

MIT Libraries
Annual Report FY 2016

Program on Information Science
by Micah Altman, Director of Research

The Libraries' pioneering decision, over fifteen years ago, to develop a general research program established its reputation for information science research and led to the development of tools and approaches used by many memory institutions and individual researchers. The *Program on Information Science* continues this tradition of applied research in information management, scholarly communication, dissemination, and preservation.

Goals

The current *Program on Information Science* was initiated as part of the Libraries general research efforts in 2012. The program has five overarching goals:

- Solve emerging research problems in information acquisition, protection, management, dissemination, communication, and stewardship that are essential to support future innovative services.
- Collaborate with faculty to develop innovative approaches to research information and research outputs.
- Future-proof planning, by incorporating the best-of-class research findings in information science, data science, and other fields into our policies, planning and practices.
- Identify, gain recognition for, and generalize the innovations that the Institute and MIT community is making to solve their own problems or to advance the information commons.
- Amplify the impact that MIT can have on the development of information science, information policy, and scholarly communication through participation in the development of standards, policy, and methods related to information science and information management,

Research Portfolio

The research program has established a portfolio in several areas of ongoing research including:

- **Data management and Privacy,**
 - This year the Program completed a collaborative project with the *United States Census Bureau* to convene a series of workshops to examine the challenges and prospects of integrating Big Data into statistical agencies. These workshops, which are funded by the Census Bureau and brought together select experts in universities, industry, and the U.S. government to apply diverse approaches to big data to use cases regularly encountered by statistical agencies. The output of this work included in three use descriptions, three white papers, and seventeen presentations – all available from the program website. These reports have led to several ongoing engagements with Census researchers and researchers at MIT and Harvard.
 - The program continues its collaboration with Harvard University and Microsoft Research in the “Privacy Tools for Research Data” project.. Which aims to develop a better understanding of the practical performance and usability of a variety of algorithms for analyzing and sharing privacy-sensitive data. This was recently awarded a research supplement to support bringing additional students and trainees. This year the project published several papers and software commentaries.

- *Data Transparency, Collection, Participation, and Reproducibility*
 - This year Program completed joint research with the *Harvard Berkman Center for Internet*, funded by an award from *Microsoft & The Berkeley Center for Law & Technology* and, culminating in an open government data symposium and special journal issue focused on recent initiatives to release large datasets of government information to the public for analysis and reuse. This resulted in the publication of two articles: . *Elements of a New Ethical and Regulatory Framework for Big Data Research*; and *Towards a Modern Approach to Privacy-Aware Government Data*; numerous invited talks; and invited testimony.
 - The program develops its collaborate with the University of Florida and ITAM on the Public Mapping Project, which uses crowd-sourced GIS information to stimulate public participation in electoral mapping and to advance research in related areas of political science, law, and public administration. As a result of this research, an article on *Open Data, Political Representation and Redistricting in Mexico* is in press, and others are a review. This work also received recognition as a finalist for the *Brown Democracy Medal*.

- *Enhancing Teaching and Learning*

- NSF continues its collaboration with TERC, and Landmark College to study how big data in education can be augmented with measurements of student attention, and how that big educational data may be managed and shared. This is funded by a \$1M+ award from the National Science Foundation. The project is using sensors to measure attention, combine this with clickstream data from online educational systems, and utilize this to provide foundational knowledge of the ways in which measures of implicit learning might be linked to explicit learning to develop educationally relevant games that are adaptive to diverse learners. This year, the project released its first working paper, and presented posters at three professional conferences.
- *Scholarly Communications*
 - Through participation on the board of ORCID, the program is actively engaging with the emerging development of new standards and practices in researcher identification and evaluation.

Outreach, Dissemination, Recognition, & Engagement

Outreach to MIT Community

In FY16 the Program Contributed to a number of initiatives supporting the MIT community:

- The program, on behalf of the library, worked with the Office of the Provost, and Information Services and Technology are working together, in coordination with OSP, ODGE, the Office of the Registrar, Human Resources, and the Office of the VPR, to implement ORCID identifiers at MIT. ORCID identifiers and creation workflow, and with MIT ID and profile systems. Over the year ORCID identifiers were successfully assigned to every member of the faculty, research staff, postdoc, and graduate student.
- The program hosted a monthly “brown-bag” seminar series cover a broad range of developments in information science relevant to libraries and scholarly communications, which included visiting presenters from Georgetown University, Simmons College, and Faculty of 1000.
- The program delivered several IAP courses in bibliometrics, grant writing, and confidential information management. This attracted students, postdocs, and research staff representing many departments across the Institute and beyond.
- The program worked with JPAL to develop new educational content for IAP and EdX courses in managing confidential data management. This work is be integrated into an upcoming EdX courses for field research methods, and related micromaster’s program.

Broad Dissemination

In FY16 the program's external outreach and dissemination activities include:

- Publishing five professional articles and related to research conducted by the program. This included articles in *Code4Lib*, *Politics y Gobierno*, *The Berkeley Law and Technology Review*, *Washington and Lee Law Review* and a chapter in an edited volume published by *Cambridge University Press*. Three additional white papers were published,, and an additional six working papers have been released (through SSRN and the program website) for comment prior to revision. These outputs both represent significant dissemination and recognition, and reflect broad collaboration with a broad set of faculty, researchers, and practitioners
- Delivering over a half dozen invited talks at, national, and international forums. These talks included delivering a keynote plenary at the University of Rhode Island SLIS; presentations at the Beijing Institute of Technology; and the Future of Privacy Forum.
- With research collaborators, the program engaged in policy commentary and standards development – submitting comments on the proposed changes to human subject regulation (the common rule); collaborating on NISO principles for library patron information privacy; and testifying on information privacy for the NCHVS.

Recognition

In FY2016 the program continued to receive national recognition for its activities:

- The Research Director joined the NMC Higher Education report expert panel for the first time, participated in the NMC Library report expert panel, and was included in *Who's Who in America* and in *Who's Who in American Politics* (for the 1rst time)
- This year the director continues to hold nominated/elected positions to a number of community organizations including: the National Digital Stewardship Alliance; the ORCID board; the Syracuse Qualitative Data Archive technical advisory board the Force 11 advisory board; and the Tides Foundation awards committee.

Organizational Engagement

And, the Libraries' research program engaged with a number of key organizations that are developing or guiding broad standards, infrastructures and community effort.

- The program serves as MIT's institutional representative to NISO. In that capacity it has participated in a Mellon/NISO project to develop principles for protecting patron privacy in libraries and library services – the results of which were released as a set of NISO Privacy Principles.
- As library representative to the *National Digital Stewardship Alliance*, and chair of the Coordination Committee, the program helped the NDSA in its transition to a new host. This year, the NDSA successfully requested proposals, selected a host organization, and transitioned to the *Digital Library Federation* as a host organization.
- The program serves as MIT's representative to the ORCID organization. And the Director was elected last year to serve on the executive committee of the ORCID board. As a part of the board and of the executive committee, the Libraries' research program continues MIT's contribution to the direction of Open Researcher & Contributor ID.
- The program serves as MIT's representative to the Digital Preservation Network (DPN), and the director sits on the policy and business steering group. This summer, DPN successfully released its first service offering.

Future Outlook

The subvention of the Libraries research program makes up a very small percentage of the library budget (typically less than 1% of the total) this subvention supports the director, quarter-time administrative assistance, and virtual research assistant assistance. The subvention allows research to be applied strategically, and to support technology and knowledge transfer from research into organization practice.

This year, the Program collaborated with a broad range of internal and external faculty and researchers (including faculty in MIT CSAIL, the MIT Media Lab, MIT Sloan School, MIT Materials Sciences, the MIT Political Science Department, Harvard University, TERC, Landmark college, U. Fl., the US Census, and ITAM) to conduct sponsored research and to develop and submit proposals for sponsored activities.

During this FY seven proposals were submitted to the Knight Foundation, NSF, CLIR; Sloan Foundation, and IMLS. One proposal with Harvard to enhance the NSF Privacytools work was funded, and will be used to support a summer student. A proposal from CLIR was funded to subvene a postdoctoral fellow in software curation. Three proposals remained under review at the end of the fiscal year.

This year, the Program successfully hosted five volunteer graduate interns to increase capacity. These interns have worked with the Program to develop computer models for preservation simulation; to review guidance on professional reputation management and alt metrics; to develop an environmental scan of software repositories for reuse and preservation; and to develop a scan of emerging virtual reality products and potential applications in library. In FY17 we will host a

undergraduate summer student in information privacy, and a postdoctoral fellow in software curation – and we plan to recruit two additional postdoctoral fellows supported by current grants.

Over the next five years, the program expects to grow its research portfolio through internal and external collaborations with library staff, faculty, partner institutions, and funders. This expanded portfolio aims to establish a Center for research in scholarly communications that would host projects in many of the following areas, as well as others: revolutionizing research information discovery; managing and disseminating "big" data, developing new methods for digital preservation; reinventing scholarly communication; developing approaches to library spaces that facilitate teaching and learning; and supporting information collection, and management and access related to online educational content. We expect to expand the research program to employ and engage additional postdocs and research staff.