

Developing a Computational Framework for Library and Archival Education

The University of Maryland seeks \$81.5K to hold a major planning symposium of library and archival educators and technologists. This symposium will identify the foundation / building blocks for an integrated library and archival Master’s educational curriculum centered on computational treatments of large complex collections. This is designed to prepare the next generation of librarians and archivists to meet the evolving needs of professionals working with digital collections. This symposium will be collocated with the March 2019 iConference meeting to be hosted at the University of Maryland.

1. Statement of Broad Need

The last two years have seen the emergence of the concept of “Collections as Data” in cultural heritage institutions:

“Combined with an increasing flow of born- digital items, digital library collections have come to represent a rich community resource for users... Yet a focus on replicating traditional ways of interacting with collections in a digital space does not meet the needs of the researcher, the student, the journalist, and others who would like to leverage computational methods and tools to treat digital library collections as data.”¹

There is a critical need to extend iSchool curricula to prepare students to manage these digital collections in order to address future needs of these researchers and others. The IMLS has funded a number of the “Collections as Data” related projects, addressing the three major gaps: the skills gap with practicing librarians and archivists, the management gap with institutional administrators, and the training gap with iSchools in particular. We will focus on the training gap with iSchools by convening the principal leaders of these IMLS-funded initiatives and other key professionals (see Project Design) at a major symposium. We seek a Laura Bush 21st Century Librarian (LB21) National Forum Grant under the National Digital Platform funding priority to develop the building blocks of an integrated library and archival educational curriculum to educate the next generation of librarians and archivists in the computational treatments of collections.

2. Project Design

For the last two years, the UMD iSchool Digital Curation Innovation Center (DCIC) has pursued a “Computational Archival Science” training and teaching agenda. We have created innovative classes that blend elements of archival and computational thinking, in project-oriented hands-on team-based environments², including the launch of a first graduate seminar in computational archival science in fall 2017. We are seeking a 12-month grant in order to design a draft integrated library and archival educational curriculum centered on computational treatments of large complex collections. We will present this curriculum at a two-day “Collections as Data” symposium at the U. of Maryland, which we will host in March 2019 co-located with the March 2019 iConference³. The invited participants will share their insights and also critique and refine the curriculum. Finally, we will publish and disseminate the vetted findings.

Richard Marciano, professor at the University of Maryland iSchool, director of the DCIC, and lead author of “Archival Records and Training in the Age of Big Data”⁴, will serve as PI. In addition, the following UMD iSchool staff will serve as Co-Investigators: **Bill Underwood**, Affiliate Faculty, **Michael Kurtz**, DCIC Assoc. Director, **Greg Jansen**, Principal Software Architect, **Adam Kriesberg**, Archives Lecturer, and **Will Thomas**, doctoral student. We will invite four categories of attendees at the symposium:

¹ Padilla, T, et al.: Always Already Computational: Library Collections as Data: <https://www.imls.gov/grants/awarded/ig-73-16-0096-16>

² Marciano, et al.: “Practical Digital Curation Skills for Archivists in the 21st Century”, presented at MARAC 2016.

³ The iConference has grown to 600 participants from 200 universities in 33 countries. It provides an interdisciplinary forum to share research on where technology and information science are heading. See: <http://ischools.org/the-iconeference/about-the-iconeference/iconeference-preview/>

⁴ Marciano, R. et al.: *Archival Records and Training in the Age of Big Data*, Chapter 9 of Re-envisioning the MLS: Perspectives on the Future of Library and Information Science Education Advances in Librarianship, Volume 44B, pp. 179–199, May 2018.

- I. iConference attendees actively contributing to the Collections as Data space, such as: (a) **Jane Greenberg** from *Drexel University*, with her IMLS-funded Library Education and Data Science – LEADS project for doctoral training, (b) **Katie Shilton** at the UMD iSchool, with her the NSF-funded PERVADE project on big data ethics, (c) **Greg Jansen** at the UMD iSchool, co-PI on the IMLS-funded DRAS-TIC Fedora project on collection infrastructure scalability.
- II. Other key US thinkers in the emerging field of Collections as Data, including: (a) **Thomas Padilla** at the *University of California Santa Barbara* on the seminal IMLS-funded project “Always Already Computational: Library Collections as Data”, (b) **John Chodacki** at the *University of California Curation Center* on the IMLS-funded “Expanding Library Carpentry” project, (c) **Chris Erdmann** at the *NCSU Libraries* part of an IMLS-funded grant headed by researchers at the University of Pittsburgh, called “The Data Scientist as the 21st Century Librarian?,” and (d) **Ceilyn Boyd, Jessica Farrell** at the *Harvard Library*, who recently hosted an unconference on Computational Archival Science in December 2017.
- III. A small set of international participants with unique expertise and knowledge that would be beneficial to the conversation and help create positive impact for US libraries: (a) **Mark Hedges**, from *King’s College London* – Chair of the last two Computational Archival Science workshops and the last three Big Humanities Data workshops all part of the IEEE Big Data Conference series, (b) **Vicki Lemieux**, from the *University of British Columbia, Canada* – Editor of *Building Trust in Information* and expert in provenance and blockchain technology for trustworthy digital repositories, and (c) **Yoichi Tomiura**, from *Kyushu University, Japan* – Dean of the University Library and expert in computational linguistics in the Library Science and Informatics department.
- IV. Prominent Washington D.C. computational and library collection leaders: (a) **Kate Zwaard** with the *Library of Congress*, Director of Digital Strategy and founder of the National Digital Initiatives “Collections as Data,” (b) **Paul Wester** with the *National Agricultural Library*, (c) experts from the *Smithsonian Institution*, including from the National Museum of Natural History – **Eric Schuettpelz**, the Office of the CIO -- staff, and National Museum of American History – **Bob Horton**, and (d) large record collections experts from the *National Archives and Records Administration (NARA)* -- **TBD**.

3. Diversity Plan

The DCIC is engaged with seminal collections related to social justice, human rights, and cultural heritage themes, including: community displacement, racial zoning, refugee narratives, citizen internment, and the legacy of slavery. In the context of exploring Collections as Data we are developing best practices for working ethically with the underlying communities those collections document as well as considering value-sensitive design. This reflects the diverse student body working with us on these projects.

4. Broad Impact

The symposium will help identify the building blocks of an integrated library and archival educational curriculum centered on computational treatments of large complex collections. We expect this will lead to the submission of a grant to implement and refine the designed curriculum. This would be tested at UMD’s iSchool in collaboration with other interested colleagues, with the goal of creating a shared, virtual, reusable, and modular educational resource for the library and archival education community at large. Our approach would include identifying canonical collections, analytical methodologies, and open source software applications, and developing end-to-end repurposable lesson plans to be disseminated and shared with the community at large.

5. Budget

We request from IMLS a total budget of \$81,542 over a 12-month period (October 1, 2018 to September 30, 2019). This includes \$59,987 in direct costs and \$21,555 in indirect costs calculated at UMD’s negotiated rate of 38.5%. Direct costs include stipends and travel for US and International experts (\$17.5K for travel, lodging, and registration); \$4K for catering symposium events over 2 days for approximately 30 participants; \$24,990 for UMD staff salaries with \$7,497 for fringe, and \$6K for conference travel for dissemination.