Investing in Open Primary Sources: an eDesiderata Forum - Summary Report

On October 18, 2018 CRL hosted the [third annual eDesiderata Forum](#), an annual online event to explore emerging challenges in the acquisition, licensing and enabling of access to databases and data services for scholarly uses. The Forums help identify priorities and shape CRL's agenda for supporting informed investment in electronic resources by CRL libraries.

Planned in cooperation with the [NorthEast Research Libraries Consortium (NERL)](#), the 2018 Forum focused on open primary sources, exploring the complex business models around open access to news, archives, historical publications, and the growing monetization of public interest and public domain content.

The Forum consisted of two sessions, *Existing Open Access and Cooperative Investment Models*; and *Emerging and Potential New Models for Open Access*; followed by a discussion panel on *Possibilities for Collective Action*.

Session 1. Existing Open Access and Cooperative Investment Models

Libby Hemphill, Director, ICPSR and University of Michigan School of Information, on the Inter-university Consortium for Political and Social Research (ICPSR). A longstanding repository of public opinion datasets sourced from researchers and organizations and curated for re-use, ICPSR provides open access to many datasets, and restricted access to others. Funding is provided by foundations and federal funding agencies, while operational support is provided by member institutions (with additional revenue from training and other data services). ICPSR is seeking funding for ambitious new initiatives: *DataLumos*, an open access crowd-sourced repository for government data; and a federated social media archive to enable users to share and replicate/find data from multiple social media platforms.

Constance Rinaldo, Harvard University, on the Biodiversity Heritage Library (BHL), a global consortium of natural history and botanical institutions, digitizing and making their own collections freely available on the web. The consortium has 20 members and a total of 80+ partner museums and libraries around the world. The Internet Archive is a key project partner and service provider. Funding is provided by annual member fees and in-kind contributions, with significant ongoing operational support from the Smithsonian Institution. With a diverse and growing number of participating institutions, BHL faces challenges of scaling up support for member institutions.

Melissa Grafe, Yale University, on the Medical Heritage Library (MHL) another, newer initiative similarly driven by a community of expertise and interest, and where overhead and indirect costs are donated by the home institution. Like the Biodiversity Heritage Library the project's support comes from a limited number of core institutions, and it has been difficult to move beyond that group.

Michael Cook, Cornell University, and **Amy Paster, Pennsylvania State University**, on Project Ceres, an effort driven and guided by a nationwide network of agricultural libraries. The project builds upon decades of preservation work and a systematic, scholarly process that targeted a discrete corpus of historical literature. The effort was launched and initially sustained by substantial funding from the National Endowment for the Humanities. When NEH discontinued funding, CRL "kept the fires burning" with financial support to augment collection digitization: since 2013, CRL has supported 37 projects at 27 institutions.

Session 2. Emerging and Potential New Models for Open Access

Peggy Glahn, Reveal Digital, on the evolving Reveal business model. Originally inspired by the Early English Books Online / Text Creation Partnership, where libraries contributed funding to create open access, the project is library-driven but enabled by the product development and distribution capabilities of a commercial organization. Since Reveal's parent company became insolvent in April 2018, Reveal has explored establishing a standalone organization, but ultimately decided to seek a new home with an existing non-profit that shares Reveal's values and open access mission. The announcement of the partnership is pending.

Deirdre Ryan, ITHAKA, on the JSTOR Forum and JSTOR Open Access Resources, reported on plans to build open access onto existing JSTOR infrastructure. Past primary source collections, like *African World Heritage Sites*, *Struggles for Freedom in Southern Africa, JSTOR Global Plants*, and *British Pamphlets* (open in UK), developed with government and philanthropic funding, combine open access discoverability with paid enhanced functionality. A new endeavor, *JSTOR Forum* adapts Artstor's existing "Shared Shelf" software to enable libraries to catalog, upload and publish their own collections on a web platform provided and hosted by JSTOR.

Bernard Reilly, CRL, on the current trend toward commercialization of public goods. Two phenomena threaten the open accessibility of public domain content and documentation and data of vital interest to citizens. Financial pressures are driving public institutions like the British Library to monetize collections through licensing of digital rights to commercial vendors. And data and documentation exposed freely on the web by "public good" organizations like the National Security Archive and International Consortium of Investigative Journalists are susceptible to loss or commercialization due to those organizations' lack of resources.

Panel Discussion: Possibilities for Collective Action.

Three experts on collection development and digital resources from the CRL community commented on the presentations and the issues raised. **Lindsay Cronk (University of Rochester)**, **Janet Hulm (Ohio University)** and **Elizabeth Kirk (Harvard University)** reflected on several issues the presentations raised: the difficulty of sustaining open access initiatives, the dearth of community-owned infrastructure for digital collections, and the complexity and lack of transparency of the open access landscape.

Risk: Panelists cited the varying degrees of risk involved in investing in open access initiatives. Access models ranged from hybrids of open and proprietary access (ICPSR, JSTOR), to fully open (BHL, Project Ceres). They cited the range of business models as well, from sponsorship and voluntary contributions (Reveal, BHL), to efforts subsidized by ancillary services (Ceres, JSTOR). Pure open access is rare and continues to be undermined by the "free rider" problem and by "donor fatigue".

Initiatives embedded in parent institutions (ICPSR and BHL) that cover overhead and indirect costs may be able to withstand those challenges. Others, like Reveal, are more vulnerable to financial duress.

Scope: The scope of open access efforts has been limited by intellectual property restrictions on content. Efforts by the Biodiversity Heritage Library, Reveal, and others to clear rights for copyrighted materials have proved extremely labor-intensive and costly. The primary focus of many of the projects, therefore has been older, public domain and cultural heritage materials.

Transparency: There is uncertainty about some open access providers and their practices. The prevalence of hybrid models (JSTOR Plants, ICPSR) where the costs of hosting and technology are subsidized by revenue from premium services that are paid for, has created questions about the extent to which these are truly open. (Some

attendees disputed the characterization of the JSTOR Plants database as an Open Access resource.) The "near-death experience" of Reveal also elevated concerns about sustainability. Janet Hulm cited the need for criteria and rubrics for identifying potential open access investments, to identify projects that will provide a return on library investment.

There was also general agreement on the need for greater accountability of open access providers to the library community. Even as the range of community-funded initiatives grow and diversify, there remain key challenges to discovery and interoperability of the resources generated. Elizabeth Kirk suggested one approach to solving that problem: "a cooperatively built platform along the lines of HathiTrust, for primary sources."

Lindsay Cronk encouraged Forum participants to "bear down together" on these issues, as CRL "is well positioned to assess and equitably distribute the burdens of the lift. CRL and NERL are equipped to sit at the heart of this large-scale collective investment as their original missions and work has always focused on making collective collections a reality."

Next Steps: Toward an Open Access Investment Agenda

The presentations, discussions and panelists' commentary suggest four specific ways CRL and NERL can move forward. These will form the basis for their strategic planning on an Open Access investment agenda.

1. **Investment:** CRL and NERL should allocate significant funding on an annual basis for investment in open access primary source resources. That funding should be devoted to supporting organizations that have existing capabilities, assets and infrastructure, and that are adequately accountable to the CRL and NERL communities.

2. **Analysis:** To achieve maximum impact and optimal return on investment, a portion of the allocated funding should be invested in the analysis and evaluation of commercial, academic and other open access primary source electronic resources and the gathering of data to inform NERL and CRL investment in open access decisions.

3. **Metrics:** CRL and NERL should establish a common set of underlying principles for open access providers, defining acceptable funding models/practices, required disclosures and mechanisms for accountability, and the baseline terms of service for resources to qualify as open access.

4. **Advocacy:** CRL and NERL should enunciate a formal policy position objecting to the privatization and commercialization of public domain, public interest and cultural heritage materials.

Report submitted 13 November 2018

Event details, participant bios and presentation slides available at:
https://www.crl.edu/events/investing-open-primary-sources-edesiderata-forum