MIT Libraries
Annual Report FY 2014

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:

• Open access publishing
  The program is collaborating with Harvard and Stanford on a “Data Sharing System for Journals,” which aims to develop data-citation and publication workflows for Open Access publications. This work is funded by an award from the Sloan Foundation.
• Data management
  o This year, the program began collaboration with BigData@CSAIL in research to understand the privacy implications of big data.
  o The program participated in a invitation-only Ideaslab to explore the potential for big data to enhance education. This lead to a collaborative proposal with TERC and Landmark college to NSF to study patterns of attention in online learning systems.
  o The program has collaborated with faculty in IMES, CSAIL and the Harvard Medical School in developing two proposals to the NIH Big Data to Knowledge program, focusing on research in data management and curation for biomedical big data.
- The program continues to collaborate 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.

- **Crowd-sourced information Collection and Management**
  - This year, the program launched pilot research into managing information and services related to 3-d fabrication in libraries. This has resulted in the development of an IAP course, and the recruitment of a volunteer graduate intern.
  - The program is continues to collaborate with George Mason University 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. This work has been partially funded by the Tides foundation.

- **Digital Preservation**
  - The program in collaboration with the National Digital Stewardship Alliance has engaged in a range of digital projects focused on digital preservation.
  - **National Agenda for Digital Stewardship**, first released in July of 2013, integrates the perspective of dozens of experts and hundreds of institutions, convened through the Library of Congress, to 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. The document highlights emerging technological trends, identifies gaps in digital stewardship capacity and provides funders and decision-makers with insight into the work needed to ensure that today’s valuable digital content remains accessible and comprehensible in the future, supporting a thriving economy, a robust democracy, and a rich cultural heritage. Thus far, even its first edition, the Agenda, published last year, has attracted unusual attention from the digital preservation community. Representatives of a wide range of funding agencies and foundations commented on the publication, and are providing review into the next version. Workshops and poster sessions on the Agenda presented at CNI, IDCC, ALA have brought discussion about a range of digital preservation to new audiences – and further dissemination efforts are planned for data curators who attend IASSIST and higher education IT leadership at NERCOMP.
  - The program collaborated in fielding the 2nd Preservation Storage Survey. Which collected information from over a 100 memory organizations on digital preservation capacity & practices.

- **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.
  - The program is engaged with a Wellcome Trust funded project to develop approaches to describing contributor roles – which recently entered the process required to become an official recommendation of NISO and CASRAI.
  - Through CODATA and FORCE11, the program participated in the development of widely recognized Joint Principles for Data Citation. These were released in January, and have been endorsed by dozens of scholarly organizations.

**Outreach, Dissemination & Recognition**
In FY14 the program’s outreach and dissemination activities include:

- Publishing six professional articles and report related to research conducted by the program. This included a publication in *Nature* on contributor roles, a book-length report on Data Citation published as a special volume of *Data Science*, and the inaugural *National Agenda for Digital Stewardship*. These outputs both represent significant dissemination and recognition, and reflect broad collaboration with a broad set of faculty, researchers, and practitioners.
- Delivering over a dozen invited talks at local workshops, national, and international forums. These talks included an invitation to a new NSF IdeasLab workshop, which was extremely competitive; and presentation on issues related to digital preservation, scholarly communication, and information privacy.
- Running 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 *Digital Science, Harvard University*, and the *US Census Bureau*.
- Developing new IAP courses in bibliometric research and analysis, and co-developed an IAP course in 3-D printing. And teaching four IAP courses in bibliometrics, grant writing, and data management. This attracted students, postdocs, and research staff representing many departments across the Institute and beyond.

In 2014 the program continued to receive national recognition for its research:

- The Director was listed in *Who’s Who in America* for the 4th time.
- For the first time, the Director was listed in *Who’s Who in American Politics*.
- The director was interviewed by Educause -- appearing on the cover of the *Educause Review Online*; and was quoted on an article on information policy for *Forbes*.
- This year the director was nominated/elected to a number of community organizations including:
  - Election as chair of the National Digital Stewardship Alliance;
  - election to the executive committee of the ORCID board;
  - nomination to the Digital Preservation Network business working group;
  - nomination to the Syracuse Qualitative Data Archive technical advisory board;
  - nomination to the Force 11 advisory board;
  - nomination to the National Academy of Sciences BRDI awards committee;
  - nomination to the Tides Foundation awards committee.

**External Engagement**

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

- The program started to serve this year as MIT’s institutional representative to NISO. In that capacity it submitted formal comments on draft standards for Open Access metadata.
- The program, in collaboration with Harvard University submitted several formal comments on information privacy, including comments to OSHA on proposed rule-making; comments to the white house for the PCAST report on Big data and privacy, and comments to the FTC on managing information for mobile data privacy.
- The program serves as MIT’s representative to the ORCID organization. This year, the
Director was elected 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. ORCID, Inc. is a non-profit organization dedicated to solving the name ambiguity problem in scholarly research; and it brings together the leaders of influential universities, funding societies, publishers and corporations from around the globe.

- 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 150 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.

- As a part of the CODATA/ICTSI/BRDI task group on data citation standards and practices, the Libraries research program is contributing to the development of data citation standards. A major report from this group was released as a special volume of the CODATA journal. This was followed by a joint effort with FORCE11 to develop a joint statement of data-citation principles, which was subsequently adopted by dozens of scholarly organizations.

- As part of an OCLC research task group, the Libraries research program is contributing to analysis of the benefits, needs, and challenges for integrated author and research identification.

**Capacity Building**

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 internal and external faculty and researchers (including faculty at CSAIL, IMES, Harvard, TERC, Landmark college, and the UK STFC, and the NAS) to develop and submit proposals for sponsored projects to NSF, NIH, CODATA, IMLS, and the EU Cooperation in Science and Technology (COST) agency.

This year, the Program successfully recruited two volunteer graduate interns to increase capacity. These interns have worked with the Program to develop computer models for preservation simulation; and to develop an environmental scan of library practices related to 3-d printing services and related data management.

**Outlook**

Over the next several years, we expect 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: managing and disseminating "big" data, developing new methods for digital preservation, and supporting information collection, and management and access related to online courses, as well as initiatives in other areas. In support of these activities, we expect to expand the research program to employ and engage additional postdocs and research staff.