Using Big Data to Visualize the Representation of Marginalized Identities

Introduction

My project aligns with the Laura Bush 21st Century Librarian Program goal 2: develop faculty, library, and archives leaders by increasing the institutional capacity of libraries, archives, and graduate programs related to library and information science. This project supports all three of the objectives of goal 2. I am requesting grant funds totaling $207,027. The purpose of this Early Career Research Development project is to support the ongoing research agenda of Dr. LaTesha Velez, Assistant Professor at the University of North Carolina at Greensboro, Department of Information, Library and Research Sciences (ILRS). Specifically, my work seeks to make visible the needs and stories of Black, Indigenous, and People of Color (BIPOC), particularly Black and Latine people. I use asset-based rather than deficit-based theories and language in my research design and writing. My research is grounded in Latine critical race theory (LatCrit) and critical race theory (CRT). We seek to answer the following research questions: R1. What are the conversations the LIS profession has about Hispanic/Latine people? R2. In what journals are these conversations occurring? R3. What are some underlying biases or assumptions in the scholarship around Latine people? R4. What are the frequency trends over time regarding Latine research? This is not human subjects research, so it is exempt from requiring IRB approval. I will lead a research team that includes Dr. Melissa Villa Nicholas, a researcher at The University of Rhode Island who has done extensive research on Latine populations, Abby Rinaldi, a free-lance data analyst earning their PhD from North Carolina State University who will write any necessary code, perform the data pulls, and create visualizations, and Michelle Rosquillo, a special librarian at the Medicus Firm who brings a much-needed practitioner perspective. The study will result in the following outcomes: the development of an ontology around Latine metadata in LIS research, evidence-based recommendations to shift cultural assumptions in the field of LIS; and an open-source code other researchers can use to conduct similar research.

Project Justification

Representation within LIS literature is important because this is one entrance into understanding how ideologies, epistemologies, and ontologies are shaped within LIS, and "as a sanctioned articulation of the field's best work, peer-reviewed journal articles shape the discourse of LIS as an academic discipline" (Hackney et al., 2018, p. 11). In conversations about Latines in Library and Information Science (LIS), qualitative content analysis methods have been traditionally used to demonstrate trends or lack of representation. This article contributes to the conversation by reviewing English-language, peer-reviewed literature about Latines in LIS through textual mining analysis. By analyzing LIS literature using the Scopus API and manual journal pulls, this study visualizes how Latines are discussed in the LIS academic and professional field in the United States. This work enters conversations about how LIS conceptualizes diverse communities in the literature en masse (Sung & Parboteeah, 2017).

We ask: What insights can be gleaned through visualization tools about the state of Latines and LIS literature? We are beginning with Latines because the Hispanic population in the United States is the fastest-growing minority in the United States, reaching nearly 58 million in 2016 (Flores, 2017). In LIS, Latines have been underrepresented as faculty, librarians, and patrons. Serving Hispanic and Latine patrons is a challenge for which LIS workers and students must prepare. The concentration of Latine publications within a textual corpus makes LIS literature uniquely suitable for discussing broader complex issues around representation, with gaps extending from the classroom to the professions.

Project Work Plan

The project team will collect articles using the Scopus scholarship database and manual pulls to fill in gaps in Scopus. Manual pulls will include: First Monday, Library & Information Science Research, Government Information Quarterly, Library Technology Reports, School Libraries Worldwide, Public Libraries, Government Information Quarterly, The Journal Of Academic Librarianship, New Review Of Academic Librarianship, Interactions: UCLA Journal Of Education and Information Studies, Urban Library Journal, Libraries: Culture, History, and Society, Journal Of Critical Library And Information Studies, Library Journal, In The Library With The Lead Pipe, Radical Librarianship, Information & Culture. We will examine all articles published over 51 years from 1970 to 2021 for a selection of 272 journals (255 from the Library and Information Sciences journal list from ScimagoJR and 17 added by the research team) in Library and Information Sciences. I chose 1970 due to the inconsistency of database items available before that date, other research noting the dirth of articles about Latine before 1970, and my own failed searches for relevant articles prior to 1970 (Güreña and Erazo, 2000). We stopped at 2021 as we began creating the code in 2022. After the API and manual pulls are completed, the files will be combined into one CSV file using a Python script. To clean the dataset, we will use a list of relevant keywords containing country names, language names, and diaspora identifiers for the Latine community, removing records that do not include identified keywords. To further refine these results, we are implementing the

following filtering parameters: at least one author among the authors needed to have a US institutional affiliation, and the article must have originally been published in English. As a final step in cleaning the data, three researchers will manually review each record to remove false positives. For example, we will look for and remove articles about the classical Latin language versus Latin as an identity marker. We will also remove articles that mention Latine only when speaking about demographic classifications of survey participants or area residents but have no other mention or discussion of Latine people. From these remaining articles, we will create visualizations to represent what journals publish about Latines, how often these articles are cited, the identity terms used, and the subject terms collocated with them. This work will be analyzed to determine the connotations of subject terms co-located with Latine identities. We will also use the article abstracts as the corpus to complete our topic modeling research. We will import our corpus into Voyant and the MALLET Topic Modeling Toolkit. From there, we will identify the topics being discussed in Latine research.

Data will be collected and cleaned during the first year, and two articles are planned. The first planned article will examine trends in patterns of study, such as the years and frequency of study, journals that publish studies about all categories of Latine people (patrons, librarians, students, etc.), the number of articles, and citation ranges. The next proposed article will use abstracts as our corpus for topic modeling to uncover trends in research conversations surrounding Latine identities, again not limited to one category such as librarians but inclusive of all we find. We will create an ontology around Latine classification during the second year, the outcome of which will help the field interrogate biases that arise within LIS classification schemas. During the second year, I will also identify future article topics based on the collected data. In both years, we will submit to conferences such as the Reforma National Conference, iConference, ACRL Conference, and the ALISE Conference. In later years, I hope to expand this research in multiple directions, including looking at other minority communities and applying our data approach to research conversations surrounding marginalized populations in other STEM fields.

Diversity Plan

This project will strengthen the field's commitment to EDI by interrogating underlying assumptions that may be present in the research we publish, disseminate to LIS workers, and teach in classrooms. Understanding the patterns of how Latines are discussed may reveal biases such as racism, sexism, gender discrimination, and classism. Before we can change deficit-based conversations, we must know if they are occurring. This recognition will help the field change this structural discrimination by shifting the narrative around *who* Latines are in the library and beyond. Such honest evaluations of invisible biases can help LIS professionals work toward a climate of cultural humility, making the profession a more welcoming and inclusive space for marginalized populations such as Latine people.

Project Results

The study will result in the following outcomes: the development of an ontology around Latine metadata in LIS research; evidence-based recommendations to shift cultural assumptions in the field of LIS; and an open-source code other researchers can use to conduct similar research. We hope to share our findings and recommendations directly through our research, presentations, and the creation of a Latine metadata website, which we will advertise in various locations such as the JESSE listserv, other listservs identified through the REFORMA resources and publications listserv page, the REFORMA newsletter, and other avenues we find.

Budget Summary. The estimated 2-year total project budget is $207,027 (d+i). Personnel: $82,286 salary and fringe for the PI and Data Analyst; Travel: $14,000 for conferences and project meetings; Supplies: $1,000 for research supplies; Contracted Services: $40,000 for research team payments and $5,000 for website creation/hosting; Direct Costs: $142,286; Indirect Costs: $64,740.

References:

Flores, Antonio. 2017. "How the U.S. Hispanic Population Is Changing." *Pew Research Center* (blog). September 18, 2017. https://www.pewresearch.org/fact-tank/2017/09/18/how-the-u-s-hispanic-population-is-changing/.

Güereña, S., & Erazo, E. 2000. "Latinos and librarianship." *Library Trends, 49*(1), 138–181.

Hackney, S. e., Handel, D., Hezekiah, B., Hochman, J., Lau, A., & Sula, C. A. 2018. "Visualizing Identities in LIS Literature." *Journal of Education for Library and Information Science, 59*(1–2), 10–32. https://doi.org/10.3138/jelis.59.1-2.04

Sung, H.-Y., & Parboteeah, P. 2017. "Diversity-related research reported in high-impact library and information science journal literature: A content analysis." *Library & Information Science Research, 39*(2), 77–84. https://doi.org/10.1016/j.lisr.2017.03.002