University at Albany, State University of New York

Data Science for the 21st Century Library and Information Professions

Statement of Broad Need

The Research Foundation for SUNY, University at Albany, in collaboration with the University of North Texas, seeks *$81,668* from the *Laura Bush 21st Century Librarian Program*, under the *Lifelong Learning* category, and proposes a 12-month ***planning project*** that aims to: (1)identify data science education models through a systematic analysis of graduate level data science programs and courses offered at LIS schools in North America, and (2)reveal the extent to which perceptions, views, and attitudes of a sample of current students, graduates, and their employers contrast and/or complement each other with respect to the skills, knowledge, values, and attitudes that the programs were meant to instill in an effective data librarian[1]. The outcome of this project is intended to inform and lay the foundation for a future project that is broader in scope, more comprehensive, and will investigate in greater detail the extent to which the demands and needs of the supply side (LIS schools) and demand side (employers) of data science are met, identify the specific demands and needs, and build a model/standard graduate data science curriculum that addresses them.

In general, LIS schools, libraries, and library professionals have adapted to changing landscapes with respect to new innovations, and the challenges and opportunities that come with it. These communities mostly acted to innovate with and integrate these opportunities into their research, teaching, and practice. LIS has been at the forefront of integrating technology and systems to solve practical problems in libraries and deliver services.

The data revolution is one example. To respond to this revolution, data science programs recently have sprouted. It is time to scrutinize these programs with respect to their ability to address and accommodate current and future workforce demands, and the relationship among employer skill demand, the acquisition of those skills, and workplace opportunities for using them. Through this ***exploratory*** project, we endeavor to examine and provide a current snapshot of data science in the LIS field, trends in education and workforce development, and perceptions of skillsets required and expected of data librarians.

Project Design

The specific goal of this project is to assess, through a variety of methods, including surveys, content analyses, and focus group discussions, the supply side (LIS schools, students, graduates, etc.) and demand side (library employers, e.g., research libraries, special libraries, etc.) of data science to: (1)determine how current data science programs are positioned in LIS; (2)identify perceived strengths and weaknesses of graduates of those programs, and employed by libraries, from the point of view of their employers/supervisors; (3)document perceptions of current students of those programs with respect to the skills, knowledge, values, and attitudes they expect to develop; and (4) identify gaps in the expectations of graduates of those programs who are employed by libraries and their current opinions with respect to the skills, knowledge, values, and attitudes required to be an effective data librarian. Table 1 is a summary of these goals, the methods used to achieve them, and the anticipated outcomes. A plan to account for any challenges due to the COVID-19 pandemic will be presented in the full proposal.

The PI (Dr. Abebe Rorissa) has over 30 years of experience as an LIS professional, educator, and scholar conducting several survey research projects and teaching research methods and statistics courses. He served as Provost's Fellow where he helped create the data and institutional repository at his institution. The Co-PI (Dr. Jeonghyun (Annie) Kim), served as director of an IMLS funded digital curation and data management graduate certificate program. Her research areas include digital libraries and archives, data management and curation, and LIS workforce development. Both will collaborate to administer the project and will contribute equally toward each element of the project, including the survey and analyses of data.

[1] Data librarian is often regarded as an ad hoc term but is defined here as a librarian who provides expert support for students, faculty, and researchers in the areas of data management, data science, data analytics, and visualization.

Table 1: Project Goals, Methods, and Outcomes

Goal #	Methods	Outcomes
1	Content analysis of current data science courses, concentrations, and programs at ALA-accredited LIS schools and iSchools in North America	A comprehensive report on the data science programs' goals/design, curriculum, course contents and structure
2	Survey of a sample of supervisors or administrators at libraries (e.g., directors of the members of Association of Research Libraries/Special Library Association) that employ data librarians through a self-administered questionnaire	A report outlining staffing patterns in data librarianship, critical activities, and the responsibilities, skills and knowledge required; Carefully designed, tested, and refined sorting & ranking activities and instruments
3	Survey of a sample of current students of data science programs at ALA-accredited LIS schools and iSchools in North America through a self-administered questionnaire	A report summarizing the skills, knowledge, values, and attitudes current students expect to develop; carefully designed, tested, and refined sorting and ranking activities and instruments
4	Focus group discussion with a sample of data librarians who are graduates of data science programs at ALA-accredited LIS schools and iSchools in North America	A comprehensive report of the gaps in terms of their expectations and current opinions with respect to the skills, knowledge, values, and attitudes needed to be an effective data librarian or professional

Diversity Plan: We will ensure that samples of survey participants are representative of the population of students, employers/supervisors, and graduates with respect to race, income, education level, gender, etc. In the absence of diversity statistics on data librarianship, reports on diversity in the library workforce by professional associations (e.g., ALA) will guide our sample design. We also plan to engage and partner with stakeholders such as the RDAP Association, ALISE, and iCaucus for our sample selection and ensure that its diversity reflects the underlying population. Cultivating the partnerships will be part of this planning project.

Broad Impact: We envision significant impact on the nature, relevance of content or focus, and goals of future data science programs at LIS schools in North America and beyond. Based on the baseline data and outcomes of this project, the next phase of our effort will be an expanded project to conduct a more comprehensive and thorough fit and gap analysis that is broader in scope while extending and utilizing survey instruments, findings, and lessons learned from the planning project as input. The ultimate goal will be to ensure that the demands and needs of both the supply side and the demand side of data science are met and to build a model/standard graduate data science curriculum with content and program goals that will: (1)serve as a basis for future data science programs and enable the LIS education community to think strategically when planning for growth in their programs while meeting employers' demands for professionals with the requisite skills, knowledge, attitudes, and values; (2)prepare LIS students for 21st century workplace roles such as data librarianship; (3)equip libraries to not only innovate their services and respond to the changing needs of their users through data-related services, but also supply them with data librarians and professionals that will help widen their scope by adding processes and services that rely on data analytics; (4)give librarians avenues for upgrading their skills and knowledge by enrolling in systematically designed and relevant data science programs, thereby providing lifelong learning and career opportunities; and (5)enable all these and other stakeholders to take advantage of the data revolution to advance their core competencies, interests, careers, and responsibilities.

Budget: Our total budget request is $81,668 and consists of the following categories: (1) Personnel costs for the PI: $4,500; (2) Fringe benefits $675; (3) Travel to conferences and conduct focus group discussions: $5,000 (4)Supplies: $2,800; (5) Subaward to University of North Texas $10,232; (6) Student support, including salary/fringe/tuition: $30,880; (7) Incentives for survey participants and other costs: $2,300; (8) Indirect Cost: $25,281 [based on 54.5% of MTDC per federally-negotiated rate agreement with DHHS].