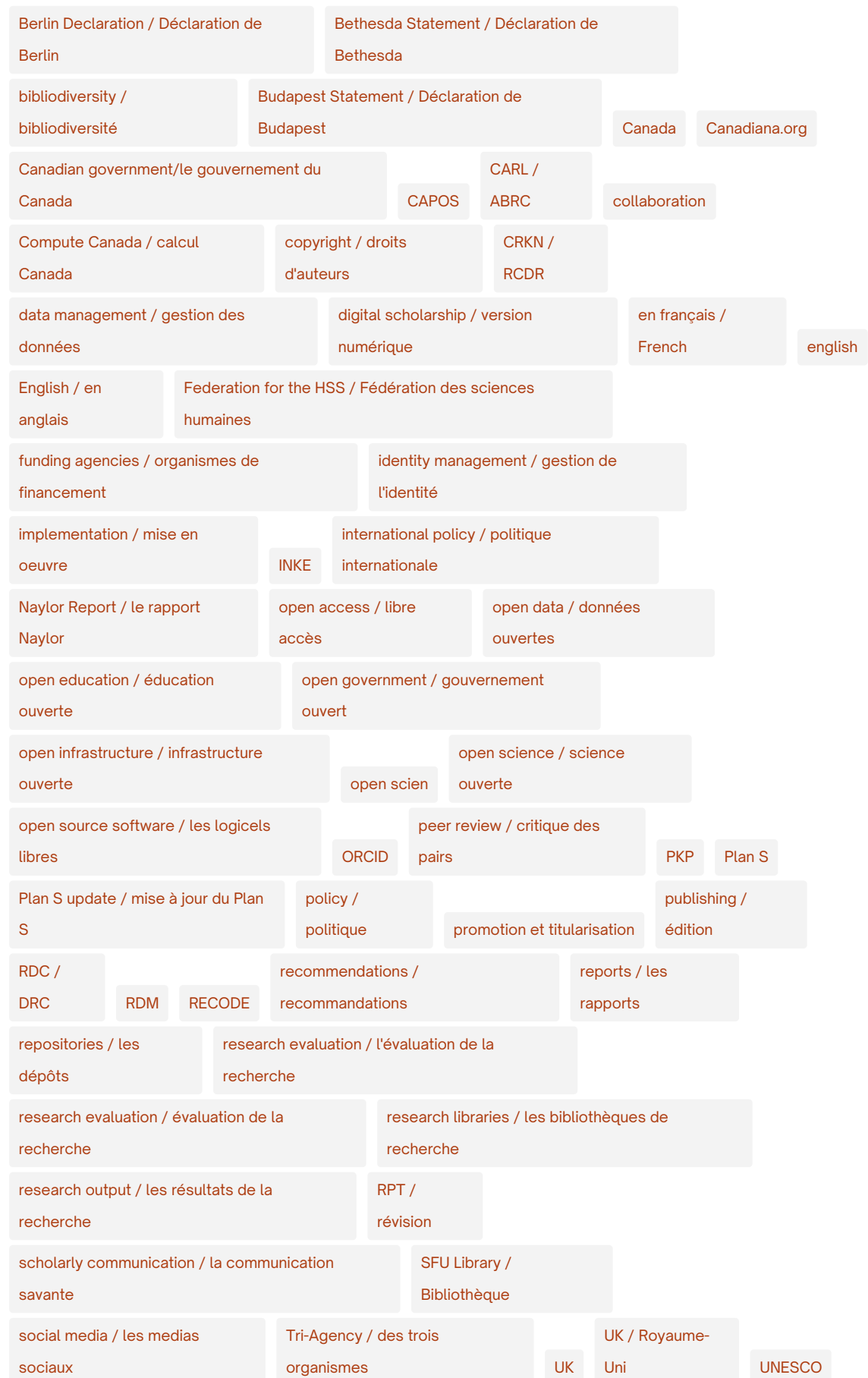
Observations

Observations and Responses

Policies

Responses

Tags





This work is licensed under a [Creative Commons Attribution-ShareAlike 4.0 International License](#).

