

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

---

Copyright, Fair Use, Scholarly Communication,  
etc.

Libraries at University of Nebraska-Lincoln

---

12-17-2018

## A Grand Challenges-Based Research Agenda for Scholarly Communication and Information Science [MIT Grand Challenge PubPub Participation Platform]

Micah Altman

*Massachusetts Institute of Technology*

Chris Bourg

*Massachusetts Institute of Technology*

Follow this and additional works at: <https://digitalcommons.unl.edu/scholcom>



Part of the [Intellectual Property Law Commons](#), [Publishing Commons](#), [Scholarly Communication Commons](#), and the [Scholarly Publishing Commons](#)

---

Altman, Micah and Bourg, Chris, "A Grand Challenges-Based Research Agenda for Scholarly Communication and Information Science [MIT Grand Challenge PubPub Participation Platform]" (2018). *Copyright, Fair Use, Scholarly Communication, etc.*. 216.  
<https://digitalcommons.unl.edu/scholcom/216>

This Article is brought to you for free and open access by the Libraries at University of Nebraska-Lincoln at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Copyright, Fair Use, Scholarly Communication, etc. by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

**MIT Grand Challenge PubPub Participation Platform**

# **A Grand Challenges-Based Research Agenda for Scholarly Communication and Information Science**

**Micah Altman, Chris Bourg**

**Published on:** Dec 17, 2018

**DOI:** 10.21428/62b3421f

**License:** [Creative Commons Attribution 4.0 International License \(CC-BY 4.0\)](https://creativecommons.org/licenses/by/4.0/)

*Contributors\* to this report, listed alphabetically: Micah Altman, Chris Bourg, Philip Cohen, G. Sayeed Choudhury, Charles Henry, Sue Kriegsman, Mary Minow, Daisy Selematsela, Anasuya Sengupta, Peter Suber, Ece Turnator, Suzanne Wallen, Trevor Owens, and David Weinberger.*

The workshop and paper are supported by a grant from The Andrew W. Mellon Foundation. Thanks to the Program Committee for scoping and mapping the original framing for the conversations: Micah Altman, Christine Borgman, Chris Bourg, G. Sayeed Choudhury, Charles (Chuck) Henry, Abby Smith Rumsey, and Ethan Zuckerman; and to the keynote speakers, whose remarks framed each discussion: Kate Zwaard, Anasuya Sengupta, and Joi Ito.

Also thanks to the external participants and library staff who participated in workshop discussion, including: Abby Smith Rumsey, Alex Chassanoff, Alex Wade, Amy Brand, Anasuya Sengupta, Bethany Nowviskie, Brewster Kahle, Charles Henry, Christine Borgman, Chris Bourg, Clifford Lynch, Daisy Selematsela, David Rosenthal, David Weinberger, Deborah Fitzgerald, Donald Waters, Douglas Armato, Ethan Zuckerman, Heather Yager, Jennifer Hansen, Karrie Peterson, Kate Zwaard, Mary Minow, Melissa Hagemann, Micah Altman, Nancy McGovern, Palagummi Sainath, Patricia Hswe, Peter Suber, Phil Bourne, Philip Cohen, Roger Mark, Safiya Noble, Sayeed Choudhury, Sue Kriegsman, Suzanne Wallen, and Trevor Owens.

Finally, we would like to acknowledge those who commented on the public draft: Gardner Campbell, Joshua Finnell, April Hathcock, Zachary Lizee, Paige Mann, Katherine Montgomery, Amy Nurnberger, Joyce Ogburn, Amanda Page, Yasmeeen Shorish, Sarah Shreeves, Vicky Steeves, Laurie Taylor, Trevor Owens, and Charles Watkinson.

---

\* *Contributor statement.* The authors describe contributions to this paper using a standard taxonomy. [1] MA and CB provided the core formulation of the paper's goals and aims, and MA and SK led the creation of the substantive topic outline. Writing for sections 1-5 was led by (respectively) MA & SK; AS, CB, MA, & SK; DW, MA, MM, GSC, & PS; DW, GSC, MA, MM, PC, & PS; and CH, ET, & MA. SW led copyediting. All contributors provided review and commentary. MA and CB led in obtaining funding for this project and served as PIs.

This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

---

## 1 INTRODUCTION

## 1.1 Preface: Identifying Grand Challenges

A global and multidisciplinary community of stakeholders came together in [March 2018](#) to identify, scope, and prioritize a common vision for specific grand research challenges related to the fields of information science and scholarly communications. The participants included domain researchers in academia, practitioners, and those who are aiming to democratize scholarship. An explicit goal of the summit was to identify research needs related to barriers in the development of scalable, interoperable, socially beneficial, and equitable systems for scholarly information; and to explore the development of non-market approaches to governing the scholarly knowledge ecosystem.

To spur discussion and exploration, grand challenge provocations were suggested by participants and framed into one of three sections: scholarly discovery, digital curation and preservation, and open scholarship. A few people participated in three segments, but most only attended discussions around a single topic.

To create the guest list of desired participants within our three workshop target areas we invited a distribution of expertise providing diversity across several facets. In addition to having expertise in the specific focus area, we aimed for the participants in each track to be diverse across sectors, disciplines, and regions of the world. Each track had approximately 20-25 people from different parts of the world—including the United States, European Union, South Africa, and India. Domain researchers brought perspectives from a range of scientific disciplines, while practitioners brought perspectives from different roles (drawn from commercial, non-profit, and governmental sectors). Notwithstanding, we were constrained by our social networks, and by the location of the workshop in Cambridge, Massachusetts— and most of the participants were affiliated with US and European institutions.

During our discussions, it quickly became clear that the grand challenges themselves cannot be neatly categorized into discovery, curation and preservation, and open scholarship—or even, for that matter, limited to library science and information sciences. Several cross-cutting themes emerged, such as a strong need to include underrepresented voices and communities outside of mainstream publishing and academic institutions, a need to identify incentives that will motivate people to make changes in their own approaches and processes toward a more open and trusted framework, and a need to identify collaborators and partners from multiple disciplines in order to build strong programs.

The discussions were full of energy, insights, and enthusiasm for inclusive participation—and concluded with a desire for a global call to action to spark changes that will enable more equitable and open scholarship. Some important and productive tensions surfaced in our discussions, particularly around the best paths forward on the challenges we identified. On many core topics, however, there was widespread agreement among participants, especially on the urgent need to address the exclusion

of knowledge production and access of so many people around the globe, and the troubling over-representation in the scholarly record of white, male, English-language voices. Ultimately, all agreed that we have an obligation to better enrich and greatly expand this space so that our communities can be catalysts for change.

## **1.2 Organization of This Report**

While the spirit and intent of the workshop is present, this report is not intended to be a summary of the March 2018 workshop discussions, nor a research agenda for a single institution. Instead, it draws attention to areas where a systematic community research agenda and coordinated leadership have the potential to create a broad impact. In doing this, we seek to catalyze the advancement of knowledge management and scholarly communications globally, and across disciplines, by charting specific challenges and by identifying innovative, interdisciplinary, transdisciplinary, and collaborative research agendas to solve them.

In particular, this report describes a vision for a more inclusive, open, equitable, and sustainable future for scholarship; characterizes the central technical, organizational, and institutional barriers to this future; describes the areas research needs to advance this future; and identifies several targeted “grand challenge” research problems for knowledge generation. These “grand challenges” are fundamental research problems with broad applications, whose solutions are potentially achievable within the next decade.

We conclude the report with recommendations for concrete actions to advance scholarship. We call for academics, funders, knowledge creators, knowledge stewards, policy makers, and educators to embrace these grand challenges, ignite changes in their own areas of research and practice to impact an information science and scholarly communications research agenda that will be: globally inclusive, open for access and participation, and promoting sustainable organizations and a durable scholarly record.

## **2 TOWARDS A MORE INCLUSIVE, OPEN, EQUITABLE, AND SUSTAINABLE SCHOLARLY KNOWLEDGE ECOSYSTEM**

### **2.1 Vision**

Despite the contested promise of internet technologies for accessibility and democratization,<sup>[2]</sup> today’s scholarly knowledge ecosystem and information sharing environments are plagued by exclusion; inequity; inefficiency; elitism; increasing costs; lack of interoperability; absence of sustainability and/or durability; promotion of commercial rather than public interests; opacity rather than transparency; hoarding rather than sharing; and myriad barriers at individual and institutional levels to access and participation. Despite, or perhaps because of, the range of perspectives represented, the

summit participants agreed that our common vision was of a global information environment that ensures durable, open,<sup>[3]</sup> equitable, and meaningful global access to knowledge consumption and creation in its many forms.

Such a vision requires the centering of knowledge-producing communities around the world into a global network of partnerships where we all work toward a more inclusive, equitable, trustworthy, and sustainable scholarly knowledge ecosystem—and a durable scholarly record and evidence base.<sup>[4]</sup> The vision is to create a powerful infrastructure to support local communities and organizations where people can create, share, evaluate, learn from, and interpret information on both small and large scales without barriers, or fear for lost knowledge, in order to support ongoing scholarship. Achieving this vision will require focusing not only on extant systems and processes of knowledge sharing and production, and on recognizing how some participants and forms of knowledge are currently privileged (see sections 3.1.1, 3.11, and 4.4), but also critically evaluating individual and institutional roles and interests that contribute to the current state.

The problems that plague our systems and prevent us from generating and utilizing wide open scholarship are fundamental and embedded in problems of social justice<sup>[5]</sup> that derive not only from the consequences of unequal distribution of knowledge, but also from trust, safety, security, and “epistemic”<sup>[6]</sup> injustice (unfairness stemming from the definition of what constitutes knowledge, who is assumed to be knowledgeable, and how knowledge is transmitted). Addressing these issues requires a recognition of the role that inequities in scholarship have played in re-enforcing discrimination against people based on their race, color, sex, sexual orientation, gender identity, class, religion, disability, age, neurodiversity, or national or ethnic origin.

This notion applies to individual people; different forms of knowledge communities and cultures; and the information, objects, and systems that support or challenge them. One could imagine that on one end of the spectrum trust, safety, and security includes how people feel with regard to job security or their role in a community, the reliability of data, unbiased and ethical algorithms, and stable networks. At the other end of the spectrum is data that disappear before they can be saved, networks that are intentionally tampered with to alter an information flow, algorithms that are opaque, and cases in which there are people who are concerned for their own personal, physical safety because of what they have learned or disseminated.

Solving these problems requires that scholarship be easier to discover, more durable, and more openly accessible—but this is not sufficient. We aim for a scholarly knowledge ecosystem embedded with core values of inclusion, equity, trustworthiness, agency, sustainability, and durability—in which people can broadly participate in both the creation and definition of scholarship, and have appropriate control over their inclusion in it.

Knowledge, how it is shared, and what other people do with it includes a wide continuum of possibilities for an improved scholarly knowledge ecosystem. We are looking at how our knowledge is learned, conveyed, interpreted, and utilized along the whole research spectrum in order to reach a more inclusive, equitable, sustainable, and trustworthy research world.

## 2.2 Broadest Impacts

Over the last 250 years, there have been unprecedented advancements in the human condition, encompassing improvements in health, longevity, life satisfaction, productivity, individual wealth, and the range of meaningful life choices. These improvements have been enabled in large part by systematic investigations to produce generalized, shared, and durable knowledge—also known as science and scholarship. (See Stephan 2012<sup>[7]</sup> for a discussion of the macroeconomic impact of science).

Despite its deep and broad social benefits, science itself remains surprisingly constricted in a number of fundamental aspects:

1. The benefits of science are unevenly distributed.<sup>[8]</sup>
2. Access to scientific data and scholarly communication, as well as STEM learning materials, has until recently been limited almost exclusively to those inside research or university environments with the ability to pay and fluency in English.<sup>[9]</sup>
3. Participation in our collective knowledge is limited to a small minority. The vast majority of research that gets into mainstream scholarly publications is conducted in elite university settings in developed countries.<sup>[10]</sup>
4. Even in those countries, participation in science is heavily skewed by gender, race, class, and language—which affect the construction and evaluation of scientific knowledge.<sup>[11]</sup>
5. The evidence base is restricted—subjects (people), behaviors, languages, even forms of knowledge, and the evidence base in many fields is shifting to new sources.<sup>[12]</sup>
6. The algorithms we use to interpret evidence in political and commercial systems embody unexamined bias.<sup>[13]</sup>

The inclusion of people belonging to several communities at once in the creation, dissemination, and use of scholarship is not only ethically imperative, but can strengthen research and scholarship globally, and increase the impacts of scholarship on the world.

The potential for broader inclusion to increase impact is apparent when one examines recent advances in social science. In the last twenty years, it has become possible to observe large groups of people and their communications in detail and over continuous periods of time. This has led to the creation of some of the largest publicly accessible collections of information about humans in history.<sup>[14]</sup> These advancements have also resulted in changes in the methods, evidence base, pace, and impact of many

disciplines in the social sciences—yielding new insights and challenging previous categorizations of people and their characteristics.<sup>[15]</sup>

However, despite this vast broadening in the evidence base, our current sources of information about people are heavily skewed to online behavior of industrialized Western populations. Even when researchers find information about groups outside of this category, they seldom gain access to the more complicated and nuanced in-group knowledge and living experience (see sections 3.1.2 and 4.4). Current systems of governance for that information raise questions of privacy, intellectual freedom, and agency—creating new opportunities to manipulate people for both profit and power.<sup>[16]</sup> The social sciences have much to gain from a globally inclusive system of evidence and knowledge, and society has much to gain from a value-driven governance of such a system.

There are shifts in the evidence base of public health and medicine that parallel the shifts in social science and offer analogous promise and perils.<sup>[17]</sup> We have only started to tap increasing gains from “citizen-science”<sup>[18]</sup> in the STEM fields. Given the scale of global problems such as climate change and refugee crises, increasing the inclusiveness of knowledge we can bring to bear on these problems is both important and urgent. Re-engineering the scholarly knowledge ecosystem has significant potential to improve people’s lives now and, ultimately, to contribute to the the health and longevity of our planet.

## 2.3 Recommendations for Broad Impact

In order to promote the broadest impacts of research in this area, in service to the vision of a more inclusive, equitable, and sustainable system of scholarship, we make the following recommendations:

- Recommendation 2-A: We recommend that researchers use rigorous and appropriately transparent methods to consider the broadest possible impact of their work and how that work could be used to improve the inclusiveness and equity of the scholarly knowledge ecosystem.
- Recommendation 2-B: We recommend that research funders include consideration of the impact on the scholarly knowledge ecosystem in their criteria for programs, and that they request applicants to describe the potential for proposed work to increase equity, inclusion, and sustainability. Funders should recognize that some impacts may be far into the future and not easily articulated in the early research stages and give appropriate weight to distant but potentially transformative impacts.
- Recommendation 2-C: We recommend that academic institutions recognize their interdependence and evolve to reflect a systemic approach to reflect the inclusive, open, equitable, and sustainable scholarly knowledge ecosystem essential to our future.

## 3 RESEARCH LANDSCAPE

### 3.1 Challenges, Threats, and Barriers

The information science and scholarly communication research community should aim for a future: which engages people across the world with true opportunities to discover, access, share, and create scholarly knowledge; in which people have agency in their interactions with knowledge systems and control over the information that derives from them; and in which scientific evidence and scholarship are abundant, durable, equitably accessible, and trustworthy. As we work towards this future, we must ensure that the infrastructures, policies, collaborations, and practices for research and scholarship that we adopt and support are informed by evidence and grounded in research-based decisions for the public good.

#### 3.1.1 Challenges to Participation in the Research Community

Most of the current scholarly knowledge ecosystem contains information produced and controlled by a small part of the world's population.<sup>[19]</sup> Scholarly outputs are similarly limited. Most discoverable scholarship is in the form of refereed journals—which are dominated by a small community of professionals and publishers. This information is rarely accessible to everyone, especially in resource-poor regions;<sup>[20]</sup> and access alone is insufficient to enable participation, or to promote the recognition of participants who are outside of the scholarly elite.<sup>[21]</sup> As a consequence, the knowledge, practices, and traditions of many communities are not discoverable, accessible, or preserved.

The potential impact of broadening participation in the creation and dissemination of scientific knowledge is substantial.<sup>[22]</sup> The substantial improvements in people's lives over the last 200 years stem largely from broader collection of and access to knowledge, and the many discoveries that knowledge enables. Broadening collection and access to knowledge increasingly depends on the meaningful participation of content creators across the world.<sup>[23]</sup>

#### 3.1.2 Restrictions on Forms of Knowledge

Current scholarly outputs are dominated by English-language journal articles,<sup>[24]</sup> and the available scholarly evidence-base is dominated by quantitative data.<sup>[25]</sup> Because of this, current scholarship captures only a small portion of the diverse forms of knowledge.<sup>[26]</sup> In many communities across the globe, knowledge is based on oral traditions, qualitative and experiential data, and other forms of knowing rarely recognized, valued, or represented in the current scholarly record.

One challenge here is to imagine new forms of scholarship that fit new forms of research in order to add new dimensions and perspectives<sup>[27]</sup> that are broader than the conventional journal article, monograph, and dataset.<sup>[28]</sup>

A second, related challenge is to work on ways to make these new genres for scholarship acceptable to research institutions, especially to hiring, promotion, and tenure committees.<sup>[29]</sup> New researchers should have the freedom to explore and present their work in a broad scope of formats and genres that are not restricted to existing norms. It is unlikely that new forms of knowledge will make their way into the envisioned ecosystem without the recognition of the voices of humanities and non-quantitative social science scholars and researchers. (See for example discussions in 4.1, 4.2, 4.3, & 5.2.)

A third, related challenge is for institutions to provide the infrastructure to support the creation and preservation of these new genres of scholarship,<sup>[30]</sup> or to pay for scholars to host them elsewhere. Scholars will not want to pour time into these works if they cannot find platforms to support them for the long term.

Many new forms for scholarship and mechanisms for recognizing them have emerged, at least as experiments. Many others have been proposed but not yet tried. Describing even the major ones would take more space than we have here, but we can point to some of the notable new properties that pioneering scholars are eager to try out.<sup>[31]</sup> Some new genres are multimedia. Some integrate texts and data, while others are interactive. Some are dynamic and offer regular or foreseeable updates; others are designed to grow indefinitely and never reach a state that could be called *finished*. Some are collaborations by dozens or hundreds of people. Some might start as projects by one person, or one group, and later expand to accept contributions from the crowd, while others start as crowd-sourced projects. Some allow conventional attribution and others do not. Some are so large that it's not feasible to download them, but only to explore them in their online habitats. Some have APIs allowing them to integrate with other works, or other sources of information, creating hybrid or compound works of scholarship. Some are closer to living libraries or to organisms than to individual works of scholarship.

As proposals for new genres become more numerous and more urgent, the research community will have to ask itself a series of hard questions. Which of these are worth trying? Which are worth encouraging and accommodating? Which are preferable to conventional genres, and for which purposes? How can contribution to conventional genres be made more inclusive (see sec. 3.11) and which genres have the potential to promote more inclusive participation? How should we evaluate them (e.g. for hiring, promotion, and tenure), especially when they are hugely collaborative or too large to explore in full, or when they focus less on offering “an argument” for new conclusions than offering new ways to organize or validate knowledge? Should research institutions take a position on whether these should use certain open licenses, reside on open-source infrastructure, or become interoperable with certain other resources? We should expect these conversations to be ongoing and lengthy.

### **3.1.3 Threats to Integrity and Trust**

Both technological advances and sustained democracy depend on the integrity of knowledge. However, formal scholarly knowledge generation is limited to small communities, and many members of the public mistrust science —directly, or implicitly.<sup>[32]</sup> Research is needed in communicating science effectively in increasingly politicized environments.<sup>[33]</sup>

Society is already wrestling with the challenges of ubiquitous fake information and disinformation—even with respect to assertions that are simplistic and relatively straightforward to verify.<sup>[34]</sup> Research is needed so that individuals, organizations, and governments can detect disinformation and faked records, and mitigate the effects of fake information on public opinion.

Furthermore, it is increasingly difficult even for scholars to evaluate the weight of the evidence that should be given to claims made in scholarly communications. Current problems are expanding as the scale of scholarly production grows, placing a strain on the mechanisms we have for peer review and quality control—which are slow, fallible, manipulable, and labor-intensive.<sup>[35]</sup> These include competing and overlapping systems of authority, including those run by corporate, state, and non-profit actors; increasing demands on the time of researchers asked to supervise and perform review and evaluation (with unclear reward systems); and threats from bad actors working at scales not previously possible (including state-level actors and automated systems).<sup>[36]</sup>

Much scientific data is not shared, and many knowledge outputs intended to be long-lasting end up as ephemeral and can be erased, changed, and removed by politics, technological change, restrictive licenses, or neglect.<sup>[37]</sup> Problems of access, integrity, and accountability all contribute to the problems of public understanding of science.

### **3.1.4 Threats to the Durability of Knowledge**

The durability of knowledge and scholarship are essential to realizing the full range of scientific discoveries and research, and to establishing the integrity of scholarly knowledge claims. Over the last several decades, widespread shifts from tangible to digital media create imminent threats to the durability of the scholarly record and scientific evidence base. Moreover, the digital traces of human behavior have expanded far more rapidly than we can collect, study, and preserve them.

The importance of digital preservation in ensuring the durability of knowledge is aptly summarized in the National Agenda for Digital Scholarship: “Effective digital preservation is vital to maintaining the authentic public records necessary for understanding and evaluating government actions; the verifiable scientific evidence base for reproducing research, and building on prior knowledge; and the integrity of the nation's cultural heritage. Substantial work is needed to ensure that today's valuable digital content remains accessible, useful, and comprehensible in the future—supporting a thriving

economy, a robust democracy, and a rich cultural heritage.”<sup>[38]</sup> This agenda and preceding work<sup>[39]</sup> have drawn attention to the challenges of particular formats, and the need for preservation infrastructure, business models, and organizational coordination among memory institutions.

Durability is not simply a challenge for memory institutions, however. Trustworthy scholarship requires that durability is designed into the evolving lifecycle of information creation and use. While the values of openness, inclusion, and durability are complementary, changes in one part of the scholarly knowledge ecosystem focused exclusively on promoting other value—such as the adoption of article-fee-based open access—have the potential to affect the infrastructure and incentives for durability.

Moreover, the lack of diversity in the scholarly knowledge ecosystem results in biases not only in what is produced and analyzed, but in what is preserved within the current scholarly knowledge ecosystem. We are losing, through neglect, much of the world’s stock of traditional, local, historical memory and tacit knowledge.<sup>[40]</sup> We are in a race against time, losing in many parts of the world the knowledge that is being generated, as well as the window of opportunity to implement solutions to global problems.

### **3.1.5 Threats to Individual Agency**

Our experiences online are heavily shaped by increasingly complex algorithms, which are often impossible for most participants to fully understand.<sup>[41]</sup> Further, ubiquitous data collection that gathers information from broad areas of society into academic and commercial research increases the need to maintain privacy, safety, and control, over information—especially an individual or group’s digital identity or footprint.<sup>[42]</sup> Moreover, as participation in scholarship is broadened, and as digitization enables access to community-generated work beyond the boundaries of the authoring communities, there will be a need to honor different community norms on access and use of information.<sup>[43]</sup>

Algorithmic discovery and analysis, while enabling many scientific advances, has the potential to amplify existing biases and to introduce new, and potentially hidden, sources of unfairness. However, despite increasing recognition of codes of ethics in software and algorithm development,<sup>[44]</sup> there is no consensus in research or practice over how to define or evaluate algorithmic transparency and fairness.

### **3.1.6 Incentives to Sustain a Scholarly Knowledge Ecosystem That Is Inclusive, Equitable, Trustworthy, Sustainable and Trustworthy**

Open scholarship has been a goal for much of the scholarly community for 20-50 years. Public policy has driven requirements for open access to journal articles and for deposit of datasets. Multiple

stakeholders<sup>[45]</sup> have invested (sometimes unevenly) in repositories to capture scholarly products in digital libraries that are open to the world. However, open scholarship is still far from achieving the goals set long ago.<sup>[46]</sup> While a focus on journals and datasets has made some inroads in open access, other formats lag far behind. The worlds of music, ebooks, and video are tightly bound in a proprietary world, with licenses and digital rights management that are generally more restrictive than copyright law.<sup>[47]</sup>

Current structures, policies, systems, and norms do not incentivize the behaviors that will lead to the imagined open scholarship future we want. (For a discussion, see sections 4.1 & 5.3.) As open access has progressed, the commercial publishing industry has challenged (and sometimes co-opted) open access through changes in business models, copyright law, acquiring smaller companies and players, and other actions.<sup>[48]</sup> At multiple levels, incentives are badly misaligned to the larger goals of scholarship and learning.

### 3.2 Grand Challenge Research Areas

The overarching question these problems pose is how to create a global scholarly knowledge ecosystem that supports participation, ensures agency, equitable access, trustworthiness, integrity, and is legally, economically, institutionally, technically, and socially sustainable. The aim of the Grand Challenges Summit and this report is to identify broad research areas and questions to be explored in order to provide an evidence base from which to answer specific aspects of that broad question.

Reaching this future state requires exploring a set of interrelated anthropological, behavioral, computational, economic, legal, policy, organizational, sociological, and technological areas. The extent of these areas of research is illustrated by the following exemplars:

- What is necessary to develop coherent, comprehensive, and empirically testable theories of the value of scholarly knowledge to society? What is the best current evidence of this value, and what does it elide? How should the measures of use and utility of scholarly outputs be adapted for different communities of use, disciplines, theories, and cultures? What methods will improve our predictions of the future value of collections of information, or enable the selection and construction of collections that will be likely to be of value in the future?
- How can we develop theories and methods that could reliably summarize the strength of evidence for scholarly knowledge claims? What are the determinants of scholarly and public trust in scholarly knowledge claims—and how do these relate to the strength of evidence? What content (e.g. workflows, data) and characteristics of (information architectures, organizations, cultures, institutions), if applied, would successfully promote trustworthiness and the ability to evaluate the strength of evidence in claims? How can the mechanisms for promoting trust and trustworthiness be

adapted to scholarly contributions by non-professional communities, and applied to non-traditional forms of knowledge?

● What extensions to legal, sociological, organizational, behavioral, and economic theory are necessary in order to create a general, coherent model of a sustainable scholarly knowledge ecosystem that is equitable, trustworthy, and efficient? What are the most basic theoretical properties that are necessary for a system to achieve these desired goals? What are the inherent constraints/trade-offs across multiple goals?

● What are the drivers for engagement and participation in scholarly knowledge creation, discovery and curation? What are the barriers to skill acquisition and transmission at the personal, organizational, disciplinary, and ecosystem levels? What interventions would lead to appropriate skills becoming pervasive? How do we address the need to be facilitative and supportive of skills development, while decolonizing knowledge and power over methods, skills, and objects of curation?

● What are forms of knowledge not represented in the current scholarly knowledge ecosystem? What approaches to describe, preserve, and transmit tacit knowledge and other non-textual knowledge can be generalized and scaled? How should the tacit knowledge that is the subject of scholarly study, or is integral to its practice, be discovered, curated, and preserved in such a way that empowers and gives agency to the communities from which the knowledge originates without enacting colonizing practices and methodologies? How valuable would this be to communities and society? To what extent would capturing tacit knowledge in the scholarly knowledge ecosystem benefit the social value of scholarship and its equitable distribution?

● What measures and algorithms are most effective for summarizing scholarly outputs at scale? What information architecture, semantic analysis, and computational infrastructure is needed to meaningfully link scholarly knowledge across sources and fields of study? How can both analysis and linkage be scaled to world knowledge, and adapted to its forms?

● What parts of the scholarly knowledge ecosystem promote the values of transparency, individual agency, participation, accountability, and fairness? How can these values be reflected in the algorithms, information architecture, and technological systems supporting the scholarly knowledge ecosystem? What principles of design and governance would be effective for embedding these values?

● What changes in the scholarly ecosystem would enable sustainable intergenerational open access to knowledge? What are the barriers and incentives against sustainable and durable open access, and what interventions could be effective in shifting laws, organizations, behaviors, and markets, to a sustainable open equilibrium?

● What are the most effective modalities for sharing knowledge across different regions and communities, and promoting mutual learning across community boundaries? How can skills in scholarly knowledge creation, curation, and preservation be shared and learned from different knowledge communities?<sup>[49]</sup> What are the existing models and traditions of preservation and curation from these broader communities, including informal and unofficial stewards? How do these traditions and their trajectories relate to the affordances of digital materials and systems, and where are adaptation and refinement needed? How can these traditions and models be integrated to transform information science and formal library and archival practice?

The list above provides a partial outline of research areas that will need to be addressed in order to overcome the major barriers to a better future for scholarly communication and information science. As the field progresses in exploring these areas, and attempting to address the barriers is discussed, new areas are likely to be identified. Even within this initial list of research areas, there are many pressing questions ripe for exploration.

### **3.3 Recommendations for Research Areas and Programs**

Based on the characterization of the research landscape above, we make the following recommendations:

- Recommendation 3-A: We recommend that funders consider developing future programs and requests for proposals to address the barriers described above.
- Recommendation 3-B: We recommend that researchers in information science and related fields strongly consider selecting problems within a grand-challenge research area as part of their research program.
- Recommendation 3-C: We recommend that reviewers and editors give particular weight to research proposals and discoveries that address these barriers or advance grand-challenge research.
- Recommendation 3-D: We recommend that the participants in the existing scholarly knowledge ecosystem—including publishers, tool builders, and platform providers—consider how the systems they build can reduce the barriers identified above.
- Recommendation 3-E: We recommend that researchers and stakeholders actively seek out new methodologies, voices, and participation in the design and conduct of research, and also challenge currently accepted ways of conducting, communicating, and evaluating research.

## 4 TARGETED RESEARCH QUESTIONS

All of the research areas described above hold great promise for exploration. In this section, we discuss in detail four targeted, individual research questions, drawn from these broad research areas. The aim is to provide a statement of the research question that can be understood by researchers and practitioners in multiple disciplines; suggest how progress toward a solution could be measured; explain how such progress could help in addressing the problems above; and identify lines of research and practice that offer potential insights into a solution. We argue that each of these questions is potentially solvable in the next seven to 10 years, and, if solved, will have a substantial impact across multiple central problem areas.

### 4.1 Research Challenge: Legal, economic, policy, and organizational design for enduring, equitable, open scholarship.

Scholarship and research are embedded within and shaped by a broader ecosystem that comprises stakeholder organizations,<sup>[50]</sup> social norms,<sup>[51]</sup> laws,<sup>[52]</sup> economic markets,<sup>[53]</sup> and political institutions.<sup>[54]</sup> This ecosystem as a whole affects how knowledge is produced, accessed, discovered, defined, and preserved. None of the major challenges to equitable, trustworthy, inclusive, and durable scholarship (discussed in this report in section 3) can be fully resolved without an improved understanding of how to design institutional and normative ecosystems, how to allocate resources within them,<sup>[55]</sup> and of what interventions are effective for moving us towards a better scholarly knowledge ecosystem.

Research on the challenge of enduring, accessible, inclusive, and open scholarship begins with an understanding of the problems exacerbated by its absence. These include weak trust in scholarly knowledge claims,<sup>[56]</sup> which remain unverifiable or opaque across research communities and among wider publics when the processes and outcomes of research are not open, and when disparate access to research knowledge exacerbates social inequalities. The pursuit of openness in scholarship, however—especially in access to published work—may manifest as a “treadmill” of increasing expenses absorbed as user fees or publisher profits that fail to lead to systemic solutions.<sup>[57]</sup> With resources devoted to these costs, investment in preservation with durable open access is threatened, even as the volume and complexity of material to be preserved in the scholarly record multiplies.<sup>[58]</sup>

Despite common recognition of this set of problems, effective incentives to drive key actors to develop and enact solutions seem to be lacking.<sup>[59]</sup> For example, scholarly societies often depend on revenue from journal subscription fees to fund various organizational and member goals and activities, thus creating a disincentive to adopting open models of dissemination that reduce or eliminate subscription revenue streams.<sup>[60]</sup> Similarly, researchers and funders, as well as universities, have incentives to see their work appear in the most prestigious publications, regardless of their public accessibility,<sup>[61]</sup> even

as scholars and society are increasingly focused on wider audiences and more reliable accumulation of knowledge.<sup>[62]</sup> Further, all of these actors are embedded in markets that do not reward the production of knowledge for social good.<sup>[63]</sup>

These incentives to publish in a narrow set of high-prestige journals may also exacerbate the challenges of science communication—which is increasingly important to the reputation and impact of science.<sup>[64]</sup> Further, there are weak incentives to develop translational work that brings technical content to wider audiences. This work is important to promoting public understanding and policy impact, but is often considered outside of the scholarly ecosystem.

Organizational and technological innovations that promote open scholarship have the potential to promote opportunities for broader engagement across research communities and broader publics,<sup>[65]</sup> and to allow the use of machine tools for analysis and dissemination of research outputs and materials.<sup>[66]</sup> However, tenure and promotion systems will need to be adapted to reward such communications. Further, such innovations pose risks, including the empowerment of bad online actors<sup>[67]</sup> at greater scale or velocity and the reduction of agency by data subjects and producers (see section 3.1.5 for a discussion).

Research on open scholarship solutions is needed to assess the scale and breadth of access,<sup>[68]</sup> the costs to actors and stakeholders at all levels, and the effects of openness on perceptions of trust and confidence in research and research organizations. Research is also needed in the intersection between open scholarship and participation, new forms of scholarship, information integrity, information durability, and information agency (see section 3.1.). This will require an assessment of the costs and returns of open scholarship at a systemic level, rather than at the level of individual institutions or actors. We also need to assess whether and under what conditions interventions directed at removing reputation and institutional barriers to collaboration promote open scholarship. Research is likewise required to document the conditions under which open scholarship reduces duplication and inefficiency, and promotes equity in the creation and use of knowledge. In addition, research should address the permeability of open scholarship systems to researchers across multiple scientific fields, and whether—and under what conditions—open scholarship enhances interdisciplinary collaboration.

## **4.2 Research Challenge: Measuring, Predicting, and Adapting to Use and Utility Across Scholarly Communities**

In order to manage information, we must value it. Systems and algorithms for discovery nominally aim to support users in finding information that is relevant to their needs—information that is of value to them in a specific context. Curation and preservation systems and strategies aim to deliver future (medium- or long-term) value to specific communities of research or practice. Assumptions are

embedded throughout the scholarly knowledge ecosystem regarding what information is (or will be) valuable, which communities will value it, and what forms of use and access will realize this valuable information. Explicit models of research information value and uses are much less common.

Search and discovery increasingly eludes expert (human) indexing and relies on algorithms—creators of search algorithms and discovery systems attempt to predict the value of specific information to a specific user within a specific context.<sup>[69]</sup> These algorithms, in turn, rely heavily on signals of broad and current use (e.g. clicks, downloads, links), and are influenced by the monetary value that can be derived from such systems (such as sales of goods or ad placements). Approaches based on these aggregate models of information value are unlikely ever to support the systematic discovery of information of value to important, but small, communities of knowledge seekers. For example, current search systems will rarely uncover previously unused material in the history of robotics, nor the most reliable software for estimating models in comparative phylogenetics—even if such materials might hold information key to future breakthroughs in the field. More generally when discovery environments are developed in ways that favor popularity and profitability, we are unlikely to discover content that is of high intellectual value—but not of high monetary value—or content which may be intensely valuable to a small community.

Researchers and curators often rely on professional judgment, manual selection, and assessment processes to decide what information to retain, how long to retain it, what effort to expend in making it accessible and understandable, and when that effort should be applied. Often these processes originate from a prior analog era when all the information on which each organization relied had to be “held” (formally acquired or created); and, in practice, it was possible to select and curate only information that was held.<sup>[70]</sup> As a result, these processes are often hyper-local and ad-hoc, based on the history of practice and the local values of the organization or community of practice making these decisions.<sup>[71]</sup> Moreover, models of value underlying our current curation processes have not been updated or adapted to fit current realities.<sup>[72]</sup> The absence of explicit models of value makes it difficult to effectively adapt these processes to non-traditional forms of evidence (e.g. software, oral testimony); for new non-traditional communities of research and practice; or for new types of use (e.g. non-consumptive data mining).<sup>[73]</sup>

The development of formal models, methods, and empirical analysis—which would lead to more rigorous, reliable, and systematic evaluation of the value of research information—constitutes an important, but challenging, set of problems. Estimating the value of information is inherently difficult. Arrow’s information paradox states that ex-ante a buyer cannot assess the value of particular information—it can only be known ex-post, at which point the buyer has limited incentive to pay for it.<sup>[74]</sup> Although assignment of intellectual property rights can address this issue to a limited extent, it is very challenging<sup>[75]</sup>—and hence markets for information goods are generally “thin,” which makes

monopoly/monopsony dominance more likely. Furthermore, intellectual property rights notwithstanding, the non-consumptive and limited excludability that is inherent in properties of information goods implies that any pure market solution will produce and distribute information at levels that are socially suboptimal.<sup>[76]</sup> Although data quality is sometimes seen as a proxy for value, no feasible, universal quality measure exists—data quality measures are notoriously varied, discipline-specific, contextual, and difficult to implement in practice.<sup>[77]</sup>

In the preservation of information, diversification of storage and representation is recognized as an essential strategy for ensuring future accessibility—and there is a well-recognized taxonomy of risk sources that guides diversification strategy. We have no equivalent strategies to diversify across the risks to information value. In economics, methods such as revealed preference analysis and contingent valuation surveys<sup>[78]</sup> are often used to measure the value of non-market goods—yet these methods have not been applied to valuing research data. Similarly, portfolio selection modeling<sup>[79]</sup> is the primary tool used in finance to diversify across risky investments but has never been applied to the “investments” in developing collections of information.<sup>[80]</sup> Solutions in this area would yield models of information valuation that could be examined, challenged, and refined; and taxonomies of uses, communities, and threats that could be used for diversification strategies.

### **4.3 Research Challenge: Designing and Governing Algorithms in the Scholarly Knowledge Ecosystem to Support Accountability, Credibility, and Agency**

Across the scholarly knowledge ecosystem automated algorithms play increasingly critical roles in discovery (e.g. relevance ranking, recommender systems);<sup>[81]</sup> in information extraction and summarization (e.g. automated abstract generation, literature mining);<sup>[82]</sup> and in the evaluation of scholars and scholarship (e.g. detection of plagiarism, image manipulation, or journal citation inflation; evaluation of collaboration impact; predicting productivity).<sup>[83]</sup> Moreover, the rapid growth in the volume of evidence, number of publications, and scale of collaboration in research<sup>[84]</sup> generates strong pressure to rely on such automated systems—the growth of scientific knowledge relies on algorithms and algorithmic systems to support knowledge discovery, evaluation, and collaboration at scale.

As their ubiquity increases, algorithms in the scholarly ecosystem are growing increasingly complex and opaque: ranging from models that, while theoretically well-defined, remain difficult to estimate and interpret (e.g. use of latent-Dirichlet allocation to extract science topics; use of network regression models to measure collaboration)<sup>[85]</sup> to the nominally transparent but effectively inscrutable (e.g. use of open deep-learning for recommender systems)<sup>[86]</sup> to algorithms that are opaque and ever-changing by design (e.g. Google’s systems for relevance ranking).<sup>[87]</sup>

The problems posed by the use of such complex algorithms are now becoming recognized in the wider public sphere. These problems include violation of human privacy or agency (e.g. recommender systems inadvertently revealing purchasing habits to others);<sup>[88]</sup> of biases and inequities in outcomes that result from algorithmic design choices (e.g. the poor performance of facial recognition algorithm for people of color);<sup>[89]</sup> the potential for algorithmic systems to aggregate and amplify human biases (e.g. substantial explicit racialization of Google search ad placement resulting from the aggregation of implicit bias in click-through behavior);<sup>[90]</sup> to the intentional adversarial manipulation of digital evidence and the creation of false records<sup>[91]</sup> and of machine-learning algorithms to game evaluation or actively harm others (e.g. adversarial attacks on image detection).<sup>[92]</sup>

Addressing this interrelated set of problems requires advances in multiple fields and at multiple levels. The design and evaluation of algorithmic bias, fairness, and manipulability is generally in the early stages. Further, in the domain of scholarly information, we have yet to identify the necessary properties of algorithms that are required to protect individual agency, facilitate collaboration, facilitate the identification of new biases, prevent gaming, and preserve trustworthiness—nor have we identified the fundamental constraints on and tradeoffs among these goals. For those few properties that have been identified as desirable, such as individual information privacy, we have limited understanding of how to successfully design and deploy algorithmic systems that satisfy these properties.<sup>[93]</sup> Even for those algorithms that are commonly in use, we have little systematic empirical evidence on their quality, manipulability, and biases.

#### **4.4 Research Challenge: Integrating Oral and Tacit Knowledge into the Scholarly Knowledge Ecosystem**

Participation in the collective knowledge of science and scholarship is currently limited to a small minority (as discussed in section 2, above). In part, this is because scholarly communication and reputation are primarily transmitted and promoted through the publication of journal articles and books. In many societies, cultural, historical, and practical knowledge is not written. Knowledge that derives from or pertains to indigenous, traditional, and local communities is often transmitted and preserved through oral histories and oral traditions.

Even within our current system of science, there is evidence that critical parts of the knowledge needed to conduct scholarly inquiry (e.g. how to perform experimental bench methods);<sup>[94]</sup> and to have a successful career as a scholar is tacit—resistant to transmission in textual form. Within scholarly communities, this is often transmitted orally and experientially through collaboration and mentoring relationships—even where such knowledge could be adequately documented. This can

have a substantial impact on both the reliability of scientific results<sup>[95]</sup> and disparities in the diversity of the academy.<sup>[96]</sup>

Neither the methods nor the systems used to represent and manage the scholarly record are well-adapted to non-textual knowledge. The result is that most knowledge in tacit or oral form remains unexamined and invisible, and is not recognized, curated, or preserved within the scholarly community.

Integrating oral and tacit knowledge into the scholarly knowledge ecosystem raises not only methodological and technical challenges, but deep conceptual challenges, as well.<sup>[97]</sup> The scholarly conceptualization of information integrity will need to be expanded, along with the mechanisms and methods we use to manage authenticity, provenance, durability, and versioning. Models of attribution, selection, authority, and trust will need to be extended to both these forms of knowledge, and to the communities that produce it. Further, the widespread dissemination of oral and tacit knowledge that is embodied in the behavior of individuals raises challenges for information agency—and for the mechanisms we use to provide consent for and control access to information.

## **5 INTEGRATING RESEARCH, PRACTICE, AND POLICY**

### **5.1 The Need for Leadership to Coordinate Research, Policy & Practice Initiatives**

Many of the opportunities for scholarship that are made possible by rapidly advancing technologies have yet to be fully realized. There are several reasons for this. As discussed above, the social, legal, technical, and organizational systems for disseminating, discovering, reusing, and communicating scholarly information have not kept pace with the technologically induced changes in the scholarly knowledge ecosystem.

Left to the market, the economics of knowledge in digital form creates both network externalities and reputation effects that are increasingly exploited by rent-seeking monopolies.<sup>[98]</sup> To avoid this market disequilibrium requires that institutions coordinate to manage scholarly knowledge—and this requires leadership. Collectively, universities and other institutions must recognize their interdependence and organize as a system to create a scholarly knowledge ecosystem that is not dominated by current market value. Further, some set of individual organizations must go beyond their local interests—and invest effort and reputation into changing the scholarly knowledge ecosystem that yield collective benefits.

At the same time, organizations should not act in isolation. Almost every institution now relies for its business, operations, and mission on large amounts of information that go beyond institutional boundaries. The amount of information is so great, and the risks so diverse, that no single organization can effectively ensure sustainable access to all the information it produces and needs.<sup>[99]</sup> At the same time, many institutions value the same pools of information. Together, these imply that collaboration is essential—institutional leaders must not only innovate, but coordinate.

Further, while research is needed to guide the design of platforms that are consistent with our values; platforms are needed that can be instrumented to evaluate these designs, and contribute to our understanding of where we are successfully promoting the objectives we seek. It is thus essential that research and practice in this area be in dialogue.

## **5.2 Role of Libraries and Archives as Advocates and Collaborators**

Research universities are among the most long-lived of human institutions. University libraries and research archives are widely trusted as the permanent stewards of the scholarly record and scientific evidence base within these institutions, and libraries and archives have highly refined expertise and infrastructures for organization, dissemination, and preservation of knowledge. Further, the grand challenges identified above will likely be solved only through a cross-disciplinary approach. Libraries are by design interdisciplinary, and in practice trusted as honest brokers of knowledge. Finally, the values of libraries and archives are deeply aligned with the values of knowledge communities—these organizations constitute themselves as being in service to scholarly communities, in contrast with commercial entities, and even in contrast to the research universities taken as a whole. As trusted brokers for information, they can advocate on behalf of the scholarly community both to the government and to commercial information providers and intermediaries, and also as a voice to enlist other change-makers.

Librarians and archivists as professionals, and libraries and archives, as institutions, can go beyond advocacy to contribute and collaborate in the grand challenge research we have described in this paper. Further, these organizations can act as direct agents of change. They can help to educate the communities that they serve about information ethics, agency, and risks—and collaborate on the development of common curricula.<sup>[100]</sup> They can collaborate to develop common open infrastructure,<sup>[101]</sup> and to develop community-based model license agreements when engaging commercial infrastructure and services. They can help to make the norms and culture of scholarship more inclusive by enabling the development of alternative metrics of scholarship<sup>[102]</sup>, and by documenting and disseminating the tacit knowledge that is part of the successful practice of scholarship—much of which is inaccessible except through direct mentoring.<sup>[103]</sup>

### 5.3 Incorporating Values of Openness, Sustainability, and Equity into Scholarly Infrastructure and practice

With respect to the practice of research, it is worth noting that many fields of scholarship, academic associations, professional groups, and societies have issued ethics statements involving integrity of the work, confidentiality of the individual, and being mindful of the direct or indirect impact that research/work outcome may have on the lives of individuals, groups, or societies. Leadership at these professional and academic organizations have the power to align “do no harm,” high-level principles with active and impactful policy implementations that set as a goal equitable, diverse, inclusive, and socially just outcomes. Universities often work under explicit policies and procedures, but defining and implementing such research outcomes requires systems in place that intentionally support the advancement of equitable and diverse societies worldwide. This remains an important challenge because it means saying no to certain funding sources, and adjusting relationships between wealthy and impactful research institutions and industries.

Much of the infrastructure for scholarship is neither owned nor designed by scholars, but has been developed by commercial entities for profit—and is controlled by a few large companies.<sup>[104]</sup> As the practice of research and publishing has accelerated, requiring more integration of information across the research lifecycle, this infrastructure has become increasingly complex, and increasingly dominated by a small number of commercial entities. Should similar ethical principles be applied to infrastructure as to practice? Does commercial dominance in infrastructure present risks to achieving the goal of open, sustainable, and equitable scholarship?

As an example of existing tensions, it has been broadly recognized that the profit-driven model of social-media companies such as Facebook and Twitter creates strong incentives to collect and monetize information about participants in this network—which is in strong tension with protecting information privacy.<sup>[105]</sup> Similarly, the reliance of Google on advertisement revenue influences both what is indexed, and how relevance is operationalized.<sup>[106]</sup> More generally, commercial entities have an incentive toward algorithmic opacity in order to protect their trade secrets and competitive advantages.<sup>[107]</sup>

The increasing prevalence of high-profile information breaches<sup>[108]</sup> and the increasing ability to re-identify individuals and their characteristics based on aggregated or nominally anonymous data<sup>[109]</sup> has led to increasingly widespread support for systems of information discovery and sharing that incorporate respect and protections for individual agency and information privacy into their core design. In some cases, values such as openness, sustainability, and equity can and should be incorporated deep into the infrastructure of new systems from the beginning. In other cases, research is needed to determine whether and how such values could be effectively expressed and enacted using

existing infrastructure that was created for very different functions and with different value propositions than those animating the creation of systems explicitly designed to support open, equitable, sustainable scholarship.

We also must critically examine and document the unintended consequences and uses of policies, practices, and infrastructures that have been explicitly developed in support of open scholarship. For example:

- How has the discovery and hosting of open-access content on proprietary infrastructure (e.g. SSRN, bepress, Google Scholar) created or mitigated barriers to accessing that content and affected long-term sustainability and durability of information?
- What creates incentives for stakeholders to use open software, interoperable standards, and APIs—particularly when hosting open access content?
- How can methods be used to design and refine open infrastructure to support reuse, extension, and adoption at the local level—while being able to function at the continually growing scale of global research output?

Addressing these questions requires integrating research with practice and infrastructure development. Research is needed to guide the design of platforms that are consistent with our values. Platforms are needed that can be instrumented to evaluate these designs, and to contribute to our understanding of where we are successfully promoting the objectives we seek. To be successful at a global scale, valuation of practice should go beyond case studies in their approach, and include replicable methods to support systematic inference, such as randomization and pre- and post-evaluation.

## 5.4 Funders, Catalysts and Coordinators

A number of organizations currently fund, coordinate, or catalyze advances in research, infrastructure, and practice, which enables open, inclusive, and durable scholarship. The US federal agencies Institute of Museum and Library Services and the National Endowment for the Humanities; the European Research Council; and The Andrew W. Mellon and Alfred P. Sloan Foundations all have long track records of supporting research, practice, and infrastructure in these areas. A number of other funders—including Wellcome Trust, National Science Foundation, National Institutes for Health, Chan Zuckerberg Initiative, The Whiting Foundation, Gates Foundation, Helmsley Foundation, Open Society Foundation, and the Gordon and Betty Moore Foundation—have supported more limited initiatives related to these areas and primarily centering on open and reproducible research. This good work notwithstanding, we argue that the problems and challenges described in this report merit

recognition by the entire spectrum of funders engaged directly or indirectly in supporting research and scholarship.

Finally, success in advancing these areas will rely on organizations to coordinate collaborative approaches to research, practice, and infrastructure—which are often intertwined. Coordination is difficult because it often has the characteristics of a public good—providing more benefits to the research community, as a whole, than to the coordinating institution (indeed, many coordinating organizations invest more than they expect to receive directly). Despite this structural challenge, organizations like the Council on Library and Information Research, the Digital Library Federation, the National Digital Stewardship Alliance, Research Data Alliance, the Scholarly Publishing and Academic Resources Coalition and CODATA have been successful in coordinating standards development and educational initiatives.<sup>[110]</sup> Organizations such as Duraspace, the Dataverse Community, Digital Preservation Network, and the Center for Open Science<sup>[111]</sup> have played vital roles in coordinating the development and support of the vital research infrastructure that supports open scholarship.

Organizations such as the Coalition for Networked Information, Association of Research Libraries, and the National Academies (primarily through the Board on Research Data and Information)—joined more recently by organizations such as Force11 and Sage Bionetworks—have established themselves as catalysts for open scholarship. They play a vital role in disseminating information on initiatives and research, convening experts, and engaging in advocacy. Over the last decade, organizations such as the National Digital Stewardship Alliance, The Long Now Foundation, and the Digital Preservation Coalition<sup>[112]</sup> have played a similar catalytic role for the issue of information durability. Only recently have organizations focused on equitable and inclusive knowledge, such as *Whose Knowledge?*,<sup>[113]</sup> and have been recognized in the scholarly community.

Progress towards a more open, equitable, trustworthy, and durable scholarly ecosystem requires that more institutions take catalyzing and coordinating roles in addressing the challenges and exploring the research areas described in section three. Further, existing organizations can help greatly by recognizing in their programs the interrelationship between openness, impact, trustworthiness, durability, and inclusivity in research and scholarship.

## 5.5 Recommendations for Integrating Research, Practice, and Policy

Summarizing the discussion of the connection across research, policy, and practice above, we make the following recommendations:

- Recommendation 5-A: We recommend that individual research institutions take public responsibility for leading and coordinating inclusive efforts to address the barriers to more equitable and inclusive systems of scholarship.

- Recommendation 5-B: We recommend that research libraries and archives promote a vision of inclusive and equitable scholarship within their institutions; that they engage in work on legislation and public policy; and that they enlist others in the scholarly community as change-makers.
- Recommendation 5-C: We recommend that those engaged in developing platforms and communities of practices actively seek new voices and participation in their design and use.
- Recommendation 5-D: We recommend that those engaged in research, practice, and advocacy in the area of open and inclusive scholarship should collaborate to develop platforms and interventions that can contribute to our understanding of what is most effective—both directly, and in advancing the broad goals of inclusion, openness, equity, and sustainability. Evaluation of practice should go beyond case studies in their approach, and include replicable methods to support systematic inference.
- Recommendation 5-E: We recommend that stakeholders give priority to resourcing programs that rigorously integrate research and practice, particularly those that systematically contribute to the overall cumulative evidence base for inclusive, equitable, and credible scholarship.

---

[1] Liz Allen, Jo Scott, Amy Brand, Marjorie Hlava & Micah Altman, “Publishing: Credit Where Credit Is Due,” *Nature* 508, (2014): 312.

[2] For thoughtful analyses of technological changes and of their potential to further revolutionize scholarship see: Atkins, D., 2003, [Revolutionizing science and engineering through cyberinfrastructure](#): Report of the National Science Foundation blue-ribbon advisory panel on cyberinfrastructure. National Science Foundation; Berman, F. and H. Brady, 2005, Workshop on Cyberinfrastructure for the Social and Behavioral Sciences: Final Report, National Science Foundation. For a discussion of the threats of these and similar technologies see for example: Raso, Filippo A., Hannah Hilligoss, Vivek Krishnamurthy, Christopher Bavitz, and Levin Kim. “Artificial Intelligence & Human Rights: Opportunities & Risks.” *Berkman Klein Center Research Publication*, 2018-6 (2018); Morozov, Evgeny. *To save everything, click here: The folly of technological solutionism*. Public Affairs, 2013; Schneier, Bruce. *Data and Goliath: The hidden battles to collect your data and control your world*. WW Norton & Company, 2015.

[3] We use the term “open” not in the restricted legal sense of “open licenses,” but in a broader sense that includes practical findability, accessibility, interoperability, and reusability. See Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas

Blomberg, et al. "The FAIR Guiding Principles for scientific data management and stewardship." *Scientific Data* 3 (2016).

[4] Throughout this paper, we use the terms "scholarship," "scholarly record," "evidence base," and "scholarly knowledge ecosystem" broadly. These denote (respectively), communities and methods of systematic inquiry aimed at contributing to new generalizable knowledge; all of the informational outputs of that system (including, but not limited to, scholarly communications); the domains of evidence that are used by these communities and methods to support knowledge claims (including, but not limited to, quantitative measures, qualitative descriptions, and texts); and the set of stakeholders, laws, policies, economic markets, organizational designs, norms, technical infrastructure, and educational systems that strongly and directly affect the scholarly record and evidence base, and/or are strongly and directly affected by it.

[5] Britz, J.J. (2008). "[Making the global information society good: a social justice perspective on the ethical dimensions of the global information society.](#)" *Journal for the American Society for Information Science and Technology*, 59(7), 1171-1183. doi:10.1002/asi.20848

[6] Fricker, Miranda. [Epistemic injustice: Power and the ethics of knowing](#). Oxford University Press, 2007.

[7] Stephan, Paula E., *How economics shapes science. Vol. 1.* (Cambridge, MA: Harvard University Press, 2012).

[8] Examples: Fausto-Sterling, Anne, *Myths of gender: Biological theories about women and men.* (Basic Books, 2008); Braun, L., Fausto-Sterling, A., Fullwiley, D., Hammonds, E.M., Nelson, A., Quivers, W., Reverby, S.M. and Shields, A.E., 2007. "Racial categories in medical practice: how useful are they?". *PLoS Medicine*, 4(9), p. 271; Fausto-Sterling, A., 2000. *Sexing the body: Gender politics and the construction of sexuality.* (Basic Books; Revised edited ed. 2000).

[9] Willinsky, John. *The access principle: The case for open access to research and scholarship.* (Cambridge, MA.: MIT Press, 2006).

[10] Azar, B. "[Are your findings 'WEIRD?'](#)" *Monitor on Psychology* 41, no. 5 (2010): 11.

[11] See, Sugimoto, Cassidy R. "[Global gender disparities in science](#)," *Nature* 504 (2013): 211-213 and supplement 1. See examples: Ho, Adrian and Patricia Hwse, "[Library publishing and diversity values: changing scholarly publishing through policy and scholarly communication education.](#)" *College and Research Libraries News* Vol. 77, no. (2016). Inefuku, Harrison W. "Globalization, open access, and the democratization of knowledge." *EDUCAUSE Review* 52, no. 4 (2017): 62.

- [12] See for example Alvarez, R. Michael, ed. *Computational social science*. (Cambridge University Press, 2016); Altman, Micah, and Marguerite Avery. "Information wants someone else to pay for it: laws of information economics and scholarly publishing." *Information Services & Use* 35, no. 1-2 (2015): 57-70; Torres, Lars Hasselblad. "[Citizen sourcing in the public interest](#)." *Knowledge Management for Development Journal* 3, no. 1 (2007): 134-145. Popejoy, Alice B., Deborah I. Ritter, Kristy Crooks, Erin Currey, Stephanie M. Fullerton, Lucia A. Hindorff, Barbara Koenig, et al. "The Clinical Imperative for Inclusivity: Race, Ethnicity, and Ancestry (REA) in Genomics." *bioRxiv* (2018): 317800.
- [13] Courtland, R. "Bias detectives: the researchers striving to make algorithms fair." *Nature* 558, no. 7710 (2018): 357.; Altman, Micah, Alexandra Wood, and Effy Vayena. "A harm-reduction framework for algorithmic fairness." *IEEE Security & Privacy* 16, no. 3 (2018): 34-45.
- [14] King, Gary. "Restructuring the social sciences: reflections from Harvard's Institute for Quantitative Social Science." *PS: Political Science & Politics* 47, no. 1 (2014): 165-172.
- [15] Lazer, David, Alex Sandy Pentland, Lada Adamic, Sinan Aral, Albert Laszlo Barabasi, Devon Brewer, Nicholas Christakis et al. "Life in the network: the coming age of computational social science." *Science* (New York, NY) 323, no. 5915 (2009): 721.
- [16] See Hilbert, Martin. "How to Measure" How Much Information"? Theoretical, Methodological, and Statistical Challenges for the Social Sciences Introduction." *International Journal of Communication* 6 (2012): 1042-1055, summarizing a special issue of IJOC on this topic.
- [17] Raghupathi, Wullianallur, and Viju Raghupathi. "Big data analytics in healthcare: promise and potential." *Health Information Science and Systems* 2, no. 1 (2014): 3; Vayena, Effy, Marcel Salathé, Lawrence C. Madoff, and John S. Brownstein. "Ethical challenges of big data in public health." *PLoS Computational Biology* 11, no. 2 (2015): e1003904.
- [18] Wiggins, Andrea, and Kevin Crowston. "From conservation to crowdsourcing: A typology of citizen science." In *System Sciences (HICSS)*, 2011 44th Hawaii international conference on, pp. 1-10. IEEE, 2011.; Levine, S.S. and Prietula, M.J., 2013. "Open collaboration for innovation: Principles and performance." *Organization Science* 25(5), pp.1414-1433; Majchrzak, Ann, and Arvind Malhotra. "Towards an information systems perspective and research agenda on crowdsourcing for innovation." *The Journal of Strategic Information Systems* 22, no. 4 (2013): 257-268.
- [19] See: Graham, Mark, Scott A. Hale, and Monica Stephens. "Geographies of the World's Knowledge." *London: Convoco*(2011).; Graham, Mark, Bernie Hogan, Ralph K. Straumann, and Ahmed Medhat. "Uneven geographies of user-generated information: Patterns of increasing informational poverty." *Annals of the Association of American Geographers* 104, no. 4 (2014): 746-764.

[20] As an example, the difficulty of accessing research information through mobile phones, which are the primary channel for accessing information, is a barrier that must be addressed: Hosman, Laura, and Elizabeth Fife. "The use of mobile phones for development in Africa: Top-down-meets-bottom-up partnering." *The Journal of Community Informatics* 8, no. 3 (2012);

[21] See for barriers to engagement in the global south: Ojanpera, Sanna, Mark Graham, Ralph K. Straumann, Stefano De Sabbata, and Matthew Zook. "Engagement in the knowledge economy: Regional patterns of content creation with a focus on Sub-Saharan Africa." *Information Technologies & International Development* (2017): 33-51. For a discussion of the citational politics of recognition, see Baildon, Michelle. "Extending the social justice mindset: Implications for scholarly communication." *College & Research Libraries News* 79, no. 4 (2018): 176.

[22] Ali-Khan SE, Jean A, MacDonald E and Gold ER. "Defining Success in Open Science." *MNI Open Res* 2018, 2:2 (doi: [10.12688/mniopenres.12780.2](https://doi.org/10.12688/mniopenres.12780.2))

[23] See for example: [European Commission on Open Science](#).

[24] Lillis, Theresa and Mary Jane Curry. *Academic writing in a global context: The politics and practices of publishing in English*. Routledge, 2013.

[25] On the importance of non-numeric information, the dearth of scientific archives providing this content, and the challenges of providing durable access, see: National Research Council. *Frontiers in massive data analysis*. National Academies Press, 2013.; Hammersley, Martyn. "Qualitative data archiving: some reflections on its prospects and problems." *Sociology* 31, no. 1 (1997): 131-142.; Elman, Colin, Diana Kapiszewski, and Lorena Vinuela. "Qualitative data archiving: Rewards and challenges." *PS: Political Science & Politics* 43, no. 1 (2010): 23-27.; Mannheimer, Sara, Amy Pienta, Dessislava Kirilova, Colin Elman, and Amber Wutich. "Qualitative Data Sharing: Data Repositories and Academic Libraries as Key Partners in Addressing Challenges." *American Behavioral Scientist* (2018).

[26] See for examples of knowledge that is not readily reducible to text: Ahn, Sun Joo, Joshua Bostick, Elise Ogle, Kristine L. Nowak, Kara T. McGillicuddy, and Jeremy N. Bailenson. "Experiencing nature: Embodying animals in immersive virtual environments increases inclusion of nature in self and involvement with nature." *Journal of Computer-Mediated Communication* 21, no. 6 (2016): 399-419; Bailenson, Jeremy. *Experience on Demand: What Virtual Reality Is, how it Works, and what it Can Do*. WW Norton & Company, 2018; McPherson, T., 2018. *Feminist in a Software Lab: Difference+ Design*. Harvard University Press; Eric Dinmore (2015) "Collecting, Curating, and Presenting '3-11' with Harvard's Digital Archive of Japan's 2011 Disaster." *Verge: Studies in Global Asias* 1(2): 37-41.

- [27] See for example: Groth, Paul, Andrew Gibson, and Jan Velterop. "The anatomy of a nanopublication." *Information Services & Use* 30, no. 1-2 (2010): 51-56; Pasquali, Matias. "Video in science: Protocol videos: the implications for research and society." *EMBO Reports* 8, no. 8 (2007): 712-716; Neylon, Cameron, Jan Aerts, C. Titus Brown, Simon J. Coles, Les Hatton, Daniel Lemire, K. Jarrod Millman, et al. "Changing computational research. The challenges ahead." *Source Code for Biology and Medicine* (2012): 2.
- [28] It is worth noting that even traditional outputs, such as datasets, have lacked consistent practices for publication and citation: Altman, Micah, Christine Borgman, Mercè Crosas, and Maryann Matone. "An introduction to the joint principles for data citation." *Bulletin of the Association for Information Science and Technology* 41, no. 3 (2015): 43-45.
- [29] See for example Andersen, Deborah Lines. *Digital Scholarship in the Tenure, Promotion and Review Process*. Routledge, 2015.; Cheverie, Joan F., Jennifer Boettcher, and John Buschman. "Digital scholarship in the university tenure and promotion process: A report on the sixth scholarly communication symposium at Georgetown University Library." *Journal of Scholarly Publishing* 40, no. 3 (2009): 219-230; and Flanders, Julia. "The productive unease of 21st-century digital scholarship." *Defining Digital Humanities: A Reader* (2013): 205-218. See also [HuMetricsHSS](https://humetricshss.org/) as an example of how to look at human based indicators that create a value-based framework for humanities research including Collegiality, Quality, Equity, Openness, and Community: <https://humetricshss.org/>.
- [30] See examples of The Andrew W. Mellon Foundation support for the transition to digital scholarship in the humanities: <https://mellon.org/resources/shared-experiences-blog/monograph-publishing-digital-age/> and "The Academic Ebook Reinvigorated." *Learned Publishing* 31, Issue 51. <https://doi.org/10.1002/leap.1185>
- [31] See (respectively) for examples of emerging mechanisms for recognition of software as a product of scholarship in the sciences and a foundational introduction to new forms of digital scholarship in the humanities: Niemeyer, Kyle E., Arfon M. Smith, and Daniel S. Katz. "The challenge and promise of software citation for credit, identification, discovery, and reuse." *Journal of Data and Information Quality (JDIQ)* 7, no. 4 (2016): 16.; Wardrip-Fruin, Noah, and Nick Montfort. *The New Media Reader, A User's Manual*. MIT Press (2003). Also, see examples of cross-disciplinary and open collaborative efforts such as team science projects on the Imagining America website: <https://imaginingamerica.org/>.
- [32] A recent poll of US adults found that while trust in scientists is relatively high compared to other groups, the public has substantially less trust in scientists to provide information about controversial

issues within their expertise, such as the health effects of genetically modified food. Moreover, regardless of trust in scientists as a class, it is equally concerning that much of what the public believes has remained unaffected by consensus scientific opinion on salient issues, see Egan, Patrick J. and Megan Mullin. "Climate change: US public opinion." *Annual Review of Political Science* 20 (2017): 209-227.

[33] See Lupia, Arthur. "Communicating science in politicized environments." *Proceedings of the National Academy of Sciences* 110, no. Supplement 3 (2013): 14048-1405.

[34] Allcott, Hunt and Matthew Gentzkow. 2017. "Social Media and Fake News in the 2016 Election." *Journal of Economic Perspectives* 31 (2): 211-36. Also see, for the verification of more complex information: Lewandowsky, Stephan, Ullrich KH Ecker, Colleen M. Seifert, Norbert Schwarz, and John Cook. "Misinformation and its correction: Continued influence and successful debiasing." *Psychological Science in the Public Interest* 13, no. 3 (2012): 106-131.

[35] Publons. 2018. "[Global State of Peer Review.](#)" [Clarivate Analytics](#); Bornmann, Lutz, Rüdiger Mutz, and Hans-Dieter Daniel. 2010. "A Reliability-Generalization Study of Journal Peer Reviews: A Multilevel Meta-Analysis of Inter-Rater Reliability and Its Determinants." *PLOS ONE* 5; Greenberg, Steven A. "How citation distortions create unfounded authority: analysis of a citation network." *BMJ* 339 (2009): b2680.; 12): e14331. <https://doi.org/10.1371/journal.pone.0014331>.; Franco, Annie, Neil Malhotra, and Gabor Simonovits. 2014. "Publication Bias in the Social Sciences: Unlocking the File Drawer." *Science* 345 (6203): 1502-5. <https://doi.org/10.1126/science.1255484>; Tennant, Jonathan P., Jonathan M. Dugan, Daniel Graziotin, Damien C. Jacques, François Waldner, Daniel Mietchen, Yehia Elkhatib, et al. "A multi-disciplinary perspective on emergent and future innovations in peer review." *F1000Research* 6 (2017).

[36] Allcott, H. and Gentzkow, M., 2017. "Social media and fake news in the 2016 election." *Journal of Economic Perspectives* 31 (2), pp.211-36; Lazer, David MJ, Matthew A. Baum, Yochai Benkler, Adam J. Berinsky, Kelly M. Greenhill, Filippo Menczer, Miriam J. Metzger et al. "The science of fake news." *Science* 359, no. 6380 (2018): 1094-1096.

[37] See: Alsheikh-Ali, A.A., Qureshi, W., Al-Mallah, MH and Ioannidis, J.P., 2011. Public availability of published research data in high-impact journals. *PloS one*, 6(9), p.e24357; Vines, Timothy H., Arianne YK Albert, Rose L. Andrew, Florence Débarre, Dan G. Bock, Michelle T. Franklin, Kimberly J. Gilbert, Jean-Sébastien Moore, Sébastien Renaut, and Diana J. Rennison. "The availability of research data declines rapidly with article age." *Current Biology* 24, no. 1 (2014): 94-97.

[38] Altman, et al., 2015, [National Agenda for Digital Stewardship](#), National Digital Stewardship Alliance; see also, ACRL Scholarly Communications Committee. "Establishing a research agenda for

scholarly communication: A call for community engagement." *Association of College and Research Libraries* (2007).

[39] Blue Ribbon Task Force on Sustainable Digital Preservation. "Sustainable economics for a digital planet: Ensuring long-term access to digital information." Final Report of the Blue Ribbon Task Force (2010). OCLC.

[40] See for example: Moseley, Christopher, ed. *Atlas of the World's Languages in Danger*. Unesco, 2010 and Stefano, Michelle L., Peter Davis, and Gerard Corsane, eds. *Safeguarding intangible cultural heritage*. Vol. 8. Boydell & Brewer Ltd, 2014.

[41] Diaz, Alejandro. *Through the Google Goggles: Sociopolitical Bias in Search Engine Design*. pp. 11-34. Springer, Berlin, Heidelberg, 2008; Lazer, D., Kennedy, R., King, G., and Vespignani, A.. 2014. "The parable of Google Flu: traps in big data analysis." *Science*, 343 (6176), pp.1203-1205.

[42] See President's Council of Advisors on Science and Technology (PCAST), "Big Data and Privacy: A Technological Perspective" (White House, Washington, DC, 2014); Narayanan, Arvind, Joanna Huey, and Edward W. Felten. "A precautionary approach to big data privacy." In *Data protection on the move*, pp. 357-385. Springer, Dordrecht, 2016; Altman, Micah, Alexandra Wood, David R. O'Brien, and Urs Gasser. "Practical approaches to big data privacy over time." *International Data Privacy Law* (2018);

[43] See for example, Robertson, Tara. "Not All Information Wants to be Free: The Case Study of On Our Backs." (2018) in: *Applying Library Values to Emerging Technology: Decision-Making in the Age of Open Access, Makerspaces, and the Ever-Changing Library* (Publications in Librarianship #72). American Library Association, pp. 225-239; more generally, see, Smith, Linda Tuhiwai. *Decolonizing methodologies: Research and indigenous peoples*. Zed Books Ltd., 2013.; and for a foundational, philosophical perspective see Nissenbaum, Helen. *Privacy in context: Technology, policy, and the integrity of social life*. Stanford University Press, 2009.

[44] See for example IEEE Global Initiative. "Ethically Aligned Design." *IEEE Standards* vol. 1 (2016); ACM, Computing Machinery. "ACM code of ethics and professional conduct." *Code of Ethics* (2018).

[45] See, for examples Callicott, Burton B., David Scherer, and Andrew Wesolek. *Making institutional repositories work*. Purdue University Press, 2017, and Lapinski, P. Scott, David Osterbur, Joshua Parker, and Alexa T. McCray. "Supporting public access to research results." *College & Research Libraries* 75, no. 1 (2014): 20-33; ACRL Scholarly Communications Committee. "Establishing a research agenda for scholarly communication: A call for community engagement." *Association of College and Research Libraries* (2007).

[46] For an overview of the current open access landscape in the US, UK, and the European Union, see: Dunn, Katherine, MIT Ad-Hoc Task Force on Open Access, 2018: "[Open Access at MIT and Beyond.](#)"

[47] For legal and empirical analysis see (respectively) Samuelson, Pamela. "DRM {and, or, vs.} the law." *Communications of the ACM* 46, no. 4 (2003): 41-45; Urban, Jennifer M. and Karaganis, Joe and Schofield, Brianna, "Notice and Takedown in Everyday Practice." (March 22, 2017). UC Berkeley Public Law Research Paper No. 2755628. <https://ssrn.com/abstract=2755628> or <http://dx.doi.org/10.2139/ssrn.2755628>

[48] See for an analysis, Altman, Micah, and Marguerite Avery. "Information wants someone else to pay for it: laws of information economics and scholarly publishing." *Information Services & Use* 35, no. 1-2 (2015): 57-70; and for a recent market description: Rob Johnson, Anthony Watkinson, Michael Mabe. *The STM report: An overview of scientific and scholarly journal publishing*, 5th edition. (2018) International Association of Scientific, Technical and Medical Publishers.

[49] Examples of sharing knowledge across communities: Library of Congress, Labs: <https://labs.loc.gov/>; Artist in the Archive podcast: <https://artistinthearchive.podbean.com/>

[50] For a discussion of relevant organizational design factors at the macro- and micro-levels see (respectively): Ostrom, Elinor. *Understanding institutional diversity*. Princeton, NJ: Princeton University Press, 1995.; Roberts, John. *The modern firm: Organizational design for performance and growth*. (Oxford University Press, 2007).

[51] See generally: Ostrom, E., 2000. "Collective action and the evolution of social norms." *Journal of Economic Perspectives*, 14(3), pp.137-158. For a survey of norms related to research and reproducibility, see Borgman, C.L., 2010. *Scholarship in the Digital Age: Information, infrastructure, and the Internet*. MIT Press.

[52] See Benkler, Yochai. *The wealth of networks: How social production transforms markets and freedom*. (Yale University Press, 2006.)

[53] See, Foray, Dominique. *Economics of Knowledge*. (Cambridge, MA: MIT Press, 2004); Altman, Micah and Marguerite Avery. "Information wants someone else to pay for it: laws of information economics and scholarly publishing." *Information Services & Use* 35, no. 1-2 (2015): 57-70.

[54] For a survey of modern approaches to institutional analysis, see: Peters, B. Guy. *Institutional theory in political science: The new institutionalism*. Bloomsbury Publishing USA, 2011.

[55] Goroff, Daniel, and Josh Greenberg. "Data and Decisions about Scholarly Knowledge." *Social Research: An International Quarterly* 84, no. 3 (2017): 733-737.

- [56] Funk, Carey. "Mixed Messages about Public Trust in Science." Pew Research Center. (December 8, 2017.) <http://www.pewinternet.org/2017/12/08/mixed-messages-about-public-trust-in-science/>.
- [57] Green, Toby. 2018. "We're Still Failing to Deliver Open Access and Solve the Serials Crisis: To Succeed We Need a Digital Transformation of Scholarly Communication Using Internet-Era Principles." Zenodo. <https://doi.org/10.5281/zenodo.1410000>
- [58] Lavoie, Brian, Eric R. Childress, Ricky Erway, Ixchel M. Faniel, Constance Malpas, Jennifer Schaffner, and Titia van der Werf. "[The Evolving Scholarly Record.](#)" June (2014).
- [59] National Academies of Sciences, Engineering, and Medicine (NASEM). 2018. "[Open Science by Design: Realizing a Vision for 21st Century Research.](#)"
- [60] Velterop, Jan. "Should scholarly societies embrace open access (or is it the kiss of death)?" *Learned Publishing* 16, no. 3 (2003): 167-169.
- [61] Nosek, Brian A., Jeffrey R. Spies, and Matt Motyl. "Scientific utopia: II. Restructuring incentives and practices to promote truth over publishability." *Perspectives on Psychological Science* 7, no. 6 (2012): 615-631.
- [62] Nosek, Brian A., Jeffrey R. Spies, and Matt Motyl. "Scientific utopia: II. Restructuring incentives and practices to promote truth over publishability." *Perspectives on Psychological Science* 7, no. 6 (2012): 615-631.
- [63] See Hess, Charlotte and Elinor Ostrom. "A Framework for Analyzing the Knowledge Commons: A Chapter from Understanding Knowledge as a Commons: from Theory to Practice." (2005) and Altman, Micah, and Marguerite Avery. "Information wants someone else to pay for it: laws of information economics and scholarly publishing." *Information Services & Use* 35, no. 1-2. For an exploration of innovated market-based alternatives, see Posner, Eric A. and E. Glen Weyl. *Radical Markets: Uprooting Capitalism and Democracy for a Just Society*. Princeton University Press, 2018.
- [64] See Lupia, Arthur. "Communicating science in politicized environments." *Proceedings of the National Academy of Sciences* 110, no. Supplement 3 (2013): 14048-14054.; Lupia, Arthur. "Now is the time: how to increase the value of social science." *Social Research: An International Quarterly* 84, no. 3 (2017): 669-694.
- [65] See for examples of broadening evaluation at different stages of the research process: Bonney, R., Shirk, J.L., Phillips, T.B., Wiggins, A., Ballard, H.L., Miller-Rushing, A.J. and Parrish, J.K., 2014. "Next steps for citizen science." *Science*, 343(6178), pp.1436-1437.; Ross-Hellauer T. "What is open peer review?"

A systematic review" [version 2; referees: 4 approved]. *F1000Research* 2017, 6:588 (doi: 10.12688/f1000research.11369.2).

[66] See for example Rodriguez-Esteban, Raul. "Biomedical text mining and its applications." *PLoS Computational Biology* 5, no. 12 (2009): e1000597.

[67] Allcott, H. and Gentzkow, M., 2017. Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, 31(2), pp.211-36; Lazer, David MJ, Matthew A. Baum, Yochai Benkler, Adam J. Berinsky, Kelly M. Greenhill, Filippo Menczer, Miriam J. Metzger, et al. "The science of fake news." *Science* 359, no. 6380 (2018): 1094-1096.

[68] For examples: Unpaywall brings together open articles from multiple sources, although it is limited to items with a DOI: <https://unpaywall.org/>; The Keepers Registry is tracking journal preservation <https://thekeepers.org/>; Piwowar H., Priem J., Larivière V., Alperin JP, Matthias L., Norlander B., Farley A., West J., Haustein S. (2018). "[The state of OA: a large-scale analysis of the prevalence and impact of Open Access articles.](#)" *PeerJ* 6:e4375.

[69] For example, recent versions of Google's search algorithm potentially incorporate hundreds of general and personalized ranking signals (measures), social-network-based ranking factors, and the context massive knowledge graph in order to customize search results to specific queries and context. See (respectively): Enge, Eric, (2018) "[Ranking Factors Session Recap from Smx 2018,](#)" *Search Engine Land*. Neystadt, Eugene John, Ron Karidi, Yitzhak Tzahi Weisfeld, Roy Varshavsky, Avigad Oron, and Kira Radinsky. "Social network based contextual ranking." U.S. Patent 9,870,424, issued January 16, 2018; Paulheim, Heiko. "Knowledge graph refinement: A survey of approaches and evaluation methods." *Semantic Web* 8, no. 3 (2017): 489-508.

[70] See: Hedstrom, Margaret. "Digital preservation: a time bomb for digital libraries." *Computers and the Humanities* 31, no. 3 (1997): 189. Neil Beagrie (2000) "The JISC digital preservation focus and the digital preservation coalition," *New Review of Academic Librarianship*, 6:1, 257-267, DOI: 10.1080/13614530009516815

[71] See for example McLure, Merinda, Allison V. Level, Catherine L. Cranston, Beth Oehlerts, and Mike Culbertson. "Data curation: a study of researcher practices and needs." *portal: Libraries and the Academy* 14, no. 2 (2014): 139-164; Jahnke, Lori M., and Andrew Asher. "The problem of data: Data management and curation practices among university researchers." L. Jahnke, A. Asher & SDC Keralis, *The problem of data* (2012): 3-31, CLIR.; Borgman, Christine L. *Big data, little data, no data: Scholarship in the networked world*. MIT Press, 2015.

[72] See Harvey, Douglas Ross. *Preserving digital materials*. Walter de Gruyter, 2008.; Altman, et al., 2015, [National Agenda for Digital Stewardship](#), National Digital Stewardship Alliance. For a discussion of the related limitations of selection and collection-building strategies in libraries, see Dempsey, Lorcan. "Library collections in the life of the user: two directions." *LIBER Quarterly* 26, no. 4 (2016).

[73] See, for example: Zeng, Jiaan, Guangchen Ruan, Alexander Crowell, Atul Prakash, and Beth Plale. "Cloud computing data capsules for non-consumptive use of texts." In *Proceedings of the 5th ACM Workshop on Scientific Cloud Computing*, pp. 9-16. ACM, 2014.

[74] Arrow, K.J., 1972. "Economic welfare and the allocation of resources for invention." In *Readings in Industrial Economics* (pp. 219-236). Palgrave, London.

[75] Gans, Joshua S., and Scott Stern. "Is there a market for ideas?" *Industrial and Corporate Change* 19, no. 3 (2010): 805-837.

[76] Hess, Charlotte and Elinor Ostrom. "A Framework for Analyzing the Knowledge Commons: a chapter from Understanding Knowledge as a Commons: from Theory to Practice." (2005).

[77] R. Price, G. Shanks, 2005. "A Semiotic Information Quality Framework: development and comparative analysis." *Journal of Information Technology* 20: 88-102. S.E. Madnick, R.Y. Wang, Y.W. Lee, H. Zhu, 2009. "Overview and Framework for Data and Information Quality Research." *ACM Journal of Data and Information Quality* 1(2) 1-22; Altman, Micah. "Mitigating threats to data quality throughout the curation lifecycle." in G. Marciano, C. Lee, and H. Bowden, *Curating for Quality: Ensuring Data Quality to Enable New Science*. National Science Foundation, Arlington County, VA (2012): 1-119.

[78] See, for an introduction: John Loomis. "Valuing Environmental and Natural Resources: The Econometrics of Non-Market Valuation." *American Journal of Agricultural Economics*, Volume 87, Issue 2, 1 May 2005, Pages 529-530.

[79] Markowitz, Harry. "Portfolio selection." *The Journal of Finance* 7, no. 1 (1952): 77-91.

[80] See Altman et al., 2015, [National Agenda for Digital Stewardship](#), National Digital Stewardship Alliance; Goroff, Daniel and Josh Greenberg. "Data and Decisions about Scholarly Knowledge." *Social Research: An International Quarterly* 84, no. 3 (2017): 733-737.

[81] See for a review of recommendation system algorithms: Park, Deuk Hee, Hyea Kyeong Kim, Il Young Choi, and Jae Kyeong Kim. "A literature review and classification of recommender systems research." *Expert Systems with Applications* 39, no. 11 (2012): 10059-10072.

[82] Andronis, Christos, Anuj Sharma, Vassilis Virvilis, Spyros Deftereos, and Aris Persidis. "Literature mining, ontologies and information visualization for drug repurposing." *Briefings in Bioinformatics* 12, no. 4 (2011): 357-368.

[83] See, for example: Farid, Hany. "Image forgery detection." *IEEE Signal Processing Magazine* 26, no. 2 (2009): 16-25.; Parrish, Debra, and Bridget Noonan. "Image manipulation as research misconduct." *Science and Engineering Ethics* 15, no. 2 (2009): 161-167; Engels, Steve, Vivek Lakshmanan, and Michelle Craig. "Plagiarism detection using feature-based neural networks." *ACM SIGCSE Bulletin* 39, no. 1 (2007): 34-38.

[84] Adams, James D., Grant C. Black, J. Roger Clemmons, and Paula E. Stephan. "Scientific teams and institutional collaborations: Evidence from US universities, 1981-1999." *Research Policy* 34, no. 3 (2005): 259-285.; Altman, Micah and Marguerite Avery. "Information wants someone else to pay for it: laws of information economics and scholarly publishing." *Information Services & Use* 35, no. 1-2 (2015): 57-70.

[85] Abbasi, Alireza, Jörn Altmann, and Liaquat Hossain. "Identifying the effects of co-authorship networks on the performance of scholars: A correlation and regression analysis of performance measures and social network analysis measures." *Journal of Informetrics* 5, no. 4 (2011): 594-607.

[86] See, for example: Wang, H., Wang, N. and Yeung, D.Y., 2015, August. "Collaborative deep learning for recommender systems." In Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 1235-1244). ACM.

[87] Diaz, Alejandro. "Through the Google goggles: Sociopolitical bias in search engine design." In *Web Search*, pp. 11-34. Springer, Berlin, Heidelberg, 2008; Lazer, D., Kennedy, R., King, G. and Vespignani, A., 2014. "The parable of Google Flu: traps in big data analysis." *Science*, 343(6176), pp.1203-1205.

[88] Ohm, Paul. "Broken promises of privacy: Responding to the surprising failure of anonymization." *Ucla L. Rev.* 57 (2009): 1701.

[89] See Phillips, P. Jonathon, Fang Jiang, Abhijit Narvekar, Julianne Ayyad, and Alice J. O'Toole. "An other-race effect for face recognition algorithms." *ACM Transactions on Applied Perception (TAP)* 8, no. 2 (2011); and for evidence of severity and ubiquity of the problem, see White, D., Dunn, JD, Schmid, AC, and Kemp, R.I., 2015. "Error rates in users of automatic face recognition software." *PLoS One*, 10(10), p.e0139827.; Garvie, Clare. "The perpetual line-up: Unregulated police face recognition in America." Georgetown Law, Center on Privacy & Technology, 2016.

[90] See, for example: Sweeney, L., 2013. "Discrimination in online ad delivery." *Queue*, 11(3), p.10; Sadler, Bess and Chris Bourg. 2015. "Feminism and the future of library discovery." *Code4Lib* 10; Noble,

Safiya Umoja. *Algorithms of Oppression: How search engines reinforce racism*. NYU Press, 2018. Eubanks, V., 2018. *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press.

[91] For the difficulties of establishing authenticity of web-based information, see: Aturban, M., Nelson, ML, and Weigle, MC, 2017. "Difficulties of Timestamping Archived Web Pages." arXiv preprint arXiv:1712.03140; and for the surprising difficulties associated with the much simpler task of verifying that numerical data has been unaltered, see: Altman, Micah. "A fingerprint method for scientific data verification." In *Advances in Computer and Information Sciences and Engineering*, pp. 311-316. Springer, Dordrecht, 2008.; For active approaches to create sophisticated false information, see: Chesney, Robert and Danielle Keats Citron. "Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security." *107 California Law Review* (2019, Forthcoming).

[92] Moosavi-Dezfooli, Seyed-Mohsen, Alhussein Fawzi, Omar Fawzi, and Pascal Frossard. "Universal adversarial perturbations." 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2017.

[93] For example, recommender systems have become ubiquitous in discovery, but generally fail to protect information privacy. Both new algorithm development and careful analysis is necessary to develop recommendation algorithms that preserve information privacy. See: McSherry, Frank and Ilya Mironov. "Differentially private recommender systems: Building privacy into the netflix prize contenders." In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 627-636. ACM, 2009. For a review of challenges in protecting privacy in algorithmic systems, and tensions with transparency and open data requirements, see Altman, Micah, Alexandra Wood, David R. O'Brien, and Urs Gasser. "Practical approaches to big data privacy over time." *International Data Privacy Law* (2018); and Altman, Micah, Alexandra Wood, David R. O'Brien, Salil Vadhan, and Urs Gasser. "Towards a modern approach to privacy-aware government data releases." *Berkeley Technology Law Journal*. 30 (2015): 1967.

[94] See, for example Pasquali, Matias. "Video in science: Protocol videos: the implications for research and society." *EMBO Reports* 8, no. 8 (2007): 712-716.; and

[95] Lithgow, Gordon J., Monica Driscoll, and Patrick Phillips. "A long journey to reproducible results." *Nature News* 548.7668 (2017): 387.

[96] Moss-Racusin, Corinne A., John F. Dovidio, Victoria L. Brescoll, Mark J. Graham, and Jo Handelsman. "Science faculty's subtle gender biases favor male students." *Proceedings of the National Academy of Sciences* 109, no. 41 (2012): 16474-16479.

[97] See, for example, an analysis of the theoretical and practical challenges of capturing oral history and “folkloric” information generally: Owens, Trevor. *The Theory and Craft of Digital Preservation*, John Hopkins University Press (2018); Kovach, Margaret. *Indigenous Methodologies: Characteristics, Conversations, and Contexts*. Toronto: University of Toronto Press (2009).

[98] See: Ostrom, E. and C. Hess, 2007, *Understanding knowledge as a commons: From theory to practice*. MIT Press.

[99] Altman et al. 2015. [National Agenda for Digital Stewardship](#), National Digital Stewardship Alliance.

[100] Becker, S. Adams, Michele Cummins, A. Davis, A. Freeman, C. Hall Giesinger, V. Ananthanarayanan, K. Langle, and N. Wolfson. NMC Horizon Report: 2017 library edition. The New Media Consortium, 2017.

[101] See for reviews: Altman, Micah. "Open source software for Libraries: from {Greenstone} to the {Virtual Data Center} and beyond." *iassist Quarterly* 25 (2002) and Lesk, Michael. "A personal history of digital libraries." *Library Hi Tech* 30, no. 4 (2012): 592-603. Also, for recent work, see the *Code4Lib* journal archives at <https://journal.code4lib.org/issues>.

[102] See for prescient identification of this opportunity: ACRL Scholarly Communications Committee. "Establishing a research agenda for scholarly communication: A call for community engagement." Association of College and Research Libraries (2007).

[103] Delany, David. "A review of the literature on effective PhD supervision." *Centre for Academic Practice and Student Learning. Trinity College, Dublin Google Scholar* (2008): Brown, Ronald T., Brian P. Daly, and Frederick TL Leong. "Mentoring in research: A developmental approach." *Professional Psychology: Research and Practice* 40, no. 3 (2009): 306. Scaffidi, Amelia K., and Judith E. Berman. "A positive postdoctoral experience is related to quality supervision and career mentoring, collaborations, networking and a nurturing research environment." *Higher Education* 62, no. 6 (2011): 685.

[104] See for an analysis, Altman, Micah, and Marguerite Avery. "Information wants someone else to pay for it: laws of information economics and scholarly publishing." *Information Services & Use* 35, no. 1-2 (2015): 57-70; and for a recent market description: Rob Johnson, Anthony Watkinson, and Michael Mabe. *The STM Report: An overview of scientific and scholarly journal publishing, 5th edition*. (2018) International Association of Scientific, Technical and Medical Publishers.

[105] Acquisti, Alessandro, Curtis Taylor, and Liad Wagman. "The economics of privacy." *Journal of Economic Literature* 54, no. 2 (2016): 442-92.

[106] Goldman, Eric. "Revisiting search engine bias." *William Mitchell Law Review* 38 (2011): 96.; Joshua G. Hazan, "Stop Being Evil: A Proposal for Unbiased Google Search," *Michigan Law Review* 111, no. 5 (March 2013): 789-820; Nathan Newman, "Search, Antitrust, and the Economics of the Control of User Data," *Yale Journal on Regulation* 31, no. 2 (Summer 2014): 401-454.

[107] See for example, Pasquale, Frank. *The black box society: The secret algorithms that control money and information*. Harvard University Press, 2015; O'Neil, Cathy. *Weapons of math destruction: How big data increases inequality and threatens democracy*. Broadway Books, 2016.

[108] Ayyagari, Ramakrishna. "An exploratory analysis of data breaches from 2005-2011: Trends and insights." *Journal of Information Privacy and Security* 8, no. 2 (2012): 33-56.

[109] See Ohm, Paul. "Broken promises of privacy: Responding to the surprising failure of anonymization." *UCLA Law Review* 57 (2009): 1701; Altman, Micah, Alexandra Wood, David R. O'Brien, and Urs Gasser. "Practical approaches to big data privacy over time." *International Data Privacy Law* (2018).

[110] More information about these organizations can be found on their websites:

<https://www.clir.org/about/> ; <https://ndsa.org/> ; <https://diglib.org/> ; <https://www.rd-alliance.org/> ; <https://sparcopen.org/> ; and <http://www.codata.org/>

[111] More information about these organizations can be found on their websites: <https://duraspace.org/> ; <https://dataverse.org/> ; <https://www.dpn.org/> ; and <https://cos.io/>

[112] See, respectively, <http://ndsa.org>, <http://longnow.org>, <http://dpconline.org>, and reports produced by these organizations, such as: N. Beagrie, M. Joves. *Digital Preservation Handbook*. Digital Preservation Coalition, 2009; Brand, Stewart. *The clock of the long now: Time and responsibility*. Basic Books, 2008.

[113] See Graham, Mark and Anasuya Sengupta. "We're all connected now, so why is the Internet so white and Western?" *The Guardian* 5 (2017).