MIT Libraries
Annual Report FY 2015

Program on Information Science
by Micah Altman, Director of Research

The Libraries' pioneering decision, over a decade 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, 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 problems in information management that are essential to support new and innovative services.
- Generate collaboration with faculty to develop innovative approaches to managing research data 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 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 Protection*
  - The United States Census Bureau awarded the program over a hundred thousand to collaborate on convening 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 hosted at MIT, will bring 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 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. The project is developing secure implementations of these algorithms, schemas, and legal instruments, which will be made publicly available and used to enable wider access to privacy- sensitive data sets. This work is partially funded by the National Science Foundation.

- o The program. in collaboration with the *National Digital Stewardship Alliance* . has engaged in a range of digital projects focused on digital preservation. *This* year, the 2015 edition of the *National Agenda for Digital Stewardship* was released, -- which integrates the perspective of dozens of experts and hundreds of institutions identify the highest-impact opportunities to advance the state of the art; the state of practice; and the state of collaboration within the next 3-5 years. For this publication, we will receive a commendation from the Society of American Archivists, at their annual meeting.

- • *Data Transparency, Collection, Participation, and Reproducibility*
  - o The Arthur P. Sloan Foundation awarded $868,954 in January, 2015 to a joint MIT-Harvard project developing easy-to-use tools for sharing confidential research data while retaining data privacy to begin in April of this year. **T**he project aims to leverage ongoing multidisciplinary collaborations and theoretical advances in computation, statistics, law, and social science to improve reuse and replicability in empirical social science, promoting archiving of and access to large-scale confidential data.
  - o An information privacy collaboration between the MIT Libraries Program on Information Science and the Harvard Berkman Center for Internet & Society is one of six projects to receive an award of $35,000 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.
  - o The program is continues to 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.
.
- • *Enhancing Teaching and Learning*
  - o NSF made an of over one million dollars to the MIT Libraries Program on Information Science, 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. The project will utilize common webcams 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.
  - o This year, the program received in-kind support from ARL to field a survey of libraries on services related to making. In conjunction with this it. delivered IAP courses on rapid fabrications; and released a website providing guidance on making.

  - *Scholarly Communications*
    - o The program successfully completed it's collaborative work with Harvard and Stanford on a "Data Sharing System for Journals," which developed data-citation and publication workflows and software for Open Access publications. This work was funded by an award from the Sloan Foundation. And resulted in the release of open source software, and in a number of publications and presentations.
    - o Research conducted in collaboration with program research affiliates resulted in the publication of *Information Wants Someone Else to Pay for It*, which identifies the affordances of, and economic barriers to, new forms of digital publication.
    - o 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.
    - o The program is engaged with a Wellcome Trust funded project to develop approaches to describing contributor roles – which was recently releases as part of the CASRAI standard framework, and is in pilot testing at a number of journals.

Outreach, Dissemination & Recognition

In FY!5 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, have now been successfully integrated into MIT ID and profile systems. By the end of this summer, every member of the faculty, research staff, postdoc, and graduate student will be supplied with their own ORCID identifier.
- 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 Harvard University, and the University of Pittsburgh.
- The program delivered four 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 will be integrated into an upcoming EdX courses for field research methods.
- The program worked with the VPR's office and the Postdoctoral Association to prepare a talk on reputation management for early-career scholars. With the assistance of graduate interns recruited by the program, this is being extende and adapted for future IAP courses.

In FY15 the program's external outreach and dissemination activities include:

- Publishing five professional articles and reports related to research conducted by the program. This included the 2015 edition of the National Agenda for Digital Stewardship, a definitive report on researcher identifiers published through OCLC, a book chapter and two journal articles. Three additional journal publications and book chapters are in press, 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 eight invited talks at, national, and international forums. These talks included delivering a keynote plenary at the annual NFAIS conference; and presentations at the United Nations and at the National Academy of Sciences.

In FY2015 the program continued to receive national recognition for its activities:

- The Director joined the NMC Library Horizons report expert panel for the first time, and was included in *Who's Who in America* for the 5th time
- The Society for American Archivists issued a special commendation for the publication of *The 2015 National Agenda for Digital Stewardship.*
- The Director was interviewed by Voice of America and by Digital Trends as an expert on digital preservation.
- 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.

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 server 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 are expected to be released this year.

- 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 technical steering group, 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 National Digital Stewardship Alliance. This year, the director was elected as chair of the Alliance. As NDSA is a collaboration of over 160 public, private, and governmental organizations committed to establish, maintain, and advance the capacity to preserve our nation's digital resources for the benefit of present and future generations.
- 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.
- The program, in collaboration with colleagues at Harvard University, U. Pittsburgh, ITAM, and U. Fl. The program submitted several formal comments on information privacy, and on open access to data, including comments to NIST on proposed information privacy frameworks.

Outlook

The subvention of the Libraries research program makes up a very small percentage of the library budget (typically less that 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 internal and external faculty and researchers (including faculty at CSAIL, Harvard, TERC, Landmark college, U. Fl. and ITAM) to conduct sponsored research and to develop and submit proposals for sponsored activities. During this FY six proposals were submitted to the Sloan Foundation, IMLS, U.S. Census, Microsoft, ARL, and the IBM Center for the Business of Government.   Based on these, and previous pending submissions, the program and collaborators received over five awards, providing over $2M in support for collaborative research projects.

This year, the Program successfully hosted six volunteer graduate interns to increase capacity. These interns have worked with the Program to develop computer models for preservation simulation; to develop reviews of approaches to early-career reputation management and university-wide information management; to identify the affordance of and economic barriers to new forms of scholarly publications; to field a survey of library practices related to 3-d printing services; and to develop a guide to online resources for "makers."

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 include 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.