The traditional scholarship products that libraries manage are now primarily in digital form. At the same time, libraries are adding support for outputs including data, thus providing new opportunities for scholars to engage with large, multifaceted, digital aggregations of content through software tools. Synthesizing content, computation, and capacity requires the development of infrastructure to enable the application of tools (including text-mining, computational analysis, machine learning, data publication, and data computation) to library collections. Such computational functionality across digital collections enables researchers to explore new capabilities and formulate new inquiries through fundamental changes in the ways scholars work—changes that increase the importance and vitality of collections across the research process.¹

Libraries need to conduct the selection and acquisition of information resources in explicit combination with decisions regarding the analytical tools and capacities to support the computational use of those resources in today’s research environment. The combination of content and computation makes the library collection a platform for exploring a host of additional research questions and applications. The benefits of networked, digital collections as a platform require investments in computational tools and access, the synthesis of content with technical and expert capacities, and the development of new campus partnerships. Well-established models for funding, acquiring, and “owning” collections become more complex. Partnerships within and between institutions must be sustained to enable the funding and scale required for computational collections.

Licensing, Infrastructure, and Partnerships

Unfortunately, the trend toward digital in library collections, accompanied as it was by the rise of the license, brought with it the side effect of closing off collections from computational analysis. Contracts were signed with only human readers in mind but were rarely renegotiated, even as computational research increased dramatically over the past decade. Research libraries, in general, have significant catching up to do:

- On the vendor and contract end, libraries need to negotiate clear, full, and reasonably unflattered access to proprietary and limited-use data for researchers who need content as data.
- On the research side, libraries need to forge and reestablish relationships with computationally intensive research teams and units that understandably have come to see the academic library as irrelevant to their work.

Libraries are a critical part of the research infrastructure within higher education and are anchors of their respective institutions. Libraries provide the content that constitutes the intellectual foundation of research across all fields, and library services (from discovery to consultation to publishing support) ensure that this content—as well as researchers’ new content that grows and builds upon it—can effectively and efficiently be disseminated to wider publics. As research is increasingly mediated through code and computation, library collections and services must be positioned as infrastructure that is no longer defined by physical, intellectual, or even directly human-oriented pursuits but, rather, extends into virtual, digital, and even machine learning and AI research modes.

Partnerships across institutions must also reflect the already present reality of computational research. On campuses, the library cannot be the sole provider of computational research data, software, hardware, or support. Rather, the library must provide these in coordination, concert, and collaboration with offices of research and information technology, research labs and centers, and other units. Libraries will need new partners beyond their individual campuses and the usual library organizations and consortia. Research libraries will need to rely on government, industry, and scholarly associations, among others.
Integrating Computational Collections and Capacity

To fully realize the potential of computational collections as part of a collections and research engagement program, libraries require significant financial, technical, and human resources:

- The costs of providing access to collections that are optimized for computational research include licensing and purchasing existing data sets, creating data sets for use by researchers, and funding open access, which might support content mining.
- The costs of providing infrastructure to support computational research across collections include the purchase and licensing of software for analysis, systems to support discovery of these data sets, and platforms that support analysis and storage.
- The costs of providing expertise to support computational research include training, retention, and potential recruitment of highly skilled, qualified experts. Expertise needs to encompass not only technical skills but also the legal and business negotiation skills that are necessary for integrating computational capacity into an overarching program of collections and research engagement.

Broadening the collections budget as a means of supporting nontraditional resources and activities—including digital scholarly communication tools, open-access fees, software applications, digital preservation, and organizational memberships—is critical. However, the reality of ongoing pressure on collections budgets cannot be ignored and might seem unsustainable and unrealistic when the challenge to “do more with less” has never been more real or more urgent.

Libraries have already become more efficient in managing collections (for example, by shifting to more on-demand monograph acquisitions models) and have diverted savings to support the increasing costs of journal packages, leaving little capacity to support modern research needs. Research libraries must identify ways to be more aggressive and creative to support computational collections—for example, by moving funding away from narrow database or journal package subscriptions or by creating co-investment and cost-sharing models with offices of research and technology. Although moving from a single oversight/ownership model to a campus-wide shared investment model leads to more complicated alliances to compile, shepherd, and sustain funding, for most institutions this shift might represent the most efficient long-term way of addressing data-intensive research.

Collections should be blended as a service, with computational capacity that spans organizational, cultural, and technical boundaries. The challenge for libraries and scholars becomes how to shift models toward a synthesis of content- and service-centricity that also incorporates software and tools such as text-mining, computational analysis, artificial intelligence, machine learning, data publication, and data computation. Libraries have the ability to build on decades of digital collections work, but they need to start investing with more intention and on a larger scale. Libraries and their partners must bring robust collections and emerging computational tools together with developing professional expertise to unlock the combined potential.

Note

Darby Orcutt is Assistant Head of Collections and Research Strategy at North Carolina State University. Hilary Davis is Head of Collections and Research Strategy at North Carolina State University. Greg Raschke is Director of Libraries at North Carolina State University.

© 2020 Darby Orcutt, Hilary Davis, and Greg Raschke. The text of this article is licensed under the Creative Commons Attribution 4.0 International License.