Mapping the Current Landscape of Research Library Engagement with Emerging Technologies in Research and Learning:
Executive Summary

By Sarah Lippincott

Edited by Mary Lee Kennedy, Clifford Lynch, and Scout Calvert

March 23, 2020
# Table of Contents

**Executive Summary** 3  
Cross-Cutting Opportunities 4  
Facilitating Information Discovery and Use 8  
Stewarding the Scholarly and Cultural Record 10  
Advancing Digital Scholarship 12  
Furthering Learning and Student Success 13  
Creating and Managing Learning and Collaboration Spaces 15  
Conclusion 16  

**Endnotes** 17
Executive Summary

The generation, dissemination, and analysis of digital information is a significant driver, and consequence, of technological change. As data and information stewards in physical and virtual space, research libraries are thoroughly entangled in the challenges presented by the Fourth Industrial Revolution:¹ a societal shift powered not by steam or electricity, but by data, and characterized by a fusion of the physical and digital worlds.² Organizing, structuring, preserving, and providing access to growing volumes of the digital data generated and required by research and industry will become a critically important function. As partners with the community of researchers and scholars, research libraries are also recognizing and adapting to the consequences of technological change in the practices of scholarship and scholarly communication.

Technologies that have emerged or become ubiquitous within the last decade have accelerated information production and have catalyzed profound changes in the ways scholars, students, and the general public create and engage with information. The production of an unprecedented volume and diversity of digital artifacts, the proliferation of machine learning (ML) technologies,³ and the emergence of data as the “world’s most valuable resource,”⁴ among other trends, present compelling opportunities for research libraries to contribute in new and significant ways to the research and learning enterprise. Librarians are all too familiar with predictions of the research library’s demise in an era when researchers have so much information at their fingertips. A growing body of evidence provides a resounding counterpoint: that the skills, experience, and values of librarians, and the persistence of libraries as an institution, will become more important than ever as researchers contend with the data deluge and the ephemerality and fragility of much digital content.

This report identifies strategic opportunities for research libraries to adopt and engage with emerging technologies,⁵ with a roughly five-year time horizon. It considers the ways in which research library
values and professional expertise inform and shape this engagement, the ways library and library worker roles will be reconceptualized, and the implication of a range of technologies on how the library fulfills its mission. The report builds on a literature review covering the last five years of published scholarship, primarily North American information science literature, and interviews with a dozen library field experts, completed in fall 2019. It begins with a discussion of four cross-cutting opportunities that permeate many or all aspects of research library services. Next, specific opportunities are identified in each of five core research library service areas: facilitating information discovery, stewarding the scholarly and cultural record, advancing digital scholarship, furthering student learning and success, and creating learning and collaboration spaces. Each section identifies key technologies shaping user behaviors and library services, and highlights exemplary initiatives.

Underlying much of the discussion in this report is the idea that “digital transformation is increasingly about change management”—that adoption of or engagement with emerging technologies must be part of a broader strategy for organizational change, for “moving emerging work from the periphery to the core,” and a broader shift in conceptualizing the research library and its services. Above all, libraries are benefitting from the ways in which emerging technologies offer opportunities to center users and move from a centralized and often siloed service model to embedded, collaborative engagement with the research and learning enterprise.

**Cross-Cutting Opportunities**

*Engage with machine learning to improve research, learning, and scholarly communication.*

Machine learning, the sub-discipline of artificial intelligence (AI) that “uses collections of examples to train software to recognize patterns, and to act on that recognition,” has demonstrated a remarkable ability to match (and outpace) human performance on certain well-
constrained but complex tasks, and is already incorporated into a range of common systems and devices. The term AI has taken on a life of its own; it is frequently invoked as an umbrella term for ML, natural language processing (NLP), expert systems, and related technologies that approximate human cognition. The casual use of the term AI often erases the distinction between substantive applications (for example, pattern and image recognition) and speculative and unproven uses (for example, prediction, reasoning, formulating original ideas). In the interests of specificity and precision, this report makes an effort to identify specific technologies (such as ML) where possible, while recognizing that some initiatives invoke AI even when the scope of their activities focuses on a specific sub-technology.

As the near-term applications of ML and related technologies shape the ways in which scholars create and engage with information, students learn and study, and communities interact with their built environments, libraries will be profoundly implicated, given their role as creators, sources, and stewards of information and as educators. Libraries can strategically engage with ML by better understanding its affordances, limitations, and risks, and by distinguishing the genuine accomplishments of ML and related technologies from AI hype. The application of ML to tasks related to classification, prediction, and pattern recognition and generation make it particularly germane to information discovery. A number of research libraries have initiatives underway that apply ML, computer vision, natural language processing, and other techniques to automate description of large-scale digital collections and enhance discovery, access, and analysis systems. A few are also leading critical discourse and educational efforts on their campuses around the implications, ethics, and future of ML. Research libraries also have opportunities for field-level collaboration. For example, libraries could assemble the large volume of validated and labelled training data that drive ML algorithms in ways that aim to recognize or mitigate bias and that are sensitive to the specific needs of cultural heritage materials.
Bolster services that recognize the centrality of data to the research enterprise.

Big or small, textual, numeric, or visual, in support of the humanities, science, or interdisciplinary research, digital data and structured knowledge have become essential and ubiquitous scholarly inputs and first-order outputs. Research libraries play a key role in data generation, dissemination, discovery, analysis, and stewardship and can contribute to realizing the vision of a FAIR (findable, accessible, interoperable, and reusable) data environment that advances open scholarship. Over the next decade, advancing FAIR data will require significant investment, creating myriad opportunities for libraries. Research libraries can contribute to FAIR data by describing structured data; building and providing access to machine-actionable and ML-ready data sets that facilitate computationally driven research; collaborating with domain experts to develop descriptive standards and ontologies that support disciplinary and multi-disciplinary research by humans and machines; and maintaining reuse-driven repository infrastructure. Research libraries are developing services that are attuned to the needs of scholars working with very large data sets as well as the long tail of smaller, heterogeneous, unique, and often labor-intensive data sets that support research across the disciplinary spectrum. In their role as educators, librarians are also well-positioned to cultivate data fluency and the technology skills required for data-centric research methods.

Integrate the library’s services and collections with the networked environment.

Researchers operate in geographically distributed, interdisciplinary, networked environments. Scholarly communication has also become diversified and disaggregated. The library’s role in information management is being reenvisioned: no longer solely a steward of a unified local collection, the library becomes the facilitator of a networked suite of open and extensible tools, resources, and services. Building local research collections will eventually
diminish in importance, while curation and facilitated access to information become critical.\textsuperscript{19} Research libraries are leveraging emerging technologies to make their services and collections interoperable and more seamlessly integrated into the lives and work of their constituents. For example, research libraries are ensuring that their unique digital collections—including digitized special collections, institutionally published content, and expert profiles—are interoperable with web-scale and federated discovery tools, by creating harvestable, machine-readable metadata, and associating them with persistent identifiers. As research praxis routinely crosses institutional and geographic boundaries, research libraries also have opportunities to act consortially or outside of their local framework to maximize their impact. Research libraries could, for example, develop coordinated models of research data stewardship in which individual institutions assume responsibility for a segment of data (such as data defined by domain or type) based on local strengths and capacity.\textsuperscript{20} Conversely, libraries could contribute their expertise to initiatives that are not affiliated with or hosted by their (or any) campus, such as specialized “data communities.”\textsuperscript{21}

\textit{Cultivate privacy awareness and privacy services.}

Emerging technologies are redefining expectations of privacy and creating tensions around the ethical use of personal data. The ease of constant surveillance is facilitated in physical space by Internet of Things (IoT) technologies that collect continuous streams of data, and in virtual space by the collection of digital analytics by campus and third-party systems. ML tools can process this data with remarkable speed and precision, making genuine data de-identification nearly impossible. As students and scholars come to expect (data-driven) personalized digital services and campuses expect to reap the benefits of large-scale data analytics, libraries will have critical choices to make. Research libraries can play a key role in helping their campus communities develop a nuanced understanding of privacy in physical and digital space. In their own work, libraries can commit to transparent data collection retention and use policies, and
conscious, thoughtful management and control of personal information. This includes negotiating vendor agreements that protect reader privacy, offering trade-offs between privacy and personalization where appropriate, and establishing boundaries around their participation in campus-wide data collection efforts.

A genuine commitment to privacy may become one of the library’s fundamental distinguishing features; many libraries are working to provide (physical and virtual) spaces that consciously minimize and make transparent the ways in which users may be tracked or their data collected. Libraries have an opportunity to position themselves as leaders in privacy education and privacy-aware approaches to personalization, learning analytics, and the use of tracking technologies on campus. A core component of user-centered library services will be positioning users at the center of discussions about the ethical use of user data and the implementation of tracking devices, algorithmic decision-making tools, and other potentially invasive technologies in libraries.

**Facilitating Information Discovery and Use**

*Invest in user-centered discovery tools.*

The widespread adoption of web-scale discovery tools, combined with a landscape of information overabundance, has upended “the notion that the library attempts to licence or provide access to all [published] material” and instead has prompted libraries to focus on creating and licensing discovery tools and services that navigate and curate content. Some of the most promising uses of emerging technologies to make search and discovery more user-centered include various kinds of enhanced search, NLP-based automated text-processing tools, recommendation systems, and personal assistants. While libraries may develop homegrown solutions, most of these tools will be commercial products, making them potentially problematic with regard to privacy. Aspirationally, these technologies expand discovery beyond simple search and retrieval, reconceptualizing it as a process of exploration and engagement with networked information.
Reveal hidden digital collections through enhanced description.

The acceleration of digitization and born-digital content creation has left libraries facing an ever-increasing backlog of resource description to drive traditional collection discovery and navigation tools and methods. As libraries place increasing value on their unique local collections, they need new ways of making those collections discoverable and navigable to internal and external audiences, both human and machine. A number of academic libraries are experimenting with technologies such as ML algorithms (including facial recognition and image recognition/classification) and natural language processing to automate metadata creation, improve discoverability of visual information, and provide unprecedented access to their rich digitized and born-digital collections.

Expose library collections beyond library systems.

As information becomes distributed, diversified, and open, researchers prefer web-scale discovery tools that aggregate resources from a range of sources over siloed library catalogs and digital asset management systems. Research libraries have a number of strategic opportunities to integrate library collections with a range of other open, digital resources, enriching the information available to users on the open web. Research libraries are meeting users where they are by implementing search engine optimization techniques; exposing metadata for harvesting by aggregators, such as the Digital Public Library of America; providing application programming interfaces (APIs) that permit new forms of computational engagement with collections; adopting interoperability standards, such as the International Image Interoperability Framework (IIIF), to facilitate discovery and reuse; and participating in linked open data (LOD) initiatives. The shift towards revealing local collections to external audiences rather than the reverse, a trend Lorcan Dempsey has called the “inside-out library” and one component of what other authors have termed the “library as platform,” is a natural consequence of an open, oversaturated, and networked information landscape.
Stewarding the Scholarly and Cultural Record

Advance open research and publishing practices.

By supporting open research practices—including the adoption of open metadata standards, creation of machine-readable publications, and deposit of outputs (including underlying data and code) in open repositories—libraries make research more discoverable, reusable, reproducible, and durable. These practices improve both the quality of scholarship itself and the quality and manageability of the scholarly record. Libraries play a critical role in achieving FAIR (findable, accessible, interoperable, and reusable) research data through their curation, education, and preservation activities.31 Realizing the vision of FAIR scholarship will be a central challenge for the research community over the next decade.

Reinforce integrity and trust in the scholarly and cultural record.

Memory institutions are built on trust: the trust that materials under their stewardship are authentic, immutable, and preserved in perpetuity or de-acquired through a transparent and well-understood process. Emerging technologies pose new challenges for fulfilling the role of trusted steward. The assurance of authenticity, for example, is threatened by the ease of manipulating and altering digital media, and the complexities of determining provenance of digital materials. Deep fakes—counterfeit video, audio, still images, and textual content created using ML—pose a particular challenge. Research libraries have a range of digital forensics tools at their disposal to authenticate digital artifacts and collections at the time of accession and throughout their life cycle. They are also identifying secure pathways—possibly involving distributed ledger technologies (such as blockchain) and public key infrastructure (PKI)—to acquire copies of digital objects from sources they trust, documenting and proving the chain of custody, and any changes that have been made to it along the way.32 After accessioning, fixity checking continually proves objects and collections do not change over time, due to degradation of the content,
or to intentional or accidental manipulation. Underlying all of these processes is the need to maintain security and integrity of computing and storage operations in the face of cyberattacks and natural disasters. Finally, librarians also help their constituents develop the skills needed to assess and critically engage with the integrity and reliability of information.

Preserve the evolving scholarly and cultural record.

The expanded scholarly and cultural record has amplified both the technical and social barriers to achieving digital preservation at scale. On the technical front, emerging technologies have led to new types of research and creative outputs that require new approaches to digital preservation, as well as an unprecedented rate of digital content creation. Software, 3-D data, dynamic web content, and the inputs and outputs of ML, among other media, push the limits of established digital preservation practices. The digital cultural and historical record—the massive volumes of digital images and video, news, social media posts, and other web-based content that constitute essential evidence for present and future scholarship—will be incompletely preserved its scale and complexity. Addressing the thorny questions of what can and should be preserved over the long term, will require deep cross-institutional coordination and cooperation. On the social front, the distributed and licensed nature of digital scholarly and cultural content presents legal, administrative, and financial barriers. Even as emerging technologies have destabilized the digital preservation environment, they have also offered new solutions and opportunities. A few libraries—and their collaborators in computer science and information technology departments—are leveraging developments in containerization, distributed ledger technologies (such as blockchain), new storage media, and automation of digital preservation practices through ML to help ensure that the expanded scholarly record remains accessible well into the future.
Advancing Digital Scholarship

*Develop data services that work for big data*\(^{16}\) *and small data across disciplines.*

Academic and research libraries are natural partners with others involved in data management activities, and many maintain robust and active research data management services. Librarians have the disciplinary, information management, and technology expertise required to manage data throughout its life cycle. The profile of library data services is being shaped by a number of forces, including the expanding emphasis on data-driven research in humanities and social sciences fields and the need for infrastructure and services that recognize data as a living asset. As they work with complex, heterogeneous, and mutable data sets, scholars need tools and education that facilitate analysis, sharing, and preservation. Emphasis on data use and reuse has profound implications for repository infrastructure, entailing a shift from infrastructure optimized for storage and retrieval to one optimized for analysis and sharing.\(^{37}\) While a few libraries have made strides in this area, most data repository services remain focused on helping scholars meet federal and funder requirements around data deposit. Research libraries also face challenges as they design data services and infrastructure that are sensitive to discovery and analysis methods that vary widely by discipline.

*Provide and sustain machine-actionable collections.*

Some of the most innovative digital scholarship work uses computational processes to derive new insights from vast troves of digital and digitized content held in library collections. Text and data mining have gained traction with many scholars in a range of disciplines as they seek more nuanced methods of discovery and analysis.\(^{38}\) Machine-actionable collections enable researchers to go beyond simple information retrieval, treating collections (including their metadata, full-text, and relationships) as the input for powerful
computational processes. Such initiatives as the Collections as Data project encourage cultural heritage institutions to thoughtfully develop digital collections (licensed, purchased, and unique) and allied services (for example, workshops, consultations, digital platforms) that support “computationally-driven research and teaching.” This means not only making digital collections available online, but providing them as structured, machine-actionable data sets. Machine-actionable collections are essential not only for human-driven computational research, but for the development of new ML tools, which rely on large quantities of structured data to become proficient at a task. Libraries can apply their “expertise and practical experience in creating and managing classification systems” to develop ML training sets that serve the needs of cultural heritage institutions.

Deliver data science education and consultation.

Data science proficiency has rapidly become a core competency for researchers and students, as scholars in many or most disciplines routinely rely on computational data analysis in their research and learning. Research libraries can cultivate the data science skill sets to sustain and expand these practices. Some research libraries have identified a niche in providing tailored educational offerings for faculty members and students outside of STEM fields, who may lack opportunities within their department or program of study. These informal educational programs can help undergraduate and graduate students in all disciplines become proficient in common data science tools (such as electronic lab notebooks), techniques (such as web scraping), research data management practices, compliance with funder and federal policies, and open science principles.

Furthering Learning and Student Success

Build digital fluency and digital scholarship skill sets.

Research libraries provide a range of informal education and consultation to impart the digital skills that contribute to the academic and professional success of undergraduates, graduate students,
and early career researchers. These include workshops that teach concrete digital scholarship and coding skills, such as programming languages, software carpentry, and data visualization; research data management and open science practices; and scholarly communications topics such as copyright, identity management, and navigating academic publishing. Longer-term cohort-based educational programs have also become popular. These programs often encourage interdisciplinary engagement with an emerging technology over the course of a semester or longer. A few research libraries have also launched formal programs that fill gaps in the academic curriculum, for example, the Temple University Libraries’ interdisciplinary cultural analytics certificate. In addition to digital scholarship skills, research libraries have opportunities to help students critically engage with and optimize their use of a new generation of productivity tools, many powered by ML, that promise to assist users in a range of tasks related to learning and study.

The ease of publishing information and misinformation on the web, the growing sophistication of counterfeit content, and the use of black box algorithms to generate and display information mean that achieving digital fluency also requires that students be able to interpret and evaluate an unprecedented array of new media formats and sources. Students need to understand not only the credibility and reliability of textual media, they need data and algorithmic literacy skills, strategies for distinguishing between genuine and manipulated or fabricated digital content, and an understanding of online data privacy. Libraries are well-positioned to deliver an expanded digital fluency curriculum in partnership with faculty members, campus IT, and other collaborators.

*Foster critical engagement with and access to emerging technologies for all students.*

As third spaces, independent from any campus department, libraries have become hubs of technology access for students in all majors. Technology-rich learning and information commons, collaboration
studios, makerspaces, and labs are now commonplace in libraries. Locating digital scholarship centers within libraries can help to democratize and de-silo access to cutting edge technologies, encouraging cross-disciplinary collaboration and discovery. These spaces provide access to specialized software and hardware for fabrication (such as 3-D printers, computer-aided design and drafting software); visualization (such as high-resolution displays); immersive reality (such as VR headsets); and other digital research and creation methods. When libraries apply their existing expertise as educators to new forms of knowledge production, they can help their communities thoughtfully and productively engage with technology in these spaces. Librarians are equally well-positioned to collaborate with faculty on the pedagogically grounded integration of technologies such as immersive reality and information visualization in the classroom.

Creating and Managing Learning and Collaboration Spaces

Create dynamic, networked spaces that promote new forms of inquiry.

While leading-edge technology is often most conspicuous in makerspaces and labs, some of its most transformative potential lies in the seamless and often invisible integration of emerging technologies into the full library-visitor experience. The use of Internet of Things technologies presents a particularly compelling opportunity for library spaces (whether in the library building or embedded in shared spaces around campus) and services to dynamically adapt to user behaviors. The ubiquitous integration of sensors and networked technologies into the library’s physical spaces could transform it into “a living-learning lab that senses and studies human dynamics, human-computer interactions, and human-building interactions.” Libraries have an opportunity to pioneer inclusive, privacy-aware approaches to this integration of sensing technologies in the public sphere. Creating networked library spaces complements the library’s role as data provider and steward, as a node for digital information discovery, and as a promoter of critical engagement with emerging technologies and the changing nature of research and information behavior.
Enhance the user experience in library spaces.

Emerging technologies offer a range of opportunities for libraries to make spaces more welcoming, navigable, interactive, comfortable, and productive. Libraries are experimenting with the Internet of Things (IoT), particularly beacon technology, to create self-guided library tours and navigational aids, build augmented reality (AR) exhibits, provide location-specific mobile alerts, help users locate materials in the library stacks, and facilitate access to bookable or restricted spaces or items. Emerging technologies can also be deployed to enhance a sense of community within library spaces. Several speculative apps propose to help users locate and connect with others in a library space who share their interests, allowing them to form study or collaboration groups on the fly. As they engage with beacons, wearables, and location-based apps, libraries are cognizant of implications around privacy and intellectual freedom, and are developing best practices for privacy-aware implementation of IoT technologies in library spaces.

Conclusion

Research libraries can bring values-based decision-making to bear as they find the right balance in their approach to adopting and experimenting with emerging technologies—the balance between agility and sustainability, convenience and privacy, transformation and persistence. As emerging technologies such as machine learning, immersive reality, and the Internet of Things change the ways researchers and students engage with information, libraries have opportunities to advance their contributions to the research and learning enterprise. As adopters of these technologies, research libraries can make information more discoverable, reusable, and durable. As educators, library workers can help their communities critically and productively engage with technology in the service of research and learning.
Endnotes


3. “ML is a subset of the larger field of artificial intelligence (AI) that ‘focuses on teaching computers how to learn without the need to be programmed for specific tasks,’ note Sujit Pal and Antonio Gulli in Deep Learning with Keras. ‘In fact, the key idea behind ML is that it is possible to create algorithms that learn from and make predictions on data.’ “—James Furbush, “Machine Learning: A Quick and Simple Definition, O’Reilly, May 3, 2018, https://www.oreilly.com/ideas/machine-learning-a-quick-and-simple-definition.


6. “‘None of This Is Really about Technology’ — Digital Transformation and Culture Change,” Jisc (blog), January 21,
Mapping the Current Landscape of Research Library Engagement: Executive Summary


8. For the purposes of this paper we use the following definition of AI from the Association for the Advancement of Artificial Intelligence (AAAI): “the scientific understanding of the mechanisms underlying thought and intelligent behavior and their embodiment in machines.”—AAAI, ”Information about AI from the News, Publications, and Conferences,” AITopics, accessed February 19, 2020, https://aitopics.org/search.


36. There are many definitions of big data. This report may be helpful to the reader: *NIST Big Data Interoperability Framework: Volume 1, Definitions*, NIST Special Publication 1500-1 (Washington, DC:


Mapping the Current Landscape of Research Library Engagement: Executive Summary


