Advancing Interdisciplinary Data Science Training for the Humanities and Social Sciences

Andrea Zeffiro, Jason J. Brodeur, Sarah Whitwell, Allison Van, Michelle Dion, Vange Holtz Schramek, Veronica Litt

McMaster University

Description of Document: This report documents the outcomes and key learnings that stemmed from the 2020-2021 SSHRC funded The Relevant Research Series. A workshop series jointly organized and hosted by the Sherman Centre for Digital Scholarship and Spark: a centre for social research innovation at McMaster University. The recommendations can inform future interdisciplinary data science training initiatives at McMaster University and beyond.

Keywords: interdisciplinary data science training; data acumen; data literacy; data fluency; research support and acceleration; Open Educational Resources; knowledge mobilization

.

This report was supported by funding from the Social Sciences and Humanities Research Council of Canada.

How to cite this paper: Zeffiro, A., Brodeur, J.J, Whitwell, S., Van, A., Dion, M., Holtz Schramek, E., Litt, V. (2022). *Advancing Interdisciplinary Data Science Training for the Humanities and Social Sciences*. McMaster University.

EXECUTIVE SUMMARY

Context

Researchers in the Humanities and Social Sciences are increasingly undertaking 'data-driven research,' a broad term describing research that collects, manages, evaluates, and analyzes datasets of many sizes and types as critical evidence to make informed decisions (Bonikowska et al., 2019; Henderson and Corry 2021; Kitchin 2014; Maass et al., 2018; Mandinach and Gummer 2013; Vanhoof et al., 2011; Wolff et al., 2016). Scholars new to data-driven research benefit from collaborating with experts from different fields to identify the methods most appropriate to support their research questions. Of course, collaboration with experts will often extend well into the life cycle of research, as scholars learn how to communicate data-driven research narratives in compelling and accessible ways. To this end, data competency training is indispensable for researchers. It allows scholars to acquire the practical knowledge and skills to work with qualitative and quantitative data critically and fluently to glean useful information and synthesize findings for academic and non-expert audiences.

In general, extra-curricular training in data-driven approaches to research at many higher education institutions (HEIs) can be dispersed among departments, institutes, centres, and central support units such as libraries and reserach computing services. In contrast, specialized intensive training opportunities can be prohibitive due to financial, physical, logistic, and temporal constraints that can make such training more difficult to access, especially when participants must be physically present. At the same time, due to workforce shortages in data literacy, building foundational and translational interdisciplinary training resources is a top priority for HEIs.

Additionally, data training is critical for student learners. Extra-curricular instruction complements formal degree training and introduces data literacy fundamentals that equip learners with data fluency to meet cross-sectoral demands. Moreover, acquiring critical data literacies enables learners to become aware of their crucial role as contributors and consumers of data in their daily lives.

To begin to address the asymmetry of formal training and the shortage of data skills across sectors of the labour force, the leading research support centres within McMaster's University Library, Faculty of Humanities, and Faculty of Social Sciences collaborated on delivering a series of complementary and interdisciplinary workshops focused on communicating data-driven research results meaningfully to a wide range of stakeholders: ***The Relevant Research Series***. Led by the *Lewis and Ruth Sherman Centre for Digital Scholarship*, and *Spark: a centre for social research innovation*, the workshops were relevant to all empirical researchers even if they had no experience or very little interest in quantitative data-driven methods. As we detail in this report, the series was an opportunity to promote greater interdisciplinarity for data science training.

In the first part of the report, we provide pertinent background on the Relevant Research workshop series and reflect on its outcomes. The second part documents the key learnings from the 2020-2021 series, which can inform future interdisciplinary data science training initiatives at McMaster University and beyond.

A note on terminology

Because we advocate for **interdisciplinary data science training** throughout this report, it is worth noting how we understand 'interdisciplinary.' Conventionally, the term is used to describe research that stems from integrating perspectives, concepts, theories, methods, information, and data from multiple disciplines or bodies of knowledge. Interdisciplinarity produces outcomes or solutions beyond the purview of a single field or area of research practice (National Science Foundation 2004). Rather than advocate or present a single vision for interdisciplinary data science training, we wish to stress how building data competencies for varying levels of fluency - from machine learning to data justice - demands continuous thinking and working across long-established and somewhat restrictive academic boundaries.

We understand **'data science'** broadly in the way Alexander et al. (2019) define it as "an interdisciplinary scientific approach that provides methods to understand and solve problems in an evidence-based manner, using data and experience" (p. 57). We use 'data science' and 'data-driven research' interchangeably throughout the report, even though we understand how many people view these terms differently. ***Data science*** is often equated with mathematical, statistical, or computational fields, or the use of tools and methods from those fields in humanistic or social science disciplines. In contrast, ***data-driven*** is perceived as a more encompassing term to describe quantitative and qualitative approaches to working with data. These terms are used interchangeably throughout the report to accentuate that whether one identifies with data science or data-driven research, the fact is that interdisciplinary knowledge is required to work with and ethically collect, use and share (primarily quantitative or 'big') data. Further, we centre the term data science to acknowledge and advocate for the contributions from researchers in the humanities and social sciences to data science domains (Christin 2020; Gillborn et al., 2020; Iliadis and Russo 2016; Neff et al., 2017; Nikunen 2021; Seaver 2017).

In an effort to describe states of preparedness, we consider the 'data competencies' that may be developed to increase levels of 'data fluency' or 'data literacy' in researchers. ***Data competencies*** are the knowledge and skills required to use the appropriate software, tools, and processes to gather, organize, analyze, manage, and communicate data (StatsCan 2020), while ***data fluency*** or ***data literacy*** is the ability to read, create, and communicate meaningful information from data (MacGillivray 2017, p. 6; Sperry 2018) and encompasses a set of shared standards, processes, tools, and terms.

Through Relevant Research, we championed interdisciplinary data science training by offering multiple points of entry to engage asymmetrically connected ways of encountering and engaging with data, including but not limited to academic research, journalism, and critical design. To this end, the Relevant Research series sought to develop and promote data science training across disciplines and levels of preparedness. In addition, some sessions, such as those on communication planning and understanding audiences, applied to a broader range of research approaches, including those based on qualitative evidence or even purely theoretical interventions.

Goals and Objectives

The Relevant Research workshop series addressed a current gap in interdisciplinary training for data-driven research methods. Workshop content introduced participants with varied disciplinary backgrounds and degrees of preparedness to some of the best practices and strategies for mobilizing data-driven research to support evidence-informed decision-making (Lyons et al., 2016; Lyons et al., 2018). The objectives of Relevant Research were two-fold.

1. The workshops brought together specialized knowledge and critical approaches to data storytelling, visualization, and audience research to expand a common skill-set and engage participants in cross-disciplinary conversations.
2. And through the workshops, the Sherman Centre and Spark continued to build a fundamental awareness of and competence in data fluency in the humanities and social sciences through training, networking and support resources.

The collaboration also emphasized the need to build pathways between the research support units at McMaster University with expertise in data science. As discussed in the latter part of the report, working as a cohesive community on training, networking, and support resources can serve the campus, local, and regional communities. However, it is not easy to appraise all the training opportunities on campus, thus making it challenging to assess currently available resources, gaps in training, and areas of overlap that could lead to fruitful collaborations. To promote and strengthen collaboration across units, we recommend a two-fold approach that begins with a gap analysis to identify the proficiencies, omissions, and potential areas for partnering in interdisciplinary data science training, followed by developing a coordinated strategy for disseminating information about training opportunities. A systematic approach at the institutional level can lead to a resilient training, networking, and support infrastructure that reflects our campus community's heterogeneous needs, interests, and competencies and paves the way for inter-institutional data science training opportunities at regional and national levels.

Description of Training

The workshop series included practical hands-on introductions to the tools and approaches increasingly crucial for effectively communicating data-driven research narratives for public engagement. Workshop topics focused on how to: develop a research communications plan (Carlson & Johnston 2015; Exner 2014; Koltay 2017, Wilson et al., 2010), shape audience receipt and interpretation of information (Gaudet 2013; Joffe 2008), approach data-driven storytelling (Dykes 2015; Tufte 1986) and data visualization (Healy 2018; Healy and Moody 2014), apply a critical design approach to participatory data collection (Sinders 2020), and understand how audiences perceive information (Gaudet 2013; Joffe 2008). The emphasis on approaches to mobilizing insights and outputs generated from data-driven research brought together a wide variety of knowledge, skills, and technologies that are often not included in disciplinary training, which tends to focus on conducting the research with less emphasis on effective knowledge mobilization. Meanwhile, effective knowledge mobilization is often informed by the types of insights into audiences and communication strategies that our series sought to promote and that can be applied to a range of disciplinary research outputs.

The Sherman Centre and Spark will continue to provide interdisciplinary data science training for participants with divergent educational backgrounds and varying familiarity with data-driven approaches to research, which in this context refers to research using quantitative or ‘big’ data (Kitchin 2014). Providing instruction to campus community members on these topics and approaches is vital to support research excellence and meet in-demand data skills for the labour force. Learning how to communicate knowledge derived from data-driven research effectively is indispensable to sharing findings and meaningfully engaging the public. Yet it is only one of many competencies important for effective data-driven research. Future series will emphasize additional competencies like data analysis and research data management to continue addressing the gaps in data fluency.

Audience Engagement

The 10-event series attracted over 500 registrants, with many registrations external to McMaster University. While all but one of the events were held synchronously, session recordings and materials were compiled to create a total of 5 self-guided modules that are shared publicly for reuse by any audiences. Due to the global pandemic, all synchronous sessions were held via Zoom, which expanded the potential audience beyond McMaster and regional university communities.

While the topical focus was on Humanities and Social Sciences research, the skills and tools introduced are relevant across all fields, both at McMaster and other higher education institutions. As we detail in the report, a priority for future workshops is

to engage and build enduring relationships with local and regional community groups, not-for-profit organizations, and industry partners.

Observations and Key Learnings

1. **Collaboration:** The Relevant Research series has formed a strong foundation for the Sherman Centre and Spark to continue collaborating on extra-curricular data competency training. Beyond the two centres, we advocate for collaboration as a key element to all research training and support at McMaster.

2. **Innovative Training:** The Relevant Research Series was ahead of the curve with its digital-by-design training opportunities. Workshops were held both synchronously and asynchronously, and recorded content was redeveloped into open educational resources (OERs).

3. **Training Resources:** Because the series led to the creation of indispensable training resources, the Sherman Centre and Spark now have an impressive roster of OERs that can be integrated into courses by instructors, adapted to meet unique pedagogical needs and reshared alike, or used for self-guided learning.

4. **Formalize Training:** The Sherman Centre and Spark have an opportunity to build on the successes of the Relevant Research workshop series and address extant training gaps beyond knowledge mobilization by working collaboratively with other campus partners to coordinate interdisciplinary data skills training.

Recommendations

1. Complete an environmental scan of interdisciplinary data science training offerings external to McMaster.

2. Conduct a gap analysis to identify the proficiencies, gaps and potential areas to partner in interdisciplinary data science training.

3. Devise a coordinated approach for disseminating information about interdisciplinary data science training opportunities.

OVERVIEW

Context

In 2019, the *Lewis and Ruth Sherman Centre for Digital Scholarship* and *Spark: a centre for social research innovation* started to discuss how each could leverage its unique strengths in providing training and support in emerging approaches to data-driven research for the McMaster community. We identified our common interests and instructional practices and our shared commitment to working collaboratively and cohesively on developing innovative training resources for students, staff, and faculty, particularly those who might not have access to similar opportunities in their departments or programs. Because of the broader recognition and prioritization of knowledge mobilization as critical to the research enterprise, we envisioned a workshop series that could build from the centres' existing strengths in data fluency training with a co-organized series on best practices and strategies to communicate data-driven research to diverse audiences.

Increasingly, researchers in the Humanities and Social Sciences are conducting data-driven research, a broad term that is used to describe research that collects, manages, evaluates, and analyzes small and large (primarily quantitative) datasets as critical evidence to make informed decisions (Bonikowska et al., 2019; Henderson and Corry 2021; Kitchin 2014; Maass et al., 2018; Mandinach and Gummer 2013; Vanhoof et al., 2011; Wolff et al., 2016). Both the Sherman Centre and Spark recognized how equipping researchers with the theoretical and practical knowledge to create compelling research narratives would be indispensable to supporting and advancing research excellence and filling gaps in existing training. Interdisciplinary data science training can provide learners with the competencies to meaningfully share research with the public and effectively advocate for social change, thus making research relevant beyond scholarly communities.

As we document in this report, developing training resources to communicate data-driven research is integral to fostering data fluency. However, long-term training on data competencies beyond knowledge mobilization will support and promote robust data acumen, what the National Academies of Sciences, Engineering, and Medicine (2018) defines as the "ability to understand data, to make good judgments about and good decisions with data, and to use data analysis tools responsibly and effectively" (p.12). In later sections of the report, we discuss how this is of urgent necessity to meet the growing need for data skills in the workforce. In 2020, for example, the number of jobs in Canada posted across the information technology (IT) sector requiring data literacy skills—those related to collecting, processing, analysing, and effectively communicating information retrieved from data models (Deja et al., 2021)--had surpassed those seeking traditional skills such as marketing and public relations,

engineering, customer support, and administration (Ramachandran & Watson 2021). Given the demand for these competencies, we argue for a sustained commitment to equity, inclusivity, diversity and belonging in training to reduce the structural barriers to data fields.

The Sherman Centre and Spark collaboration also presents a unique and transformative opportunity to reappraise where and how data fluency training happens at higher education institutions (HEIs). Rather than siloed training in faculties, departments, or programs, we consider how research centres and support units – like the Sherman Centre and Spark – can support and advocate for the ongoing development of interdisciplinary training and support resources to fill institutional gaps. This is even more crucial because, for reasons such as budgetary and curriculum constraints, departments and programs have been slow to integrate the practical resources needed to train students and staff in interdisciplinary data-driven approaches. Extra-curricular training in data-driven approaches at many HEIs is dispersed among departments, institutes, centres, computing support and library units, and specialized intensive training opportunities can be prohibitive due to financial, physical, logistic, and temporal restraints. Additionally, organizations outside of the institution offer many high-quality resources and training opportunities; identifying and sharing information about these resources would further help researchers find the training they need while assisting local support units in prioritizing areas of development.

The collaboration between the Sherman Centre and Spark emphasized a need to build pathways between the research support units at McMaster University with expertise in data science. As discussed in the latter part of the report, working as a cohesive community on training, networking, and support resources can serve the campus, local, and regional communities. Yet it is not easy to appraise all the training opportunities within and beyond campus, thus making it challenging to assess currently available resources, gaps in training, and areas of overlap that could lead to fruitful collaborations.

To promote and strengthen partnerships across units, we recommend a three-fold approach that begins with an environmental scan of data science training offerings external to McMaster, followed by a gap analysis to identify the proficiencies, gaps, and potential areas for collaboration in interdisciplinary data science training at McMaster, and culminating in the development of a coordinated strategy for disseminating information about training opportunities. Collaboration between institutional units can lead to the growth of innovative training resources for learners from diverse disciplinary backgrounds with varying levels of preparedness and at different stages in their educational and professional trajectories. Further, a coordinated approach at the institutional level for sharing information about data science training opportunities can also foster communication and transparency and build pathways between the research support units on campus with areas of expertise in data science. The concerted effort to

partner on interdisciplinary data science training institutionally can lead to a resilient training, networking, and support infrastructure that reflects our campus community's heterogeneous needs, interests, and competencies and pave the way for inter-institutional data science training opportunities at regional and national levels.

Finally, mobilizing training for and co-creating resources with local and regional community members can address ongoing social inequities and exclusions by equipping individuals and groups with the competencies needed to enact agency and autonomy over the collection, analysis, and ownership of data (Kukutai & Taylo 2016; Quinless 2022; Walter & Andersen 2013). While our workshop series was originally envisioned as a series of in-person events, the global pandemic required that we reconsider our plans and move to online delivery. This, too, had the benefit of expanding participation beyond the local McMaster community. Similarly, building enduring industry partnerships can ensure that training and education at higher education institutions are well matched to the requirements of the rapidly evolving workforce (National Academies of Sciences, Engineering, and Medicine 2018).

GOALS AND OBJECTIVES

Addressing Training Gaps and Barriers

In Canada, the digital economy continues to surge despite the impacts of the pandemic. In August 2021, the Information and Communications Technology Council (ICTC) reported that jobs in Canada's technology sector and technology roles in all other sectors jumped more than 11 percent from 9.5 percent before the pandemic (Ivus & Kotak 2021). The ICTC predicts that by 2025, 2.26 million people will work in digitally skilled positions in high-demand roles, including data engineer, data analyst, and data scientist (Ivus & Kotak 2021, pp.49-51).

Because data has become a major driving force of global economies, the demand for data competencies to access, interpret, critically assess, manage, and ethically use data and transform it into information and actionable knowledge has rapidly accelerated (Koltay 2016b; UK Government 2021). Yet, despite the growth in data science training at higher education institutions (Havill 2019; Salloum et al., 2021; Schwartz 2019), there remains a shortage of data skills in the labour force (Fain 2017).

In the last few years, industry and academic reports and news articles regularly accentuate how demand for data skills across numerous sectors has outstripped supply (Lindzon 2022; Lyons et al., 2021). For example, in January 2022, the Royal Bank of Canada's CEO revealed that banks are struggling to hire enough data scientists, artificial intelligence specialists, and engineers due to competition from other sectors (Odeh 2022). This talent shortage and cross-sector competition is compounded in part

by the fact that recent graduates entering the labour force with data skills lack industry experience (UK Government 2021).

A May 2021 policy report prepared by the United Kingdom's Department for Digital, Culture, Media & Sport corroborates the increase in demand for data competencies across industries. The report, which assesses the supply of data skills in relation to the rapidly increasing demand, estimates that UK businesses in need of 'hard data skills' are recruiting for approximately 215,000 to 234,000 new roles. As the report details, over the last few years, about 48% of businesses in the UK recruited for positions that require 'hard data' skills, which include basic IT skills, information management, analysis skills, data ethics, programming, database management, data processing, data literacy, data visualization, advanced statistics, machine learning, data communication skills, and the knowledge of emerging technologies and solutions (UK Government 2021). However, almost half (46%) of businesses have struggled to recruit for roles that require these skills (UK Government 2021). The report examines two critical sources of data skills supply: higher education institutions (HEIs) and workplace upskilling. Despite growth in new talent in the form of recent graduates of HEIs entering the workforce, the report highlights how it is impractical, especially in the short term, for HEIs to fill the talent gap alone. Therefore, it is essential to seek opportunities for upskilling the current workforce to fill the gap in data skills (UK Government 2021).

The dearth of in-demand data literacy skills is aggravated by the lack of diversity and inclusion in data science domains (Cowgill et al., 2020; Paxton 2020). Initiatives like Toronto Womxn in Data Science seek to bridge gaps by supporting members of equity-seeking groups to network, learn, find mentors, and gain traction in data science (Toronto Womxn in Data Science 2022). Barriers to access might be partially mitigated by robust data fluency training and support infrastructures. At the same time, we recognize the need for a culture shift. Along with 'hard data skills,' interdisciplinary data science training can thoughtfully integrate critical considerations of data-driven approaches and outcomes. For example, this training might build awareness about the social and ethical consequences of algorithms trained with biased data sets or discuss how social and cultural assumptions and norms can influence data collection, analysis and reporting. While we agree that training in 'hard data skills' is needed to fill cross-sectoral talent shortages, we feel that it is equally significant to foster critical data acumen. In short, reducing barriers to data science training is partially solved by 'hard data skills' training. At the same time, we advocate for approaches to training that question some of the core principles, norms, and values of 'big data' paradigms that can unintentionally exacerbate existing inequities by continuing to centre normative world views and lived experiences.

The Relevant Research Series sought to fill current training gaps by leveraging interdisciplinary and industry expertise in communicating data-driven research by virtually hosting ten intensive workshops. The sessions effectively combined data

storytelling (Dykes 2015; Hullman & Diakopoulos 2011; Lee et al., 2015; Segel & Heer 2010; Tufte 1986), visualization (Correll 2019; D'Ignazio & Klein 2016; Healy 2018; Healy & Moody 2014; Mahmud et al., 2017), audience research (Athique 2018; Bail 2014), and knowledge mobilization planning.

We were motivated to collaborate on a workshop series focusing on communicating data-driven research for three reasons. First, because of the enthusiasm from research communities and industry to work with new digital sources of evidence and methods of analysis, and because competencies in approaches to communicating data-driven research to general audiences are still limited, we felt this focus would immediately benefit our communities (Noei et al., 2019). Second, because academic research is expected to serve more purposes and audiences, extending its reach to new areas and stakeholders, the knowledge of and practices for disseminating data-driven research are indispensable to advancing research excellence in support of evidence-informed decision making. Existing literature indicates that knowledge mobilization is generally viewed as an essential part of the work of postsecondary institutions (Côté & Darling 2018; Erwin 2015; Jordan & Weller 2018; Klar et al., 2020; Lavis 2006; Mandinach & Gummer 2013; Phipps et al., 2012; Sá et al., 2011). However, disciplinary research training seldom formally addresses the why and how of communicating research to broad, non-academic audiences. Finally, as a topic, knowledge mobilization for data-driven research can foster the critical data acumen necessary for consumers, producers, and users of digital data to refine their awareness of how perfunctory or biased design decisions can adversely impact how information is received and perceived and (mis)understood.

Series Objectives

At the time of writing the grant, the workshop series aimed to bring together specialized knowledge and approaches to data storytelling, data visualization, and audience research from various disciplines to expand a common skill-set and cross-disciplinary conversation among highly qualified personnel, including faculty, staff, and students.

Upon reflecting on the outcomes of the workshop series, it is evident that the series achieved these goals, as well as addressed a profound need for interdisciplinary data science training. This is an essential step toward more intensive and long-term collaboration among the research support centres at McMaster, other colleges and universities, and scholarly and research support organizations at the regional, national, and international levels to collectively fill gaps in interdisciplinary data science training in Canada.

DESCRIPTION OF THE SERIES

Core Areas and Competencies

The workshops focused on how data-driven research is communicated to general audiences, introducing these insights in a critical manner through best practices and strategies in support of evidence-based decision--making.

The series included practical hands-on introductions to the tools and approaches that are increasingly important for effectively communicating data-driven research narratives for public engagement (Carlson & Johnston 2015; Exner 2014; Koltay 2017, Wilson et al., 2010), data-driven storytelling (Dykes 2015; Tufte 1986), data visualization (Healy 2018; Healy and Moody 2014), and the role of visual primacy and understanding how audiences perceive information (Gaudet 2013; Joffe 2008).The emphasis on interdisciplinarity brought together heterogeneous knowledge, skills, and technologies that participants, unless they have actively studied all of these approaches and available tools, are not directly trained in.

Shift to Remote Delivery

The Relevant Research Series consisted of five workshops hosted virtually by the Sherman Centre and Spark during the 2020-21 academic year. When the series was proposed, each workshop was to be hosted at the Sherman Centre in a face-to-face format with an hour of introductory remarks followed by six hours of content. Within two weeks of each workshop, the Sherman Centre and Spark intended to host practice sessions to allow participants to apply their newfound knowledge.

These plans, however, were revised at an early stage due to the COVID-19 pandemic. Unable to host the workshops on campus due to health and safety concerns, the Relevant Research Series was redesigned before launch for remote delivery, making the project an early digital-by-design learning opportunity. The plan for remote delivery, coupled with the development of self-guided Open Educational Resources (OERs), became a vital component of the series.

For the most part, existing online training falls primarily into two categories: asynchronous or synchronous. Both have their advantages and disadvantages. Online asynchronous training allows participants to work through content at their own pace. Such activity is flexible and easily accessible with little more than an internet connection. But online asynchronous training can lack opportunities for engagement, especially when a facilitator does not support participants. This can hinder many participants who want assistance as they work through new material.

In contrast, synchronous training excels at supporting participants, as they are typically working closely with a facilitator and other participants. But sometimes, logistic

and temporal constraints can make such training more difficult to access. The Relevant Research Series effectively takes the best of both models.

THE WORKSHOPS

The Relevant Research Series brought together international experts and made data-driven knowledge mobilization training accessible to wider audiences by offering it entirely online with a blend of asynchronous and synchronous components. Each workshop offered a minimum of one synchronous session during which participants could engage with the facilitator, ask questions, and gain hands-on experience. This is especially fitting for a workshop series that is about knowledge mobilization and research dissemination.

The workshop series included a range of knowledge mobilization research topics, including:

- How to **develop a communications plan** (November)
- How cognition shapes **audience receipt and interpretation of information** (December)
- How to approach **data-driven storytelling** (January)
- How to engage with **data visualization** for research exploration and dissemination (March)
- How to apply a **critical design approach to community engagement as a means of disseminating knowledge** (April).

While each workshop exists as a standalone event, the lessons are cumulative. They deepen and expand the conversation beyond the technical skills required to produce powerful visualizations or a digital story to look closely at the complexities that challenge knowledge translation and exchange, including uptake of new information and the complexity of research stories for public engagement. The series was envisioned so that each workshop covered a gap in formal training in Humanities and Social Sciences approaches to data-driven knowledge mobilization and fit into the larger whole. Below, we include the workshop descriptions used to advertise the workshops to potential participants.

1. Communicating Research Beyond Academia

Date(s): 13 and 20 November, 2020
Facilitator: Mihaela Gruia, Director, Research Retold

Link to workshop

Workshop Description: Mihaela Gruia, Director of Research Retold, delivered a two-part workshop series on knowledge mobilization. The workshop led participants through considerations for how to share research findings with broader, non-academic audiences. Gruia explained how to create a communications plan to aid researchers in translating long, complex research reports into documents that spark conservation and compel change.

By the end of this workshop series, participants were be able to:

1. Explain the importance of knowledge mobilization
2. Apply knowledge mobilization concepts to individual contexts
3. Describe the key components of a communications plan
4. Develop a personalized communications plan

Facilitator Bio: Mihaela Gruia is the Founder and Director of **Research Retold**, a consulting firm that works with ten universities in the UK and many outside organizations to develop and implement research communications. She received a Master's degree in Data Science from University of Sheffield in 2017 and went on to establish Research Retold, which has been operating for three years and received significant recognition from University partners in the UK.

2. Transcending Mental Models

Dates: 4 and 10 December, 2020
Facilitator: Brian Southwell, Senior Director, Science in the Public Sphere, Research Triangle Institute

Link to workshop

Workshop Description: Brian Southwell, Senior Director of Science in the Public Sphere at the Research Triangle Institute, delivered a two-part workshop series on translating research for audiences beyond the university. In particular, Southwell focused on understanding how social and political opinions are formed and reformed. Audiences are not blank slates, but bring their own mental models – thoughts, ideas, opinions, experiences – to the conversation.

By the end of this workshop series, participants were able to:

1. Explain how people form social and political opinions, and the implications of these mental models on knowledge mobilization
2. Identify potential partnerships with media professionals
3. Identify and address misinformation or misperceptions about their research

Facilitator Bio: **Brian Southwell** is the Senior Director of the Science in the Public Sphere program at **Research Triangle Institute**, a U.S.-based nonprofit research institute that advances social progress. Southwell is a Duke/RTI scholar at Duke University Medical School, an Adjunct Professor at Duke's Social Science Research Institute, and an Adjunct Associate Professor in Media and Journalism and Health Behavior at the University of North Carolina. Southwell's award-winning research and theoretical contributions appear in more than 100 journal articles and chapters. In 2013, he published *Social Networks and Popular Understanding of Science and Health*. He has served as senior editor for Health Communication and as a member of numerous other editorial boards, including Communication Research and Public Opinion Quarterly. He also published *Innovations in Home Energy Use: A Sourcebook for Behavior Change* in 2016. Southwell's latest book, *Misinformation and Mass Audiences*, which he coedited, appeared in January 2018. He is regarded as one of the world's leading public opinion and messaging scholars to induce behaviour change.

3. Telling Data-Driven Stories: Lessons from Data Journalism

Dates: 11 and 28 January 2021
Facilitator: Roberto Rocha, Data Journalist, CBC/Radio Canada

Link to workshop

Workshop Description: Roberto Rocha, Data Journalist for CBC/Radio Canada, delivered a workshop series on the marriage of data analysis and data visualization with public-interest storytelling. Using examples from Canadian data journalism, this workshop demonstrated how principles from data journalism can help academics create compelling data-driven stories about their own research outcomes.

By the end of this workshop series, participants were able to:

1. Explain and implement the steps for creating a data-driven story (from obtaining data, to making sense of data, to writing an article outline)
2. Reflect on the importance of storytelling for sharing data with broader audiences

Facilitator Bio: Roberto Rocha is a digital journalist and educator who specializes in data-driven reporting and storytelling. Mr. Rocha has worked with CBC Montréal since 2015, and formerly the Montréal Gazette from 2005 to 2015. He is a former lecturer of Advanced Digital Journalism at the University of Ottawa. His recent data journalism work includes a number of stories covering the use of data and statistics in communicating and managing the COVID-19 pandemic.

4. Design Principles for Data Visualization

Dates: 5 and 12 March, 2021
Facilitator: Juan Velasco, 5W Infographics

Link to workshop

Workshop Description: Juan Velasco, Founder and Creative Director of 5W Infographics, delivered a workshop series on data visualization. In the first session, Velsaco explained the importance of design in data visualization. This included an introduction to visual perception principles through the creation of accessible and visually appealing data visualizations. In the second session, participants had the opportunity to practice applying their newfound design skills using visualization platforms like Tableau, Datawrapper, and Flourish. As part of the workshop series, Velasco reviewed data visualizations and provided feedback. Participants were provided with an overview of additional data visualization platforms and tools that do not require programming skills.

By the end of this workshop series, participants were able to:

1. Apply design principles to the creation of data visualizations
2. Use data visualization platforms (Tableau, Datawrapper, Flourish, etc.) to create visually appealing and accessible visualizations

Facilitator Bio: Juan Velasco is a visual journalist with a long career focused in visualizing complex research and scientific information to make it understandable and engaging for large audiences. He was the Art Director of National Geographic magazine from 2005 to 2014, and the Graphics Art Director of The New York Times from 1996 to 2001. In 2001 he founded his own company, 5W Infographics (based in Washington DC, New York City and Madrid) which provides graphics, data visualization, motion graphics, and information design training and consulting for organizations all over the world. Juan has won over fifty Society of News Design and Infographics Malofiej awards – including the Best of Show Award in 2013 – as well as awards from the Society of Publication Design and the American Institute of Graphic Arts. He was also a Pulitzer

Prize finalist as part of a team of writers and visual journalists working for the New York Times.

5. Research Driven Art and Building Feminist Data

Dates: 9 and 16 April, 2021
Facilitator: Caroline Sinders

Link to workshop: https://scds.github.io/building-feminist-data/

Workshop Description: Caroline Sinders, an artist and researcher, explores the intersections of critical design, trauma, data, and artificial intelligence as art. In this workshop series, Sinders introduced the methodology she created to guide both her art and research practice: research driven art. Inspired by photojournalism, critical design, and open source software, this methodology is both a process and an approach to knowledge mobilization. In the second session, Sinders explored data collection as both art and protest. Participants collected feminist data after a brief introduction to machine learning, data, and design thinking. By treating data collection as a collaborative process, participants created a feminist data set from the ground up, while also learning how data collection can be used as an artistic practice and a collaborative, community practice.

By the end of this workshop series, participants were able to:

1. Explain research driven art as a methodology and approach to knowledge mobilization
2. Engage in data collection as an artistic practice and as community practice
3. Create a dataset as part of a Feminist AI

Facilitator Bio: Caroline Sinders is a machine learning design researcher and artist. She is the founder of Convocation Design + Research, a design and research agency focusing on the intersections of machine learning, user research, designing for public good, and solving difficult communication problems. As a designer and researcher, she's worked with groups like Amnesty International, Intel, IBM Watson, the Wikimedia Foundation as well as others. She is also a research fellow with Harvard University's Kennedy School of Government and Policy.

AUDIENCE ENGAGEMENT

The Relevant Research Series brought together international experts and made knowledge mobilization training accessible to wider audiences by offering it entirely online with a blend of asynchronous and synchronous components.

Table 1: Event Registration for Synchronous Workshops

The event attendance averaged approximately 60% of registration numbers across all series and events. This conversion rate is greater than those reported generally for virtual events in 2021 (which ranged between 45% and 50%), and rates typically achieved for in-person events prior to COVID.

WORKSHOP NAME	REGISTERED	WAITLISTED
Part 1: Communicating Research Beyond Academia	**40**	**12**
Part 2: Communicating Research Beyond Academia	**40**	**12**
Part 1: Transcending Mental Models	**35**	
Part 2: Transcending Mental Models	**15**	
Part 1: Telling Data Driven Stories	**Released asynchronously**	
Part 2: Telling Data Driven Stories	**16**	
Part 1: Design Principles for Data Visualization	**107**	
Part 2: Design Principles for Data Visualization	**97**	
Part 1: Research Driven Art & Building Feminist Data	**89**	
Part 2: Research Driven Art & Building Feminist Data	**60**	**33**
TOTALS	**499**	**57**

Engagement Summary

Combined, the 10-event series attracted over 500 registrants, with many registrations external to McMaster University, while as of October 2022, the asynchronous module has received over 800 views by more than 200 unique users. While all but one of the events were held synchronously, session recordings and

materials were compiled to create 5 self-guided modules shared publicly for reuse by any audience. Since their release, this series has had over 1200 views by nearly 300 distinct users. While the topical focus was on Humanities and Social Sciences research, the skills and tools introduced are relevant across all fields, both at McMaster and other Canadian colleges and universities.

OBSERVATIONS AND KEY LEARNINGS

Collaboration

The Sherman Centre and Spark have formed a strong foundation for longer-term collaboration on interdisciplinary data science training. Our commitment to interdisciplinarity is a natural outgrowth of the interdisciplinary nature of our team, which includes faculty, librarians, information specialists, and research staff with a range of backgrounds, including Communication Studies, Information Science, Public Policy, and Political Science. Our diverse disciplinary backgrounds helped us identify core topics that would be most likely to appeal to a broad interdisciplinary audience. Our efforts to bring together participants from across fields, institutions, and industries were combined with an intense focus on engaged, applied learning and active discussion. Together, these workshops propelled discussions and engagement that have continued beyond the series, both among and between participants, the Sherman Centre, and Spark.

The centres will continue to offer adaptive training programs to foster interdisciplinary and intersectoral exchange. By providing workshops and creating self-guided learning resources, the centres will continue to support participants with varied disciplinary backgrounds and degrees of preparedness to acquire the requisite competencies to build data acumen. While Relevant Research focuses on mobilizing data-driven research to aid policy and critical practice, future workshops will concentrate on additional core competencies like data analysis to continue to address the talent gap in data skills.

In addition to bolstering collaborations between the Sherman Centre and Spark, we aim to strengthen ties with other research support units on campus to partner on future training initiatives and foster collaboration as an indispensable facet to all research training and support at McMaster. Along with building long-lasting campus partnerships through training, both centres engage with members of local and regional communities. Promoting training resources with groups off-campus is one way to support broader data competency upskilling and reskilling. Furthermore, sharing resources can also be an avenue for developing new partnerships. Rather than provide training to communities, we aim to co-develop resources that resonate with local and regional groups. We recognize how these partnerships take time to build. Since

Relevant Research, the Sherman Centre and Spark have independently started to initiate conversations with community-based and private organizations about leveraging each group's resources in mutually beneficial ways to build enduring relationships.

Innovative Training

In December 2020, the Ontario Ministry of Colleges and Universities (MCU) invested $50 million to drive growth and advancement in virtual learning across the province's postsecondary institutions. The purpose of this funding was to encourage learning through the accelerated use of both online and hybrid learning, recognizing that high-quality virtual learning is essential both in the current and post-pandemic contexts. The funding was in response to a dearth of digital-by-design learning opportunities, which makes the Relevant Research Series stand out as a unique experience for learners when many similar training opportunities were being cancelled due to pandemic restrictions. At the time this funding was announced, the Relevant Research Series had already hosted its first two events ("Communicating Research Beyond Academia" and "Transcending Mental Models.").

In this respect, Relevant Research was ahead of the curve. The workshops were designed for remote delivery, and OERs were developed from the recorded workshop content to support self-guided learning beyond the workshop series. Relevant Research, in turn, offers a model that could be adapted for future interdisciplinary data science training at McMaster and beyond.

Training Resources

The necessity to host the Relevant Research Series online made it possible to reach audiences who might otherwise not have been able to attend the workshops. No longer restrained by geography, both participants and workshop facilitators could attend from across the globe. The series has led to the creation of indispensable open educational resources (OERs). Whether a workshop was held synchronously or asynchronously, recordings were produced, and the Sherman Centre and Spark now have an impressive roster of OERs.

Combined, the series attracted hundreds of registrants, with many registrations external to McMaster University. The combination of skill development workshops stands out because while many other groups and training initiatives (see: Appendix A) teach data visualization, it is less common to do so in the context of a knowledge mobilization plan and audience research. Moreover, the focus was not so much on the how but on choosing what to visualize to have the impact sought, or on how to move beyond information-sharing to artistic and community practices that immerse people in the insights of the research.

The focus on data-driven mobilization and the digital-by-design approach to workshop development differentiates Relevant Research from other series. The fact that

the OERs can be integrated into the curriculum or used as self-guided learning is a strength. At the same time, we view the workshop series as complementary to and in conversation with other training initiatives.

Formalize Training

In higher education, micro-credentials are increasingly being explored around the world. eCampusOntario, for example, as of summer 2020, was supporting 22 pilot micro-credential initiatives that demonstrated support and positive alignment between academic and workplace partners. While evaluation of these projects is still ongoing, there is a clear desire for new ways to promote and formalize skill development.

Since 2019, the Sherman Centre has piloted a certificate program with uptakeprimarily by graduate students. Spark is also developing certificate offerings, with a focus on those working in the social sector. In other words, both centres have been considering formalizing the training they offer for some time. The organizers of Relevant Research will leverage the insights gleaned through the workshop series as they continue to engage in the work they have always done, which is to train, support and engage with learners from multiple disciplines with varying data competencies.

We anticipate ongoing collaborations between the two centres to work jointly with other support units on campus. These partnerships can lead to robust and innovative short-term, flexible interdisciplinary data science micro-credentials. Learners can boost new data skills for research and become job-ready by quickly upskilling and reskilling to enter the workforce or pivot in their careers.

Throughout the report, we have highlighted the demand for data skills in the workforce. Formalizing interdisciplinary data science training through micro-credentials cannot entirely solve the shortage of skills in the talent pool. However, rapid training of this kind can undoubtedly mitigate supply issues. More importantly, this type of training can serve recent graduates who wish to complement their degree training with data skills, and professionals who seek training to bolster new career opportunities.

Finally, we will continue to stress interdisciplinary data science because interdisciplinarity balances the development of rudimentary data skills with the ability to assess and think critically about data-driven methods and approaches, including the affordances and limitations of these approaches in representing the complexities of lived experiences. Training of this kind, which aims to build critical data acumen, will complement and enrich any field of study and job sector. And beyond employability, such training can equip learners with the critical data competencies to understand their roles as consumers and producers of data in daily life.

RECOMMENDATIONS

Based on the insights gleaned from the Relevant Research workshop series and observations collected from many years of data competency training offered at both the Sherman Centre and Spark, we put forward the following recommendations for the next steps for expanding interdisciplinary data science training at McMaster.

1. Complete an environmental scan of data science training offerings beyond McMaster

As a precursor to planning, developing, and coordinating the delivery of data science training opportunities at McMaster, we recommend that a comprehensive review of external offerings is conducted. Compiling information on competencies covered by these opportunities, their depth of coverage, availability, and cost would help identify areas for potential development at McMaster. Furthermore, such an inventory could be shared with the McMaster research community to improve discovery of and access to training. Much of this work has been completed in preparing this report (see Appendix A). However, we recommend that scans be conducted regularly, given that training opportunities and their details change frequently.

2. Conduct a gap analysis to identify the proficiencies, gaps and potential areas for collaboration in interdisciplinary data science training.

We wish to reiterate that both for-credit and extracurricular data science training opportunities exist on campus. However, it isn't easy to systematically assess the range of training available because this information is dispersed across departments, faculties, labs, centres, institutes, and support units. This approach has worked well in the short term because it has enabled each group to develop innovative programming to meet the needs and interests of the stakeholders they serve. Still, a unit-by-unit or group-by-group approach has made it difficult to scale and innovate training across campus in a coordinated manner.

For research support units whose mandates are to advance research excellence through training, consultations, and resource development, this dearth of institutional knowledge can present challenges to the work they do. The absence of information can create gaps and redundancies simultaneously. Indeed, there is sufficient demand for data skills training that even if overlaps exist, community members will attend and be served well. Nonetheless, having readily accessible information about which unit offers and specializes in what can lead to synergies between groups, resulting in the development of innovative interdisciplinary training.

In the short term, a gap analysis can clarify the current state of data science training and HQP competencies on campus and mitigate some of the challenges to the

training addressed above. This information would also benefit the research support units – like the Sherman Centre and Spark – that are enthusiastic about partnering with other campus groups to advance interdisciplinary data science. For the longer term, the gap analysis can help define ways of allocating training resources, identify the modalities of training that are institutionally appropriate based on factors such as institutional strengths, resources and demand, and the additional skills needed to close training gaps, all of which can support a desired future direction of interdisciplinary data science training and research at McMaster.

3. Devise a coordinated approach for disseminating information about data science training opportunities.

As increasingly more faculty, staff, students, and local and regional community members seek out data skills training, a scarcity of information will only fragment offerings to the detriment of the strategic development of interdisciplinary data science training and research at McMaster. Using the external environmental scan and gap analysis data, information about data science training (within and beyond McMaster) and campus experts' proficiencies could be corralled and shared publicly with researchers across disciplines and institutions.

A concerted approach to disseminating information about training opportunities on campus could strengthen training and research in multiple ways. First, readily accessible information about training opportunities would enable research support units to reach stakeholders on campus and beyond equally. Second, systematically sharing information about training opportunities could lead to greater transparency on campus by making information about the groups specializing in data science training visible to one another. In turn, groups already providing training could use this information to coordinate training efforts and partner on initiatives, thus mobilizing pathways between units. Third, this information could assist research support units with knowledge brokering. When a single team might lack the capacity or expertise to support a training or consultation request, they can direct researchers to other units, or researchers can access this information and go directly to a campus expert. In turn, communicating data science training services will identify campus experts' proficiencies, notably research support staff, which can be difficult to find otherwise, especially for researchers.

CONCLUSION

This report reflects on and documents observations from the 2020-2021 SSHRC-funded workshop series, The Relevant Research Series. The Sherman Centre and Spark learned immensely from the successes and challenges of co-organizing an

interdisciplinary data science training workshop series. As we detail in this report, the Sherman Centre and Spark will continue to advance interdisciplinary data science training, networking, and support for our campus community by offering training on additional core data competencies to address the talent gap in data skills. However, we are equally invested in mobilizing training for and co-creating resources with local and regional community members to prepare individuals and groups with the data acumen to enact autonomy over the collection, management, analysis, and ownership of data, just as we are keen to establish enduring industry partnerships that will ensure that some of our training is well matched to the requirements of the rapidly evolving workforce.

At the same time, as we have noted, any sustained effort to expand on existing training efforts will benefit from a systematic approach at the institutional level to identify areas for potential development at McMaster and a cooperative approach to disseminating information about training opportunities. This report aims to contribute to these ongoing efforts to build a durable training, networking and support infrastructure that responds to our campus community's wide-ranging needs, interests, and competencies and lays the foundation for inter-institutional interdisciplinary data science training opportunities at regional and national levels.

Finally, we take seriously McMaster's commitment to "teaching and modeling a sense of positive global citizenship, engaging in discussions about public policy, data and democracy, and fostering the skills of critical thinking and comprehension needed to evaluate information and seek truth and accuracy" (Office of the President 2021, p.14). To this end, the Sherman Centre and Spark are committed to working with other support units on campus, members of the campus, local and regional communities, along with our colleagues at other higher education institutions, to promote a sustained culture shift in data science domains. As we discussed in the report, this includes developing training that fosters the interdisciplinarity required to work with and ethically collect, use, and share data while prioritizing training and support resources that balance the development of rudimentary data skills with the ability to assess and think critically about the affordances and limitations of data-driven research methods in representing the complexities of lived experiences.

REFERENCES

Alexander, R., Lyons, K., Alexopoulos, M., & Austin, L. (2019, November 4-6). *Workshop on barriers to data science adoption: Why existing frameworks aren't working* [Conference presentation]. 29th Annual International Conference on Computer Science and Software Engineering (CASCON '19), Toronto, ON, Canada. https://dl.acm.org/doi/proceedings/10.5555/3370272

Athique, A. (2018). The dynamics and potentials of big data for audience research. *Media, Culture & Society, 40*(1),*(*59-74.

Bail, C. (2014). The cultural environment: Measuring culture with big data. *Theory and Society, 43*, 465-482.

Bonikowska, A., Sanmartin, C., & Frenette, M. (2019). *Data literacy: What it is and how to measure it in the public service*. Statistics Canada. https://www150.statcan.gc.ca/n1/pub/11-633-x/11-633-x2019003-eng.htm

Carlson, J., & Johnson, L.R. (2015). *Data information literacy: Libraries, data, and the education of a new generation of researchers.* Purdue University Press.

Christin, A. (2020). The ethnographer and the algorithm: Beyond the black box. *Theory and Society*, *49*, 897-918.

Correll, M. (2019, May 4-9). *Ethical dimensions of visualization research* [Conference presentation]. CHI 19: Proceedings of the 2019 Conference on Human Factors in Computing Systems, Glasgow, Scotland, UK. https://doi.org/10.1145/3290605.3300418

Côté, I.M., & Darling, E.S. (2018). Scientists on Twitter: Preaching to the choir or singing from the rooftops? *Facets, 3*(1). https://doi.org/10.1139/facets-2018-000

Cowgill, B., Dell'Acqua, F., Deng, S., Hsu, D., Verma, N., & Chaintreau, A. (2020, July 13-24). *Biased programmers? Or biased data? A field experiment in operationalizing AI ethics* [Conference presentation]. 21st ACM Conference on Economics and Computation, virtual. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3615404#

Deja, M., Januszko-Szakiel, A., Korycińska, P., & Deja, P. (2021). The impact of basic data literacy skills on work-related empowerment: The alumni perspective. *College & Research Libraries, 82*(5). https://doi.org/10.5860/crl.82.5.708

D'Ignazio, C., & Klein, L. F. (2016, October 24). *Feminist data visualization* [Workshop]. Visualization for the Digital Humanities (VIS4DH 1st Edition), Baltimore, Maryland, USA. https://dspace.ceid.org.tr/xmlui/bitstream/handle/1/955/Feminist_Data_Visualization.pdf

Dykes, B. (2015). Data storytelling: What it is and how it can be used to effectively communicate analysis results. *Applied Marketing Analytics, 1*(4), 299-313.

Erwin, Jr., R. W. (2015). Data literacy: Real-world learning through problem-solving with data sets. *American Secondary Education, 43*(2),18-26.

Exner, N. (2014). Research information literacy: Addressing original researchers' needs. *The Journal of Academic Librarianship, 40*(5). https://doi.org/10.1016/j.acalib.2014.06.006

Fain, P. (2017, March 30). Data science disconnect. *Inside Higher Ed*. https://www.insidehighered.com/news/2017/03/30/report-urges-data-science-course-work-all-undergraduates-close-growing-skills-gap

Gaudet, J. (2013). It takes two to tango: Knowledge mobilization and ignorance mobilization in science research and innovation. *Prometheus: Critical Studies in Innovation, 31*(3). https://doi.org/10.1080/08109028.2013.847604

Gillborn, D., Warmington, P., & Demack, S. (2018). Quantcrit: Rectifying quantitative methods through critical race theory. *Race Ethnicity and Education, 21*(2), 158–179.

Havill, J. (2019, February 22). *Embracing the liberal arts in an interdisciplinary data analytics program* [Conference presentation]. Proceedings of the 50th ACM Technical Symposium on Computer Science Education (SIGCSE '19), Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3287324.3287436

Healy, K. (2018). *Data visualization: A practical introduction.* Princeton University Press.

Healy, K., & Moody, J. (2014). Data visualization in sociology. *Annual Review of Sociology, 40*,105-128.

Henderson, J., & Corry, M. (2021). Data literacy training and use for educational professionals. *Journal of Research in Innovative Teaching & Learning, 14*(2). https://doi.org/10.1108/JRIT-11-2019-0074

Hullman, J., & Diakopoulos, N. (2011). Visualization rhetoric: Framing effects in narrative visualization. *IEEE Transactions on Visualization and Computer Graphics, 17*(12), 2231-2240.

Iliadis, A., & Russo, F. (2016). Critical data studies: An introduction. *Big Data & Society, 3*(2). https://doi.org/10.1177/2053951716674238

Ivus, M., & Kotak. A. (2021). *Onwards and upwards: Digital talent outlook 2025.* Information and Communications Technology Council, Government of Canada. https://www.ictc-ctic.ca/wp-content/uploads/2021/08/digital-talent-outlook-for-2025.pdf

Joffe, H. (2008). The power of visual material: Persuasion, emotion and identification. *Diogenes, 55*(1), 84-93. https://doi.org/10.1177%2F0392192107087919

Jordan, K., & Weller, M. (2018). Academics and social networking sites: Benefits, problems and tensions in professional engagement with online networking. *Journal of Interactive Media in Education, 1*(1), 1–9.

Kitchin, R. (2014). Big data, new epistemologies and paradigm shifts. *Big Data & Society, 1*(1), 1–12.

Klar, S., Krupnikov, Y., Ryan, J. B., Searles, K., & Shmargad, Y. (2020). Using social media to promote academic research: Identifying the benefits of Twitter for sharing academic work. *PLoS ONE, 15*(4). https://doi.org/10.1371/journal.pone.0229446

Koltay, T. (2017). Data literacy for researchers and data librarians. *Journal of Librarianship and Information Science, 49*(1), 3-14.

Koltay, T. (2016b). Data governance, data literacy and the management of data quality. *IFLA Journal*, *42*(4). https://doi.org/10.1177/0340035216672238

Kukutai, T., & Taylo, J. (Eds.). (2016). *Indigenous data sovereignty: Toward an agenda*. Australian National University.

Lavis, J. (2006). Research, public policymaking, and knowledge translation processes: Canadian efforts to build bridges. *Journal of Continuing Education in the Health Professions, 26*(1), 37-45.

Lee, B., Riche, N.H., Isenberg, P., & Carpendale, S. (2015). More than telling a story: Transforming data into visually shared stories. *IEEE Computer Graphics and Applications, 35*(5), 84-90.

Lindzon, J. (2022, March 4). Canada struggles to prepare its workforce for changing digital economy. *The Globe and Mail*. https://www.highereddive.com/news/6-universities-making-big-investments-in-data-science/550048/

Lyons, K., Alexopoulos, M., & Egwaikhide, A. (2021, November 22-25). *What will the transition to a digital economy look like? Exploring future skills, jobs, and policies needed post COVID* [Conference presentation]. Proceedings of the 31st Annual International Conference on Computer Science and Software Engineering (CASCON '21), Toronto, Ontario, Canada. https://dl.acm.org/doi/abs/10.5555/3507788.3507849

Lyons, K., Stroulia, E., Luo, D., Miller, R., & Onut, V. (2016, October 31- November 2). *Data-driven knowledge mobilization* [Conference presentation]. Proceedings of the 26th Annual International Conference on Computer Science and Software Engineering (CASCON '16), Toronto, Ontario, Canada. https://dl.acm.org/doi/abs/10.5555/3049877.3049908

Lyons, K., Stroulia, E., & Mindel, M. (2018, October 29-31). *Third annual workshop on data-driven knowledge mobilization* [Conference presentation]. Proceedings of the 28th Annual International Conference on Computer Science and Software Engineering (CASCON '18), Markham, Ontario, Canada. https://dl.acm.org/doi/abs/10.5555/3291291.3291353

Maass, W., Parsons, J., Purao, S., Storey, V. C., & Woo, C. (2018). Data-driven meets theory-driven research in the era of big data: Opportunities and challenges for information systems research. *Journal of the Association for Information Systems, 19*(12), 1253-1273.

MacGillivray, H. (2017, January 16). *Data literacy: What, why and how? Data and statistics: The sciences, the literacies and collaboration* [Conference presentation]. UN World Data Forum, Cape Town, South Africa.

Mahmud, A., Hogan, M., Zeffiro, A., & Hemphill, L. (2017). Teaching students how (NOT) to lie, manipulate, and mislead with information visualization. In S. A. Matei, N. Jullien & S. P. Goggins (Eds.), *Big Data factories: Collaborative approaches* (pp. 101-114). Springer.

Mandinach, E., & Gummer, E. (2013). A systemic view of implementing data literacy in educator preparation. *Educational Researcher, 42*(1), 30-37.

National Academies of Sciences, Engineering, and Medicine. (2018). *Data science for undergraduates: Opportunities and options*. The National Academies Press. https://doi.org/10.17226/25104

National Science Foundation. (2004). *What is interdisciplinary research?* https://www.nsf.gov/od/oia/additional_resources/interdisciplinary_research/definition.jsp

Neff, G., Tanweer, A., Fiore-Gartland, B., & Osburn, L. (2017). Critique and contribute: A practice-based framework for improving critical data studies and data science. *Big data*, *5*(2). https://doi.org/10.1089/big.2016.0050

Nikunen, K. (2021). Ghosts of white methods? The challenges of Big Data research in exploring racism in digital context. *Big Data & Society, 8*(2). https://doi.org/10.1177/20539517211048964

Noei, E., Lyons, K., Stroulia, E., & Andritsos, P. (2019, November 4-6). *Fourth annual workshop on data-driven knowledge mobilization* [Conference presentation]. Proceedings of the 29th Annual International Conference on Computer Science and Software Engineering (CASCON '19), Toronto, Ontario, Canada. https://dl.acm.org/doi/abs/10.5555/3370272.3370325

Odeh, L. (2022, January 18). Shortage of ESG skills hurting Canada's finance sector, study says. *The Financial Post*. https://financialpost.com/fp-finance/a-shortage-of-esg-skills-is-hurting-canadas-financial-firms

Office of the President, McMaster University. (2021). *Institutional priorities and strategic framework 2021-2024*. https://president.mcmaster.ca/app/uploads/2021/05/Institutional-Priorities-and-Strategic-Framework_FINAL_5May21.pdf

Paxton, L. (2020, August 27). Crunching the numbers on diversity in data science: Events & resources to foster inclusion. *STEM and Culture Chronicle*. https://medium.com/stem-and-culture-chronicle/crunching-the-numbers-on-diversity-in-data-science-events-resources-to-foster-inclusion-5dc81d2ab52

Phipps, D., Jensen, K., Johnny, M., & Myers, G. (2012). Field note: Research summaries improve access and utility of academic research. *Scholarly and Research Communication*, *4*(1), 1-17.

Phipps, D., & Shapson, S. (2009). Knowledge mobilization builds local research collaborations for social innovation. *Evidence & Policy, 5*(3), 211-227.

Quinless, J. M. (2022). *Decolonizing data: Unsettling conversations about social research problems*. University of Toronto Press.

Ramachandran, K., & Watson, J. (2021, March 10). Tech looks to analytics skills to bolster its workforce. *Deloitte Insights*. https://www2.deloitte.com/us/en/insights/industry/technology/data-analytics-skills-shortage.html

Sá, C., Li, S., & Faubert, B. (2011). Faculties of education and institutional strategies for knowledge mobilization: An exploratory study. *Higher Education, 61*(5), 501-512.

Salloum, M., Jeske, D., Ma, W., Papalexakis, V., Shelton, C., Tsotras, V., & Zhou, S. (2021, March 13-20). *Developing an interdisciplinary data science program* [Conference presentation]. Proceedings of the 52nd ACM Technical Symposium on Computer Science Education (SIGCSE '21), Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3408877.3432454

Schwartz, N. (2019, March 8). 6 universities making big investments in data science. *Higher Ed Dive*. https://www.highereddive.com/news/6-universities-making-big-investments-in-data-science/550048/

Seaver, N. (2017). Algorithms as culture: Some tactics for the ethnography of algorithmic systems. *Big Data & Society, 4*(2). https://doi.org/10.1177/2053951717738104

Segel, E., & Heer, J. (2010). Narrative visualization: Telling stories with data. *IEEE Transactions on Visualization and Computer Graphics, 16*(6), 1139-1148.

Sinders, C. (2020). *Feminist data set*. The Clinic for Open Source Arts, University of Denver. https://carolinesinders.com/wp-content/uploads/2020/05/Feminist-Data-Set-Final-Draft-2020-0517.pdf

Sperry, J. (2018). *Data literacy: Exploring economic data*. 2018 Economic Programs Webinar Series, U.S. Census Bureau.

Statistics Canada. (2020). *Data literacy competencies.* Government of Canada. https://www.statcan.gc.ca/en/wtc/data-literacy/compentencies

Toronto Womxn in Data Science. (2022). *Homepage*. https://www.womxndsto.com/

Tufte, E. (1986). *The visual display of quantitative information.* Graphics Press.

United Kingdom Government, Department for Digital, Culture, Media & Sport. (2021, May 18). *Quantifying the UK data skills gap*. https://www.gov.uk/government/publications/quantifying-the-uk-data-skills-gap/quantifying-the-uk-data-skills-gap-full-report#fn:13

Vanhoof, J., Verhaeghe, G., Verhaeghe, J. P., Valcke, M., & Van Petegem, P. (2011). The influence of competences and support on school performance feedback use. *Educational Studies, 37*(2). https://doi.org/10.1080/03055698.2010.482771

Walter, M., & Andersen, C. (2013). *Indigenous statistics: A quantitative research methodology*. Routledge.

Wilson, P. M., Petticrew, M., Clanan, M. W., & Nazareth, I. (2010). Disseminating research findings: What should researchers do? A systematic scoping review of conceptual frameworks. *Implementation Science, 5*(91). https://doi.org/10.1186/1748-5908-5-91

Wolff, A., Gooch, D., Cavero Montaner, J. J., Rashid, U., & Kortuem G. (2016). Creating an understanding of data literacy for a data-driven society. *The Journal of Community Informatics, 12*(3). https://doi.org/10.15353/joci.v12i3.3275

APPENDIX A: EXISTING KNOWLEDGE MOBILIZATION AND DATA LITERACY TRAINING

Title	Institution	Organizers/Partners	Country	Discipline	Content	Link
4T Virtual Conferences on Data Literacy	University of Michigan	School of Information, School of Education, Inter-University Consortium For POlitical and Social Research, Institute of Museum and Library Services	United States	Interdisciplinary	Archives webinars	http://datalit.sites.uofmhosting.net/conference/
A Culture of Data Literacy	Qlik	Qlik	Global	Interdisciplinary	Course (Online)	https://learning.qlik.com/course/view.php?id=723
ACRL Roadshows, Research Data Management	American Library Association	Association of College & Research Libraries	United States	Interdisciplinary	Workshops	http://www.ala.org/acrl/conferences/roadshows/rdmroadshow
CCGHR Workshop for Knowledge Translation	Queen's University	Canadian Coalition for Global Health Research	Canada	Health Science	Workshop	https://www.ccghr.ca/wp-content/uploads/2014/04/Two-Day-Skills-Workshop.pdf

Certificate in Knowledge Mobilization	University of Guelph	Open Learning and Educational Support, and the College of Social and Applied Human Sciences.	Canada	Interdisciplinary	Course(s)	https://courses.opened.uoguelph.ca/public/category/courseCategoryCertificateProfile.do?method=load&certificateId=453705
Data Analytics and Visualization in Health Care	edX (founded by Harvard and MIT)	Rochester Institute of Technology	United States	Health Science	Course (Online)	https://www.edx.org/course/data-analytics-and-visualization-in-health-care
Data Literacy Primer	National Library of Medicine	National Network of Libraries of Medicine (NN/LM)	United States	Science	Course materials, guides, reference material, research	https://nnlm.gov/data/guides/data-literacy
Data Literacy Project	The Data Literacy Project	Qlik, Cognizant, Experian & Pluralsight	Global	Interdisciplinary	Online assessments, e-learning and educational resources, and community of practice	https://thedataliteracyproject.org/
Data Literacy Training	Government of	Statistics Canada	Canada	Interdisciplinary	Online workshops	https://ww

	Canada					w.statcan.gc.ca/eng/wtc/data-literacy/catalogue
Data Management and Visualization	Wesleyan University	Wesleyan University	United States	Interdisciplinary	Course (Online)	https://www.coursera.org/learn/data-visualization
Data Management Short Course for Scientists	ESIP Federation	ESIP Federation, NOAA, and the Data Conservancy	United States	Science	Course(s) (Online)	https://commons.esipfed.org/datamanagementshortcourse
DataCamp	DataCamp	1600+ companies, 3000 academic organizations	Global	Interdisciplinary	Course(s) (Online)	https://www.datacamp.com/
DH@Guelph	University of Guelph	College of Arts	Canada	Humanities	Workshops, lectures, presentations	https://www.uoguelph.ca/arts/dhguelph
Digital Humanities Summer Institute (DHSI)	University of Victoria	30+ institutions and academic organizations	Canada	Humanities	Workshops, lectures, presentations	https://dhsi.org/about-dhsi/
Enhancing Data and Science Literacy (Data Literacy Workshop Series)	Rutgers University	Rutgers Dataspire	United States	Science	Workshops, training, consultations, resources	https://datascienceliteracy.rutgers.edu/opportunities/cor

						e-programs/data-literacy-series/
Fundamentals of Research Data Management	University of Washington	University of Washington Libraries	United States	Interdisciplinary	Online Module(s)	https://canvas.uw.edu/courses/889213
Healthcare Data Literacy	University of California at Davis	UC Davis Continuing and Professional Education	United States	Health Science	Course (Online)	https://www.coursera.org/learn/healthcare-data-literacy
HELDIG Digital Humanities Forum	University of Helsinki	Helsinki Centre for Digital Scholarship	Finland	Interdisciplinary (Humanities and Social Sciences)	Workshops, lectures, presentations	https://www2.helsinki.fi/en/helsinki-centre-for-digital-humanities/heldig-digital-humanities-forum
Knowledge Translation Workshop	University of Waterloo	Concept by Velocity	Canada	Interdisciplinary	Workshop	https://uwaterloo.ca/games-institute/blog/post/we-teamed-concept-velocity-

						deliver-knowledge-translation
MANTRA	University of Edinburgh	Information Services, Institute for Academic Development	Ireland	Interdisciplinary	Course (Online)	https://mantra.edina.ac.uk/about.html
Research Data Management and Sharing	University of North Carolina at Chapel Hill & University of Edinburgh	Curating Research Assets and Data Using Lifecycle Education (CRADLE) Project in collaboration with EDINA at the University of Edinburgh	United States; Ireland	Interdisciplinary	Course	https://www.coursera.org/learn/data-management
Research Data Management for Clinical Research	Vanderbilt University	Vanderbilt University	United States	Science	Course (Online)	https://www.coursera.org/learn/clinical-data-management
Research Impact Academy	Research Impact Academy	Research Impact Academic (formerly Knowledge Translation Australia)	Australia	Interdisciplinary	Online and face-to-face workshops, consultations	https://researchimpactacademy.com/about/
Specialist Knowledge Translation Training	The Hospital for Sick Children (SickKids)	The Hospital for Sick Children (SickKids) - Knowledge Translation Program	Canada	Health Science	Workshops	https://www.sickkids.ca/en/learning/continuing-professional-

						development/knowledge-translation-training/
The KT Canada National Seminar Series	Knowledge Translation Canada	Knowledge Translation Canada	Canada	Health Science	Seminars	https://ktcanada.org/education-2/
The KT Canada Summer Institute	Knowledge Translation Canada	Knowledge Translation Canada	Canada	Health Science	Lectures	https://ktcanada.org/education-2/
The STIHR Training Program	Knowledge Translation Canada	Knowledge Translation Canada	Canada	Health Science	Workshops, training	https://ktcanada.org/education-2/
What is Data? What is Data Literacy? (Tutorial)	Eastern Michigan University	Eastern Michigan University Library	United States	Interdisciplinary	Video tutorial	https://www.emich.edu/library/help/tutorials/assets/data_intro/story_html5.html
Project Tier	Project TIER	Haverford College	United States	Social Sciences	Workshops	https://www.projecttier.org/tier-classroom/course-materials/?