College of Computing & Informatics, Drexel University (Drexel CCI)
LIS Education And Data Science-4-the National Digital Platform (LEADS-4-NDP)

Drexel University is submitting this proposal to the Laura Bush 21st Century Librarian Program for Doctoral-level education, focusing on the National Digital Platform (NDP). In collaboration with several confirmed national NDP organizations (identified below), we propose the **LIS Education And Data Science-4-the National Digital Platform** (LEADS-4-NDP) project, a unique summer experience to enhance standard LIS doctoral curricula. The program, a boot camp combined with an immersive internship experience, will be open to LIS doctoral students across the country. The initiative will build a national cohort among the next generation of LIS faculty and bring necessary knowledge and skills in data science to LIS education.

1. Statement of Need: As purveyors and stewards of information, library scientists and archivists have important and growing responsibilities to actively and meaningfully participate in our data-driven world, and apply data science techniques to improve library services and operations. It is essential, thus, to educate the next generation of LIS faculty in tools, skills, and service needed for the National Digital Platform (NLP) and other data-driven services. Despite this need, LIS data-focused graduate programs primarily address digital curation, with limited attention to analytical skills.[1] This is not surprising, given that only 3% (11 of 330) schools on Swanson's Awesome Data Science Colleges list[2] partner information and library science programs.

The NDP's growth necessitates infrastructure development and increased capacity; however, the NDP's sustainability requires the application of data science methodological frameworks, processes, and tools to spot trends, generate new knowledge, and improve services and operations. To this end, the next generation of LIS faculty need to develop strong analytical skills; although many LIS programs across nation lack the resources required to prepare their doctoral students in data-intensive areas (analytics, curation, metadata, Big Data management, text mining, & visualization). Furthermore, it is prohibitively expensive to keep up with the rapid changes in data science. Our proposed, innovative summer program, LEADS-4-NDP, presents a cost-effective way to address these challenges and leverage the data science expertise of Drexel's doctoral-level data science track, and enhance LIS doctoral education across the country.

2. Project Design: LEADS-4-NDP will provide intensive exposure to the most recent data science advancements. Each summer, 9 doctoral students will participate in an on-site data science camp at Drexel, and have a first-hand learning experiences working with an NDP project. LEADS-4-NDP outcomes include:
• Two national cohorts of IMLS LEADS Fellows, from ALA accredited LIS programs, with data science expertise.
• A sustainable educative model that can be expanded to include additional NDP partners and LIS programs.
• Course materials and tool sets accessible to LIS educators, including NDP targeted lesson plans.
To achieve these outcomes, the project will be organized into four phrases:

Phase 1-Curriculum development & Recruitment *(November 2017-April 2018)*. Project PIs prepare LEADS-4-NDP curriculum modules covering: 1) Introduction to data science, 2) Digital curation and metadata for data science, 3) Analytical algorithms and data analytics systems, 4) Data Visualization, 5) NDP projects and services, and 6) Data science LIS lesson plans. NDP partners will contribute to the curriculum design, including case studies and internship projects. Recruitment materials will be developed collaboratively with NDP partners, reviewed by the advisory board, and distributed nation-wide.

Phase 2-LEADS-4-NDP Summer Program 1 *(May 2018-September 2018)*. *May/early summer:* LEADS-4-NDP students (fellows) will be invited to Blackboard (an online learning platform) and complete a short pre-camp course. Students will then participate in a 3-day intensive data science/NDP bootcamp at Drexel, and meet project mentors. *Summer:* Students will complete a 10-week summer internship. Most internships will be virtual, following the RDA (Research Data Alliance) and DataONE datanet model. (Lead PI, Greenberg, has had substantive involvement with both efforts.) During the internship period, Drexel faculty will mentor and guide the cohort, with periodical check-ins. Project PIs will also assess the boot camp and summer experience at strategic intervals, utilizing the IMLS evaluation framework.[3] The assessment approach will include surveys, interviews, and focus groups with students and NDP project mentors.

Phase 3-LEADS-4-NDP Summer Program 2 *(Oct. 2018-August 2019).* PIs will review project assessments and revise the LEADS-4-NDP program accordingly for a second iteration. Students will again be recruited and participate in a similar, improved LEADS-4-NDP summer camp and NDP internships. During this phase, participants from Summer Program 1 will be invited t ointeract with Summer 2 students and encouraged to disseminate results and learning experiences via ALA, CNI, DLF, and other relevant venues.

Phase 4-A sustainable LEADS-4-NDP model *(Oct. 2018-Oct. 2019).* Throughout year two, we will seek advisory board input, and develop the 'LEADS-4-NDP sustainable model' that is transferable to other LIS programs and NDP partners. We will standardize and publish all the digital products (curriculum, lesson plans, internship model, and assessment data), and enable access for the national and international LIS community. We will also share our experience in relevant LIS conferences and journals.

3. Diversity Plan: The LEADS-4-NDP has a three-pronged diversity plan: 1) Targeted recruitment of doctoral students from under-represented populations. We will work through the ALA Office for Diversity, Literacy, and Outreach Services to target doctoral students from under represented populations. Drexel's Office of Equality and Diversity will also offer guidance for our recruitment approach. 2) LEADS-4-NDP student showcase (virtual and in-person) and a career goals roundtable at the Dornsife Center, West Philadelphia Obama Administration Promise Zone. 3) The NDP partnership—Mapping Inequality project, and data science lesson plan development documenting minority issues.

4. Project Team: Lead PIs: Drs. Jane Greenberg and Xia Lin, Dept. of Information Science, Drexel University (IS/Drexel). Greenberg has expertise in digital data and knowledge organization; she chaired Drexel's IS doctoral program revision incorporating data science; Lin has expertise in data analytics, information visualization, and digital libraries. **Co-PIs:** IS/Drexel faculty, Dr. Il-Yeol Song has database design expertise and directs the UG data science program; Dr. Weimao Ke has expertise in information retrieval, data cleansing, and visualization. Lead PIs will manage the program, and all PIs will develop and deliver the curriculum and work with NDP partners. **Confirmed NDP partners/mentors:** Biodiversity Heritage Library (BHL), Smithsonian (Martin Kalfatovic), California Digital Library (CDL) (John Kunze), Digital Curation Innovation Center (DCIC), Univ. of MD (Michael Kurtz), Digital Public Library of America (DPLA) (Emily Gore), Historical Society of Pennsylvania (HSP) (Heather Willever-Farr, John Houser), OCLC (Roy Tennant), Philadelphia Free Public Library (John Meier), U.Penn Libraries (Laurie Allen). **Confirmed advisory board members:** Christopher Erdmann (Chief Strategist for Research Collaboration, NCSU Libraries), Margaret L. Hedstrom (Prof., School of Information, Univ. of Michigan), Cliff Lynch (Exec. Dir., Coalition for Networked Information), Richard Marciano (Dir., DCIC & Prof., College of Info. Studies, Univ. of MD), Erik Mitchell (Assoc. CIO, Univ. Libraries, UC Berkeley & Sr. Fellow, Berkeley Inst. for Data Science).

5. National Impact and Sustainability: LEADS-4-NDP aligns with IMLS/LB21 goals and the NDP's need for a digital library workforce having an expanded skill-set[4]. This program will advance doctoral student learning capacity, knowledge, and skill on a national scale via cross-training, hands-on experience, and collaboration. LEADS-4-NDP will have national impact in four key areas: 1) enhancing current LIS doctoral curricula, 2) forming a national cohort of of IMLS LEADS Fellows, as future LIS faculty, who will bring data science knowledge and skills to LIS education on a national scale, 3) building an open educational infrastructure (course materials, lesson plans and shared datasets), and 4) developing a sustainable, educative model encompassing diversity, and which can be expanded to include additional NDP partners and LIS programs.

7. Budget. Total budget request to IMLS: $313,344.00, with total budget of $466,882.00. The 1:1 cost share less student support costs will be met with Drexel commitment of $153,553.

---

[1] http://doi.org/10.3233/EFI-160977.

[2] https://github.com/ryanswanstrom/awesome-datascience-colleges/blob/master/data_science_colleges.csv.

[3] https://www.imls.gov/performance-measure-statements-and-information-learning-and-community-projects.

[4] https://www.imls.gov/publications/imls-focus-summary-report-national-digital-platform.