LG-250130-OLS-21, Massachusetts Institute of Technology

Title: Community Tracking Indicators for Open and Inclusive Scholarship

Summary: The Massachusetts Institute of Technology requests $314,443 to establish a set of trusted, open, standardized indicators of openness and inclusion in the scholarly ecosystem.

Statement of National Need: In order to make reliable progress toward a socially-desirable scholarly ecosystem the research community requires ongoing, systematic, and trusted measures of the inclusivity, equity, durability, and sustainability of the system. Recent environmental scans such as the Mellon Foundation supported Grand Challenges Summit and ACRL report on Open and Equitable Scholarly Communications of research and practice have drawn attention to the need to measure and integrate equity and inclusion into the system. There is convincing evidence that scholarly processes and outputs have substantial bias and/or create barriers to inclusion[1] and that more openness in science and scholarly communication is needed.

IMLS's analysis of the National Digital Platform for Libraries (NDP) (Owens et al, 2017, 2018) draws attention to the need for systematic measurement and evaluation of the scholarly ecosystem with a focus that reflects library values including diversity as laid out by the American Library Association and IMLS.[2] This need is not being met by the major players that produce statistics on scholarship. For example, the primary source of national statistics on science and engineering, the NCSES, tracks participation in the workforce by gender and minority status, but does not track participation in scholarly communications. And, while it is routine to use publisher-produced indicators of the citation 'impact' of scholarly communications for research, institutional decision making, and research policy -- there is currently no comparable public data that summarizes diversity in who is citing, producing, or accessing the same communications. And ironically, despite recent advances in making scholarly communication more openly available, there are few systematic measures that can be used to track, compare, or evaluate diversity and inclusion in open scholarship. The project will develop open, reliable, comparable, standardized indicators that will go beyond measures of 'overall impact' to advance the understanding of who is, and who is not, participating in open scholarship.

Project Design: The project will create measurements of scholarly knowledge production in the United States over time to address three questions: What is the prevalence of members of different groups in open-scholarship and open-science initiatives and outputs? Where are open-scholarship and open-science outputs that are produced with and by group members used in the scholarly ecosystem? How does group prevalence in open-scholarship and science and the use of Open Access products vary within the scholarly ecosystem?

Three modes of data collection will be employed. First, as the evidential basis for national-level measures we will deploy automated, repeatable data-science pipelines to retrieve, clean, link, and normalize data from a set of targeted public data sources including: the ORCID annual public data release, I40C open citation database, DOAJ journal database, ROARMAP policy database, PLOS article metrics, and OSF.io preprint metadata. Second, a panel-based design (repeated measures of the same units over time) will provide specific and comparable evidence of changes occurring at the level of individual institutions. Second, we will seek input from diverse voices in the research community to identify key institutional stakeholders in the open science and open access field; and use social media mining and web-mining approaches to extract information about targeted open-access and open-science initiatives. Finally, we will develop additional limited-time sets of community-selected measures based on the approach pioneered by the TESS project and American National

[1] See for example Lee, Carole J., Cassidy R. Sugimoto, Guo Zhang, and Blaise Cronin. "Bias in peer review." Journal of the American Society for Information Science and Technology 64, no. 1 (2013): 2-17.

[2] American Library Association "Core Values of Librarianship" http://www.ala.org/aboutala/governance/policymanual/updatedpolicymanual/section2/40corevalues and IMLDigital Infrastctucutes that Embody Library Principles https://www.imls.gov/sites/default/files/publications/documents/applying-library-values-emerging-technologychapter-5.pdf

election study. This will involve soliciting proposals from diverse voices in the open-science and open-access communities to identify additional research questions that focus on diversity, inclusion, and accessibility,

The project encompasses four work packages: development of open-data based participation and inclusion indicators; development of salience indicators using web and social media mining; piloting community-based extensions; and dissemination/outreach. These packages will be phased-in over the first half of the project, starting with development of a core data processing pipeline; then proceed in parallel for the project. By the completion of the funding period we aim to have automated data production in order to continue to produce regular updates of core indicators over an extended period.

National Impact: The project will address two critical gaps in the understanding of inclusive scholarship, scholarly practice, and will produce standardized indicators to describe the volume and types of open science output systematically over time. Communities will increase their ability to track changes in scholarly practice, understand the mechanisms of change, and evaluate the impact of new practices; systematic, comprehensive measures of participation in open access and open science will be put into place to inform future program development; and they will have the ability to expand and track systematic measurements of inclusion and diversity in research and scholarly communications.

Project Team: Micah Altman, PhD, Research Scientist, MIT Libraries will be co-PI and direct the overall scientific design and execution of the project. Altman has authored 4 books and over 90 articles in the fields of information science, social science and computer science. Chris Bourg, PhD, Director of Libraries at MIT, will be co-PI and provide expert guidance on research design as well as lead recruitment of and coordination of the Center for Research in Equitable and Open Scholarship (CREOS) advisory board. Sue Kriegsman, CREOS Deputy Director, will operate as the program manager and will be responsible for the overall functioning and success of the program. external advisory group will be essential to ensuring this work and other CREOS activities are aligned with and responsive to the communities we intend to serve. In general we will seek to include scholarly experts working in information science, sociology of knowledge, science of science, information policy, and/or other related fields; practitioners with deep experience in publishing, libraries, open science and open scholarship, and diversity and equity work in higher education and scholarship; and administrators with experience running similar research centers. CREOS is committed to ensuring the group includes individuals who represent and deeply understand the needs and perspectives of communities who have been historically and radically underserved by current scholarly communications infrastructure and practices.

Outcomes & dissemination. The primary output of the project will be a set of standardized indicators and reports describing the state of open access and open scholarship. These will be derived from an open database, a set of replicable data pipelines sufficient to reproduce these indicators; and will be accompanied by community-selected measurement modules. The database of indicators, standardized reports, replication data and code will be disseminated under open licenses through the CREOS website. Reports, data, and code will be archived through DSPACE@MIT, MIT Libraries Dataverse, and actively developed code will be published through Github. Results from the research (e.g, environmental scan, methodology, trend analysis) will be disseminated through conference presentations, conference workshop sessions aimed at training researchers on using the indicators, journal publications, policy-briefs to inform institutional administrators on indicator availability and use, and publicly available project documentation.

Budget Overview: The total budget requested from IMLS will be roughly $325K . This is based on a 3-year project timeline; 25% research staff FTE, 3% annual increase, plus fringe benefits. Undergraduate research assistance (10 hours week); computing and publication; domestic travel for outreach and dissemination; plus indirect costs. MIT will also offer a targeted 25% cost-share, consisting of staff salary, EB, etc. on 10% FTE staff costs.