Toward Data Quality Assurance Infrastructure for Research Data Repositories

1. Project Justification

1.1 The IMLS NLG goal and objective addressed by the project

Florida State University (FSU), the lead applicant, in partnership with Texas A&M University (TAMU), is requesting $92,922 of IMLS funds to complete a collaborative 18 month applied exploratory research project. The project will contribute towards the IMLS NLG program Goal 3 and Objective 3.2: "Support innovative approaches to digital collection management." This exploratory research project will identify and inventory the data quality assurance practices (DQA) of research data repositories. In particular, the study will identify the types of data quality problems and incidents; models, standards, and strategies used; the division of labor and the roles played; challenges and barriers to evaluating and maintaining data quality, and skills and competencies needed. In addition, the project will use the study's findings to develop DQA design scenarios. DQA scenarios will describe specific types of data quality assurance problems or incidents and associated DQA actions. Finally, we will develop a model and a metadata vocabulary of the DQA work of research data curators and repository managers. We will encode them as an RDF/OWL ontology. The ontology will represent relationships among the types of data use activities, data quality problem types, and DQA actions in a structured, machine-processable format. We will share the study's findings and products, including scenarios and ontology, with library and academic communities. They can inform data curators' DQA work, help them design and manage their DQA workflows, identify data quality problems and address those problems.

1.2 The need addressed by the project

The problem addressed by this exploratory research project is the lack of empirical studies of DQA practices of research data repository managers and curators and the lack of generalized use cases and design recommendations for the DQA work of that occupation that are grounded in the data quality literature. Quality is defined as "fitness for use" [21]. A DQA process comprises data quality conceptualization, measurement, and intervention activities [37]. *Data quality, along with privacy and access, are critical ethical aspects of data use. In the era of big data and a flood of research data and publications, the old dictum "garbage in garbage out" is as relevant as ever. Quality of data determines the quality of research findings, teaching, business decisions, policies, and ultimately, it affects human lives* [26,34]. Universities are presently making significant investments in developing trustworthy and secure infrastructure to curate digital research

datasets produced and/or used by their faculty and students. These efforts are motivated by faculty's need to preserve and share their research data [28,39]; state and federal funding agencies requiring their grantees to share data with open access to benefit taxpayers, to enable its reuse in research and teaching, and to enhance research reproducibility and replicability [27,28]; national and state laws that require ensuring the quality of data and preserving individuals' privacy [4,41], and universities' desire to enhance their visibility and reputation by providing open access to data produced by their faculty and students as public goods to benefit their states and society in general [40]. Furthermore, universities are interested in tracking and measuring the impact of those datasets, including for faculty promotion and tenure evaluation [25]. ***One of the main inhibitors of data sharing and reuse, however, is concern about the quality of data.*** Data owners can be concerned about the quality and/or documentation of their data and its potential misuse or misinterpretation by others [37]. The users, on the other hand, need useful, valid, and trustworthy data that represents the phenomena they are interested in, not just big data [7,29]. They may not have access to and/or knowledge of the process that generated and/or manipulated the data to evaluate its quality. Hence, they may mistrust and not use the data. Furthermore, data is often incomplete and contains biases and inaccuracies. Data creators usually collect or assemble datasets for specific purposes or uses. If data is not properly documented, understanding those purposes is often a challenge and a barrier to data reuse [38]. Finally, data and DQA are not free. There is always a cost to DQA that someone has to shoulder. ***Digital data repositories need to find cost-effective, efficient ways to evaluate, maintain, and communicate the quality of the data they share to facilitate its ethical reuse***.

1.3 The target groups and beneficiaries of the project

The ultimate beneficiaries of this applied exploratory research project are research data repository managers and research data curators at the institutions of higher education in the USA and worldwide. The study's findings, DQA design scenarios, success stories, and the DQA ontology will inform data curators' DQA practices, including the design of their services, data stewardship metadata, and policies. Other important indirect beneficiary groups of this project are the users of data curated by research data repositories that include faculty, students, entrepreneurs, policymakers, and the public in general. The project will inform the design and construction of digital research data curation infrastructure components on university campuses that aim to provide access not just to big data but reusable, trustworthy data - data that could be used with confidence in research, teaching, policymaking, and developing services for consumers and society in general. The study's outcomes will also inform the data curation and data science training and education curricula of LIS schools and communities of practice.

1.4 How this proposed project differs from, complement, or build upon existing theory, scholarship, and practice

The following subsections discuss how this exploratory research project will complement or build upon existing theory, scholarship, and practice.

1.4.1 Lack of empirical studies that examine and interpret DQA practices in research data repositories through a lens of the data quality literature

This applied *exploratory* research project will have significant implications to both the theory and practice of data curation and DQA. First, it will address the problem of the lack of empirical studies that survey ***DQA practices and activities in research data repositories and analyze and interpret those findings through a lens of the data quality literature.*** DQA activities may range from quality evaluation and improvement actions performed by data providers and repository staff to data cleaning performed by students as part of their class assignment or DQA hackathons and research reproducibility challenges[1]; evaluating the quality of datasets for training AI models, and/or making policy and business decisions[2] [15,33]. There have been several general quality assurance standards and approaches used in the industry (e.g., Six Sigma, ISO 9000, ISO 19157). Likewise, there is a significant body of literature on and a few models of data curation [e.g.,3,8,17,23,24]. There is a renewed interest in data quality and making datasets FAIR (i.e., findable, accessible, interoperable, and reusable) in data curation communities of practice. They develop and share very valuable data cleaning, normalization, linking and disambiguation procedures and scripts [42,1,12,20,31]. At the same time, however, attempts to operationalize the FAIR framework have been largely fragmentary, situational and lack a strong grounding in the metadata and information quality literature. That limits their generalizability.

1.4.2 Little research on how research data repositories select and prioritize their DQA targets

There have been conceptualizations of research data quality and studies of researcher perceptions and priorities for data quality [e.g., 10,16,18,34]. The perception of what constitutes quality and useful data and/or when the data becomes useful may vary within the same process, discipline, and across different processes within disciplines [17]. Researchers may rely on different properties and cues of data to assess its relevance, quality, value, and reusability [13,37]. ***There is little research, however, on how DQA practices of research data curators and repository managers are aligned with researchers' needs for and ways to evaluate and perceive data quality***. This exploratory study will contribute to filling that gap by comparing and contrasting the data quality needs and evaluation strategies of dataset users from the r/Datasets Reddit community of data professionals and enthusiasts to the DQA practices of research data curators and repository managers.

A concept related to data quality is data value. The value of data is shaped by its informativeness: what questions it can answer or what and how many concepts and relations it represents, and how novel and

[1] https://paperswithcode.com/rc2021

[2] https://www.geekwire.com/2022/commentary-how-homeowners-defeated-zillows-ai-ultimately-leading-to-zillow-offers-demise/

frequently sought those questions and concepts are [45,37,34]. The amount or scale of data can be used as a predictor of its value. For instance, big consumer data usually translates into a higher value and a larger market share for the company that owns the data [46]. Another concept related to value is cost. The cost of data creation is often used for evaluating its value [45]. The higher value of big data also comes with a higher cost of its curation. Furthermore, the value of preserving and curating an extant dataset can be evaluated relative to the cost of its re-creation from scratch as needed [e.g., DNA sequencing data or simulated data; 47]. Data quality and value can be used to identify and prioritize DQA targets [6,36,48]. Although repository managers cannot eradicate all quality problems in their data repositories, they can still address data quality problems that matter. The question then is how they determine what data quality problems matter. ***There is little research on how institutional data repositories evaluate the value of datasets and how they prioritize their DQA targets***. This exploratory research will contribute to addressing that need.

1.4.3 Lack of a systematic study/review of DQA infrastructure components used in university research data repositories

Another relevant literature that informs this study is the digital data curation literature [e.g., 8,17,23,24]. Although general infrastructure components of digital data curation are shared across different disciplines, the research project tasks, the types of data and digital objects produced, and the norms followed in managing, sharing, and evaluating data may vary [49,50,18,37]. Furthermore, DQA work requires access to the appropriate infrastructure for data quality evaluation, monitoring, intervention, and annotation. There are general data cleaning tools developed by industry, such as Open Refine. In addition, data curation consortia and communities of practice (e.g., DataOne, Data Curation Network, California Digital Library), and individual data repositories develop their own data cleaning and normalization modules [1,12,20,31]. ***To the best of our knowledge, there has not been a systematic study/review of DQA infrastructure components used in university data repositories***. ***Our study will address this gap***. Furthermore, the data curation literature includes studies that examined research data management and curation skills of repository managers [e.g., 23] and researchers [e.g., 18,43]. ***There is, however, a lack of studies that focus on the DQA skills and competencies of research data repository managers. Our targeted examination of those skills and competencies can inform the DQA and data science curriculum and training in library schools and communities of practice.***

2. Project Work Plan

2.1 The study design: research questions, methods, and theoretical framing

This applied exploratory research study will address the following research questions:

1. What are the DQA practices of data curators and repository managers?
    a. What data quality problems, challenges, and incidents do data curators / repository managers encounter in their work?
    b. What DQA methodologies, standards, workflows, metrics, metadata, and tools do they use to evaluate and ensure the quality of datasets they curate and to communicate that quality level/status to users?
    c. How do they prioritize their DQA actions?
    d. How is the DQA work divided? What roles are played? Does the DQA practice involve the original contributors and reusers of data?
2. What skills and competencies are needed for successful DQA?

We will use a case study design [44] to examine these questions (see Fig. 1). The study will be guided by a theoretical framework that comprises activity theory [22], information quality and information credibility frameworks [9,36], and self-determination theory [32]. We will use activity theory to conceptualize general structures of DQA activities and problems. The information quality and credibility evaluation frameworks will help us model the structure of data quality and credibility evaluation and relations among data use activities and data quality problems. Finally, we will use self-determination theory to develop interview protocol questions related to motivations for contributing to DQA. To illustrate how the research questions can be operationalized in a single data collection activity, we have included a draft interview protocol in the Supportingdoc2 attachment.

We will begin with constructing a sample of university data repositories. We will sample 50 data repositories run by universities or university consortiums/systems in the USA. We will use the re3data.org registry, DataOne, Data Curation Network, and the r/Datasets subreddit (an online community of dataset users and enthusiasts) to assemble that sample. In particular, we will assemble a list of university web domains referenced by members of the subreddit as data sources. Next, we will use this list to search the re3data.org registry, the membership lists of DataOne and DCN, and the web to assemble a sample of 50 data repositories run by research universities or universities consortia/systems. This selection procedure will ensure that we will have access to instances of data reuse and data quality problems experienced by repository users. Furthermore, when

assembling the sample, we will use the Carnegie Classification of Institutions of Higher Education[3] to ensure that the sample comprise both research-intensive and undergraduate universities. As a result, the set of data cases will reflect both high and low-end DQA infrastructure, problems, and challenges found in the field.

The study will use three types of sources for collecting data for each case: the websites/portals of data repositories, the r/Datasets subreddit, and interviews with repository managers. Guided by the theoretical framework, we will analyze repositories' data curation service descriptions and guides, policies, data use/reuse agreements, and metadata schemas and vocabularies for DQA themes. In addition, we will analyze the instances of data reuse in the Datasets subreddit community for the types of quality problems experienced by users. The subreddit was created on October 8, 2009, and at the time writing this proposal it has more than 150,000 registered members[4]. Members of the community ask and answer dataset related questions, share information about datasets, and seek the community's feedback and assistance with their data projects [34].

We will use findings of this documentary analysis in selecting interview participants and refine the interview protocol. Next, we will conduct semi-structured interviews with 30 data repository managers and curators. Potential participants will be identified from the analysis of the repository websites and by using a snowball sampling approach. Since some of the managers of the repositories may decline to participate in our study, we will sample a higher number of data repositories (i.e., 50) than the number of participants we plan to recruit (i.e., 30). Thus, we plan to have 30 cases representing the DQA practices of 30 universities. Some smaller universities might not have a dedicated data repository and might use their institutional repositories or general repositories to curate datasets of their stakeholders. As a consequence, we will have a higher number of cases (i.e., 30) than suggested by the literature [i.e., 12 cases; 11] to build a sound theoretical model of the DQA of university research data repositories.

We will use thematic content analysis to analyze the documentary and interview data. We will analyze them for both a priori themes defined by the research questions for the study and themes that emerged from the data. To ensure the validity and reliability of the study's findings we will evaluate the validity and readability of the interview protocol before its use. In addition, each dataset used in this study will be coded by two members of the project team. We will evaluate the intercoder reliability of coding, identify, discuss and resolve coding disagreements, revise the coding schema as needed, and then recode the data. Our case study design comprises multiple theories and multiple sources of data/evidence. That will allow us to triangulate (i.e., map, validate, and converge) results of the documentary and interview analysis by the theory and data sources to further enhance the validity and reliability of the study's findings [19,30,44].

Next, we will apply scenario-based task analysis [14] to the findings of the documentary and interview data analysis to develop DQA design scenarios. In particular, we will interpret and amplify the descriptions of individual and/or institutional DQA practices obtained from the data by our theoretical framework and an

[3] https://carnegieclassifications.iu.edu/
[4] https://www.reddit.com/r/datasets/

iterative process of scenario exploration and abstraction. That will help us articulate a DQA model accompanied with design scenarios [5,14]. The design scenarios will be concrete stories that illustrate and put into context the components of the DQA model, such as the types of data quality problems, their structures, and associated DQA relations and actions grounded in the data quality literature. Finally, we will encode the concepts and relations of the DQA model as an OWL/DRF ontology. The ontology can be used in assembling context-specific DQA metadata models/profiles and a machine-actionable reference source to design and support DQA workflows.

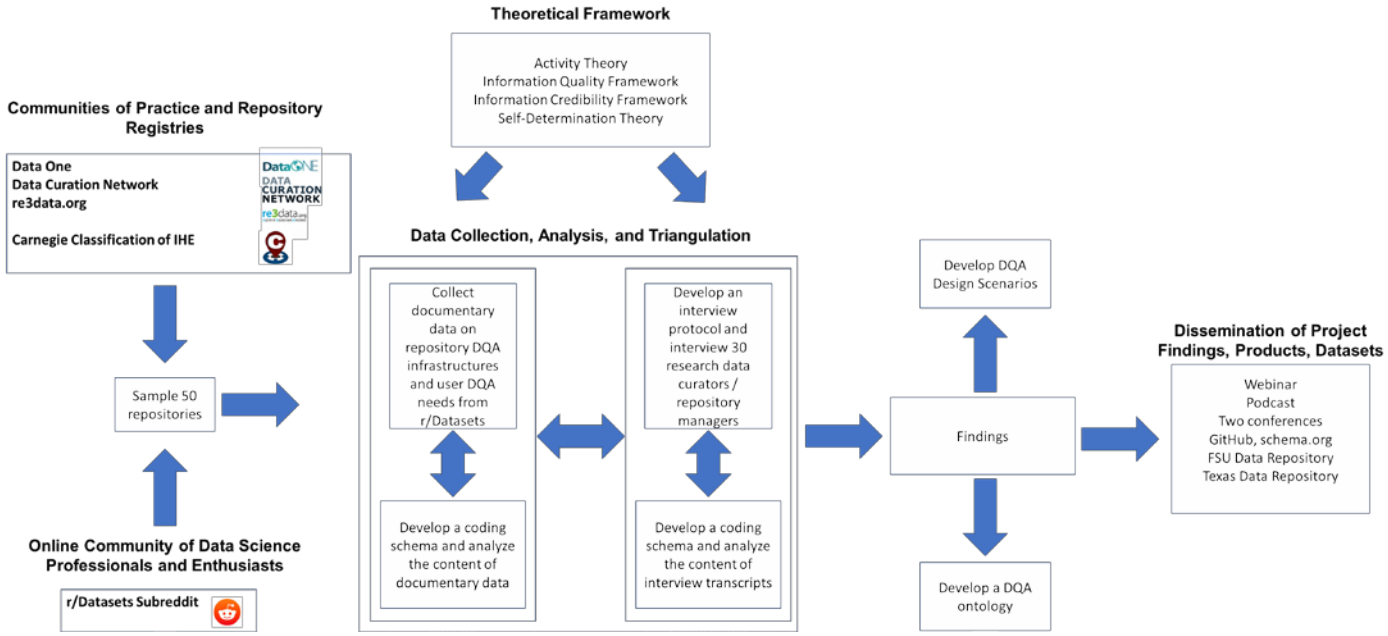

Fig. 1. The design of the study.

2.2 Dissemination plan

We will disseminate our study's findings and products at two library conferences (e.g., RDAP 2024, Open Repositories 2024) as well as using peer-reviewed journal publications. In addition to these traditional genres of scholarly communication, we will organize a webinar that are grounded in project outcomes. The webinar will help make project findings more accessible and usable in practice. We will work with the communities of practices such as RDAP, Open Repositories, Data One, and DCN to organize the webinar and reach the target groups and audiences of this project. Project publications, presentations, data, and products, including DQA scenarios and the DQA ontology, will be posted on the project's GitHub site and deposited to the FSU and TAMU data repositories. We will also publicize them in research data curation and management communities through community listservs. In addition, we will deposit the DQA ontology to schema.org to make it findable by and accessible to the broader data management audience - educators, students, and the public.

2.3 Project management; financial, personnel, and other resources

The length of the proposed applied exploratory research project is 18 months. It will start on August 1, 2022, and end on January 31, 2024. The total requested amount from IMLS is $92,922 ($67,749 excluding student support). In addition, FSU commits $61,408 as cost share in the form of 15% of Stvilia's time for Fall 2022 and Spring 2023.

Dr. Besiki Stvilia, a Professor in the School of Information at FSU, will serve as the Project P.I. and Director. He brings to the project expertise in the areas of data and information quality and data curation. He has taught information organization and data curation-related classes for more than two decades to thousands of students. Before joining academia, Stvilia was an information practitioner with a decade of work experience as a systems librarian, database administrator, and information manager. Dr. Stvilia will lead the overall effort to conduct the proposed research, assemble DQA design scenarios and the DQA ontology. He will be supported by a doctoral research assistant (RA). Dr. Dong Joon (D.J.) Lee, a TAMU Associate Professor and Research Information Systems Librarian, will serve as the Project Co-PI. He brings to the project expertise in research information management and research data curation. He will lead on the content analysis of documentary data collected from research data repository portals and the r/Datasets subreddit.

Stvilia will manage the project, and the project team will use the Teamwork project management software to track their progress toward achieving the project's objectives. In addition, the project team will be assisted by a grant analyst at Florida State University who will monitor and prepare monthly budget reports for the project. The study has been approved by the Human Subjects Committee of Florida State University (FSU HSC Number STUDY00002947). Before participants are interviewed, participants will be given a consent form approved by the FSU IRB. The form contains information about the project, including information about potential risks associated with participation in the data collection.

3. Project Results

Section 1.4 of the proposal provides a detailed discussion of how this exploratory research project will complement existing theory, scholarship, and practice. This section summarizes the expected results of the project and their implications for research and practice.

This exploratory research project will identify and inventory the practices of DQA of university research data repositories. In particular, it will identify the types of data quality problems and incidents; models, standards, and strategies used; the division of labor and the roles played; challenges and barriers to evaluating and

maintaining data quality, and skills and competencies needed. We will also examine how data curators and repository managers' DQA practices are aligned with the data quality needs and quality evaluation practices of data users. The study will investigate how research data repositories prioritize their DQA targets and how they communicate/signal the quality of their datasets to users (e.g., what dataset quality scoring schemas are used). In addition, the project will use the study's findings to develop DQA design scenarios. DQA scenarios will describe specific types of data quality assurance problems or incidents and associated DQA actions. Finally, we will develop a model and a metadata vocabulary of the DQA work of research data curators and repository managers. We will encode them as an RDF/OWL ontology.

This applied exploratory research project will advance the state of the art of the digital data curation literature. It will utilize a theoretical framework that integrates concepts and models from two theories and two frameworks to identify and categorize the types and antecedents of data quality problems found in university data repositories. In addition, it will provide a qualitative theory of DQA work of research data curators and repository managers. As such our study will contribute to the development of a common theoretical framework and methodology for studying DQA issues and problems in research data repositories.

This project will also have substantial practical implications. We will translate the study's findings into a set of DQA design scenarios and an ontology representing DQA concepts and relationships. There have been data quality assessment models and vocabularies proposed in the literature and used in practice (e.g., Six Sigma, ISO 9000, ISO 19157, W3C Data Quality Vocabulary, ISO 25012). However, to the best of our knowledge, none of the existing DQA standards provide DQA design scenarios specifying data quality problem-related DQA workflow requirements in a user-centered way. Furthermore, the current standards focus on data quality assessment and do not include models and vocabularies for data quality intervention. Finally, they do not capture the contexts of research data curation in US university ecosystems. When assembling our DQA ontology, we will build on and extend the ISO 19157 and ISO 25012 data quality models and W3C Data Quality Vocabulary [52-54]. Research data curators and librarians can use the knowledge products provided by this study to design their data quality conceptualization, assessment, monitoring, and intervention strategies, actions, and tools. The study's outcomes will also inform the data curation and data science training and education curricula of LIS schools and communities of practice.

3.1 Generalizability and adaptability of the project's results

We will ensure our project outcomes' adaptability, generalizability, and usability by making our data cases diverse and recruiting study participants from both research-intensive and undergraduate universities. In addition, will deploy quality checks throughout the project's lifecycle to ensure the validity and reliability of its findings. These will include evaluating the validity and readability of the interview protocol and evaluating the reliability of thematic coding.

Furthermore, to enhance the research findings' practical usefulness, usability, and accessibility, we will disseminate them using both traditional and non-traditional genres of scholarly communication such as a webinar. In addition, we will translate the study's findings into practical products such as DQA problem scenarios and associated design recommendations and a DQA ontology. These products can be used by library practitioners as well as information and data professionals outside the library and university ecosystems in the design and development of DQA services and software.

The study's strong theoretical and practical implications and broad, accessible and equitable distribution of its findings and products will benefit society. The project will help make DQA practices in institutional data repositories more systematic. The project will also make them more effective in recognizing and meeting the dynamic, evolving, and more equitable understandings of data quality (e.g., who's data counts as high quality or credible[5]), and data quality metadata needs and requirements of their users.

[5] https://t.co/EhuyA9iqk8

YEAR 1

	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12
A1	**Sample 50 research data/institutional repositories** 1. Extract a list of university web domains referenced as data sources by r/Datasets subReddit members 2. Triangulate the list of web domains with re3data.org, the membership lists of DataOne and DCN 3. Stratify the list by institution type using the Carnegie Classification of Institutions of Higher Education											
A2		**Data collection from documentary sources** 1. Collect documentary data on repository DQA infrastructures and user DQA needs from the r/datasets subReddit 2. Develop a coding schema and analyze the content of documentary data										
A3				**Data collection from human subjects** 1. Develop an interview protocol and interview 30 research data curators / repository managers 2. Develop a coding schema and analyze the content of interview transcripts								
A4							**Merge and triangulate data** Merge and triangulate the findings of documentary data and interview analyses					
A5								**Develop DQA scenarios** Develop DQA design scenarios and an associated guide/cookbook				
A6									**Develop a DQA model and ontology** Develop a DQA model and encode it as a RDF/OWL ontology			
A7											**Document and share DQA ontology and design scenarios** 1. Document the DQA ontology and design scenarios for public distribution 2. Deposit the DQA ontology to Schema.org 3. Post the DQA design scenarios along with an associated guide/cookbook to the project GitHub site	
A10								**Disseminate project results** Disseminate and publicize project outcomes and products: present at conferences, write peer-reviewed journal papers, create a podcast, and organize a webinar				

Table Legend: M=Month; A=Activity

YEAR 2

	M13	M14	M15	M16	M17	M18
A8	**Document and deposit data** 1. Anonymize and document research datasets 2. Deposit the datasets in the FSU Data Repository and Texas Data Repository					
A9				**Write a project summative report**		
A10	**Disseminate project results** Disseminate and publicize project outcomes and products: present at conferences, write peer-reviewed journal papers, create a podcast, and organize a webinar					

Table Legend: M=Month; A=Activity

The project will generate two research datasets. One dataset will comprise policy documents, dataset documentation guide and submission forms, and any other kinds of DQA related pages, software descriptions collected from 50 websites/portals of university research data repositories and institutional repositories. The dataset will be in the Nvivo file/database format. The database will also include the coding schema used to analyze the data.

Type: a Nvivo data file/database.

Availability: We will deposit the file to the FSU and Texas research data repositories.

Access: The file will be openly accessible in the repositories without any restriction.

Sustainability: The repositories will ensure long term access to the dataset.

The second research dataset will comprise transcripts of 30 interview recordings along with the coding schema used to analyze the data.

Type: a Nvivo file/database.

Availability: We will deposit a fully anonymized version of the database to the FSU and Texas research data repositories.

Access: The file will be openly accessible in the repositories without any restriction.

Sustainability: The repositories will ensure long term access to the dataset.

The project will create DQA design scenarios to illustrate specific types of data quality problems and associated design recommendations.

Type: a text file

Availability: We will deposit the file to the FSU and Texas research data repositories.

Access: The file will be openly accessible in the repositories without any restriction.

Sustainability: The repositories will ensure long term access to the file.

The project will create a DQA ontology that will encode a DQA model/theory produced by the study.

Type: an OWL file

Availability: We will deposit the file to the FSU and Texas Digital Data Repository.

Access: The file will be openly accessible in the repositories without any restriction.

Sustainability: The repositories will ensure long term access to the file.

The project will create a video podcast of project results

Type: an MPEG-4 file

Availability: We will deposit the file to the FSU and Texas research data repositories.

Access: The file will be openly accessible in the repositories without any restriction.

Sustainability: The repositories will ensure long term access to the file.

The project will produce peer-reviewed journal publications.

Type: PDF files.

Availability: Preprints of the publications will be deposited to the FSU institutional repository.

Access: The files will be openly accessible in the repository without any restriction.

Sustainability: The repository will ensure long term access to the files.

The project will produce conference presentations.

Type: PDF files.

Availability: Copies of the presentations will be deposited to the FSU institutional repository.

Access: The files will be openly accessible in the repository without any restriction.

Sustainability: The repository will ensure long term access to the file.

The project will generate two research datasets. One dataset will comprise policy documents, dataset documentation guides, and submission forms, and other kinds of data quality assurance (DQA) related pages and software descriptions collected from 50 websites/portals of university research data repositories and institutional repositories. The dataset will be generated manually and will be stored in the Nvivo file/database format. The database will also include the coding schema used to analyze the data. The dataset will be collected from August 1, 202 to November 30, 2023. The data will be analyzed using a combination of a priori codes and emerging codes related to the study's research questions.

Type: a Nvivo data file/database.

Availability: We will deposit the file to the FSU and Texas research data repositories.

Access: The file will be openly accessible in the repositories without any restriction. The dataset will be documented using the repositories' descriptive metadata profiles. In addition, we will link the dataset to publications and presentations that use the dataset. We will use persistent identifiers such DOIs to make that linkage. That way, the publications and presentations will serve as additional documentation or 'data papers' for the dataset.

Sustainability: The repositories will ensure long-term access to the dataset.

The second research dataset will comprise transcripts of 30 interview recordings along with the coding schema used to analyze the data. The data will be collected from October 1, 2022 to January 31, 2023. The dataset will be used to address the study's research questions. The data will be analyzed using a combination of a priori codes and emerging codes related to the study's research questions. Before depositing the dataset to the FSU and Texas research data repositories, we will anonymize it by removing any information that might identify the study's participants. The study has been approved by the Human Subjects Committee of Florida State University (FSU HSC Number STUDY00002947). Before participants are interviewed, participants will be given a consent form approved by the FSU IRB. The form contains information about the project, including information about potential risks associated with participation in the data collection.

Type: a Nvivo file/database.

Availability: We will deposit fully anonymized file to the FSU and Texas research data repositories.

Access: The file will be openly accessible in the repositories without any restriction.

Sustainability: The repositories will ensure long term access to the dataset.

Organizational Profile
Florida State University
College of Communication and Information
School of Information

MISSION

Access to and use of information technology, services, and products by people in all their diversity throughout their lives is of profound individual and societal importance. Students educated at Florida State University's School of Information Studies work to ensure information access for all people.

Our instructional programs are concerned with recordable information and knowledge, and the services and technologies to facilitate their management and use encompassing information Architecture; Digital Libraries; Information Management and Policy; Knowledge Organization; Information Technology, Information Behavior, and Information Needs of Youth.

Through gifted teaching, significant research, and proactive outreach, the faculty supported by a talented and dedicated staff and engaged alumni, sustain the School's commitment to empowering people through assuring access to relevant information.

As part of a comprehensive research university, the School has tripartite goals relating to instruction, research and public service.

PLACEMENT WITHIN PARENT ORGANIZATION (Service Area)

Founded in 1947, the School of Information has been preparing professionals to make vital connections between people and information ever since. The School's undergraduate and graduate studies are guided by the Director of the School of Information, who reports to the Dean of the College of Communication and Information. The project responsibility will be assigned to the Project Director and faculty members in the School of Information.

DEGREE PROGRAMS OFFERED

The flagship program of the School of Information is its ALA accredited Master of Science in Information. The School also offers a Bachelor of Science in Information Technology, a Master's of Science in Information Technology, a post Master's Specialist and a Ph.D. in Information.

Sources used

FSU School of Information (2019). *iSchool Mission*. Retrieved from https://ischool.cci.fsu.edu/about/mission/)
FSU School of Information (2016). *2016 Narrative Report to American Library Association*.