
Retention of BIPOC academic librarians in Canada and the United States

A Data Management Plan created using DMP Assistant

Creator: Maha Kumaran

Principal Investigator: Maha Kumaran

Data Manager: Maha Kumaran

Affiliation: University of Saskatchewan

Funder: Portage Network

Template: Portage Template

Project abstract:

All researchers in this project are participants of the Leadership and Career Development Program (LCDP) of the Association of Research Libraries. This year-long program aims to prepare mid-career librarians from underrepresented racial and ethnic groups for leadership positions. The recent census data from Canada and the United States (CAPAL Census 2018; Mian 2020) show that racial minority and Indigenous librarians are underrepresented in academic libraries. We will contact librarians from current and previous LCDP cohorts to understand their perspectives about why they stayed in the profession (Retention). This project will be conducted in two phases. The first is a survey done using Qualtrics. The link to the survey will be sent to previous cohorts of the LCDP. The survey will ask participants to provide their contact information if they are interested in participating in the second phase, the virtual interview. For the second phase, the virtual interview, we will use the WebEx software from the PI's institution. Through these phases, we will investigate the retention efforts of current and previous LCDP participants to identify challenges and gaps in this area. Our study will propose best practices, recommendations, and pathways towards the retention of racial and ethnic minority librarians.

Identifier: 6859

Last modified: 01-03-2022

Copyright information:

The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customize it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal

Retention of BIPOC academic librarians in Canada and the United States

Data Collection

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

This project will use data from an online survey and one-on-one online interviews. This data management plan (DMP) describes the data collection process, data storage, and long term preservation requirements.

We will conduct an online survey using Qualtrics from the Massachusetts Institute of Technology (MIT). Survey data will be exported.

We will also be conducting semi-structured interviews using WebEx from the University of Saskatchewan (U of S) and produce digital audio and text transcriptions based data.

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?

This project will use data from an online survey and one-on-one online interviews. This data management plan (DMP) describes the data collection process, data storage, and long term preservation requirements.

We will conduct an online survey using Qualtrics from the Massachusetts Institute of Technology (MIT). Survey data will be exported.

We will also be conducting semi-structured interviews using WebEx from the University of Saskatchewan (U of S) and produce digital audio and text transcriptions based data.

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?

Survey

Upon exporting the survey data from Qualtrics, the data will be named as "raw" data. The raw data will be de-identified and named as "de-identified" data for analysis. From the de-identified survey data files, a variety of analytic data and visualization files will be created. A file naming convention developed by the team for the process of managing and analyzing data will be used and documented to name the various data files.

Interview

We will follow a similar process for interviews. Digital audio files will be named "raw" audio and their verbatim transcripts will be called "raw" transcripts. Once the transcripts are processed and cleaned, the resulting data will be named "processed" transcript. Copies of the "processed" transcripts will be de-identified and saved as "de-identified" transcripts. The "de-identified" transcripts will be used for analysis purposes and more data analytics and visualization files will be created. The research team will create and document a guide for naming conventions for each of these data files.

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, processed, and de-identified in the two phases, and the dates of all data collection. We will also provide details on how the analysis was performed.

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

We will have written protocols for documenting the survey and interview methodology, and process.

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

During the data analysis stages, we will create a data dictionary to define different data fields, data type, and explanation of data coding.

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?

We expect to have less than 5 gigabytes of interview audio, transcription and analysis files. For the survey data, we expect to have less than 2 - 3 gigabytes data.

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

Survey results on the Qualtrics survey will be stored in the Qualtrics server (based in the US). All exported survey results and interview recordings will be stored in the PIs (Maha Kumaran, University of Saskatchewan) institutional OneDrive that can be accessed through password protected computers by researchers. All de-identified data will be stored in Nvivo for coding, categorizing, analysis, interpretation and presentation purposes. The identifiable survey results and interview recordings will be safely destroyed five years post publication following the PI's institutional protocol.

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

Collaborators will access the Qualtrics survey results via password-based accounts. Collaborators will access the exported survey results and audio recordings on the PIs OneDrive in the cloud only via password protected accounts and they will not download the recordings to their local computers. The access, modifications, and contributions to the transcripts will happen in Nvivo or Word documents.

Preservation

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

The PIs institution does not have a data repository. We plan to deposit the data in open data repositories that accept data from other institutions such as the [Dataverse of University of Alberta](#) or a discipline-specific repository for social science studies, such as [ICPSR](#).

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.

Where possible we will convert the data into .csv format for tabulated data, simple text format or markdown format for texts, and .svg format for graphs for long-term preservation. Any accompanying documentation will be converted into .txt format or markdown format to help other researchers understand and reuse the data.

Sharing and Reuse

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

Analyzed and final data and depending on the data, de-identified data may also be shared. Since our research is focused on minority librarians and there are few of them in both countries, some of the de-identified data may still be identifiable. Hence, we will know more about sharing de-identified data at the completion of the project.

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

We will use Creative Commons BY 4.0 ([CC BY 4.0](#)) for our dataset.

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

We will link our dataset to the publications and presentations arising from this study. We also hope to have a DOI associated with the data for permanent and persistent identification to the data through the repositories we deposit them to.

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.

Responsibilities will be shared amongst all researchers. Major responsibilities for this project's data management tasks are: collecting data, developing documentation, organizing data files, using Nvivo for de-identified data, inspecting data accuracy, converting and depositing data.

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?

There are multiple researchers in this project and more than one researcher is knowledgeable in handling DMP. If changes happen one of the other researchers will step in.

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

OneDrive is provided through the institution at no extra cost. Since we plan to deposit data in open data repositories that are free of charge, we do not anticipate any costs for preservation purposes.

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?

PI has control of the OneDrive and will only provide access to researchers in this project. Researchers will not save any raw data on their computers.

If applicable, what strategies will you undertake to address secondary uses of sensitive data?

At this time, we do not anticipate sharing identifiable data with other researchers for secondary use.

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

N/A

