Draft Recommendations of the MIT Ad Hoc Faculty Task Force on Open Access[1] to MIT's Research

March 16, 2019

Introduction

"Openness increases transparency and reliability, facilitates more effective collaboration, accelerates the pace of discovery, and fosters broader and more equitable access to scientific knowledge and to the research process itself."
-National Academies of Sciences, Engineering, and Medicine[2]

The open sharing of research outputs promises to quicken the accumulation of knowledge and insight and enhance opportunities for collaboration. It also aligns with MIT's mission. At MIT, we are "committed to generating, disseminating, and preserving knowledge, and to working with others to bring this knowledge to bear on the world's great challenges."[3] We currently manifest that mission via the open sharing of educational materials through OCW and MITx, and by openly sharing faculty research via the MIT Faculty Open Access Policy. In addition, as MIT makes bold moves to address the challenges and opportunities presented by the prevalence of computing and the rapid advances in artificial intelligence, our efforts in these areas will depend on the open availability of large, diverse, and inclusive sets of data in all formats.

The Task Force on Open Access to MIT's Research has been charged with developing recommendations to further support and enhance the open sharing of MIT research and educational materials and to contribute to the global transition to open science. Recommended as part of the 2016 report from the Institute-wide Task Force on the Future of Libraries, the task force is intended to address the large proportions of MIT's research and teaching outputs that are not yet available for open dissemination. This includes the vast majority of faculty journal articles published before the adoption of the Faculty Open Access Policy in 2009, and over 50% of faculty articles published since then.[4]

These bold, vital aims must, however, be considered in the context of complex changes in distribution and publication processes, as they evolve to harness the potential of the digital age to enhance and facilitate the sharing of science and scholarship so that research output can have maximum impact. We offer these recommendations amid signs of growing pains in this transition: at a time when proprietary

---

[1] For our working definitions of open access for publications, code, data, and educational materials, please see https://open-access.mit.edu/about-open-access

[2] National Academies of Sciences, Engineering, and Medicine. 2018. Open Science by Design: Realizing a Vision for 21st Century Research. Washington, DC: The National Academies Press. https://doi.org/10.17226/25116.

[3] MIT Facts, Mission: http://mit.edu/about/#mission

[4] Institute-wide Task Force on the Future of Libraries: Preliminary Report. https://future-of-libraries.mit.edu/sites/default/files/FutureLibraries-PrelimReport-Final.pdf

and open systems and services for sharing data, code, and all forms of publication are proliferating; when the economic models for these new approaches are still being developed, debated, and tested; and when practices and policies around openness vary in different parts of the globe. In this time of transition, many publishers are struggling to implement successful open access business models and to meet new requirements from public and private research funders for more open access to scholarly articles and data. Researchers stretch to simultaneously act upon their wish to share their work broadly while meeting expectations for the kind of publication and credentialing that will advance their careers; some—such as a system-wide group at the University of California[5] — are leading bold initiatives to assert their principles regarding the scholarly communication system and insist that publishers manifest them.

New opportunities for data sharing are accompanied by new concerns for appropriate protection of data, including private or sensitive information, datasets created with substantial resource and career investment, and those built from proprietary sources. While new capabilities for sharing data provide the opportunity to support robust validation and replication of research—a core aim—broad sharing also raises questions about how to maintain appropriate levels of privacy for sensitive human subject data, and appropriate security for other types of sensitive, classified, or proprietary data. With these complex questions and tensions as yet unresolved, data continues to proliferate, as do new forms of scholarly publications, including public scholarship (e.g. blogs, social media engagement, and podcasts) and interactive, multimedia publications.

While these changes are occurring at an uneven pace across disciplines, it is abundantly clear that access to large datasets is increasingly critical to research work and scientific advancement. This can be seen in large efforts such as MIT's Quest for Intelligence and in the creation of the Stephen A. Schwarzman College of Computing as well as in specific research projects, such as the application of artificial intelligence to improve early breast cancer detection or the use of machine learning to identify patterns in materials "recipes" from large corpora of journal articles. The increasing importance and reliance on big data is accompanied by new risks and concerns. These data can also be used to craft unjust societal policies or be leveraged by nations to surveil their populace. As we seek to make data more accessible, we must be aware of these dangers and of the possibility of non-cooperative sharing behavior as we forge partnerships, choose our research directions, and develop the infrastructure to support data sharing.

These dangers are receiving attention at the highest levels of US government today. There is a widespread and bipartisan sense of unresolved tension between a belief in and push for openness of research outputs (especially data, including a recent Executive Order on Artificial Intelligence) and a desire to protect the US's leadership in technology and AI research by protecting access to our data. At the federal level, the difficulty in reconciling these opposing objectives is leading to movement in both directions—open and closed—at once.

Recognizing the core challenges and tensions in this environment, our recommendations are offered in the spirit of advancing MIT's mission to share its scholarly work as openly and widely as possible, while remaining grounded in the complex realities of the current landscape. We likewise were guided by our collective understanding of and commitment to scientific and scholarly ethics and best practices. Our

---

[5] The faculty-driven process has included the University of California's Systemwide Library and Scholarly Information Advisory Committee (SLASIAC), in partnership with the university libraries and the system-wide academic senate's Committee on Library and Scholarly Communication (UCOLASC).

recommendations reflect the principle that openness is not only compatible with, but can and should be built on, the foundational value of academic freedom and a belief in the importance of scholarly attribution. Our approach to openness also seeks to respect and accommodate relevant disciplinary differences.

The Open Access Task Force has been guided by MIT's vision: that science and knowledge progress more quickly and can more readily be applied to solving the world's biggest challenges when shared openly. In our approach, we have aimed to reflect President L. Rafael Reif's vision for working on global challenges related to growth, innovation, cooperation, and sharing. He has made clear his view that in defending "fair international competition and America's strategic and commercial interests," we must resist the urge "to try to double-lock all our doors."[6]

Rather, we should double down on responsible ways to manifest MIT's foundational belief in the value of open sharing. This is the aim of our recommendations.

Statement of principles

We recommend that the Faculty Policy Committee, in close coordination with the Faculty Committee on the Library System and representatives from this task force, review and ratify a set of principles for open science and open scholarship. These principles should provide guidance for individual faculty, staff, and students in making intentional decisions about communicating their work and participating in scholarly communications activities such as editing and reviewing. They should also be broadly useful in providing the basis for Institute-wide negotiations with publishers and others who provide services and tools in support of scholarly communications, including MIT Libraries' negotiations with commercial publishers.

By ratifying a set of shared principles for open science and open scholarship, the MIT community affirms that control of scholarly communications should reside with scholars and their institutions.

Such principles might include:

1. Scholarly authors should retain copyright in their own work and full rights to reuse their work.
2. Scholarly outputs should be openly available to readers everywhere, regardless of institutional affiliation or individual ability to pay.
3. Data, code, and other supporting materials necessary to validate and/or replicate scholarly work should be openly available.
4. Scholarly work should be openly available to computational analysis, and to algorithmic and machine learning applications and uses.
5. The full life cycle of research should be part of the scholarly record, and therefore scholars should have the right to openly share early versions of articles and other publications in open preprint servers, institutional repositories, and/or open platforms, with no restrictions on subsequent publication choices.

---

[6] L. Rafael Reif, "China's Challenge is America's Opportunity," New York Times, August 8, 2018.
https://www.nytimes.com/2018/08/08/opinion/china-technology-trade-united-states.html

Recommendations

The Task Force offers recommendations organized around three strategies for supporting the open dissemination of MIT research and educational outputs:

1. Policy recommendations
2. Infrastructure and resource recommendations
3. Advocacy and awareness recommendations

Policy recommendations

1. Adopt an all-campus open access policy, granting MIT non-exclusive permission to openly disseminate scholarly articles[7] written by any MIT author. Such a policy would apply to scholarly articles written by undergraduate and graduate students, staff, post-doctoral fellows, research scientists, and other MIT community members who produce scholarly research articles while employed and/or enrolled at MIT, and would be modeled on the existing MIT Faculty Open Access Policy. This is not intended to apply to work that students produce solely for courses; it would apply only to scholarly articles that students (undergraduate and graduate) publish in journals, conference proceedings, or other venues.

   ○ This policy should include an option for any MIT author to elect to opt out of the policy for a given scholarly article.
   ○ We suggest that implementation be overseen by the Provost's Office and administered by the Libraries.[8]

2. Adopt an open access policy for monographs, granting MIT non-exclusive permission to openly disseminate digital versions of scholarly monographs[9] written by any MIT scholar. Following the model of the current faculty OA policy, this policy would include a per-monograph opt-out provision. To support MIT authors whose publishers require a subvention to offset publication costs, MIT should establish an Open Monograph Fund.[10] To assist MIT authors who wish to disseminate manuscript versions of their monographs ("green" open access, including cases

---

[7] Scholarly articles are defined as in the MIT Faculty Open Access Policy: Scholarly articles: Faculty's scholarly articles are articles that describe the fruits of their research and that they give to the world for the sake of inquiry and knowledge without expectation of payment. Such articles are typically presented in peer-reviewed scholarly journals and conference proceedings.

[8] Resource implications: If we expand the open access policy to all MIT authors, the pool of articles under an MIT open access policy would more than double. While we are already acquiring a large portion of these through existing workflows, if we applied the systematic and supported-deposit approach we currently take for faculty-authored articles (which we believe to be critical to a successful deposit rate) and applied it to all MIT-authored articles, we would need to expand library staffing by .5 FTE, with an approximate annual cost of $50,000 in salary and benefits.

[9] Monographs are defined as single authored or coauthored scholarly books, excluding edited volumes and textbooks: "A monograph is a specialist work of writing...on a single subject or an aspect of a subject, often by a single author, and usually on a scholarly subject." from: https://en.wikipedia.org/wiki/Monograph

[10] Resource implications: For books, a typical subvention rate with a university press to make a scholarly monograph open access is $15K. Of the approximately 100 books published annually by MIT authors, less than ⅓ would be appropriate and eligible for open access subventions. We recommend an Open Monograph fund not to exceed $500,000 annually.

where a publisher does not offer a platform or business model for offering an open version), this new policy will provide a legal mechanism for such sharing, modeled on the faculty OA policy.

- ○ We recommend that the MIT Libraries administer this policy and develop an outreach program for working with monograph authors and their publishers to maximize the open dissemination of MIT authored scholarly books.
- ○ We recommend that MIT authored open monographs be made available under an appropriate [Creative Commons](#) license.

3. To support MIT's existing commitment to and practice of open sharing of theses, we recommend that the policies regarding holds/embargoes be clarified, consolidated, and published on the [MIT's Policies](#) website. The policy should specify that requests for a hold/embargo of longer than three months must be approved by the Vice President for Research, in consultation with the Technology Licensing Office and/or the Office of Graduate Education, and must be supported by evidence of a pending patent application, a book contract requiring an embargo, and/or evidence of extenuating circumstances related to safety, confidentiality, or national security interests.

   The aim is to support MIT's existing practice of making theses as open as possible by reaffirming and clarifying that thesis holds should not be granted except in exceptional circumstances, and only for short periods of time, and that the holds should require approval by the Vice President for Research.

4. Following the lead of the [Gates Foundation](#), [cOAlition S](#), and other research funders, MIT should consider adding an open access requirement to all existing and new internal MIT research grant programs that establishes immediate open access publishing as an expectation placed upon grant recipients.

   - ○ As an initial implementation of this recommendation, grant recipients of the [Professor Amar G. Bose Research Grants](#) should be required to make all publications and associated original data, code, and other shareable research outputs openly available via either publication in an open access journal or via deposit in a trusted open repository (e.g., [DSpace@MIT](#)) immediately upon publication. (While we intend this new requirement to apply only to new Bose recipients, we recommend that the Libraries work with past awardees to deposit their Bose research outputs into a new Bose collection in MIT's open access repository, DSpace@MIT.)

Infrastructure and resource recommendations

5. Although major fields such as genetics already benefit from established open source standards and repositories for often-used data types such as Single Nucleotide Polymorphisms ([SNPs](#)) in the genetics field, most fields and topics reside in the "long tail" of academic research. In the long tail, open access standards and data repositories have yet to be established. In many fields, fledgling standards and open data repositories are the result of the efforts of individual scholars.

   MIT should support such efforts among its own faculty, students, and staff. Therefore, to support the open sharing of data created and originating at MIT (not pre-existing data or data

acquired from external sources) the Institute should create an Open Data Fund to provide seed funding and resources to form, build, or sustain:

- New data repositories, where lack of one is limiting the ability of MIT scholars to share their data.
- Researcher-led efforts to create discipline-appropriate open databases.[11]

6. To support the long-term sustainability and discoverability of open data, the MIT Libraries should investigate the potential to develop and maintain a data registry, indexing MIT-created data, and linking data to publications, grants, and other outputs and supporting materials.[12]

- Full support for open data also requires infrastructure to support data storage and computation, and wide adoption of open data standards which facilitate data access and reuse. We recommend that appropriate entities at MIT (including the College of Computing and IS&T) work together to develop clarity about roles, responsibilities, and approaches to data across MIT.
- We further recommend that MIT convene a cross-institutional working group on open data to develop best practices and to consider collaborative and coordinated approaches for policy and for long-term access, storage, preservation, and support for open data.

7. The Institute should provide funding for open access initiatives on two critical paths: support for open access work and initiatives spearheaded by MIT scholars; and support for open access infrastructure, such as tools and services that reduce the burden on and/or create incentives for authors in openly sharing their work. Examples of such funding might include:

- An Open Textbook Fund to support the development, adaptation, and use of open textbooks.
- An Open Access Infrastructure Fund to support MIT contributions to collectively funded infrastructure projects such as the Public Access Submission System (PASS), arXiv and other open preprint servers, PubPub, and Open Science Framework.[13]

---

[11] Resource implications: Costs of implementation would vary depending on whether we relied on existing resources, provided enhanced resources to meet peers, or wanted to provide funding sufficient to take a leadership position. The resources required to maintain a currently operating research database may be extensive, and are greater than those for supporting a research repository. Required resources increase with the number of maintained databases. With existing resources we would be relying on DSpace@MIT (MIT Libraries) and Dataverse (Harvard University IQSS). DSpace is limited in its features to support data discovery and reuse. Dataverse offers greater functionality, but is managed by an institutional partner. With enhanced funding, MIT could meet peer efforts such as Purdue University Research Repository (PURR), University of Minnesota Data Repository DRUM, University of California Data Publication & Repository DASH.)

[12] Resource implications: The MIT Libraries can absorb the costs of creating and maintaining a data index in their existing library technology staffing and budget.

[13] Resource implications: Costs of implementation would vary depending on the scale of what was undertaken. A moderate approach could involve: a pilot with a limited pool of funds available to faculty to support them in efforts to build open textbooks (for example, hiring temporary staff to assist in identifying and formatting materials), and some modest support added to the Libraries' budget to expand their contributions to open access infrastructure projects.

8. To increase the impact of open educational content, MIT should:

   ○ Adopt and promote a publication standard (addressing file types, access controls, etc.), such as the standard currently being developed by MITx/edX for educational materials, that ensures open access and interoperability.
   ○ Optimize the design of current and future MIT learning management systems (including Stellar) so that it is easy for creators to make materials as open as possible. One example would be making class materials available by default to all MIT students, not merely those enrolled in the class, while instructors retain the ability to select sharing options other than this default.
   ○ Specify approved open access licenses and licensing procedures for MIT publication of educational materials, recognizing that these materials may incorporate content, code, and interactive elements.
   ○ Provide streamlined paths for publishing all educational material under open licenses on platforms like MIT OpenCourseWare.

9. For the purpose of encouraging more open sharing of code and to reduce the potential negative impact of the proliferation of software patents on entrepreneurship and innovation,[14] MIT should:

   ○ Develop a set of recommended open licenses for software produced by MIT.
   ○ Create and publicize guidelines, policies, and practices for publishing code under open-source licenses.
   ○ Review software licensing practices to ensure they promote innovation, and encourage MIT authors who wish to distribute code openly under popular open-source licenses.

Advocacy and awareness recommendations

10. The Provost should direct each department head to develop a plan to encourage and support their faculty, students, and staff in openly sharing their research, as appropriate for their discipline. The MIT Libraries will work closely with departments to develop these plans. Some sample plan elements are provided below, but each department will have considerable leeway to develop plans appropriate for their discipline.

    Sample plan elements:

    ○ Department-level affirmation and education about ways in which open access can enhance the quality and impact of publications, and about peer-reviewed open access publications and platforms available to communicate faculty and student work.
    ○ Department-level encouragement and support for the use of the wide range of opportunities for sharing work openly at many stages, including preprint servers such as arXiv and bioRxiv and discipline-specific open access repositories such as PubMedCentral.

---

[14] See Gonzalez, A.G. (2006). The Software patent debate. Journal of Intellectual Property & Law Practice, 1(3), 196-206. https://doi.org/10.1093/jiplp/jpi046

- ○ Commitment to providing education to faculty and students about open publishing options and author rights.
- ○ Mechanism for highlighting openly available research and teaching materials in annual reviews and in tenure and promotion packets.
- ○ Mechanism for alerting external reviewers that MIT places high value on open sharing of research for tenure and promotion, and for referring letter writers to the Institute's open repository (currently [DSpace@MIT](#)) to access a candidate's work.
- ○ Encouragement to faculty to ensure all eligible work is deposited in the Institute's open repository as part of the tenure and promotion preparation.

11. MIT should advocate for greater recognition and credit to researchers who share data including making data ready to usefully share. Part of this effort should include developing and adopting metrics for assessing impact of non-traditional research outputs, such as data, software, and educational materials. As one step, we recommend MIT create an annual award to give recognition to MIT scholars for demonstrated dedication and contribution to open science.

12. MIT should continue to consistently advocate with the federal government in support of policies and legislation that advance MIT's commitment to open access to research as part of its mission.[15]

13. We recommend that the [ad hoc committee](#) on international engagements include the topic of open access and open science in their deliberations, including the complex issues and tensions between MIT's commitment to and policies regarding open sharing of research outputs and a desire to maintain intellectual and competitive control.

Conclusion

Our task force was [charged](#) with taking up the question of whether MIT should strengthen its activities in support of providing open access to the research and educational contributions of the MIT community. We were also asked to coordinate an Institute-wide discussion on this topic. We offer these draft recommendations after wide consultation across the Institute, and with the intent of sparking an even wider conversation across the MIT community and other stakeholder communities.

To join that conversation and offer input on these recommendations, please contact us at [openaccesstaskforce@mit.edu](#), submit your ideas to our [Idea Bank](#), comment on the [recommendations on PubPub](#), and/or attend our upcoming open forum on April 10, 3-4:30 pm in 56-114. For updates, go to the [MIT Open Access Task Force](#) website.

---

[15] For example, advocate in support of bills that would codify into law the [White House Directive](#) on Expanding Public access to the Results of Federally Funded Research.