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

Facilitating Information Discovery and Use

By Sarah Lippincott

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

April 14, 2020

ASSOCIATION OF RESEARCH LIBRARIES

born-digital
RESEARCH + CONSULTING

cni
Coalition for Networked Information

EDUCAUSE

Table of Contents

This is the third installment of a forthcoming report, *Mapping the Current Landscape of Research Library Engagement with Emerging Technologies in Research and Learning*, that will be published in its entirety by late spring 2020.

The following installments are being published as they become available at https://doi.org/10.29242/report.emergingtech2020.landscape:

Executive Summary [published March 26, 2020]
Introduction, Methodology, and Cross-Cutting Opportunities [published April 2, 2020]
Facilitating Information Discovery and Use [published April 14, 2020]
Stewarding the Scholarly and Cultural Record
Advancing Digital Scholarship
Furthering Learning and Student Success
Building and Managing Learning and Collaboration Spaces

Landscape Overview

The library's role as connector between researchers and information has evolved over hundreds of years. Historically, libraries amassed and disseminated broad and deep collections of print and digital resources to their local communities. To many constituents, this remains the primary perceived function of libraries today. Libraries continue to invest significant portions of their annual budgets to license and purchase information resources, and continue to use collection size as a primary metric of quality and value.[1] Academic libraries are adept at managing discrete publications: negotiating licenses and purchasing agreements, making content "discoverable via institutional systems populated with hand-crafted metadata,"[2] and ensuring long-term preservation. However, this model is being rapidly disrupted and displaced by a "greatly expanded scholarly record—one that is less dependent on papers and articles, and that is increasingly expressed in terms of networks of links and associations among diverse research artifacts."[3] The expanded scholarly record has engendered three interrelated challenges for library discovery and access.

1. **The types of information researchers seek is changing.** Researchers increasingly require access to information resources outside the traditional scope of library collections, from massive data sets, to visualizations, three-dimensional objects, and computer models. Many researchers work outside of and across traditional disciplinary boundaries and require information sources from a range of fields of study. For some researchers, metadata, rather than published content, may be the primary object of study.
2. **What researchers intend to do with that information is changing.** Researchers increasingly expect to mine, process, and analyze content. With knowledge production rapidly outpacing human processing capacity, researchers will increasingly rely on machines to parse and interpret information. For example, experiments in unsupervised text mining of the scientific

literature have demonstrated that the data in the existing published scientific literature contains a wealth of unrecognized discoveries.[4] Only by analyzing this content at scale can scholars identify the overlooked patterns and connections embedded in the scholarly record.

3. **How researchers go about looking for that information is changing.** Researchers increasingly expect search and discovery interfaces that support a range of inputs and outputs. For example, new math-aware search engines allow users to enter mathematical equations as search terms and return results based on similarities in either the structure or meaning of the equation.[5] The Dig That Lick project searches its large-scale corpus of jazz recordings for pattern similarities based on a user's input on a virtual keyboard.[6] In addition to accepting non-textual inputs, researchers increasingly expect searches to return personalized, context-aware results. As search practices vary widely by discipline, scholars desire discovery tools that align with their field's research methods and expectations.

Together, these changes in scholarly expectations signal a future in which the library catalog and other local discovery systems will diminish in value, in favor of web-scale discovery. The library's role in discovery is undoubtedly shifting, a trend accelerated by emerging technologies such as machine learning (ML). One expert interviewed for this report remarked that "the internet has put us [libraries] on a collision course with the world," threatening to disintermediate the library in the discovery process.[7] Some experts have suggested that commercial web-scale search may entirely replace local academic library discovery systems.[8]

Much of the literature on the future of discovery in libraries, along with the expert interviews conducted for this report, provides a resounding counterpoint. Authors and interviewees suggest that the networked environment presents a number of strategic opportunities for libraries, specifically related to helping researchers optimize their use of ML-enhanced search applications, text-mining tools, and other approaches

to sifting through the data deluge;[9] making unique digital collections available and discoverable at an unprecedented scale; and meeting users where they are by making unique local resources available in web-scale discovery environments.[10] Key emerging technologies with an impact on discovery include ML, natural language processing (NLP), and computer vision.

The following sections detail these opportunities and highlight examples of academic and research library engagement with the range of emerging technologies that are driving and responding to changes in how scholars discover, use, and create information.

Strategic Opportunities

Invest in user-centered discovery tools

The widespread adoption of web-scale discovery tools, combined with a landscape of information overabundance, may "completely upend the notion that the library attempts to licence or provide access to all [published] material" and instead prompt libraries to focus on licensing (ML-powered) tools and services that navigate and curate content.[11]

An increasing emphasis on user-centered discovery positions the user, rather than the collection, as the organizing principle within a discovery environment.[12] Manifestations of this user focus include expanding functionality beyond "search and retrieval" to enable users to actively engage with, interact with, and supplement library collections.[13] Known-item and exploratory search can be supplemented with "current awareness" tools, that is, mechanisms that help scholars keep up with developments in their field;[14] automated text-processing tools that provide just-in-time article summaries; visualizations of the connections between different resources; the ability to create and curate personal collections that include library-held and external resources; or scholarly profiles that showcase a researcher's work and allow them to set up a personalized feed of newly published research based on their interests.

Some of the most promising uses of emerging technologies to make search and discovery more user-centered include ML-enhanced search, automated text-processing tools, recommendation systems, and virtual assistants. The following sections discuss each in more detail, including several examples of academic library adoption or engagement in each area.

ML-enhanced search

Many academic library search interfaces primarily rely on keyword matching against the full-text of a publication or its metadata record. This approach to information retrieval can be onerous for users, who must experiment with different search terms and combinations, contend with incomplete metadata, and sift through large volumes of search results. As one expert interviewed for this report noted, keyword search makes interdisciplinary research particularly difficult, as it often fails to bring together "parallel conversations."[15]

A range of new search and discovery tools are challenging the centrality of simple keyword search, or enhancing its power through machine learning. The options available to libraries and scholars include several tools tailored to academic literature discovery, including Yewno,[16] Iris.ai,[17] Dimensions,[18] and Semantic Scholar,[19] among others, which rely on NLP and other machine learning to enhance search results.[20] These new tools tout semantic search capabilities, which attempt to return results based on a query's meaning, rather than specific keywords. These and other search tools, which understand the semantic meaning of queries and can build associations between different discipline-specific terms for the same concept, will significantly lower barriers for scholars aiming to discover literature across domains.

Some next-generation discovery tools also aim to produce a more serendipitous search experience, one in which users can discover unlikely sources and unexpected connections. Google's Talk to Books experiment, for example, uses NLP to return potentially relevant book passages based on a user's query.[21] Users are encouraged to ask questions rather than enter search terms (that is, topics or entities).

The Talk to Books algorithm then returns search results based on predictions of likely response statements. While Talk to Books does not purport to be a rigorous search tool, it may point to a redefinition of user expectations for information retrieval.

Next-generation search and discovery tools are also improving upon and pushing the boundaries of the traditional search results list. Yewno's underlying technology, for example, produces conceptual units from its vast corpus of literature using a deep learning network to extract and group topics, allowing searchers to explore a complex network of interrelated literature. The biomedical literature search tool PubMed, from the National Library of Medicine (NLM) combines a "state-of-the-art machine-learning algorithm trained on past user search history" with other indicators, such as an article's popularity and publication date, to attempt to deliver the most germane results and sort them by relevance.[22]

Librarians have much to bring to the table in designing, enhancing, and selecting appropriate ML-powered search tools. Librarians' specialized skill sets in managing information could be redirected towards automating processes that remain largely manual. For example, librarians' expertise working with controlled vocabularies and mapping ontologies could be productively applied to training ML models that facilitate interdisciplinary search. Their information literacy and search expertise can help scholars productively select appropriate search tools depending on their goals (for example, a comprehensive literature review versus getting quickly caught up on a topic). Libraries can help ensure that scholars and students understand the limitations and downsides of ML-enhanced search, reminding them that "[b]lindly using any research engine doesn't answer every question automatically."[23]

Perhaps more significantly, libraries can offer their attention to the values of transparency and integrity in the scholarly research process. "Explainable" or "human-centered AI" have emerged as the bywords for transparency and integrity in algorithm-based information

tools, and are cited as a crucial feature of the services that libraries acquire, license, or otherwise support.[24] In the context of search and discovery, human-centered AI reveals the "thought process" behind the algorithm, making it clear to the user why they are seeing a certain set of search results, and gives the user some level of control over the algorithm. For example, transparent discovery interfaces might allow users to "adjust the parameters of an algorithm being applied to a collection."[25] One of the experts interviewed for this report underscored the risk of "black box" algorithms to the integrity of the research process, explaining that "once we're in the bot-driven world, it would be trivial for businesses running those bots to tweak algorithms to privilege research from their own publications, and there would be incentives for them to do that."[26]

The promise of ML to enhance discovery goes beyond search tools. Scholars are also turning to a range of emerging technologies that, in the words of one expert interviewed for this report, "distill an overwhelming amount of content into something meaningful and manageable."[27] These include automated text-processing technologies, recommendation systems, and virtual assistants and conversational agents.

Automated text processing

ML tools can generate increasingly accurate content summaries using techniques that are extractive (in which the model abridges text by distinguishing relevant and irrelevant passages) and abstractive (in which the model attempts to interpret and paraphrase content). Google's TensorFlow machine-learning library can perform both types of summarization with high accuracy,[28] and commercial services like Scholarcy have emerged to allow non-computer scientists to take advantage of automated text summarization.[29]

The applications of such tools are clear for scholars striving to keep up with recent publications in their fields. Automated text summarization, perhaps to a greater extent than a human-generated abstract, can help them digest more content at a superficial level and determine which

content demands a closer read. The applications for digital libraries are also apparent. At Virginia Tech (VT), for example, the University Libraries and the Digital Library Research Laboratory partnered with a computer science course in fall 2018 to experiment with deep learning models to generate chapter-level summaries for a corpus of VT's electronic theses and dissertations (ETDs).[30] Automated generation of text summaries has the potential to greatly enhance discovery of textual materials in digital libraries and save countless hours of human labor.

Beyond summarization, automated text processing can help researchers discover new meaning and hidden connections in existing texts. For instance, a team of researchers ran a corpus of abstracts in materials science through the Word2vec unsupervised machine-learning algorithm. By associating and clustering related terms, the algorithm replicated existing categories in the domain materials science without human intervention.[31] Next, the researchers successfully trained the algorithm to predict which of a set of materials was most likely to have thermoelectric properties based on its semantic relationships in the corpus. The end goal is to develop a method for scientists to generate hypotheses and glean new insights based on existing literature.

Automated text processing can also be used to make research more accessible to heterogeneous user communities. Researchers at MIT have developed a tool that uses NLP to "read scientific papers and produce a short summary in plain English,"[32] which may be particularly useful to scholars conducting cross-disciplinary research. Get the Research, a project from Impactstory, aims to use NLP to generate plain-language summaries of research for the general public.[33] Machine translation, which has become reliable enough that it can be used for "translating non-English medical studies into English for the systematic reviews that health-care decisions are based on," could be used to make critical research available in the languages of communities that can use it.[34]

Automated processing of scholarly literature will also impact the ways in which research is evaluated. Publishers and publishing-service providers are increasingly exploring the potential of automated text processing to streamline operations, improve discoverability, and add value to their products. Meta Bibliometric Intelligence, for example, uses machine learning to extract likely topics from a submitted manuscript, gauge its relevance to the journal, and predict its impact, all in the name of streamlining editorial workflows and decision-making. An ML-powered tool developed by Scite.ai "automatically detects whether an article's citing papers were written in support or contradiction of the cited article claims."[35] As tools like these demonstrate proficiency, they might be incorporated into researcher evaluation systems, tenure and promotion decisions, and other determinants of scholarly merit. As with most ML tools, this presents both tremendous opportunities and risks. On the one hand, ML tools could provide a more accurate and nuanced understanding of a work's reception in the scholarly community. On the other, they can replicate and amplify biases, be prone to error or manipulation, and further alienate human judgment from critical decisions that affect a scholar's career.

Approaches to machine-**generated** text have also come a long way in recent years. An October 2019 New Yorker article used a predictive text algorithm to co-author an article on the future of writing in a post-AI world;[36] in early 2019 Springer Nature published a proof-of-concept machine-generated book that used abstractive text summarization to peruse a corpus of articles on lithium-ion batteries and produce a general overview of the topic.[37] In the near future, a machine may author the first draft of a researcher's manuscript, automating the rote work of describing materials and methodology. Manuscript Writer,[38] an AI-based tool from the company SciNote, has already proven successful at drafting the introduction, methodology, results, and references sections of a scientific article, liberating the researcher to focus on interpreting the results and writing the discussion section.[39]

Recommendation systems

One strength of ML algorithms is their ability to dynamically adjust and adapt as they receive new inputs. ML enables digital services that tailor themselves to their users; rather than mass produced and generic, ML allows web content to be "customized based on individual users' personas, needs, wishes, and traits—an approach known as mass personalization."[40]

Recommendation systems are one manifestation of mass personalization. ML-powered recommenders can suggest resources based on a user's query or based on the system's understanding of a user's preferences and interests. Such systems have proliferated in the context of e-commerce, streaming media, and social media sites. They seem particularly well suited for library discovery systems, given that researchers are frequently looking for all available content that relates to their research interests. Search platforms for academic literature increasingly incorporate recommendation systems as a complementary discovery tool (for example, Mendeley and Ex Libris's bX Article Recommender). Stand-alone applications, like Meta (backed by the Chan Zuckerberg Initiative)[41] and the recently released Scitrus platform,[42] provide a curated feed of content based on the system's evolving understanding of the user's interests.

While recommendation systems hold promise for streamlining the research process and enhancing serendipitous discovery, they rely on intensive collection and analysis of user data, which can compromise user privacy in ways that are anathema to most libraries. Specifically, recommendation engines, and other discovery systems that rely on personal data, can be perceived as compromising libraries' commitment to open inquiry, which requires the searcher to feel unconstrained by surveillance, and to have agency in the discovery process.[43] Linked data infrastructure, on the other hand, can embed the same types of "meaningful relationships" as recommendation engines, but in a way that "reflects some level of systematic thought and consensus within and among domains of knowledge."[44] Research has

shown that students have a complicated relationship with algorithm-driven platforms, including discovery systems, and express a mixture of discomfort and resignation to the idea of being tracked online.[45]

Despite these risks, Clifford Lynch cautions libraries against "taking an absolutist approach to information collection, as opposed to more nuanced, transparent, and opt-in collection of data about user activities and interests," arguing that a refusal to provide convenient and sophisticated search tools may only serve to drive users away.[46] Instead, libraries can develop and advocate for discovery systems that leverage the power and convenience of recommendation engines and other forms of personalization in ways that respect user privacy and facilitate open inquiry. Libraries are already undertaking projects that aim to provide such privacy-aware alternatives. For example, librarians at the University of Illinois at Urbana-Champaign developed an open source plug-in for the VuFind library discovery system that uses anonymized borrowing data to cluster related items and provide recommendations to users. Rather than tracking an individual user's history and habits, the system infers associations based on items checked out in a single transaction.[47] Libraries have also come up with creative recommendation engines that encourage information literacy and robust research skills. At the University of Tsukuba in Japan, for example, the libraries are developing a recommendation engine that will be installed as a browser plug-in on the library's computers and will suggest library materials based on Wikipedia articles the user has accessed.[48] The system uses a convolutional neural network to automatically classify Wikipedia articles and identify related content in the library's collections.

Libraries have an opportunity to contribute approaches to personalization that provide convenience and support information literacy while minimizing and disclosing risks to user privacy, providing transparent opt-in mechanisms, and prioritizing strong cybersecurity practices.

Virtual assistants and conversational user interfaces

The ways that researchers seek information are being shaped by the prevalence of conversational user interfaces and voice-controlled virtual assistants. Virtual assistants have rapidly become ubiquitous in homes and offices, and on the web. Smart devices like phones and speakers come equipped with voice-activated virtual assistants that can perform basic information retrieval tasks, interact with other smart devices like light switches and thermostats, and communicate with other web-based services. Chatbots embedded in websites proactively offer information and assistance. This class of tools, known as virtual assistants, chatbots, or conversational agents, among other terms, gives and receives information in the form of conversational speech, simulating interaction with a human.

Libraries have been experimenting with chatbots since at least the early 2000s.[49] Contemporary chatbots tend to manifest as a pop-up instant-message window in the corner of the library website. Chatbots can answer many fact-based reference questions, and may even be adept at answering more complex queries. A team of liaison librarians at McGill University, for example, has been exploring the effectiveness of commercial voice assistants (Siri, Google Assistant, and Alexa) at providing front-line research assistance.[50] Other libraries are also experimenting with leveraging commercially available virtual assistants to perform library-specific tasks. For example, the University of Oklahoma has developed an Alexa skill that "allows library users to perform a voice search of LibGuides or Primo using vendor APIs."[51]

While virtual assistants do not obviate human-to-human interaction, they can make it easier to provide individualized, point-of-need service to library users at scale; ease the anxiety some students may feel when approaching a librarian or initiating a research task;[52] and function as a digital triage system, automatically directing users to appropriate services and resources. Thus, a proactive virtual assistant invites engagement and provides a gateway for more substantive interactions with human librarians. Jeff Steely, dean of Georgia State University

Library, invoked chatbots as an example of an emerging technology that can make library services more student-centered, advising that "engagement with a chatbot is really about starting the conversation."[53]

Given well-structured and accurate source data, chatbots can rapidly and precisely answer transactional questions about library hours, the status of loans, or the location of a call number range at any time of day or night, from any location. However, they require significant up-front investment, both in developing their functionality and populating them with information. After all, "At its core, a chatbot is a library of answers that are organised to respond to the goals of its user. Poor organisation of the library of responses will negatively impact the responses the chatbot chooses."[54] Chatbots cannot currently approach human proficiency in making inferences, asking clarifying questions, or interpreting ambiguity. At this stage in their maturity, voice-controlled virtual assistants such as Google Assistant, Siri, and Alexa, provide poor user experience, especially beyond very basic queries.[55]

Given their limitations, chatbots are typically offered alongside conventional visual interfaces. That could eventually change. As conversational user interfaces become increasingly sophisticated, they may completely supplant visual interfaces. In this scenario, instead of visiting Google (or a library catalog) and entering a text-based query, a user might instead encounter a proactive chatbot that asks what the user is looking for. The chatbot processes a natural language statement (such as, "three or four references for an article I'm writing on Anglo-Saxon literature, specifically in Wessex") and asks follow-up questions to refine the search (such as, "Do you require only articles or other types of content? Do the articles need to be peer reviewed?").[56] Libraries will have a role refining and maintaining these conversational agents as well as in educating users to optimize their use.

Highlighted initiative

PubMed
National Library of Medicine
https://pubmed.ncbi.nlm.nih.gov/
PubMed's biomedical literature search tool combines a "state-of-the-art machine-learning algorithm trained on past user search history" with other indicators, such as an article's popularity and publication date, to attempt to deliver the most germane results and sort them by relevance, rather than recency.[57] Text snippets for each search result expose the algorithm's logic and make it easy for researchers to identify the most pertinent articles.

Reveal hidden digital collections through enhanced description

The acceleration of digitization and born-digital content creation has left libraries facing an ever-growing backlog of resource description. As libraries place increasing value on their unique local collections, they need new ways of making those collections discoverable to internal and external audiences, both human and machine. Accurate and comprehensive metadata are essential to the discovery, use, and preservation of digital collections, yet libraries lack the human resources to catalog content at the rate it is being created. Machine-learning approaches to automated metadata generation have shown promising results, opening up new possibilities for libraries to describe digitized collections of text, audio, and still and moving images at scale.

Discovery of textual materials has benefited greatly from advances in optical character recognition (OCR), which enables full-text search. However, structured metadata remains essential to discovery, making it easier for users to systematically identify pertinent items and enabling search aggregators to efficiently harvest and index content. To produce structured metadata at scale for large corpora of digitized texts, libraries are turning to NLP and named-entity recognition (NER) tools. At Northern Illinois University (NIU), the library is using NLP to extract topics from and generate subject headings for a collection of

tens of thousands of dime novels.[58] These materials would otherwise require intensive human effort to productively catalog. A similar project is underway at the Koninklijke Bibliotheek, the National Library of the Netherlands, where an NLP algorithm is being trained to apply subject tags to a collection of electronic dissertations.[59] At Singapore's National Library Board (NLB), an experimental initiative utilized NER to populate metadata records across several digital collections.[60] The NLB's NER system extracts the names of places, people, and organizations from a full-text document and compares them against a controlled vocabulary supplied by subject experts. Entities recognized by the system can then be added to an object's metadata record. The project has enriched the metadata of collections that had little to no prior cataloging, and has bolstered cross-collection discovery.

While many efforts focus on text processing, machine learning also has significant implications for processing collections of still and moving images and audio. The British Library Machine Learning Experiment site, launched in 2015 as a test bed for the library's digital research team, is using open source software and public-image recognition APIs to automatically process and tag a collection of over a million public domain images.[61] Japan's National Diet Library (NDL), under the auspices of its Next Digital Library project, has created an illustration search tool to automatically extract images and diagrams from its 30,000 digitized publications, and group similar images across the collection.[62] The Center for Open Data in the Humanities is using a deep-learning-based classification algorithm to extract images, and recognize facial expressions from its collection of digitized Japanese manuscripts.[63] In this instance, the research team chose deep learning (as distinct from machine learning) in order to allow the machine to identify patterns independently.

A collaborative initiative from the Indiana University Bloomington Libraries, the University of Texas at Austin, New York Public Library, and digital consultant AVP, funded by a grant from the Andrew W. Mellon Foundation, also aims to create metadata-generation

mechanisms for audiovisual content through an open source Audiovisual Metadata Platform (AMP).[64] To date, the project has piloted the application of "speech-to-text, named entity recognition, video OCR, speaker diarization, and speech/music/silence detection"[65] to a sample collection. Future work will include genre detection and instrument identification for digitized music and object detection for video. The National Library of Norway's Nancy initiative explores several vectors of machine learning for its cultural heritage collections, including a speech-to-text initiative that promises to make thousands of hours of radio broadcasts deeply searchable for the first time.[66]

Machine-learning approaches to metadata generation have been experimental since at least the 1980s, but the computing resources and technical expertise required to implement them presented significant barriers to wide adoption. Improvements in commercially available hardware, containerization technologies, the availability of public APIs and open source code, and the availability of high-speed networking on many university campuses have made it possible to implement machine-learning tools at scale. Using modern tools and computing resources equivalent to a standard laptop computer, a team of researchers indexed the 57 million pages of unstructured digitized text in the Biodiversity Heritage Library in 14 hours, an operation that previously took 45 days.[67]

The growth in available commercial machine-learning services can also lower barriers to entry in this space. Several of the initiatives described in this section rely on commercial cloud-based services for data processing. Amazon and Google both offer machine-learning services, as do dedicated vendors like Clarifai and Machine Box (which provides a containerized machine-learning environment). Microsoft has partnered with the Library of Congress and Israel's Ben-Gurion University of the Negev to apply machine learning to massive troves of digitized manuscripts.[68] The team behind the Audiovisual Metadata Platform (AMP) cautions that commercial machine-learning services lack transparency (using "black-box" algorithms to process data) and that vendor terms of service often require users to proactively opt-out

of allowing data reuse.[69] Further, they warn, commercial tools may not be suitable for library use cases without considerable modification.

Indeed, many of the projects referenced above have noted the considerable effort involved in producing machine-generated metadata that matches human accuracy and precision. Significant human intervention is still required in the form of tweaking algorithms, supplying pertinent training data, and performing quality control.[70] The NLB in Singapore undertook multiple rounds of iteration before it was confident in the performance of its NER tool. The University of Utah, which recently received a grant to develop and test a machine-learning tool for its historical image collection, will rely on nearly a half-million digitized images with existing, detailed, human-created metadata as a training corpus.[71] Well-resourced libraries could collectively develop "gold-standard" training data sets that could be broadly shared within the cultural heritage community as a step towards making this technology accessible to institutions of all sizes.[72]

Machine-**assisted** cataloging may be a productive middle ground in the near term. The NIU dime-novel project, for example, will "aggregate unusual keywords into different top-level dime-novel genres, like seafaring, Westerns, and romance," allowing human catalogers to make educated inferences about a novel and complete the catalog record.[73] Western Washington University (WWU) is using a commercial service, Clarifai, for machine-assisted description of photographs and videos in its Islandora digital repository.[74] During the ingest process, images are sent to the Clarifai server for processing. They are returned with a set of suggested tags (and their confidence intervals). Human repository administrators can add or remove suggested tags before publishing the content.

As libraries grapple with the thorny technical challenges of automated resource description, they will also face critical questions about policy and implementation. Poor-quality metadata can undermine researchers' confidence in the search process; overly broad subject tags, for example, could exacerbate rather than mitigate the problem of an

overabundance of material. Inaccurate metadata concerning locations, identities, or other factual information could have serious implications for research. Responsible approaches to integrating machine-generated metadata will therefore require clear indications to users. The British Library's machine-learning-powered search interface illustrates one approach: each metadata record includes a set of hand-created metadata fields and a clearly designated set of machine-generated tags with their corresponding confidence interval.

Perhaps more importantly, libraries will face ethical and privacy issues as they apply ML algorithms to their digital collections. Algorithms are prone to adopt and amplify biases, and are only as good as their training data.[75] Facial recognition and NER present even more significant concerns. Thoughtful policies about when and how ML is applied to library collections, and under what conditions it may be removed, can help libraries move forward on solid footing (for example, takedown notices for machine-generated metadata, particularly any metadata derived from facial recognition or NER, which might inappropriately identify living people, perpetuate biases, or expose sensitive information). ML techniques can also be applied to bolster data privacy (for example, using algorithms to automatically identify suspected Social Security numbers or other sensitive information in troves of digitized documents).

At this stage of maturity, automated metadata generation may be particularly advantageous as a "good-enough" tool for describing resources that might otherwise remain uncataloged. Though the quality and precision of machine-generated metadata may not yet match human-created metadata, its potential to describe collections at scale, to provide a minimum level of description for digitized objects that would otherwise remain hidden, represents a watershed moment for cultural heritage organizations. This is an opportunity for reflection on the ethical and privacy implications of machine processing massive volumes of digitized material.

Visual information has proliferated over the past several decades, from mass digitization of historical image collections, to the millions of digital photos and videos uploaded each day from personal electronic devices. Computer-vision technologies, often powered by convolutional neural networks, provide new ways of processing and exploring this deluge of information. Computer vision is an umbrella term that encompasses attempts to computationally replicate the human visual system and automate visual tasks, such as pattern and known-entity recognition.[76] Computer vision is already being used to detect cancer and other illnesses, identify wildlife whose images are caught on trail cameras, guide self-driving vehicles, and inspect food quality, among other experimental uses. Within the cultural heritage sector, computer vision can enable a range of novel approaches to visual-resource description, analysis, and discovery, giving researchers a range of options beyond text-based search (lexical or semantic). Libraries can apply these techniques to their own collections, enhancing broad discovery of visual materials, and support faculty projects that aim to process digital images at scale.

As discussed in the section on automated resource description, ML models have shown promise for identifying objects and known entities in visual materials, retrieving or grouping similar images, and generating topical or thematic metadata. Computer-vision techniques can be applied to digitized still images, moving images, textual documents that contain embedded figures, and even collections of 3D data, which will benefit from shape-based retrieval mechanisms that identify similar objects.[77] A number of notable projects are successfully using computer-vision techniques to engage with library collections.

As part of the Mellon-funded Collections as Data: Part to Whole project, a team at Harvard University and the University of Richmond will implement computer-vision techniques to analyze born-digital ephemera relating to the rise of nationalist and anti-immigrant movements in Europe.[78] The project's goals include "expanding the processing of digital images and subsequent algorithmic discovery of connections across collections." Notably, the project also explicitly aims

to "illustrate how distant viewing can offer a paradigm for addressing the social and ethical challenges of using machine learning with images, particularly of sensitive topics."

At Yale University Library's Digital Humanities Laboratory, Doug Duhaime, Monica Ong Reed, and Peter Leonard, have used a convolutional neural network to analyze images from the Meserve-Kunhardt Collection of 19th-century photography at the Beinecke Rare Book and Manuscript Library.[79] While the typical end-result of this process would be a text-based caption or description of the image, in this case the researchers were interested in the penultimate level of interpretation, which clusters similar images together. They present the results in a visual interface that allows visual exploration of the photographs in a dynamic website. The related PixPlot tool, also developed at the Yale Digital Humanities Lab, offers an alternative visualization of the entire collection as a dynamic map of content, plotted based upon similarity, which allows pattern recognition at a glance.[80]

At Dartmouth College, researchers are working with a collection of films held by the library and the Internet Archive to develop a tool that allows users to search within moving images just like they would search for keywords in a document. The tool "takes search queries expressed in textual form and automatically translates them into image recognition models that can identify the desired segments in the film."[81]

In addition to digitized and born-digital special collections content, computer vision also has applications for digging into the published literature. Scientists have used computer vision to analyze diagrams, visualizations, and images embedded in scientific papers, for the purposes of enabling new discovery and engaging in viziometrics research, or the study of the "organization and presentation of visual information in the scientific literature."[82]

So far, the deep neural networks (DNN) that underlie computer-vision technology remain fragile and easy to fool. Researchers have shown that changing a few select pixels can cause a DNN to interpret an image

of a lion as an image of a library, for example.[83] And computer-vision models, like other ML tools, are not optimized for use with cultural heritage materials. In collaboration with other cultural heritage institutions, and possibly with industry, libraries have an opportunity to contribute to building more appropriate training corpora, refining and testing models, and exploring the ethical and policy implications of broadly applying computer vision to their collections.

While the experiments described above are being run on carefully selected corpora by small groups of researchers, this type of functionality may eventually become commonplace in discovery and digital-asset management systems at scale. Libraries have a dual opportunity, supporting innovative, one-of-a-kind projects while generalizing the most promising methodologies and making them broadly available to researchers.

Highlighted initiatives

Audiovisual Metadata Platform (AMP)
Indiana University Libraries, the University of Texas at Austin, New York Public Library
https://wiki.dlib.indiana.edu/display/AMP
The collaborative AMP initiative, funded by a grant from the Andrew W. Mellon Foundation, aims to create metadata-generation mechanisms for audiovisual content. To date, the project has piloted the application of speech-to-text; named-entity recognition; video OCR; speaker diarization; and speech, music, and silence detection to a sample corpus.

Image Analysis for Archival Discovery (Aida)
University of Nebraska–Lincoln and University of Virginia
http://projectaida.org/
The Aida project explores the application of neural networks to digitized library collections, particularly historic newspapers. The project has demonstrated success in identifying poetry from digitized

newspaper images. The team's proof-of-concept suggests that libraries could eventually provide just-in-time, dynamically extracted content from their digitized collections.

Neural Neighbors
Yale University Library Digital Humanities Lab
https://dhlab.yale.edu/projects/neural-neighbors/
The Neural Neighbors project applies machine-vision techniques to a rich collection of 19th-century photographs to identify patterns and similarities, enabling new approaches to visual information discovery and analysis.

Sheeko
The University of Utah
https://sheeko.org/
Sheeko provides a suite of pre-trained ML models for automating image description as well as tools for users to automate the training of their own models.

Expose library collections and services beyond library systems

As information becomes distributed, diversified, and open, many researchers prefer web-scale discovery tools that aggregate resources from a range of sources over siloed library catalogs and digital-asset management systems.[84] 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 (SEO) techniques; exposing metadata for harvesting by aggregators, such as the Digital Public Library of America; providing APIs that permit new forms of computational engagement with collections; adopting interoperability standards, such as the International Image Interoperability Framework (IIIF),[85] 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"[86] and one component of what other authors have termed the "library as platform,"[87] is a natural consequence of an open, oversaturated, and networked information landscape. The library's role in content management is being reenvisioned: no longer the steward of a unified collection, the library becomes the facilitator of a networked suite of open and extensible tools, resources, and services. Homegrown and manually maintained discovery systems may become less desirable to maintain as users increasingly turn to web-scale services and as emerging technologies enable more sophisticated discovery mechanisms. The academic library's facilitation services and interactions may supersede its role as a local content collector. Among the core functions of this role is advancing interoperability. Research library collaboration with interoperable repositories of data, preprints, and publications ensures that local troves of knowledge become discoverable at scale. Expertise in metadata and standards development can be contributed to maintaining and enhancing interoperability standards. Librarians' relationships with faculty and students on campus position them well to encourage adoption of persistent identifiers like ORCID IDs that help power interoperable discovery infrastructure, and the use of interoperable metadata schemas in faculty research.

In this vision of academic library services, the library no longer represents a "portal we go through on occasion, but...infrastructure that is as ubiquitous and persistent as the streets and sidewalks of a town."[88] The end users of this infrastructure will increasingly include both humans and machines.[89] A less institutionally driven approach to discovery might include working with vendor-supplied APIs to develop shared discovery layers, contributing to large-scale linked open data initiatives, or collectively developing systems that fill gaps in the discovery ecosystem, such as discovery of open access content. Academic libraries' existing expertise in standards and interoperability will be crucial as they participate in and enhance the "broader scholarly ecosystem, which only works through these frameworks."[90]

Highlighted initiative

Enslaved: Peoples of the Historic Slave Trade
Matrix, the Center for Digital Humanities and Social Sciences at Michigan State University
http://enslaved.org/
The Enslaved project uses linked data to aggregate materials related to the transatlantic slave trade from a distributed network of library and archives partners. Bringing together disparate resources through linked data creates unprecedented opportunities for scholarly discovery and analysis, and brings light to the histories of underrepresented individuals and issues.[91]

Key Takeaways

1. **Libraries will retain a critical role in information discovery and facilitated access, even as locally acquired collections[92] diminish in importance.** The experts interviewed for this report overwhelmingly asserted that discovery will remain core to the identity and service model of the academic and research library, albeit in different and expanded ways.
2. **ML and NLP technologies will facilitate new forms of search, discovery, and academic inquiry.** At best, these technologies create exciting new modes of inquiry, facilitate cross-disciplinary discovery, and make research more efficient and productive. However, they have the potential to suppress human agency in the research process, amplify biases, and expose users to data-privacy violations.
3. **Library expertise can be effectively redirected towards creating and maintaining computationally ready digital collections that facilitate discovery, analysis, and use.** Libraries' expertise in creating and managing structured data can be effectively utilized to make local collections discoverable in web-scale discovery systems through more widespread adoption of APIs and linked open data. That expertise can also be used to

make digital assets more discoverable through the application of ML tools to resource description. Resources formerly invested in maintaining local catalogs might be repurposed into the purchase, licensing, or development of ML-enhanced search, discovery, and recommendation systems; compiling relevant training data sets for ML models; training virtual research assistants; and enabling other novel approaches to information retrieval and processing.

Endnotes

1. Lorcan Dempsey and Constance Malpas, "Academic Library Futures in a Diversified University System," in *Higher Education in the Era of the Fourth Industrial Revolution*, ed. Nancy W. Gleason, 65–89 (Singapore: Springer, 2018), https://doi.org/10.1007/978-981-13-0194-0_4.

2. Stephen Pinfield, Andrew M. Cox, and Sophie Rutter, *Mapping the Future of Academic Libraries: A Report for SCONUL* (London: SCONUL, 2017), https://sconul.ac.uk/publication/mapping-the-future-of-academic-libraries.

3. Anna Gold, "Cyberinfrastructure, Data, and Libraries, Part 2," *D-Lib Magazine* 13, no. 9–10 (September–October 2007), https://doi.org/10.1045/july20september-gold-pt2.

4. Michael Ridley, "Explainable Artificial Intelligence," *Research Library Issues*, no. 299 (2019): 28–46, https://doi.org/10.29242/rli.299.3.

5. Deanna C. Pineau, "Math-Aware Search Engines: Physics Applications and Overview," preprint, submitted September 8, 2016, http://arxiv.org/abs/1609.03457.

6. Dig That Lick website, accessed April 8, 2020, http://dig-that-lick.eecs.qmul.ac.uk/.

7. Kristin Antelman, interview by author, November 15, 2019.

8. Roger C. Schonfeld, "Does Discovery Still Happen in the Library? Roles and Strategies for a Shifting Reality," *Ithaka S+R Blog*, September 24, 2014, https://sr.ithaka.org/blog/does-discovery-still-happen-in-the-library-roles-and-strategies-for-a-shifting-reality/.

9. Andrew M. Cox, Stephen Pinfield, and Sophie Rutter, "The Intelligent Library: Thought Leaders' Views on the Likely Impact of Artificial Intelligence on Academic Libraries," *Library Hi Tech* 37, no. 3 (2019): 418–35, https://doi.org/10.1108/LHT-08-2018-0105.

10. Lorcan Dempsey, "Libraries and the Informational Future: Some Notes," *Information Services & Use* 32, no. 3–4 (2012): 203–14, https://doi.org/10.3233/ISU-2012-0670.

11. Cox, Pinfield, and Rutter, "The Intelligent Library."

12. Gwen Evans and Roger C. Schonfeld, *It's Not What Libraries Hold; It's Who Libraries Serve: Seeking a User-Centered Future for Academic Libraries*, Issue Brief (Columbus, OH, and New York, NY: OhioLINK and Ithaka S+R, January 23, 2020), https://doi.org/10.18665/sr.312608.

13. Marshall Breeding, *The Future of Library Resource Discovery: A White Paper Commissioned by the NISO Discovery to Delivery (D2D) Topic Committee* (Baltimore: National Information Standards Organization, February 2015), https://www.niso.org/publications/future-library-resource-discovery.

14. Schonfeld, "Does Discovery Still Happen in the Library?"

15. Eszter Hargittai, interview by author, November 27, 2019.

16. Yewno website, accessed April 10, 2020, https://www.yewno.com/.

17. Iris.ai website, accessed April 10, 2020, https://iris.ai/.

18. Dimensions website, accessed April 10, 2020, https://www.dimensions.ai/.

19. Semantic Scholar website, accessed April 10, 2020, https://www.semanticscholar.org/.

20. Andy Extance, "How AI Technology Can Tame the Scientific Literature," *Nature* 561 (September 2018): 273–74, https://doi.org/10.1038/d41586-018-06617-5.

21. Talk to Books website, accessed April 10, 2020, https://books.google.com/talktobooks/.

22. Nicolas Fiorini et al., "PubMed Labs: An Experimental System for Improving Biomedical Literature Search," *Database: The Journal*

of Biological Databases and Curation 2018 (September 18, 2018), https://doi.org/10.1093/database/bay094.

23. Extance, "How AI Technology Can Tame the Scientific Literature."
24. Ridley, "Explainable Artificial Intelligence."
25. Thomas Padilla, *Responsible Operations: DataScience, Machine Learning, and AI in Libraries* (Dublin, OH: OCLC Research, 2019), https://doi.org/10.25333/xk7z-9g97.
26. Antelman, interview by author.
27. Keith Webster, interview by author, November 19, 2019.
28. Peter Liu and Xin Pan, "Text Summarization with TensorFlow," *Google AI Blog*, August 24, 2016, http://ai.googleblog.com/2016/08/text-summarization-with-tensorflow.html.
29. Scholarcy website, accessed April 10, 2020, https://www.scholarcy.com/.
30. Naman Ahuja et al., "Big Data Text Summarization: Using Deep Learning to Summarize Theses and Dissertations," VTechWorks, December 5, 2018, http://hdl.handle.net/10919/86406.
31. Olexandr Isayev, "Text Mining Facilitates Materials Discovery," *Nature* 571 (July 2019): 42–43, https://doi.org/10.1038/d41586-019-01978-x.
32. Elliot Jones, Nicolina Kalantery, and Ben Glover, *Research 4.0: Interim Report* (London: Demos, October 2019), https://demos.co.uk/wp-content/uploads/2019/10/Jisc-OCT-2019-2.pdf.
33. Rick Anderson, "Get The Research: Impactstory Announces a New Science-Finding Tool for the General Public," *Scholarly Kitchen*, November 12, 2018, https://scholarlykitchen.sspnet.org/2018/11/12/get-the-research-impactstory-announces-a-new-science-finding-tool-for-the-general-public/.

34. John Seabrook, "The Next Word: Where Will Predictive Text Take Us?," A Reporter at Large, *New Yorker*, October 14, 2019, https://www.newyorker.com/magazine/2019/10/14/can-a-machine-learn-to-write-for-the-new-yorker.

35. Arthur "A.J." Boston, "What Do You Mean? Research in the Age of Machines," *College & Research Libraries News* 80, no. 10 (November 2019): 565–68, https://doi.org/10.5860/crln.80.10.565.

36. Seabrook, "The Next Word."

37. Lettie Y. Conrad, "The Robots Are Writing: Will Machine-Generated Books Accelerate Our Consumption of Scholarly Literature?," *Scholarly Kitchen*, June 25, 2019, https://scholarlykitchen.sspnet.org/2019/06/25/the-robots-are-writing-will-machine-generated-books-accelerate-our-consumption-of-scholarly-literature/.

38. "Manuscript Writer by SciNote," SciNote, accessed April 10, 2020, https://www.scinote.net/manuscript-writer/.

39. Jones, Kalantery, and Glover, *Research 4.0: Interim Report*.

40. Nitin Mittal and Dave Kuder, "AI-Fueled Organizations," in *Tech Trends 2019: Beyond the Digital Frontier*, ed. Bill Briggs and Scott Buchholz, Deloitte Insights (Deloitte Development, 2019), 21–39, https://www2.deloitte.com/be/en/pages/technology/articles/tech-trends-2019-beyond-the-digital-frontier.html.

41. Meta website, accessed April 10, 2020, https://www.meta.org/.

42. Scitrus website, accessed April 10, 2020, https://www.scitrus.com/.

43. D. Grant Campbell and Scott R. Cowan, "The Paradox of Privacy: Revisiting a Core Library Value in an Age of Big Data and Linked Data," *Library Trends* 64, no. 3 (Winter 2016): 492–511, https://muse.jhu.edu/article/613920.

44. Campbell and Cowan, "The Paradox of Privacy."

45. James F. Hahn, "User Perspectives on Personalized Account-Based Recommender Systems" (paper presented at ACRL 2019 Conference, Cleveland, OH, April 10–13, 2019), http://hdl.handle.net/2142/102364; Alison J. Head, Barbara Fister, and Margy MacMillan, *Information Literacy in the Age of Algorithms: Student Experiences with News and Information, and the Need for Change* (Project Information Literacy Research Institute, January 15, 2020), https://www.projectinfolit.org/uploads/2/7/5/4/27541717/algoreport.pdf.

46. Clifford A. Lynch, "Reader Privacy: The New Shape of the Threat," *Research Library Issues*, no. 297 (2019): 7–14, https://doi.org/10.29242/rli.297.2.

47. Jim Hahn and Courtney McDonald, "Account-Based Recommenders in Open Discovery Environments," *Digital Library Perspectives* 34, no. 1 (2018): 70–76, https://doi.org/10.1108/DLP-07-2017-0022.

48. Keita Tsuji, "Book Recommender System for Wikipedia Article Readers in a University Library," in *2019 8th International Congress on Advanced Applied Informatics* (IIAI-AAI) (IEEE, 2019): 121–26, https://doi.org/10.1109/IIAI-AAI.2019.00034.

49. Victoria L. Rubin, Yimin Chen, and Lynne Marie Thorimbert, "Artificially Intelligent Conversational Agents in Libraries," *Library Hi Tech* 28, no. 4 (2010): 496–522, https://doi.org/10.1108/07378831011096196.

50. Amanda Wheatley and Sandy Hervieux, "Need Research? Ask Siri: Evaluating the Impact of Virtual Assistant AI on the Future of the Research Process" (slides presented at NFAIS 2019 AI Conference, Alexandria, VA, May 16, 2019), https://nfais.memberclicks.net/assets/AI2019/Amanda%20and%20Sandy.pdf.

51. Twila Camp and Tim Smith, "Ready or Not: Here Comes Voice Search" (presented by Carl Grant, CNI Fall 2019 Membership Meeting, Washington, DC, December 9, 2019), https://www.cni.

org/topics/information-access-retrieval/ready-or-not-here-comes-voice-search.

52. Indra Ayu Susan Mckie and Bhuva Narayan, "Enhancing the Academic Library Experience with Chatbots: An Exploration of Research and Implications for Practice," *Journal of the Australian Library and Information Association* 68, no. 3 (2019): 268–77, https://doi.org/10.1080/24750158.2019.1611694.

53. Joan Lippincott et al., "Library Perspectives on the EDUCAUSE 2019 Top 10 IT Issues," *EDUCAUSE Review*, February 11, 2019, https://er.educause.edu/articles/2019/2/library-perspectives-on-the-educause-2019-top-10-it-issues.

54. Mckie and Narayan, "Enhancing the Academic Library Experience with Chatbots."

55. Raluca Budiu and Page Laubheimer, "Intelligent Assistants Have Poor Usability: A User Study of Alexa, Google Assistant, and Siri," *Nielsen Norman Group* (blog), July 22, 2018, https://www.nngroup.com/articles/intelligent-assistant-usability/.

56. Lisa Janicke Hinchliffe, Jason Griffey, Emily King, and Michael Schofield, "Is the Researcher Human? Is the Librarian? Bots, Conversational User Interfaces, and Virtual Research Assistants" (presentation, CNI Spring 2017 Membership Meeting, Albuquerque, New Mexico, April 3, 2017), https://www.cni.org/topics/information-access-retrieval/is-the-researcher-human-is-the-librarian-bots-conversational-user-interfaces-and-virtual-research-assistants.

57. Fiorini et al., "PubMed Labs."

58. Matthew Short, "Text Mining and Subject Analysis for Fiction; or, Using Machine Learning and Information Extraction to Assign Subject Headings to Dime Novels," *Cataloging & Classification Quarterly* 57, no. 5 (2019): 315–36, https://doi.org/10.1080/01639374.2019.1653413.

59. Martijn Kleppe et al., *Exploration Possibilities: Automated Generation of Metadata* (The Hague: National Library of the Netherlands, August 23, 2019), https://doi.org/10.5281/zenodo.3375192.

60. Rachael Goh, "Using Named Entity Recognition for Automatic Indexing" (paper presented at IFLA WLIC 2018: "Transform Libraries, Transform Societies," Kuala Lumpur, Malaysia, 2018), http://library.ifla.org/id/eprint/2214.

61. British Library Machine Learning Experiment website, accessed April 10, 2020, http://blbigdata.herokuapp.com/.

62. Next Digital Library website, accessed April 10, 2020, https://lab.ndl.go.jp/dl/.

63. Asanobu Kitamoto, "Facial Collection (Face Collection)," Center for Open Data in the Humanities, accessed April 9, 2020, http://codh.rois.ac.jp/face/.

64. AMPPD (Audiovisual Metadata Platform Pilot Development) website, last modified November 20, 2019, https://wiki.dlib.indiana.edu/pages/viewpage.action?pageId=531699941.

65. Jon Dunn and Shawn Averkamp, "Commercial ML Tools in Metadata Production" (slides presented at Machine Learning + Libraries Summit, Washington, DC, September 20, 2019), https://wiki.dlib.indiana.edu/display/AMP/AMP+Presentations?preview=%2F549127083%2F549127086%2FAMP+LC+ML%2BLibraries+2019-09-20.pdf.

66. Svein Arne Brygfjeld, "AI: Lessons from the National Library of Norway" (slides presented at SCONUL Summer Conference 2019, Manchester, UK, June 12, 2019), https://www.slideshare.net/secret/x2eTpP3OTUHzYi.

67. Dmitry Mozzherin, Alexander A. Myltsev, and David Patterson, "Finding Scientific Names in Biodiversity Heritage Library, or How to Shrink Big Data," *Biodiversity Information Science and Standards* 3 (2019), https://doi.org/10.3897/biss.3.35353.

68. Zachary Keyser, "Microsoft Implementing AI in Creating Archive of Ben-Gurion's Handwritten Works," Israel News, *Jerusalem Post*, November 6, 2019, https://www.jpost.com/Israel-News/Microsoft-implementing-AI-in-creating-archive-of-Ben-Gurions-handwritten-works-607013.

69. Dunn and Averkamp, "Commercial ML Tools in Metadata Production."

70. Goh, "Using Named Entity Recognition"; Clifford A. Lynch, "Machine Learning, Archives and Special Collections: A High Level View," *ICA Blog*, International Council on Archives, October 2, 2019, https://blog-ica.org/2019/10/02/machine-learning-archives-and-special-collections-a-high-level-view/.

71. Valeri Craigle, "Law Libraries Embracing AI," in *Law Librarianship in the Age of AI*, ed. Ellyssa Kroski (Chicago: American Library Association, 2019), http://dx.doi.org/10.2139/ssrn.3381798.

72. Padilla, *Responsible Operations*; Michael Ridley, "Training Datasets, Classification, and the LIS Field," *Library AI* (blog), September 26, 2019, https://libraryai.blog.ryerson.ca/2019/09/26/training-datasets-classification-and-the-lis-field/.

73. Short, "Text Mining and Subject Analysis for Fiction."

74. "IBU (Islandora Batch Uploader)," Western Washington University, accessed April 9, 2020, https://mabel.wwu.edu/ibu.

75. Safiya Umoja Noble, *Algorithms of Oppression: How Search Engines Reinforce Racism* (New York: NYU Press, 2018).

76. Thomas S. Huang, "Computer Vision: Evolution and Promise," in *19th CERN School of Computing: Proceedings*, ed. Carlo E. Vandoni (Geneva : CERN, 1996), 21–25, https://doi.org/10.5170/CERN-1996-008.21.

77. David Koller, Bernard Frischer, and Greg Humphreys, "Research Challenges for Digital Archives of 3D Cultural Heritage Models,"

Journal on Computing and Cultural Heritage 2, no. 3 (December 2009): 7:1–7:17, https://doi.org/10.1145/1658346.1658347.

78. Collections as Data—Part to Whole, "Announcing Collections as Data Cohort 2," January 6, 2020, https://collectionsasdata.github.io/part2whole/cohort/.

79. Peter Leonard, "Neural Networks: Machine Vision for the Visual Archive" (presentation at CNI Spring 2018 Membership Meeting, San Diego, CA, March 13, 2018), https://www.cni.org/topics/special-collections/neural-networks-machine-vision-for-the-visual-archive.

80. PixPlot project webpage, Yale University Library Digital Humanities Laboratory, accessed April 9, 2020, https://dhlab.yale.edu/projects/pixplot/.

81. *Digital Humanities 2018: Puentes—Bridges: Book of Abstracts/Libro de resúmenes* (Mexico City: Red de Humanidades Digitales, 2018), https://dh2018.adho.org/abstracts.

82. Po-Shen Lee, Jevin D. West, and Bill Howe, "Viziometrics: Analyzing Visual Information in the Scientific Literature," *IEEE Transactions on Big Data* 4, no. 1 (March 2018): 117–29, https://doi.org/10.1109/TBDATA.2017.2689038.

83. Douglas Heaven, "Why Deep-Learning AIs Are So Easy to Fool," *Nature* 574 (October 2019): 163–66, https://doi.org/10.1038/d41586-019-03013-5.

84. David Attis and Colin Koproske, "Thirty Trends Shaping the Future of Academic Libraries," *Learned Publishing* 26, no. 1 (January 2013): 18–23, https://doi.org/10.1087/20130104; John Akeroyd, "Discovery Systems: Are They Now the Library?," *Learned Publishing* 30, no. 1 (January 2017): 87–89, https://doi.org/10.1002/leap.1085.

85. Stuart Snydman, Robert Sanderson, and Tom Cramer, "The International Image Interoperability Framework (IIIF): A

Community & Technology Approach for Web-Based Images," in *Archiving Conference*, vol. 2015 (Springfield, VA: Society for Imaging Science and Technology, 2015), 16–21.

86. Dempsey, "Libraries and the Informational Future."

87. Schonfeld, "Does Discovery Still Happen in the Library?"; David Weinberger, "Library as Platform," *Library Journal*, September 4, 2012, https://www.libraryjournal.com?detailStory=by-david-weinberger.

88. Weinberger, "Library as Platform."

89. Chris Bourg, "What Happens to Libraries and Librarians When Machines Can Read All the Books?," *Feral Librarian* (blog), March 16, 2017, https://chrisbourg.wordpress.com/2017/03/16/what-happens-to-libraries-and-librarians-when-machines-can-read-all-the-books/; Cox, Pinfield, and Rutter, "The Intelligent Library."

90. Carole Palmer, interview by author, October 30, 2019.

91. Amy Crawford, "A Massive New Database Will Connect Billions of Historic Records to Tell the Full Story of American Slavery," *Smithsonian Magazine*, January 2020, https://www.smithsonianmag.com/history/massive-new-database-connect-billions-historic-records-tell-full-story-american-slavery-180973721/.

92. Lorcan Dempsey, "Library Collections in the Life of the User: Two Directions," *LIBER Quarterly* 26, no. 4 (October 11, 2016): 338–59, https://doi.org/10.18352/lq.10170.