Data Management Plan for Usage of Academic Profile Websites (v. 1.0)

Abstract

This project will collect usage data of librarian researchers on four academic profile websites: Google Scholar Citations, ResearchGate, Academia.edu and ORCID (Open Researcher and Contributor ID). This data management plan describes the data collection process, data storage and long term preservation requirements, and potential data-related issues that may rise in the research process.

Principal investigators:

Maha Kumaran and Li Zhang

Administrative Details

Project Name:

Investigating the Usage of Academic Profile Websites among Librarian Researchers

Principal Investigator / Researcher:

Maha Kumaran and Li Zhang

Project Data Contact:

maha.kumaran@usask.ca; li.zhang@usask.ca

Description:

Academic profile websites are online platforms where researchers create their professional profiles, and organize and showcase their research outputs. Examples of academic profiles websites include Google Scholar Citations, ResearchGate, Academia.edu, and ORCID (Open Researcher and Contributor ID). Recent years have seen an increasing adoption of these websites among researchers, mainly because of the perceived value for facilitating collaboration and boosting research impact. While librarians are providing assistance to researchers in developing profiles on these platforms, it is unknown how librarians themselves, as researchers, are using these tools. This project will investigate the presence on the academic profile websites amongst librarian researchers (i.e., librarians whose professional responsibilities include the requirement of conducting research) in the United States and Canada. The results from this project will provide a more complete understanding of academic profile websites, identify new

directions for improving the functionalities of these sites, and discover opportunities for providing new library services to the research community.

Data Collection

What types of data will you collect, create, link to, acquire and/or record?

Both textual and numerical and data will be collected in this project. The specific textual data include the names of librarian researchers, rank, and institutions, their postings, and questions asked and answered on the academic profile websites. The numerical data collected will include publications listed, reads, views, followers, followings, citations, h-index, and RG Score (a metric provided by ResearchGate). Because not all of the four academic profile websites provide the same information, the data listed above may be only applicable to certain websites.

We will choose academic librarian researchers from Canada and the United States as the study objects. These librarian researchers must belong to Canadian Association of Research Libraries (CARL) or Association of Research Libraries (ARL) institutions that require them to publish for tenure, promotion, or continuing appointments. Academic Profile Websites will be chosen from ResearchGate, Academia.edu, Google Scholar Citations, and ORCID - the four major academic profile websites used by librarian researchers in Canada and the United States. Information about academic librarian researchers will be gathered from CARL and ARL websites.

What file formats will your data be collected in? Will these formats allow for data re-use, sharing and long-term access to the data?

MS Excel spreadsheet and MS Access database will be used to collect the information. The data from the database will be transferred to SPSS for analysis. After the project is completed, we will convert the data format to a non-proprietary format, such as .csv (comma-separated values) format, to facilitate data reuse, sharing, and preservation.

Proprietary file formats (e.g., MS Excel, MS Access) are associated with a particular software, which may not be available in the long term. Therefore, it is better to save files to a non-proprietary format (e.g., .csv format for database and tabular data) for long term preservation.

Data Management Plan for Usage of Academic Profile Websites (v. 1.0)

What conventions and procedures will you use to structure, name and version-control your files to help you and others better understand how your data are organized?

Each file will be named with a short description/acronym to reflect its content, followed by the date of creation. To record different versions, we will add a version number in the file name. For example, file name GSC_20200608_v01.xls represents the data about Google Scholar Citations acquired on June 8, 2020, the 1st version.

We will create a document to detail file naming conventions and provide a list of explanations of the short descriptions/acronyms used in file names.

Documentation and Metadata

What documentation will be needed for the data to be read and interpreted correctly in the future?

We will provide a brief description of the project, a detailed methodology on how the data is collected, date of data collection, analysis performed, and details of who performed each task to accompany the data.

A data dictionary will be created to define different data fields, data type, and explanation of data coding.

How will you make sure that documentation is created or captured consistently throughout your project?

The researchers will have ongoing meetings to discuss any issues/problems that might occur during the research process: for example, how to deal with name variations, and how to deal with inaccurate/incomplete data on academic profile websites. Any decision made will be recorded in the methodology document to ensure it captures the changes in the data collection process.

> *Documenting the decisions and changes made in the data collection process will also save you time when writing the manuscript later.*

Data Management Plan for Usage of Academic Profile Websites (v. 1.0)

If you are using a metadata standard and/or tools to document and describe your data, please list here.

Dublin Core Metadata Elements will be used to document and describe the data. It is one of the most widely used metadata schemas in disciplinary and institutional repositories, and includes 15 metadata elements. The main elements of Dublin Core we plan to use include Title, Subject, Description, Creator, Date, Type, Format, and Identifier (e.g. DOI).

Storage and Backup

What are the anticipated storage requirements for your project, in terms of storage space (in megabytes, gigabytes, terabytes, etc.) and the length of time you will be storing it?

Based on the data collected in a previous study focusing on researchers in other fields, we estimated the overall size of the raw data for this project will be 20-50 Megabytes. This does not take into account file versions. The final storage requirement will be about 500 Megabytes when multiple file versions are considered.

> *You will often need to save multiple versions of the data collected. Data storage requirements will therefore be at least several times more than the size of the raw data.*

Because academic profile websites are revolving rapidly, we plan to retain the data acquired including raw data and those generated from the raw data in OneDrive for a minimum of 7 years, aligning with the requirements from common research funders.

> *Note that the length of time for storage refers to how long you intend to keep the data on your own storage space (OneDrive for this project), and it is different from long term preservation in a data repository.*

Data Management Plan for Usage of Academic Profile Websites (v. 1.0)

How and where will your data be stored and backed up during your research project?

The data will be stored in the University of Saskatchewan's institutional OneDrive, a secure cloud based storage developed by Microsoft. It can be easily accessed from different types of devices and can be shared by the research team. Further, it is backed up automatically to prevent data loss.

How will the research team and other collaborators access, modify, and contribute data throughout the project?

Access to the research data on OneDrive will be limited to the members of the research team. Using OneDrive minimizes the efforts needed for data transferring. We anticipate that the initial data collection work will be divided between the two researchers. When the initial data collection work is completed, the data will be checked by the other researcher to ensure its accuracy and completeness. After final inspection, the data collected by each researcher will be combined into a master file.

Preservation

Where will you deposit your data for long-term preservation and access at the end of your research project?

While our institution currently does not have a data repository, we plan to deposit the data acquired in a Canadian open data repository that accepts data from other institutions, such as Dataverse of University of Alberta. Dataverse is an open source web-based data repository application for depositing, sharing, preserving, discovering, and citing research data. Data deposited in Dataverse of the University of Alberta can be discovered from other repository platforms such as FRDR (Federated Research Data Repository), therefore it increases the discoverability of our data.

> *If you are not familiar with data repositories, contact your local librarian to find one that is appropriate to your discipline or to your institution.*

Data Management Plan for Usage of Academic Profile Websites (v. 1.0)

Indicate how you will ensure your data is preservation ready. Consider preservation-friendly file formats, ensuring file integrity, anonymization and de-identification, inclusion of supporting documentation.

Because the data collected will be in MS Excel and MS Access formats, we will convert them into .csv format for long term preservation. The accompanying documentation will be converted into .txt format to help future researchers to understand and reuse the data.

Although this study does not include sensitive data as all the information collected will be publicly available, we will anonymize the names of the study objects as we do not expect this information will be of interest to future data users.

Sharing and Reuse

What data will you be sharing and in what form? (e.g. raw, processed, analyzed, final).

Processed data will be shared after de-identifying the names of the study objects. The format for data sharing will be .csv format.

> *Sometime it may not be necessary to share raw data for various reasons (e.g., confidentiality, size of raw data). Think about how future researchers may use your data and decide what form of data to share.*

Have you considered what type of end-user license to include with your data?

We will have the Creative Commons Attribution CC BY license for the data, which allows others to distribute, reuse, adapt, and build upon the data as long as the original data creators are credited.

Data Management Plan for Usage of Academic Profile Websites (v. 1.0)

What steps will be taken to help the research community know that your data exists?

We will present the project at conferences, publish the results in a peer-reviewed journal with links to the data, and promote the data's existence through social media and word-of-mouth. We hope to have a DOI (Digital Object Identifier) assigned to the data. A DOI provides persistent and permanent identification to the data to which it is associated, thus making it more discoverable by search engines and by other researchers.

The citation of the deposited data will be included in the journal publication, and the publication citation will be included in the deposited data so that the information can be discovered from both ways.

Responsibilities and Resources

Identify who will be responsible for managing this project's data during and after the project and the major data management tasks for which they will be responsible.

Data managing responsibilities will be equally shared between the researchers due to the relatively small scale of the project. Major data management tasks include: identify study objects, develop documentation, build MS Database for data collection, collect data on four academic profile websites, inspect data accuracy and completeness, analyze data, convert data, and deposit data to an open data repository.

How will responsibilities for managing data activities be handled if substantive changes happen in the personnel overseeing the project's data, including a change of Principal Investigator?

If one PI is no longer able to participate in the project, the other PI is intended to continue the project. In this case, the Social Science Research Laboratory (SSRL) at the institution may be asked to develop a script to automatically harvest data on academic profile websites, although data collected by this approach may not be as reliable as manual collection.

Another alternative is to conduct the research on a smaller scale, for example, studying librarian researchers in Canada only, or selecting two academic profile websites instead of four.

Data Management Plan for Usage of Academic Profile Websites (v. 1.0)

What resources will you require to implement your data management plan? What do you estimate the overall cost for data management to be?

The data storage (OneDrive) during the active phase of the project is provided by the institution, and there is no extra cost. Because we plan to deposit the data in an open data repository, we also do not anticipate any direct cost for long term data preservation.

If we need to use SSRL to develop a script for data harvesting, the anticipated cost would be around $2000. In this case, we will apply for research funding to support the work. If not successful, we also have access to other funding that we may use for this project.

Consider in advance if additional resources/funding might be needed for data management.

Ethics and Legal Compliance

If your research project includes sensitive data, how will you ensure that it is securely managed and accessible only to approved members of the project?

Because all of the information needed for this project is publicly available, there is no sensitive data involved.
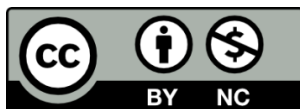
If applicable, what strategies will you undertake to address secondary uses of sensitive data?

Although there is no sensitive data involved in this project, we will have names, rank, and institution information of the study objects. Releasing the individual names publicly may cause unnecessary concerns. Therefore, we will replace the names by study object number, such as Librarian Researcher 1, 2, 3, etc., when depositing data in a repository when the project is completed.

Data Management Plan for Usage of Academic Profile Websites (v. 1.0)

How will you manage legal, ethical, and intellectual property issues?

Because there is no sensitive data included in this study, we do not expect there will be any legal, ethical, or intellectual property issues involved.