Strengthening Public Libraries' Information Literacy Services Through an Understanding of Knowledge Brokers' Assessment of Technical and Scientific Information

A. Statement of Broad Need

Access to information alone is no longer sufficient to support civic decision-making. Although a glut of information is available to anyone with Internet access, a web browser, and basic reading skills, access does not guarantee understanding (Perry, 2020). Civic decision-making requires not just access to but also sensemaking of diverse, high quality information (Dervin, 1994; Webster, 1999). Misinformation and conspiracy theories have increased in prevalence in recent years. This poses challenges to ALA Core Values of Librarianship (2019) such as Access, Democracy, The Public Good, Social Responsibility, and Sustainability. Misinformation has been linked to consumption of social media, as well as to deliberate media manipulation, including disinformation campaigns, strategic amplification and framing, and hyperpartisan news sources (Garrett, 2017; Marwick & Lewis, 2017; Wardle & Derakhshan, 2017). Yet belief in misinformation has also been associated with traditional media. For example, misinformation on COVID-19 and disinformation on climate change have been disseminated by some mainstream news sources (Hall Jamieson & Albarracín, 2020; McCormick, 2021; Pew Research Center, 2020). Researchers have also raised concerns about how the media might influence public understanding of the risk posed to the future of human labor by artificial intelligence (Scheufele & Krause, 2019). Although misinformation impacts multiple domains, including scientific and technical information, it has been studied largely in the political sphere, and library discussions of misinformation have primarily focused on political misinformation (e.g., Cooke, 2018; Lenker, 2016; Sullivan, 2019).

Complex scientific and technical information is highly relevant to the average person. It affects policy, legislation, and choices people make in their day-to-day lives. Scientific misinformation and pseudoscience have a significant impact on public deliberation. For instance, Hartley & Minh Khuong (2020), highlight protesters' challenges to "the scientific community's largely consensus views about the need for social distancing to limit the spread of COVID-19." Consumers are at risk of misunderstanding scientific information, or of being swayed by misrepresentations. For instance, factual information is woven into conspiracy theories, making them more sticky (Introne et al., 2018). The ability to unweave and track the sources, and particularly the conflicts of interest of creators, is important for consumers; empowering users to apply critical thinking to their own concerns is a key part of critical information literacy (Hall, 2010). In fact, consumers' health information seeking has received significant attention (e.g., Philbin et al., 2019); in the past, the Public Library Association partnered with the National Network of Libraries of Medicine to create a toolkit for libraries to support consumer health information seeking (the Healthy Community Tools for Public Libraries, https://publiclibrary.health)(2021)).

Far less attention has been paid to developing tools for librarians brokering scientific and technical information to the average person. For libraries the opportunity is two-fold: (1) Develop better services for non-library knowledge brokers who mediate public access to information, which amplifies librarians' impact, because one patron who is a knowledge broker can reach many other information consumers; and (2) Identify strategies for managing and provisioning scientific and technical information, which can support interactions with any patron. This is particularly important because it is an area not well-covered in preprofessional training or in most librarians' backgrounds.

This project builds on librarians' role as knowledge brokers, i.e.,"people whose job it is to move knowledge around and create connections between researchers and their various audiences" (Meyer, 2010). In addition to librarians, other knowledge brokers include journalists, Wikipedia editors, and activists. Henceforth, we refer to librarians as a distinct group and use the term "knowledge brokers" when speaking about other types of

knowledge brokers. Understanding the needs, behaviors, and sensemaking strategies of knowledge brokers can be valuable in designing services. Library-based services to knowledge brokers have the potential to change the amount of misinformation circulating, especially misinformation that draws on scientific and technical information. Controversies within science are often amplified and distorted as they move from expert communities into communities of practitioners and citizens. For example, by presenting a corporate agenda as "an expression of popular will" by a "seemingly independent organization" with hidden ties to an industry group (a practice known as "astroturfing") (Apollonio & Bero, 2007), corporate-sponsored advocacy has been used to avoid regulation and to gain drug approvals (Apollonio & Bero, 2007; Moynihan & Bero, 2017).

The information seeking behaviors of knowledge brokers have an outsized impact on civic decision-making. Vocal segments of the populous have a significant influence on how expert knowledge is applied to public policy. However, the average person relies on knowledge brokers to translate expert knowledge for use in decision-making. When public perception of scientific knowledge conflicts with expert views, some policymakers will give greater weight to public opinion. Even when policymakers follow the experts, they can struggle to obtain public support if people's understanding of the science is skewed. In the past, even renowned journalists have been deceived by manufactured controversy, as was the case with past findings regarding tobacco smoke (Oreskes & Conway, 2011, p. 19).

My long-term research agenda analyzes persistent controversies applying science to public policy; how knowledge brokers influence citizens; and whether controversies are sustained by citizens' disparate interpretations of scientific evidence and its quality. This is a natural expansion of my current work, which has focused on the information quality of scientific information and on controversies within science. For example, my work has described how scientists synthesize scientific evidence for use by practitioners and by other researchers (Grizzle et al., 2019; Hoang & Schneider, 2018; Jackson & Schneider, 2018; Schneider & Jackson, 2018) and analyzed persistent disagreements within science (Hsiao et al., 2020). I have suggested that digital libraries take a knowledge maintenance perspective, to go beyond providing access to individual scientific papers (Fu & Schneider, 2020). Much of my work has focused on misinformation within science, particularly the continued citation and use of retracted research (Schneider et al., 2020).

Currently the Project Director (PD) is engaged in a stakeholder consultation, funded by the Alfred P. Sloan Foundation, that is focused on developing practical recommendations to stop inadvertent citation of retracted papers (Schneider & The RISRS Team, 2021). My team gathered 70 stakeholders from across scholarly publishing for a series of one-on-one interviews and issue-focused online workshops, leading to draft recommendations (Schneider et al., 2021). Stakeholder partners and I have already disseminated the draft recommendations at NISOPlus 2021 to encourage follow-up actions by the standards community (*NISO Plus 2021*, 2021). Moving research into practical implementations requires partnership and co-development with stakeholders, which is a core approach of my research.

Public libraries are uniquely positioned to reach both the knowledge brokers through whom information is filtered as well as the general public. This project will help public libraries leverage that role to improve the quality of technical and scientific information available to consumers and to support well-informed civic decision making.

B. Project Design
The ultimate goal of this project is to provide libraries with actionable information to help them support citizens in finding trustworthy sources for societally relevant technical and scientific knowledge. To do this, we will seek to better understand the information literacy of knowledge brokers as well as how knowledge brokers influence individuals in assessing the trustworthiness of information sources, and to translate our findings into a practical toolkit of services and guidelines for public libraries.

We start with three driving questions:
1.   Where, how, and for what purpose do knowledge brokers access technical and scientific information?
2.   What conceptual model describes their sensemaking and use of this information?
3.   How do knowledge brokers assess the quality and utility of technical and scientific data and documents? Specifically, which features impact knowledge brokers' quality and utility evaluations of technical and scientific data and documents?

In the final 18 months of the project, we will focus on disseminating sample service information to public librarians, iteratively improving materials about sample services, and designing a pilot test for these services, to be submitted as a IMLS proposal for a follow-on project. Our dissemination work with public libraries will use an iterative co-design methodology, analogous to design of the PD's current Sloan-funded consultation. Through our partnership with the Public Library Association (see LoS) we will recruit 5 co-design partners. Co-design partners, who will help design and test sample services, will be public services staff such as librarians, non-MSLIS library staff at small libraries, and managers with responsibility for overseeing public services. We will use the initial conceptual model that we develop from the first 18 months of the project to identify the most salient aspects of diversity for the 5 co-design partners (currently envisioned as library size and geographic location).

Methods
This project will utilize three case studies, one per year, to address the three questions above, with each case study centering on public communication of scientific information. While the topics will be finalized based on recent events, currently the case studies are planned as:

Year 1: Public health and public policy decision-making around the COVID-19 pandemic.

Year 2: Artificial intelligence and the future of human labor.

Year 3: Climate change and its mitigation.

For each case study, the PD and Graduate Research Assistants (RAs) will:

(1) collect and analyze a purposive sample of about 250 public-facing documents and multimedia, including news (e.g., online print outlets), Wikipedia pages, membership-based online forums, documentaries, and data visualizations, that report, quote, or analyze research papers and/or other scientific products; and

(2) conduct and analyze 40 semi-structured interviews with document authors (e.g., journalists, Wikipedia editors, activists/advocates), librarians, and other relevant informants to be identified based on input from the Advisory Board, as well as preliminary findings under IRB-approval using the critical incident technique (Urquhart et al., 2003) to focus on a recent instance of brokering scientific information to the public. RAs will code themes (Braun & Clarke, 2006) from interview transcript data using the Atlas.ti qualitative data analysis software.

The resulting model that we develop could describe dimensions of information activities, aspects of the process, and situations in which information is used (Thivant, 2005; Widén-Wulff et al., 2008; Zhang & Soergel, 2016). For instance, knowledge brokers' approaches could be different for different activities such as finding experts or documents on a topic versus verifying their credibility. Knowledge brokers might use more robust validation procedures when directly quoting or using information as compared to background information that does not directly appear in a report. Variation between brokers is part of the model. Diversity in the 40 knowledge brokers recruited will help us determine the extent to which knowledge brokers' approaches tend to differ based on personal attributes, values, or perceptions, for instance related to their lived experience, their perception of their audience, or their perception of the medium. Some variable attributes identified in previous studies of science and health journalists include "scientific training, professional experience (tenure), and audience served (local, regional, or national)" (Gesualdo et al., 2020). Research on science journalists has also uncovered various roles they play, as conduit, public intellectual, agenda-setter, watchdog, civic educator, curator, convener, and advocate (Fahy & Nisbet, 2011). Diversity in the three domains from which case studies are chosen (health, technology, environment) will help us determine the extent to which differences in these domains may influence how knowledge brokers approach these areas. This variation, if found, will be part of our mode: to be useful to librarians serving knowledge brokers, the ultimate findings need to hold across different types of knowledge brokers and across different types of topics.

Document Analysis
The project will use an iterative process to identify items for the document analysis:

(1) <u>Identify science that has been highly discussed in the public.</u> We will use altmetrics data sources (likely Paperbuzz, a free open tool from OurResearch, but possibly either altmetric.com or PlumX, both of which offer free access for scientometrics researchers by application).
(2) <u>Consider items that have had wide circulation.</u> This will be based on the Reuters Institute Digital News Report (Newman, 2020), Wikipedia pageview statistics (*Topviews Analysis*, n.d.), published lists of popular online forums, data visualizations, and widely distributed documentaries.

The overall framework for document analysis will be Musi & Aakhus (2019)'s approach for large scale analysis of argumentative polylogues (defined as "multiple different argumentative parties defending their distinct positions" (Aakhus & Lewiński, 2017)). Identification of topics, players, and places will be aided by use of topic modeling, which has been found to correlate with more time-intensive grounded theory investigations (Baumer et al., 2017). This will allow us to select a sample for more intensive manual analysis of positions and argumentative moves. For the manual analysis we will oversample documents presenting positions that are contrary to established scientific consensus, in order to more closely analyze their logic and rhetoric.

Interviews
The project will interview 120 informants, 40 each year for three years. For each case study we will select 10 each of the following:

- Librarians from socio-economically diverse rural, urban and suburban areas
- Journalists from a diverse array of publications with a broad range of editorial policies
- Wikipedia editors of different ages and genders
- Activists and policy advocates representing a broad range of views and concerns.

The PD will recruit interviewees through professional societies, organizations, and published documents, with particular attention to diversity (see below). This will enable us to consider whether a person's lived experience affects their approach to assessing the reliability of information sources.

Interviews will be semi-structured, remote via phone or video. The protocol will ask interviewees to describe the last time they wrote about/helped a patron with scientific or technical information and will include questions such as:

- What was the topic?
- What kind of information did you need?
- Where did you find relevant information?
- How did you interpret the information you found?
- What was challenging in this process?
- Was the information trustworthy? How did you know?

The PD will seek approval from her institution's Office for the Protection of Research Subjects during the pre-award period.

Toolkit Co-Dedevelopment
In the final 18 months of the project, we will disseminate project findings by co-designing services in partnership with 5 public libraries. First the team will develop (1) draft materials to support public libraries in providing services to support citizens' functional literacies in science and health information online; and (2) services for knowledge brokers to support their dissemination of high quality information. Using feedback from our 5 library partners, we will then test and refine these draft materials. Our testing process will focus on assessing the effectiveness of the services and determining potentially relevant factors for scaling up. This will inform the design of a national-scale project, which the PD will submit as a IMLS proposal "Strengthening public libraries' information literacy services of technical and scientific information - Scale up," with the goal of assessing the replicability and sustainability of the proposed services in libraries of different sizes and located in rural, urban, and suburban communities with different demographics.

Project Staff
The PD will be supported by two Graduate Research Assistants (RAs) who will conduct the interviews and document analysis in parallel, thus enabling us to complete three case studies in three years. This will also enable us to approach prospective interviewees based on documents they authored (e.g., to identify journalists whose public work contradicts scientific consensus) as well as to incorporate interviewees' suggestions about relevant documents as we go along. Doing so will add to the diversity of the interviewees and ensure that our choice of document case studies via heuristics does not miss important items.

In addition, diverse methods require different skill sets and it is the rare individual who possesses the full range of skills required by this project: qualitative synthesis, data science, and document analysis. Computational work can run into unexpected bumps, which will require that an assistant devote significant time to troubleshooting and developing pipelines, particularly at the beginning of the project. The pragmatic outlook of an MSLIS student, in grounding ideas to current practice, can be particularly valuable for this project since it seeks to derive actionable insights for libraries. Meanwhile, the project provides practical hands-on work related to information literacy, which is relevant to students interested in careers in public librarianship, science librarianship, medical librarianship, consumer health information, government librarianship, communications librarianship or even metadata librarianship and digital humanities librarianship.

Master's students have been RAs for several of my past projects, drawing on the strong data science opportunities in the Illinois MSLIS program as well as students' interest in the information quality topics in my lab. Past Master's RAs have gained a deeper understanding of research and are becoming future library leaders who develop and evaluate innovative services, and whose work keeps me connected to practice.

Advisory Board The project includes an Advisory Board of 5 people from public libraries, Wikipedia, and schools of library & information science, to ensure relevance for these groups. The Advisory Board consists of Jessamyn West, library consultant working at the intersection of technology and civic engagement, who is a long-time Wikipedia editor and Dr. Carole Palmer, Professor and Associate Dean for Research, Information School, University of Washington, who is an expert in data curation and digital research collections (see letters of support, attached), and three members to be named, who will be recruited before the beginning of the project. Two will be public library practitioners or thought leaders with expertise in reference services in small and medium public libraries and/or STEM programming who have broad personal networks connecting them to public libraries serving racially and socioeconomically diverse groups in different areas of the country. The third will have experience in providing or receiving reference services for journalists or activists/advocates. At least one of the Advisory Board members will be a person of color and at least one will self-identify as ideologically conservative. Advisory Board members will contribute to the further development of the research methods and help ensure the project's relevance to practice. They will advise the PD on recruiting a diverse and representative group of informants to interview, and on any considerations for types of documents and groups that they know well (e.g. Wikipedia editors). Access to certain groups, such as Wikipedia editors, will be based in part on the PD's status as a community insider, as well as on introductions and assistance from that group's representative on the Advisory Board. Regular input from the public librarian practitioners on how to apply the research results in practice will be particularly valuable.

Outcomes

In the final 18 months of the project, we will focus on disseminating sample service information to public libraries and knowledge brokers, iteratively improving materials about sample services, and designing a national-scale pilot test for these services, to be submitted as a IMLS proposal for a follow-on project. In partnership with 5 public libraries, we will use the project findings to develop, test, and refine a toolkit for libraries, particularly public libraries, providing services to both consumers and to other knowledge brokers about science and technical information. This will inform the design of a national-scale project, which the PD will submit as a IMLS proposal with the goal of assessing the replicability and sustainability of the proposed services in libraries of different sizes and located in rural, urban, and suburban communities with different demographics.

To reach a broad audience of practitioners, the PD and RAs will present preliminary results at conferences for practitioners (currently planned as Public Library Association, American Library Association, and Association for Rural and Small Libraries); in the PD's blog, which is aggregated on Planet Code4Lib; and in an annual public webinar coordinated with organizations such as the Public Library Association, the Science and Technology Section of the Association of College and Research Libraries, the National Network of Libraries of Medicine, and the Wikimedia Foundation. The PD and RAs will present in research venues (currently planned as ASIS&T and iConference) and publish in scholarly journals (e.g., JASIST, Journal of Documentation). An open access version of publications, presentations, and models will be made available on the PD's public website and via UIUC's digital repository, the Illinois Digital Environment for Access to Learning and Scholarship (IDEALS) and data will be housed in the Illinois Data Bank. The project's approach has applicability to the design of repositories, databases, and digital libraries, and will inform future research on public access, information literacy, and understanding of science.

Ideas from the project will be incorporated into a required course for MSLIS students, Information Organization & Access, as well as into advanced courses aimed at MSLIS students.

<u>Evaluation and Assessment Plan</u>
1.   *Effectiveness* will be assessed by successful collection of data, by whether the project yields a workable model, and by the dissemination of a public library toolkit.
2.   *Efficiency* of the budget and staffing will be assessed every 6 months by the PD and discussed with the PD's Associate Dean for Research and pre-tenure review group.
3.   *Quality* will be assessed through the project's success in presenting conference papers and publishing journal articles, and by surveys of the public library toolkit users.
4.   *Timeliness* will be assessed monthly by the PD, and discussed every 6 months with the Advisory Board.

C. Diversity Plan
Diversity has been considered in terms of public impact, diversifying student populations, and recruiting diverse Advisory Board members and informants (i.e., librarians and knowledge brokers to be interviewed). This research directly impacts diversity, equity, and inclusion by improving access to quality information for all. The PD has a demonstrated personal commitment to increase research participation of women, students from low-income families, and other underrepresented students across intersections of gender, class, race, and disability, drawing on her background as a first-generation college student. This project will also involve 1 PhD and 1 MSLIS student in the PD's Information Quality Laboratory, who will be mentored in qualitative methods, document analysis, use of technology in information sciences research, and the research field of scholarly communication. To increase the diversity of the pool of potential RAs, the PD will advertise the iSchool PhD program to Historically Black Colleges and Universities, Hispanic-serving institutions, ALA Spectrum Scholarship recipients, and via the Coalition for Diversity and Inclusion in Scholarly Communications.

The Advisory Board is being recruited with diversity in mind, with the aim of recruiting at least one member who is Black, Indigenous or a Person of Color, as well as advisors with experience in public libraries, journalism, Wikipedia, activism/advocacy, and library & information science research.

Our recruitment of librarian and other knowledge broker informants will also emphasize diversity. The PD will seek to include black, Indigenous and people of color as well as people across the political spectrum (for instance by using sources such as AllSides.com to identify political leaning). To reach a diverse group of journalists, the PD will use organizations found in the Journalism Diversity, Equity, & Inclusion Tracker as a resource (*Journalism DEI Tracker*, n.d.). For instance, the PD will reach out to membership organizations such as the Native American Journalists Association, National Association of Hispanic Journalists, National Association of Black Journalists, and the Association of Gay and Lesbian Journalists; diversity commitees of professional societies such as the National Association of Science Writers, the News Leaders Association, the Association of Health Care Journalists, and the Society of Environmental Journalists; and capacity building programs such as the Knight Diversity Initiative, AAAS Diverse Voices in Science Journalism Internship, and Arizona State University's bilingual journalism program. To identify a diverse group of Wikipedia editors, the PD will use the list of previous Wikipedians in residence ("Category," 2019; *Wikipedian in Residence*, n.d.), mailing lists (e.g. LGBT affinity group Gendergap, which focuses on gender equity); and WikiProjects on Medicine, Computing, and Climate Change. To identify a diverse group of activists and advocates, the PD will use the Guidestar Directory of Charities and Nonprofit Organizations, which lists relevant groups such as the Indigenous Environmental Network, LatinX in AI, Black in AI; as well as landowners and miners' rights groups and the rural poor in predominantly white regions such as Appalachia. The PD will recruit from across the ideological spectrum, reaching think tanks such as the Heritage Foundation, the Freedom Foundation, and the Sutherland Institute as well as the Institute for Policy Studies and the Urban Institute. The PD will particularly

recruit groups and people whose public position contradicts scientific consensus (e.g., groups on the Global Warming Skeptic Organizations list from The Union of Concerned Scientists; organizations posting false COVID-19 information per the CoronaVirusFacts Alliance database; using the document analysis to identify journalists whose articles contradict scientific consensus). In recruiting librarians, the PD will pay particular attention to librarians reaching underserved populations, as well as to geographic, socio-economic diversity, and diversity of the library type.

D. Broad Impact

Identifying strategies for managing and provisioning scientific and technical information that can support interactions with any patron is particularly important because it is an area not well-covered in preprofessional training or in most librarians' backgrounds. The project will benefit public libraries by illuminating their role as knowledge brokers, both to other knowledge brokers such as journalists, Wikipedia editors, activists and advocates, and to the public at large. This project will also give libraries the tools to increase the information literacy of both groups (knowledge brokers and the public) around scientific and technical information. The project will directly train 2 students in research and impact 100 Master's students per year who take the PDs classes. This project will advance the research agenda of the PD, an early career scholar, leading to new questions in information literacy and information quality research related to science and technical information.

By utilizing the toolkit developed in this project, public libraries will be able to help consumers of information to become more discerning about the sources they deem reliable as well as more skilled at processing information for themselves. In the case of knowledge brokers like journalists, Wikipedia editors, and activists/advocates, we expect this to lead to a decrease in the dissemination of misinformation, whether deliberate or accidental. In the case of the general public, it would mean less susceptibility to conspiracy theories and pseudoscience and also a concurrent decrease in the dissemination of misinformation traveling from individual to individual (e.g., over social media or by word of mouth). Ultimately this will enable people to make fully-informed choices and raise the level of public discourse around civic decision making.

SCHEDULE OF COMPLETION

ACTIVITIES	PREAWARD			YEAR ONE												YEAR TWO												YEAR THREE											
	MY	JN	JL	AG	SE	OC	NV	DE	JA	FE	MR	AP	MY	JN	JL	AG	SE	OC	NV	DE	JA	FE	MR	AP	MY	JN	JL	AG	SE	OC	NV	DE	JA	FE	MR	AP	MY	JN	JL
MANAGEMENT																																							
IRB approval (preproposal)																																							
Recruit and onboard PhD student RA & hourly MSLIS student																																							
Meet with advisory board, planning & quality improvement																																							
Recruit partner libraries via Public Library Assocation																																							
CASE STUDIES 1-3 - INTERVIEW WORK																																							
Recruit interviewees																																							
Conduct interviews																																							
Recruit additional interviewees found via document analysis																																							
Qualitative coding & analysis of transcripts																																							
CASE STUDIES 1-3 - DOCUMENT ANALYSIS																																							
Develop & refine computational workflow																																							
Develop initial document sample																																							
Topic modeling of initial document sample																																							
Refine document sample based on interview feedback																																							
Qualitative coding to refine topic modeling																																							
DEVELOP & REFINE CONCEPTUAL MODEL																																							
Develop & refine conceptual model																																							
TOOLKIT CO-DEVELOPMENT WITH 5 PARTNER LIBRARIES																																							
Develop & refine citizen science literacy materials																																							
Develop & refine knowledge broker services																																							
Plan scale-up project to assess replicability and sustainability																																							
DISSEMINATION & TRAINING																																							
Present at practitioner & research conferences																																							
Write & publish journal articles																																							
Incorporate methods & findings into required MSLIS course																																							
Present annual project webinar																																							
Disseminate sample service toolkit beyond partner libraries																																							

DIGITAL PRODUCT FORM

INTRODUCTION

The Institute of Museum and Library Services (IMLS) is committed to expanding public access to digital products that are created using federal funds. This includes (1) digitized and born-digital content, resources, or assets; (2) software; and (3) research data (see below for more specific examples). Excluded are preliminary analyses, drafts of papers, plans for future research, peer-review assessments, and communications with colleagues.

The digital products you create with IMLS funding require effective stewardship to protect and enhance their value, and they should be freely and readily available for use and reuse by libraries, archives, museums, and the public. Because technology is dynamic and because we do not want to inhibit innovation, we do not want to prescribe set standards and practices that could become quickly outdated. Instead, we ask that you answer questions that address specific aspects of creating and managing digital products. Like all components of your IMLS application, your answers will be used by IMLS staff and by expert peer reviewers to evaluate your application, and they will be important in determining whether your project will be funded.

INSTRUCTIONS

If you propose to create digital products in the course of your IMLS-funded project, you must first provide answers to the questions in **SECTION I: INTELLECTUAL PROPERTY RIGHTS AND PERMISSIONS.** Then consider which of the following types of digital products you will create in your project, and complete each section of the form that is applicable.

> **SECTION II: DIGITAL CONTENT, RESOURCES, OR ASSETS**
> Complete this section if your project will create digital content, resources, or assets. These include both digitized and born-digital products created by individuals, project teams, or through community gatherings during your project. Examples include, but are not limited to, still images, audio files, moving images, microfilm, object inventories, object catalogs, artworks, books, posters, curricula, field books, maps, notebooks, scientific labels, metadata schema, charts, tables, drawings, workflows, and teacher toolkits. Your project may involve making these materials available through public or access-controlled websites, kiosks, or live or recorded programs.
>
> **SECTION III: SOFTWARE**
> Complete this section if your project will create software, including any source code, algorithms, applications, and digital tools plus the accompanying documentation created by you during your project.
>
> **SECTION IV: RESEARCH DATA**
> Complete this section if your project will create research data, including recorded factual information and supporting documentation, commonly accepted as relevant to validating research findings and to supporting scholarly publications.

SECTION I: INTELLECTUAL PROPERTY RIGHTS AND PERMISSIONS

A.1 We expect applicants seeking federal funds for developing or creating digital products to release these files under open-source licenses to maximize access and promote reuse. What will be the intellectual property status of the digital products (i.e., digital content, resources, or assets; software; research data) you intend to create? What ownership rights will your organization assert over the files you intend to create, and what conditions will you impose on their access and use? Who will hold the copyright(s)? Explain and justify your licensing selections. Identify and explain the license under which you will release the files (e.g., a non-restrictive license such as BSD, GNU, MIT, Creative Commons licenses; RightsStatements.org statements). Explain and justify any prohibitive terms or conditions of use or access, and detail how you will notify potential users about relevant terms and conditions.

The University of Illinois will hold rights and ownership to the resulting intellectual property but will not use them for commercial gain. Digital products will be released under a Creative Commons BY license, to facilitate reuse. An open access version of publications, presentations, data, and models will be made available on the PD's public website and via UIUC's digital repository, the Illinois Digital Environment for Access to Learning and Scholarship (IDEALS) (http://ideals.illinois.edu) and the Illinois Data Bank (https://databank.illinois.edu). In addition to the use of 'green' open access through the IDEALS repository for all publications, key research publications will additionally use 'gold' open access to facilitate the ease with which the public and worldwide research community can find and reuse them at scholarly publishers' websites.

A.2 What ownership rights will your organization assert over the new digital products and what conditions will you impose on access and use? Explain and justify any terms of access and conditions of use and detail how you will notify potential users about relevant terms or conditions.

Project resources will be provided openly and freely in the IDEALS repository and the Illinois Data Bank for anyone to use without any access restrictions. The authors of publications, presentations, data, and models will hold copyright (CC BY license).

A.3 If you will create any products that may involve privacy concerns, require obtaining permissions or rights, or raise any cultural sensitivities, describe the issues and how you plan to address them.

Research interview data we generate will be governed by Institutional Research Board approval, gathered under a formal consent process, and stored under IRB-approved UIUC Box digital storage. Research interviews will be confidential materials used internally for model building, and will not be released beyond the project team.

SECTION II: DIGITAL CONTENT, RESOURCES, OR ASSETS

A.1 Describe the digital content, resources, or assets you will create or collect, the quantities of each type, and the format(s) you will use.

Research data
1. Interview data
    a. MP3 and MP4 – Recordings of interviews with 120 participants (40 per year for 3 years)
    b. DOCX and TXT – Transcripts of interviews with 120 participants (40 per year for 3 years)
    c. atlproj – Qualitative coding of the interview transcripts (3 ATLAS.ti projects, 1 per year)

       d.   Excel, PDF, Word – Selected exports from ATLAS.ti qualitative coding, e.g. codebook

       e.   DOCX, PDF, Box.com Note – Project minutes, working files, draft conceptual model

2. <u>Document analysis data</u>
   a. HTML, DOCX, MP4, TXT, PDF – A corpus of 750 documents and multimedia (250 per year for 3 years) will include news (e.g., online print outlets), Wikipedia pages, membership-based online forums, documentaries, and data visualizations.
   b. TXT – text derivatives of the corpus as input to topic modeling
   c. TXT, R, Python – analysis scripts and script output (e.g., topic modeling)
   d. TXT formatted as Markdown (MD) – documentation for scripts
   e. TXT, DOCX, PDF – manual argumentation analysis
   f. JSON, PNG, HTML, XML using Argument Interchange Format schema – computer visualization of argumentation analysis in Online Visualisation of Argument software (http://ova.arg-tech.org/)
   g. XML, svg – diagrams of workflows created in draw.io (http://draw.io)
   h. DOCX, PDF, Box.com Note – Project minutes, working files, draft conceptual model

3. <u>Iterative design of services and guidelines for public libraries</u> Design methods will be used to draft and refine materials to support public libraries in providing services to support citizens' functional literacies in science and health information online; and services for knowledge brokers to support their dissemination of high quality information. Data will include draft materials and communications with the 5 co-design partners from public libraries who will help design and test sample services.

Public facing products

1. Publications and presentation materials
   a. PDF – journal articles, papers in conference proceedings
   b. PDF, PPTX – conference presentations slides, conference posters
   c. MP3, MP4 – recordings of conference talks

2. Webinars produced by the project team
   a. MP3, MP4 – recordings of annual and end-of-project webinars (4 recordings)
   b. PDF, PPTX – presentations slides and handouts

3. Toolkit
   a. DOCX, PDF, HTML– descriptions of potential services for public libraries to adopt; success stories of implementations in co-design libraries; lists of suggested resources

A.2 List the equipment, software, and supplies that you will use to create the digital content, resources, or assets, or the name of the service provider that will perform the work.

Project staff will use existing server space and software (Microsoft Office, web editing suite, Zoom, etc.), ATLAS.ti qualitative software, open source scripting tools (Python, R, Rstudio), open source topic modeling packages, open source drawing tools (e.g. draw.io), freely available argumentation analysis support tools (e.g. Online Visualisation of Argument software) Laptops will be purchased for Graduate Research Assistants to ensure sufficient computing power and hard drive space for project resources, and to enable purchase of university licenses for ATLAS.ti. Digital recorders will be purchased as a backup for Zoom for collection of research data interviews.

A.3 List all the digital file formats (e.g., XML, TIFF, MPEG, OBJ, DOC, PDF) you plan to use. If digitizing content, describe the quality standards (e.g., resolution, sampling rate, pixel dimensions) you will use for the files you will create.

Main digital file formats will be DOCX, HTML, Markdown (MD), Box.com note, PDF, TXT. Research materials will also include Atlas.ti project files, Python and R scripts, and exports from drawing software such as draw.io and Online Visualisation of Argument software (JSON, PNG, HTML, XML, svg), and a multimedia corpus (e.g. HTML, DOCX, MP4, TXT, PDF and perhaps other formats). Public webinars will be recorded in MP4 using Zoom default settings (which includes the export of an MP3 audio track). Interviews will be recorded in MP4 using Zoom default settings, with a secondary backup recording from a handheld digital audio recording device.

Workflow and Asset Maintenance/Preservation

B.1 Describe your quality control plan. How will you monitor and evaluate your workflow and products?

The Project Director will meet weekly with the Research Assistants to develop and check project workflows. At least once a semester, the Project Director will formally audit the workflow and the products.

B.2 Describe your plan for preserving and maintaining digital assets during and after the award period. Your plan should address storage systems, shared repositories, technical documentation, migration planning, and commitment of organizational funding for these purposes. Please note: You may charge the federal award before closeout for the costs of publication or sharing of research results if the costs are not incurred during the period of performance of the federal award (see 2 C.F.R. § 200.461).

During the award period, project staff will use existing server space and software (Microsoft Office, web editing suite, Zoom, etc.) for this project. UIUC provided Box.com will be used for IRB-restricted materials, corpora, and project notes. The Project Director's InfoQualityLab GitHub.com space will be used for research scripts.

After the award period, material will be retained on existing web server space, Box.com, GitHub.com. IRB-restricted materials will be retained for a time-limited period.

Metadata

C.1 Describe how you will produce any and all technical, descriptive, administrative, or preservation metadata or linked data. Specify which standards or data models you will use for the metadata structure (e.g., RDF, BIBFRAME, Dublin Core, Encoded Archival Description, PBCore, PREMIS) and metadata content (e.g., thesauri).

Metadata will be produced in the IDEALS and Illinois Data Bank repository interfaces, and will rely on their underlying metadata structures, which are both based on Dublin Core. The Illinois Data Bank complies with DataCite standards.

C.2 Explain your strategy for preserving and maintaining metadata created or collected during and after the award period of performance.

Both IDEALS and Illinois Data Bank will support metadata beyond the award period of performance, for long-term preservation and access.

C.3 Explain what metadata sharing and/or other strategies you will use to facilitate widespread discovery and use of the digital content, resources, or assets created during your project (e.g., an API [Application Programming Interface], contributions to a digital platform, or other ways you might enable batch queries and retrieval of metadata).

Discoverability in IDEALS is facilitated by the Open Archives Initiative Protocol for Metadata Harvesting (http://www.ideals.illinois.edu/dspace-oai/request?verb=Identify) and the use of DublinCore metadata (see https://wiki.illinois.edu/wiki/display/IDEALS/Metadata+Policy). Datasets in the Illinois Data Bank are each assigned a Digital Object Identifier (DOI) and are discoverable through the DataCite (https://www.datacite.org/) global registry of resources.

Access and Use

D.1 Describe how you will make the digital content, resources, or assets available to the public. Include details such as the delivery strategy (e.g., openly available online, available to specified audiences) and underlying hardware/software platforms and infrastructure (e.g., specific digital repository software or leased services, accessibility via standard web browsers, requirements for special software tools in order to use the content, delivery enabled by IIIF specifications).

Materials will be openly available online. WebJunction (http://webjunction.org) will be a dissemination partner to raise the profile of public facing products, particularly the toolkit aimed at public library practitioners.

D.2. Provide the name(s) and URL(s) (Universal Resource Locator), DOI (Digital Object Identifier), or other persistent identifier for any examples of previous digital content, resources, or assets your organization has created.

RISRS project webpage: https://infoqualitylab.org/projects/risrs2020/
IDEALS collection for the RISRS project: https://www.ideals.illinois.edu/handle/2142/108359
Project Director's publications webpage: http://jodischneider.com/jodi.html
Datasets in Illinois Data Bank:
https://databank.illinois.edu/datasets?q=Schneider%2C+jodi&per_page=25

SECTION III: SOFTWARE

General Information

A.1 Describe the software you intend to create, including a summary of the major functions it will perform and the intended primary audience(s) it will serve.

N/A

A.2 List other existing software that wholly or partially performs the same or similar functions, and explain how the software you intend to create is different, and justify why those differences are significant and necessary.

N/A

Technical Information

B.1 List the programming languages, platforms, frameworks, software, or other applications you will use to create your software and explain why you chose them.

N/A

B.2 Describe how the software you intend to create will extend or interoperate with relevant existing software.

N/A

B.3 Describe any underlying additional software or system dependencies necessary to run the software you intend to create.

N/A

B.4 Describe the processes you will use for development, documentation, and for maintaining and updating documentation for users of the software.

N/A

B.5 Provide the name(s), URL(s), and/or code repository locations for examples of any previous software your organization has created.

N/A

Access and Use

C.1 Describe how you will make the software and source code available to the public and/or its intended users.

N/A

C.2 Identify where you will deposit the source code for the software you intend to develop:

Name of publicly accessible source code repository: N/A

URL: N/A

SECTION IV: RESEARCH DATA

As part of the federal government's commitment to increase access to federally funded research data, Section IV represents the Data Management Plan (DMP) for research proposals and should reflect data management, dissemination, and preservation best practices in the applicant's area of research appropriate to the data that the project will generate.

A.1 Identify the type(s) of data you plan to collect or generate, and the purpose or intended use(s) to which you expect them to be put. Describe the method(s) you will use, the proposed scope and scale, and the approximate dates or intervals at which you will collect or generate data.

1. <u>Interview data</u> Interviews with 120 participants (40 per year for 3 years) will generate recordings and transcripts. Qualitative coding of the interview transcripts will be used for the purpose of generating and refining a conceptual model.
2. <u>Document analysis data</u> A corpus of 750 documents and multimedia (250 per year for 3 years) will include news (e.g., online print outlets), Wikipedia pages, membership-based online forums, documentaries, and data visualizations. The corpus will be selected using free altmetrics data sources and from referrals by participants. Analysis approaches will include topic modeling and manual argumentation analysis, which will also contribute to generating and refining a conceptual model.
3. <u>Iterative design of services and guidelines for public libraries</u> Design methods will be used to draft and refine materials to support public libraries in providing services to support citizens' functional literacies in science and health information online; and services for knowledge brokers to support their dissemination of high quality information. Data will include draft materials and communications with the 5 co-design partners from public libraries who will help design and test sample services.

All data collected will contribute to scholarly publications and presentations, dissemination to practitioners, and toolkit development.

A.2 Does the proposed data collection or research activity require approval by any internal review panel or institutional review board (IRB)? If so, has the proposed research activity been approved? If not, what is your plan for securing approval?

IRB-approval is required. The Project Director will seek this approval from her institution's Office for the Protection of Research Subjects during the pre-award period.

A.3 Will you collect any sensitive information? This may include personally identifiable information (PII), confidential information (e.g., trade secrets), or proprietary information. If so, detail the specific steps you will take to protect the information while you prepare it for public release (e.g., anonymizing individual identifiers, data aggregation). If the data will not be released publicly, explain why the data cannot be shared due to the protection of privacy, confidentiality, security, intellectual property, and other rights or requirements.

1. <u>Interview data</u> Full interview transcripts will not be shared in order to protect the privacy of our interviewees, and to promote frank discussions with interviewees.
2. <u>Document analysis data</u> We cannot release the text of copyrighted documents. We will deposit a list of URLs or other identifiers for documents and multimedia in the Illinois Data Bank. We will also test the use of archiving tools to reduce the impact of linkrot (for instance, the Internet Archive's Save Page Now, for sites allowing crawlers).
3. <u>Iterative design of services and guidelines for public libraries</u> Feedback provided in emails and one-on-one conversations with individuals from public libraries will not be shared publicly, but will be used to inform the design of subsequent versions.

A.4 What technical (hardware and/or software) requirements or dependencies would be necessary for understanding retrieving, displaying, processing, or otherwise reusing the data?

Aside from commonly used office software (e.g. Microsoft Office, Zoom), the only proprietary software we anticipate using is ATLAS.ti.

Reusing the corpus will require crawling a list of URLs due to copyrighted materials, and this will be subject to the availability of those URLs in the future. Multimedia included in the corpus may assume the use of common web technologies, which may not necessarily have long-term availability far beyond the award period.

A.5 What documentation (e.g., consent agreements, data documentation, codebooks, metadata, and analytical and procedural information) will you capture or create along with the data? Where will the documentation be stored and in what format(s)? How will you permanently associate and manage the documentation with the data it describes to enable future reuse?

In-process materials will be stored in Illinois Box for the use of the team. We will deposit any shareable data sets and data products for long term preservation and dissemination, in a suitable repository for each type of documentation:
- Interview procedures (e.g. recruiting email templates and consent agreement templates without identifying information; semi-structured interview scripts) will be shared in IDEALS.
- Codebooks derived from the analysis of interview transcripts will be stored in the Illinois Data Bank, with minimal excerpts from transcripts, and no personally identifying information.
- Document analysis scripts will be shared in the InfoQuality Lab GitHub site (https://github.com/infoqualitylab). We will generate versioned DOIs via Zenodo (https://zenodo.org), a public repository maintained by OpenAIRE and CERN, which is commonly used to facilitate long-term software citation.
- Document analysis completed on Online Visualisation of Argument (OVA) software http://ova.arg-tech.org/ will be stored in Argument Interchange Format (AIF) in the community searchable database AIFDB: http://www.aifdb.org/search. JSON copies of the analysis, which can be loaded into OVA for viewing and editing, will be shared via the Illinois Data Bank.
- Document analysis research protocols and workflows will be shared via the Illinois Data Bank.
- Iterative design protocols will be shared via the Illinois Data Bank.

A.6 What is your plan for managing, disseminating, and preserving data after the completion of the award-funded project?

1. Interview data: Interviewee recordings and transcripts are governed by the IRB. They are not intended for dissemination or for long-term preservation beyond the award period.
2. Document analysis: Copyrighted data will be stored as a corpus in Illinois Box storage beyond the award period. A list of identifiers for items in the corpus will be stored in the Illinois Data Bank for long-term preservation beyond the award period.
3. Iterative design of services and guidelines for public libraries: Protocols will be stored in the Illinois Data Bank for long-term preservation beyond the award period. The final toolkit will be widely disseminated as a formal digital product.

A.7 Identify where you will deposit the data:

Name of repository: Illinois Data Bank

URL: http://databank.illinois.edu

A.8 When and how frequently will you review this data management plan? How will the implementation be monitored?

The Project Director will review this data management plan at the start of the project and annually during the award period.