

#3

COMPLETE

First Name: David
Last Name: Weinberger
Custom Data: scholarly discovery

Page 4: Candidate Grand Challenge Problems

Q4 What is the most significant grand challenge problem in {{ contact.custom1 }}? Why it is significant and to whom? Identify the stakeholder groups in the problem and their relationship to it. Briefly identify which aspects of the problem are technical, social, political, economic, ethical, or institutional.

Building an open, public knowledge graph ("Library Graph") that includes everything that libraries know...for starters! For this we need access to metadata and content, interoperable standards, distributed infrastructure, designers, builders, maintainers, a front end that is usable by content researchers and enticing to casual explorers, resources for digital humanities researchers doing computationally intense work, an open API...and more.

Libraries, the Internet Archive's Open Library, Google Books, OCLC, the DPLA and Linked Data for Libraries are potential contributors (as are others I am not remembering or don't know about -- so many great projects roughly in this space).

Ideally this would be built and maintained primarily by a distributed network of libraries.

Q5 How will we know we've solved the problem? Criteria of success may include quantitative metrics such economic impact and number of adopters; and qualitative changes in individual or organizational behavior, professional practice, and changing incentives.

A growing Library Graph will exist. Users will come to its front door to do research. Other sites, apps, and services will start to incorporate its content into their own. Not to mention that people will at last know what to link to when referring to a book!

Q6 Who is in the best position to address those challenges? Identify domains, stakeholders, institutions, funding agencies, particular individuals, and other parties as appropriate. In particular, who is addressing this grand challenge most effectively? How have they made a positive difference? Based on their experience, which components of the problem appear the most tractable and which the least? What other aspects of the challenge are most urgent to address?

Libraries, the Internet Archive's Open Library, Google Books, OCLC, the DPLA and Linked Data for Libraries are potential contributors (as are others I am not remembering or don't know about -- so many great projects roughly in this space).

Q7 If you like, please upload a document which answers the questions above or supplements your answers. **Respondent skipped this question**

MIT Libraries Grand Challenges Summit -- Problem Survey

Q8 If we quote from your response, should we use your name for attribution? **Yes**

#4

COMPLETE

First Name: Richard
Last Name: Whitt
Email: whitt@google.com
Custom Data: digital curation and preservation

Page 4: Candidate Grand Challenge Problems

Q4 What is the most significant grand challenge problem in {{ contact.custom1 }}? Why it is significant and to whom? Identify the stakeholder groups in the problem and their relationship to it. Briefly identify which aspects of the problem are technical, social, political, economic, ethical, or institutional.

Business models to pay for the technology.

Q5 How will we know we've solved the problem? Criteria of success may include quantitative metrics such economic impact and number of adopters; and qualitative changes in individual or organizational behavior, professional practice, and changing incentives.

When digital preservation becomes a viable commercial business.

Q6 Who is in the best position to address those challenges? Identify domains, stakeholders, institutions, funding agencies, particular individuals, and other parties as appropriate. In particular, who is addressing this grand challenge most effectively? How have they made a positive difference? Based on their experience, which components of the problem appear the most tractable and which the least? What other aspects of the challenge are most urgent to address?

Government policymakers, archivists, NGOs, and technology companies are best situated to address these challenges. Some smaller companies are trying to launch alternative business models, but the early returns are unclear.

Q7 If you like, please upload a document which answers the questions above or supplements your answers.

Whitt, Through A Glass Darkly, Technical Policy and Financial Actions to Avert the Coming Digital Dark Ages 2017.pdf (1.4MB)

Q8 If we quote from your response, should we use your name for attribution? Yes

#5

COMPLETE

First Name: Peter
Last Name: Suber
Custom Data: open scholarship

Page 4: Candidate Grand Challenge Problems

Q4 What is the most significant grand challenge problem in {{ contact.custom1 }}? Why it is significant and to whom? Identify the stakeholder groups in the problem and their relationship to it. Briefly identify which aspects of the problem are technical, social, political, economic, ethical, or institutional.

Persuading researchers themselves to make their own work OA. It could be OA through repositories (green) or OA through journals (gold); for this purpose that doesn't matter. It should be OA somehow. Researchers should understand the benefits, understand their OA options, want OA for their own work, and take the steps necessary to make it happen -- even if armies of helpers are standing by to make it easier for them. No aspect of this challenge is technical, although some tools could make the job easier. Some part is economic; the green OA infrastructure needs adequate funding; OA journals need revenue to cover their expenses. Some part is institutional; there are many ways that institutions could create incentives for authors to make their work OA.

No part of the challenge **need** be ethical. Some authors think it's an ethical duty to make their work OA. Good. I agree with them, and this sense of duty certainly works as a motive. But even authors who don't think it's an ethical duty will benefit by making their work OA and can be persuaded to do so.

Q5 How will we know we've solved the problem? Criteria of success may include quantitative metrics such economic impact and number of adopters; and qualitative changes in individual or organizational behavior, professional practice, and changing incentives.

I'd follow the percentage of new articles that are OA from birth. The numbers differ by discipline today and may always differ by discipline. So I'd track this number in as many fields as possible. It would help to break the numbers down by method or venue of OA (green v. gold) and by the type of license (from all-rights-reserved copyrights to CC-BY and CC0).

Another possible metric: If 80% of the articles are published by 20% of the authors, then identify the 20% (field by field), and track how many of those authors routinely make their new work OA. The first metric above tracks articles, while this one tracks authors.

MIT Libraries Grand Challenges Summit -- Problem Survey

Q6 Who is in the best position to address those challenges? Identify domains, stakeholders, institutions, funding agencies, particular individuals, and other parties as appropriate. In particular, who is addressing this grand challenge most effectively? How have they made a positive difference? Based on their experience, which components of the problem appear the most tractable and which the least? What other aspects of the challenge are most urgent to address?

Authors are in the best position to decide whether to make their own new work OA. (They decide whether to submit it to OA journals, whether to deposit it in OA repositories, and how to manage their copyrights.) Funders and universities are in the best position to influence author decisions, through policies and incentives.

Even though most funders and universities have not yet adopted OA policies, adopting good policies is easier than implementing them. One challenge is to persuade more funders and universities to adopt good policies. Another challenge is to implement these policies effectively (this includes adequate funding for implementation). A third challenge common to the first two is to avoid antagonizing authors, and to depend on expectations, education, incentives, and assistance, not coercion. To respect academic freedom, university policies must usually stop short of exceptionless mandates. But funder policies can make OA an exceptionless condition of accepting a grant.

A related challenge is to adopt "good" policies and learn from the experience of institutions that have gone before. Apologies for the selfie, but for university policies I recommend the guide to good practices for university OA policies that I maintain with Stuart Shieber <<http://bit.ly/goodoa>>. We're working on a companion guide for funder OA policies.

Q7 If you like, please upload a document which answers the questions above or supplements your answers.

Respondent skipped this question

Q8 If we quote from your response, should we use your name for attribution?

Yes

#6

COMPLETE

First Name: Amy
Last Name: Brand
Custom Data: open scholarship

Page 4: Candidate Grand Challenge Problems

Q4 What is the most significant grand challenge problem in {{ contact.custom1 }}? Why it is significant and to whom? Identify the stakeholder groups in the problem and their relationship to it. Briefly identify which aspects of the problem are technical, social, political, economic, ethical, or institutional.

Lack of institutional investment in and ownership of scholarly communication ecosystem, and the stranglehold that a handful of commercial entities maintain over not only the markets for research information, but also on academic reputation systems, publishing technologies, and digital innovation in scholarly communication. Concerns all stakeholder groups, but mainly has to do with university leadership recognizing the problem and the need to coordinate internally and externally, with other institutions. Has to do with university technical infrastructure, and with entrenched researcher workflows that can only be superseded if new efficiencies are provided to scholars in managing and sharing their research.

Q5 How will we know we've solved the problem? Criteria of success may include quantitative metrics such economic impact and number of adopters; and qualitative changes in individual or organizational behavior, professional practice, and changing incentives.

When this problem is solved, universities will have embraced academy-owned scholarly communication; university presses and other university-based publishing will be embraced on campus; researchers will experience much more streamlined workflows for their scholarship and the management of the metadata, or record, of that scholarship, with integrated systems for institutional FT deposit, reputation and productivity tracking, and CVs or profiles; academic libraries will no longer need to buy back as much of that scholarly record or the data about it; there will be greater cross-university integration, data sharing, and discovery.

MIT Libraries Grand Challenges Summit -- Problem Survey

Q6 Who is in the best position to address those challenges? Identify domains, stakeholders, institutions, funding agencies, particular individuals, and other parties as appropriate. In particular, who is addressing this grand challenge most effectively? How have they made a positive difference? Based on their experience, which components of the problem appear the most tractable and which the least? What other aspects of the challenge are most urgent to address?

Leadership needs to recognize the constellation of issues at a high level and agree to invest and disintermediate accordingly -- to insource more, outsource less, and to coordinate with other institutions. Libraries need to takeover some of the functions that are sometimes associated with institutional research or faculty administration offices -- namely, managing the metadata about outputs and productivity, not just the full-text within institutional repositories. In academy-owned scholarly communication, there is a more central role of university presses and other university-based publishers of scholarly works. Non-profit publishers have shown how to make research outputs affordable and/or OA. One complex challenge concerns the availability of tools and technologies that researchers and mission-driven publishers can use to participate directly, and in lieu of proprietary, expensive third-party models and platforms. Another complex problem is the reputation economy. We need to grow our means of capturing and assessing academic prestige and impact.

Q7 If you like, please upload a document which answers the questions above or supplements your answers.

Respondent skipped this question

Q8 If we quote from your response, should we use your name for attribution?

Yes

#7

COMPLETE

First Name: Roger
Last Name: Mark
Custom Data: digital curation and preservation

Page 4: Candidate Grand Challenge Problems

Q4 What is the most significant grand challenge problem in {{ contact.custom1 }}? Why it is significant and to whom? Identify the stakeholder groups in the problem and their relationship to it. Briefly identify which aspects of the problem are technical, social, political, economic, ethical, or institutional.

Investigators at universities such as MIT generate enormous amounts of research data, on the order of petabytes per year, in a wide variety of forms (time series, waveforms, images, videos, documents, spreadsheets, etc.). The challenge is to preserve these data to support verification and extension of published studies, and to make expensive research data available for re-use in new studies.

These data sets should be annotated, preserved and shared openly and efficiently with the research community. The process for achieving that goal is currently inadequate, and much valuable research data is not discoverable and opportunities for re-use are gone forever. The university library is the ideal institution to provide accessible, convenient (local), and affordable systems for research data management, archiving, and dissemination to the broad research community. Libraries have highly developed expertise and infrastructures for archiving, cataloging and disseminating information. However, the deluge of digital research data and associated software has far outpaced their existing infrastructure.

The grand challenge for the MIT libraries is to develop the necessary methods, algorithms, protocols and financing to acquire, manage, share and publish the research data and software produced by the Institute's investigators.

Q5 How will we know we've solved the problem? Criteria of success may include quantitative metrics such economic impact and number of adopters; and qualitative changes in individual or organizational behavior, professional practice, and changing incentives.

When each research lab has an individualized and tailored interface to the library's research data archive that feels supportive and comfortable to use. When the research data is made find-able and accessible to the community inside and outside of the institution

Q6 Who is in the best position to address those challenges? Identify domains, stakeholders, institutions, funding agencies, particular individuals, and other parties as appropriate. In particular, who is addressing this grand challenge most effectively? How have they made a positive difference? Based on their experience, which components of the problem appear the most tractable and which the least? What other aspects of the challenge are most urgent to address?

University library systems are a good place to start because they are close to the investigators and are more likely to succeed in working collaboratively with the investigators to acquire the data and to provide technical support in annotating and archiving them.

MIT Libraries Grand Challenges Summit -- Problem Survey

Q7 If you like, please upload a document which answers the questions above or supplements your answers.

Respondent skipped this question

Q8 If we quote from your response, should we use your name for attribution?

Yes

#8

COMPLETE

First Name: Alison
Last Name: Mudditt
Custom Data: open scholarship

Page 4: Candidate Grand Challenge Problems

Q4 What is the most significant grand challenge problem in {{ contact.custom1 }}? Why it is significant and to whom? Identify the stakeholder groups in the problem and their relationship to it. Briefly identify which aspects of the problem are technical, social, political, economic, ethical, or institutional.

That of perverse incentives that drive and reward the wrong behavior. At its heart, this is a cultural problem within science and academia: research is collaborative but it is also highly competitive. It's a system driven in large part by the culture of prestige, and the currency of that system is publication in a prestigious (= high impact) journal. Adding to that, it's far easier to publish exciting, novel results and so the temptation for researchers to rush into print, cut corners, exaggerate their findings, and overstate the significance of their work is magnified.

This problem is significant for all actors in the system, and is a major barrier to changing behavior among all but the idealists and existing supporters of change. Funders have started to chip away at this, but ultimately change is going to require a big lift on the part of institutions.

Q5 How will we know we've solved the problem? Criteria of success may include quantitative metrics such economic impact and number of adopters; and qualitative changes in individual or organizational behavior, professional practice, and changing incentives.

Changing researcher behavior requires a different system of rewards and incentives and ultimately, that means a change in the ways that institutions assess faculty. Success metrics here could include:

- The number of institutions implementing (and not just signing) DORA and its recommendations;
- Growth in submission to and publication in open publication outlets (whether at publishers like PLOS, through institutions/libraries, or funders' open platforms);
- Number of traditional, closed publishers/journals moving to open models;
- Development and adoption of new, article-level metrics

MIT Libraries Grand Challenges Summit -- Problem Survey

Q6 Who is in the best position to address those challenges? Identify domains, stakeholders, institutions, funding agencies, particular individuals, and other parties as appropriate. In particular, who is addressing this grand challenge most effectively? How have they made a positive difference? Based on their experience, which components of the problem appear the most tractable and which the least? What other aspects of the challenge are most urgent to address?

Many stakeholders have been chipping away at this issue, but the entrenched culture of prestige in academia is hard to shift. Progress ultimately requires concerted action from institutions, but many stakeholders can and are playing a role:

- Scholarly and scientific societies have a critical role to play: cultures are shaped by disciplinary communities and societies shape those community norms. In psychology, APS has been actively reshaping norms by adopting the TOP guidelines, broadening incentives through badging, and persistently and prominently covering these issues.

- Collaborative, cross-community efforts are also essential - culture/behavior change also requires the right infrastructure to support it. So, initiatives such as DORA, the CRediT taxonomy, ORCID, TOP guidelines provide alternative structures, and begin to create alternative workflows (although for some period of time, the linking of all the infrastructures we have created – some of which is well over 20 years old – will need to rely on standards to link common denominators together).

- Funders have made a difference in requiring OA publication and, increasingly, publication of underlying data sets. Although their ability to drive change is less of a carrot and more of a stick - but that is obviously effective.

- In Europe at least, governments have played a positive role in both mandating and funding openness, and building the necessary infrastructure.

Ultimately, this kind of culture change requires multiple actors working at multiple levels in the system. Brian Nosek of COS has a neat model that summarizes this (see attached PP file).

Q7 If you like, please upload a document which answers the questions above or supplements your answers.

Nosek.CultureChange.2018.02.20[2].pdf (2MB)

Q8 If we quote from your response, should we use your name for attribution? **Yes**

#9

COMPLETE

First Name: Sayeed
Last Name: Choudhury
Custom Data: any of the summit focus areas

Page 4: Candidate Grand Challenge Problems

Q4 What is the most significant grand challenge problem in {{ contact.custom1 }}? Why it is significant and to whom? Identify the stakeholder groups in the problem and their relationship to it. Briefly identify which aspects of the problem are technical, social, political, economic, ethical, or institutional.

The most significant grand challenge problem is the development of a linked data, resource-centric (as opposed to repository-centric) open scholarly infrastructure. It is significant because it affects every aspect of the research workflow at both an individual and institutional level. Perhaps equally important, there are developments by for-profit entities (e.g., Elsevier) that are noteworthy but may result in closed systems (especially as they relate to access to and sharing of data). With our current trajectory, it is not unreasonable to assume that university central administration (e.g., Provost, Vice President for Research) might choose to adopt one of these closed systems given the associated efficiency and integration. Individual researchers may not care about the issue of open or closed given their desire for seamless workflows throughout the entire research lifecycle.

While there are technical problems to solve, the main problems relate to a collective capacity challenge with developing and choosing robust software in a collective, meritocratic manner. While this challenge relates to issues of governance, it also reflects the challenge for higher education of knowing when to lead and when to follow.

Q5 How will we know we've solved the problem? Criteria of success may include quantitative metrics such economic impact and number of adopters; and qualitative changes in individual or organizational behavior, professional practice, and changing incentives.

The most important criterion for success needs to be adoption by individual researchers and their respective university administration. More to the point, the *voluntary* adoption of this open scholarly infrastructure (and associated services) needs to reach the point of "If you remove this, I will complain..." level of engagement. It needs to be viewed as important as other research infrastructure that are becoming critical such as high-performance computing.

Additional metrics for success include seamless integration between components of the infrastructure that result in greater efficiency with the research process (e.g., simultaneous submission of articles to publishers and funding agencies, single authentication and authorization). This class of metrics relates more to the functionality of the open scholarly infrastructure which, in turn, should permit evaluation in terms of what is required for greater adoption and engagement.

Q6 Who is in the best position to address those challenges? Identify domains, stakeholders, institutions, funding agencies, particular individuals, and other parties as appropriate. In particular, who is addressing this grand challenge most effectively? How have they made a positive difference? Based on their experience, which components of the problem appear the most tractable and which the least? What other aspects of the challenge are most urgent to address?

While some funding agencies have attempted to develop such community infrastructure, there are various inhibitors or blockers with their standard processes or procedures that make it difficult (though not impossible) to achieve this goal. It seems on too many occasions, grant funding is viewed as an accomplishment unto itself rather than a means to accomplish a broader goal. The recent 2.5% initiative attempts to engage institutions more directly. While this may be necessary, it is important to first identify the overall needs (e.g., use cases, requirements) and then seek appropriate levels of funding rather than define a benchmark for funding in advance. If one considers the development of scientific instruments (e.g., a telescope), it is often the case that academics define the needs in a clear, concise manner that is conveyed to a contractor (typically for-profit) that builds the instrument according to these well-defined specifications. It may be necessary to adopt this type of partnership without worrying about whether it is "open" -- rather the resulting system should be open in the sense that data can be imported, exported, migrated, etc. without encumbrances or barriers (subject to appropriate policies).

Our researchers have already chosen tools, services, etc. from for-profit entities (ranging from Google office suite to Elsevier's evolving suite) because there is not a viable or useful academic alternative. It will be increasingly challenging to ask (much less require) our researchers to move away from using these existing tools, which is another reason to consider what type of partnerships may be necessary.

Q7 If you like, please upload a document which answers the questions above or supplements your answers.

Respondent skipped this question

Q8 If we quote from your response, should we use your name for attribution?

Yes

#10

COMPLETE

Page 4: Candidate Grand Challenge Problems

Q4 What is the most significant grand challenge problem in {{ contact.custom1 }}? Why it is significant and to whom? Identify the stakeholder groups in the problem and their relationship to it. Briefly identify which aspects of the problem are technical, social, political, economic, ethical, or institutional.

What do you mean by scholarly discovery? advances in science, advances in human understanding, etc? The challenges are:

1. knowing what others have already done. Barriers are the fragmentation of knowledge by paywalls, by social and economic inequality, and by insufficient linking of dispersed knowledge assets. Linking requires both an infrastructure and metadata.
2. thinking creatively about research. Out of the box thinking requires cultures which take risks, diversity of teams and ideas, and a focus on the problem.

stakeholders are knowledge creators, publishers, tech experts, readers, administrators, and the government

Q5 How will we know we've solved the problem? Criteria of success may include quantitative metrics such economic impact and number of adopters; and qualitative changes in individual or organizational behavior, professional practice, and changing incentives.

1. solutions will be adopted and implemented, and we'll move on to the next challenges.
 2. all papers and data will be OA
 3. Publishers will have imploded.
 4. Authors will have a one button way to share their work.
 5. There will not be institutional repositories, but rather there will be nodes (disciplinary?) which present economical ways of storing and sharing information.
 6. There will need to be someone to promote the shared knowledge. It won't be enough to just deposit it. There need to be ways to make it visible and useful for others. AI may help.
 7. Do ROI on having knowledge be open. If ideas travel faster and generate advances (lives saved, hunger sated, better health and welfare for people), are there also financial benefits (fewer hospitalizations, more durable goods sold to more prosperous people, less money devoted to warfare and maiming
-

MIT Libraries Grand Challenges Summit -- Problem Survey

Q6 Who is in the best position to address those challenges? Identify domains, stakeholders, institutions, funding agencies, particular individuals, and other parties as appropriate. In particular, who is addressing this grand challenge most effectively? How have they made a positive difference? Based on their experience, which components of the problem appear the most tractable and which the least? What other aspects of the challenge are most urgent to address?

Authors and researchers are key to change. Could one get all Nobel Laureates to promote access to their work? Could you get the Nobel Prize Committee to only award to people who freely disseminate and share their knowledge? Can one get private foundations to require a plan for open dissemination?

I'm not sanguine about the role of government. In the UK, the Finch report resulted in higher costs to leading research universities and the enrichment of publishers. Governments can regulate, but really can't lead change.

Q7 If you like, please upload a document which answers the questions above or supplements your answers.

Respondent skipped this question

Q8 If we quote from your response, should we use your name for attribution?

No

#11

COMPLETE

First Name: David
Last Name: Rosenthal

Page 4: Candidate Grand Challenge Problems

Q4 What is the most significant grand challenge problem in {{ contact.custom1 }}? Why it is significant and to whom? Identify the stakeholder groups in the problem and their relationship to it. Briefly identify which aspects of the problem are technical, social, political, economic, ethical, or institutional.

We know how to curate and preserve digital information. We can't afford to do it at the scale required. The problem is economic.

Q5 How will we know we've solved the problem? Criteria of success may include quantitative metrics such economic impact and number of adopters; and qualitative changes in individual or organizational behavior, professional practice, and changing incentives.

Given the rapid evolution of digital technology, the problem is on-going. It will never be "solved". A suitable metric of progress would be an increase in the probability of content being preserved.

Q6 Who is in the best position to address those challenges? Identify domains, stakeholders, institutions, funding agencies, particular individuals, and other parties as appropriate. In particular, who is addressing this grand challenge most effectively? How have they made a positive difference? Based on their experience, which components of the problem appear the most tractable and which the least? What other aspects of the challenge are most urgent to address?

The Internet Archive is the only institution operating at anything like the necessary scale.

Q7 If you like, please upload a document which answers the questions above or supplements your answers.

RosenthalGrandChallenges.pdf (374.3KB)

Q8 If we quote from your response, should we use your name for attribution? **Yes**

#12

COMPLETE

First Name: Nancy
Last Name: McGovern

Page 4: Candidate Grand Challenge Problems

Q4 What is the most significant grand challenge problem in {{ contact.custom1 }}? Why it is significant and to whom? Identify the stakeholder groups in the problem and their relationship to it. Briefly identify which aspects of the problem are technical, social, political, economic, ethical, or institutional.

In each generation of computing technology, computers struggle with scalability and complexity. For the latter, leveraging current technologies to address complexity is an opportunity and a challenge. Sustainable digital preservation requires a combination of strategies and we have not yet harnessed newer technologies (e.g., that enable machine learning and AI) and other technologies that could be out to work for preservation strategies.

This is significant to any organization (aka repository) that is responsible for the long-term preservation of complex digital content (e.g., multi-part objects, extensive metadata associated with large masses of content, extensive relationships that need to be linked and preserved to ensure preservation).

Technical: Technologists often preference emulation as a strategy, but there are limitations to emulation, especially in addressing complexity in preserving content (complex digital objects) for the long haul. Preservationists increasingly agree that emulation is more access strategy than preservation strategy and at best is a bridge approach that enables near to mid-term access.

Institutional: it is particularly challenging to devise and implement long-term strategies for complex digital content when the content curators with long-term care responsibilities are not able to work with creators or producers to effectively document content before essential information for understanding and using the content now and in the future is lost.

Economic: devising strategies like these requires support for practice-based research (or research-based practice) to support collaborative efforts across a range of domains. Information about the costs of implement these kinds of strategies is emerging – we need longitudinal data about our digital curation practices.

Q5 How will we know we've solved the problem? Criteria of success may include quantitative metrics such economic impact and number of adopters; and qualitative changes in individual or organizational behavior, professional practice, and changing incentives.

One indicator of success might be that we will have identified a research into practice program for continually tackling these kinds of complex digital preservation strategies. Success criteria: preservation strategies that are suited to complex digital content will be in use by a growing number of organizations, take less time to devise and implement, have an increasing number of researchers and practitioners who are able to work on these problems, and the cumulative effect will be feasible strategies because the costs will decrease to a manageable range as these strategies go into production.

One idea has been that the conceptual construct of a GIS system that enables users to navigate an accumulation of content and metadata with supporting services is an example of the kind of environment that might enable something like storytelling as a digital preservation strategy.

Q6 Who is in the best position to address those challenges? Identify domains, stakeholders, institutions, funding agencies, particular individuals, and other parties as appropriate. In particular, who is addressing this grand challenge most effectively? How have they made a positive difference? Based on their experience, which components of the problem appear the most tractable and which the least? What other aspects of the challenge are most urgent to address?

Storytelling as a digital preservation strategy emerged from the Digital Sustainability Lab at MIT Libraries, an example of the kind of team that could be a focus for this kind of work supplemented by researchers and practitioners who are creators, users, curators, and maintainers of complex digital content. This kind of research-based practice/practice-based research should be in scope for most funders who support or contribute to long-term digital collection management by individuals and organizations. An example of a project that is getting underway within the DSLab is storytelling as a digital preservation strategy with a focus on the Highrise project, a fascinating example of complex digital content. Preliminary discussions with the National Film Board of Canada, the Open Documentary Lab at MIT and others are helping to scope the problem and initial projects. Challenges include being able to generalize the approach by applying it to multiple use cases and getting the benefit of a GIS-like or similar environment without unintentionally complicating the preservation of the content. Normalization is a preservation strategy that appeals to preservationists (and has lots of untapped possibilities) because it reduces software dependency. How might the principles of normalization (in the digital preservation sense) be applied to storytelling as a digital preservation strategy?

Q7 If you like, please upload a document which answers the questions above or supplements your answers.

GrandChallenges-DCP_March2018.pdf (1.1MB)

Q8 If we quote from your response, should we use your name for attribution? **Yes**

#13

COMPLETE

First Name: Greta
Last Name: Suiter
Email: gsuiter@mit.edu
Custom Data: digital curation and preservation

Page 4: Candidate Grand Challenge Problems

Q4 What is the most significant grand challenge problem in {{ contact.custom1 }}? Why it is significant and to whom? Identify the stakeholder groups in the problem and their relationship to it. Briefly identify which aspects of the problem are technical, social, political, economic, ethical, or institutional.

We are trying to impose traditional notions of curation and preservation onto a digital landscape that is very different than the analog world especially when it comes to records and documents. In the physical world items (very broadly defined as anything created by man or machine) can easily be removed from the context of their creation making curation and preservation relatively easy. In the digital world items are intrinsically linked to the context of their creation (which often involves a third party application/software/hardware not owned by the creator) so they are more difficult to access, curate, and preserve.

Technical - how to remove items and put in new contexts for curation and preservation

Social - digital documents are much more collaborative in nature meaning more owners and more complications with rights

Political - many people don't understand their rights when it comes to digital assets / records

Economic - who is making money? who is doing curation and preservation better than libraries and non-profits?

Ethical - what is ownership in a digital world?

Institutional - there is a lack of control over digital records that is of great concern to institutions

Q5 How will we know we've solved the problem? Criteria of success may include quantitative metrics such economic impact and number of adopters; and qualitative changes in individual or organizational behavior, professional practice, and changing incentives.

When people have a neat, tidy, and comprehensive package of their life's work when they die. Something that is easily exportable and consists of all of their digital output from multiple sources and platforms.

Q6 Who is in the best position to address those challenges? Identify domains, stakeholders, institutions, funding agencies, particular individuals, and other parties as appropriate. In particular, who is addressing this grand challenge most effectively? How have they made a positive difference? Based on their experience, which components of the problem appear the most tractable and which the least? What other aspects of the challenge are most urgent to address?

Creators of applications/software/hardware are in a position to help with these challenges. A collaboration between technology creators and collectors (curators/preservationists/non-profits/educational institutions) could help to address the challenge but collectors need to shift their thinking from a world based on discreet items to one based on interconnected contexts. Funding from a third party would be needed.

MIT Libraries Grand Challenges Summit -- Problem Survey

Q7 If you like, please upload a document which answers the questions above or supplements your answers.

Respondent skipped this question

Q8 If we quote from your response, should we use your name for attribution?

Yes

#14

COMPLETE

First Name: Melissa
Last Name: Hagemann
Email: melissa.hagemann@opensocietyfoundations.org
Custom Data: open scholarship

Page 4: Candidate Grand Challenge Problems

Q4 What is the most significant grand challenge problem in {{ contact.custom1 }}? Why it is significant and to whom? Identify the stakeholder groups in the problem and their relationship to it. Briefly identify which aspects of the problem are technical, social, political, economic, ethical, or institutional.

Realigning incentives to promote open scholarship. If we want authors to share their research, then sharing has to count for promotion and tenure.

Stakeholder groups include researchers, promotion and tenure committees, university administration, and funders.

Solutions and potential activities:

1. Develop guidelines for reforming academic evaluations to reward authors who make their work open.
2. Develop quantitative and qualitative metrics to assess impact of shared works.
3. Seek community feedback on best ways to adapt to different disciplines or working environments (teaching vs. research units)
4. Build grassroots efforts, train local champions to encourage adoption of new incentive schemes in their institutions.

The main barriers are institutional and social.

Q5 How will we know we've solved the problem? Criteria of success may include quantitative metrics such economic impact and number of adopters; and qualitative changes in individual or organizational behavior, professional practice, and changing incentives.

What does success look like?

1-2 years: Experts convened, guidelines drafted, and community feedback received. By the end of period, guidelines have been developed.

3-5 years: Adoption and testing of new guidelines by at least five pilot institutions, hopefully including at one high profile university, such as MIT.

10 years: Adoption of guidelines (with some customization) in multiple institutions in three or more countries. Possible adoption of guidelines by at least one national-level research-level institution.

MIT Libraries Grand Challenges Summit -- Problem Survey

Q6 Who is in the best position to address those challenges? Identify domains, stakeholders, institutions, funding agencies, particular individuals, and other parties as appropriate. In particular, who is addressing this grand challenge most effectively? How have they made a positive difference? Based on their experience, which components of the problem appear the most tractable and which the least? What other aspects of the challenge are most urgent to address?

Potential partners include:

MIT
SPARC
FORCE11
EIFL
National Academy of Sciences
Gates Foundation
Wellcome Trust
Open Society Foundations

Q7 If you like, please upload a document which answers the questions above or supplements your answers.

Respondent skipped this question

Q8 If we quote from your response, should we use your name for attribution?

Yes

#15

COMPLETE

First Name: Phil
Last Name: Bourne
Email: peb6a@Virginia.EDU
Custom Data: scholarly discovery

Page 4: Candidate Grand Challenge Problems

Q4 What is the most significant grand challenge problem in {{ contact.custom1 }}? Why it is significant and to whom? Identify the stakeholder groups in the problem and their relationship to it. Briefly identify which aspects of the problem are technical, social, political, economic, ethical, or institutional.

The business model is wrong which leads to a lack of accessibility but also prevents sufficient experimentation in new modes of digital delivery and discovery that are accepted by the community.

Q5 How will we know we've solved the problem? Criteria of success may include quantitative metrics such economic impact and number of adopters; and qualitative changes in individual or organizational behavior, professional practice, and changing incentives.

Researchers are not insistent on publishing in one word journals and we can demonstrate faster dissemination and discovery.

Q6 Who is in the best position to address those challenges? Identify domains, stakeholders, institutions, funding agencies, particular individuals, and other parties as appropriate. In particular, who is addressing this grand challenge most effectively? How have they made a positive difference? Based on their experience, which components of the problem appear the most tractable and which the least? What other aspects of the challenge are most urgent to address?

No one is being truly effective. Funding agencies are too risk adverse to take bold steps and publishers are too profitable as the major levers to control the system. Scholars are not well organized or as yet motivated enough to make a difference. Libraries don't have enough support from their faculty and administration to make a difference.

Data are heading the same way as the knowledge derived from that data. This could still be addressed by a collective effort to define a sustainable business model before the private sector does it for us and drives the prices up.

Q7 If you like, please upload a document which answers the questions above or supplements your answers.

Respondent skipped this question

Q8 If we quote from your response, should we use your name for attribution?

Yes

#16

COMPLETE

First Name: Phillip
Last Name: Cohen
Email: pnc@umd.edu
Custom Data: open scholarship

Page 4: Candidate Grand Challenge Problems

Q4 What is the most significant grand challenge problem in {{ contact.custom1 }}? Why it is significant and to whom? Identify the stakeholder groups in the problem and their relationship to it. Briefly identify which aspects of the problem are technical, social, political, economic, ethical, or institutional.

Developing a culture of openness among scholars.

Open scholarship will only work when it is normative and expected in the research community. This is a social problem, but can only be overcome with interventions in the political, economic, technical, and institutional arenas.

Technical: We need systems to facilitate open scholarship, which will at once improve research, increase scholarly communication efficiency, enable new models of knowledge production, review, and dissemination.

Economic: We need to divert resources from publishing companies to university libraries and societies, so they can play their historical roles of facilitating the review, dissemination, and archiving of research and its products.

Political: We need universities and funding agencies to make open scholarship a policy priority, to increase the return on investment in research, and in turn to justify its expense to government and public actors.

Institutional: We need universities, libraries, and scholarly societies to implement standards of open scholarship, including reforming reward systems (tenure and promotion) to incentivize openness.

Ethical: I don't know about ethics!

Researchers ultimately will have to expect (for themselves), and demand (from others), open scholarship. But this is unlikely to happen in the absence of these political, economic, and institutional changes.

Q5 How will we know we've solved the problem? Criteria of success may include quantitative metrics such economic impact and number of adopters; and qualitative changes in individual or organizational behavior, professional practice, and changing incentives.

It's likely we will see gradual progress up to some tipping point, after which open scholarship will be normatively expected. I don't know what that point will be. Maybe there is something in the history of ArXiv and its core disciplines that could help provide that answer. When did it become automatic to use ArXiv in math and physics?

The indicators we need to track involve the prevalence of open access journals (and their relative status in the disciplines), preprint sharing and acceptance, reproducibility practices, open peer review, and promotion and tenure policies.

Q6 Who is in the best position to address those challenges? Identify domains, stakeholders, institutions, funding agencies, particular individuals, and other parties as appropriate. In particular, who is addressing this grand challenge most effectively? How have they made a positive difference? Based on their experience, which components of the problem appear the most tractable and which the least? What other aspects of the challenge are most urgent to address?

The transition to a culture of open scholarship is likely to be led -- or appear to be led -- by high-status actors and institutions, such as elite universities, big research libraries, national disciplinary societies, and high-status researchers. The standards and practices of these actors are the models for those below them in the academic status hierarchy.

A core movement among researchers, research libraries, and disciplinary societies may be the nucleus of this movement, which eventually pulls along universities more broadly. This will also involve non-profit organizations, and progressive foundations, especially for designing and building infrastructure and fielding innovative programs.

Q7 If you like, please upload a document which answers the questions above or supplements your answers.

Respondent skipped this question

Q8 If we quote from your response, should we use your name for attribution?

Yes

#17

COMPLETE

First Name: Douglas
Last Name: Armato
Email: armat001@umn.edu
Custom Data: digital curation and preservation

Page 4: Candidate Grand Challenge Problems

Q4 What is the most significant grand challenge problem in {{ contact.custom1 }}? Why it is significant and to whom? Identify the stakeholder groups in the problem and their relationship to it. Briefly identify which aspects of the problem are technical, social, political, economic, ethical, or institutional.

In the next phase of the transition to a digital ecosystem for scholarly communication the key challenge is achieving interoperability of scholarly works (monographic and journal) and the archival sources and collections they build upon. Along with this, the individual archives that scholars assemble in the course of their work should be preserved and made accessible to readers and researchers.

Pursuing this goal, requires a conceptual change from seeing scholarship as delimited works (e.g., containers) and instead viewing them as research projects, the outcome of which may include individual published works. To do so effectively would almost certainly necessitate a shift to Open Access publication but also a streamlining of the process for providing access to third-party materials, particularly those in archives. It would also suggest the movement toward a standardized database model of scholarly activity rather than the haphazard self-archiving practices currently in force.

Q5 How will we know we've solved the problem? Criteria of success may include quantitative metrics such economic impact and number of adopters; and qualitative changes in individual or organizational behavior, professional practice, and changing incentives.

It would best begin with agreement among stakeholders about the parameters of the challenge and the potential for unified action -- a good model for such an approach might be the recent TOME (Toward an Open Monograph Ecosystem) initiative on the part of the Association of American Universities (AAU), Association of Research Libraries (ARL), and Association of University Presses (AUPresses), which begins with fifteen pilot institutions. Later stages would include selection of development of data infrastructure and assessment of legal (intellectual property) frameworks.

Q6 Who is in the best position to address those challenges? Identify domains, stakeholders, institutions, funding agencies, particular individuals, and other parties as appropriate. In particular, who is addressing this grand challenge most effectively? How have they made a positive difference? Based on their experience, which components of the problem appear the most tractable and which the least? What other aspects of the challenge are most urgent to address?

The Andrew W. Mellon Foundation has been, of course, an active and informed participant in scholarly communication discussions from the outset of the digital era. Also, as in my response to question #4, the alliance of the Association of American Universities (AAU), Association of Research Libraries (ARL), and Association of University Presses (AUPresses) could form the core of any new initiative.

MIT Libraries Grand Challenges Summit -- Problem Survey

Q7 If you like, please upload a document which answers the questions above or supplements your answers.

Respondent skipped this question

Q8 If we quote from your response, should we use your name for attribution?

Yes

#18

COMPLETE

First Name: Alex
Last Name: Wade
Email: awade@chan Zuckerberg.com
Custom Data: open scholarship

Page 4: Candidate Grand Challenge Problems

Q4 What is the most significant grand challenge problem in {{ contact.custom1 }}? Why it is significant and to whom? Identify the stakeholder groups in the problem and their relationship to it. Briefly identify which aspects of the problem are technical, social, political, economic, ethical, or institutional.

The most significant problems faced by open scholarship center on how to create and sustain a trusted, distributed, and open knowledge infrastructure that supports the needs of the research community, reduces friction, and helps to advance the process and record of human knowledge. The most critical stakeholder group should be the researchers themselves, but secondary stakeholders are the general public, libraries, archives, publishers, and other service providers.

Q5 How will we know we've solved the problem? Criteria of success may include quantitative metrics such economic impact and number of adopters; and qualitative changes in individual or organizational behavior, professional practice, and changing incentives.

This is not a problem that will ever be "solved". But it will be a significant milestone when the infrastructure becomes transparent and when researcher workflows of pre-registration, data collection, protocols, methods, collaboration, publication, etc. are built/layered on top of a set of open and inter-operable services and systems. Where the researchers no longer need to consider where the data are stored or where to publish or where to archive, but that these operations simply 'happen' by way of the software and services that researchers use in their daily work.

Q6 Who is in the best position to address those challenges? Identify domains, stakeholders, institutions, funding agencies, particular individuals, and other parties as appropriate. In particular, who is addressing this grand challenge most effectively? How have they made a positive difference? Based on their experience, which components of the problem appear the most tractable and which the least? What other aspects of the challenge are most urgent to address?

Funding/sustaining of such an open infrastructure should be shared between research institutions (public and private) and the agencies that fund the research.

Q7 If you like, please upload a document which answers the questions above or supplements your answers.

Respondent skipped this question

Q8 If we quote from your response, should we use your name for attribution?

Yes

#19

COMPLETE

First Name: Trevor
Last Name: Owens
Custom Data: digital curation and preservation

Page 4: Candidate Grand Challenge Problems

Q4 What is the most significant grand challenge problem in {{ contact.custom1 }}? Why it is significant and to whom? Identify the stakeholder groups in the problem and their relationship to it. Briefly identify which aspects of the problem are technical, social, political, economic, ethical, or institutional.

A comprehensive solution to enduring access to software based works. I would put much of scientific research, architecture and design, electronic records, video games, and a wide range of files that require software to authentically use them in the category of content that requires this solution. The work that Yale is leading on around Emulation as a Service, and that the Software Preservation Network is engaged in, is very promising in these areas. However, the solutions in this space are going to require a sizable investment of all the institutions interested in long term access to this content and it's also going to require significant engagement with the software companies.

Q5 How will we know we've solved the problem? Criteria of success may include quantitative metrics such economic impact and number of adopters; and qualitative changes in individual or organizational behavior, professional practice, and changing incentives.

When software preservation/curation and use of emulation and virtualization are both common practices at a wide range of cultural heritage institutions and when users have come to expect these tools as common modes for interacting with historical content.

Q6 Who is in the best position to address those challenges? Identify domains, stakeholders, institutions, funding agencies, particular individuals, and other parties as appropriate. In particular, who is addressing this grand challenge most effectively? How have they made a positive difference? Based on their experience, which components of the problem appear the most tractable and which the least? What other aspects of the challenge are most urgent to address?

The ARL libraries, National Libraries, IMLS, Sloan, NEH, Software Preservation Network, Digital Legal teams at Harvard and Brandon Butler at UVA, Microsoft, Autodesk, other major software companies.

My sense is that significant movement has occurred in this space over the last decade, but that the next 5-7 years is going to be particularly critical period for some of those early explorations to converge into a national movement.

Q7 If you like, please upload a document which answers the questions above or supplements your answers. **Respondent skipped this question**

Q8 If we quote from your response, should we use your name for attribution? **Yes**

#20

COMPLETE

First Name: Bethany
Last Name: Nowviskie
Email: bnowviskie@clir.org
Custom Data: digital curation and preservation

Page 4: Candidate Grand Challenge Problems

Q4 What is the most significant grand challenge problem in {{ contact.custom1 }}? Why it is significant and to whom? Identify the stakeholder groups in the problem and their relationship to it. Briefly identify which aspects of the problem are technical, social, political, economic, ethical, or institutional.

Libraries and archives—along with the home institutions, educational/pipeline programs, professional and membership organizations, and commercial entities that support them—are not preparing adequately for the consequences of climate change. These include not only obvious and direct technical challenges to digital infrastructure and to care for physical collections that remain un- or inadequately digitized or which are fixed in media more subject to degradation in warmer or wetter conditions—but also more subtle and unpredictable social and political implications: constrained economies and altered/re-prioritized markets; increasing domestic strife and international hostilities stemming from mass human migration and competition for resources; changing professional demographics due to increasing social inequality and even factors like disease... [I got this far and felt too depressed to go on — but you get the idea: climate change.]

Q5 How will we know we've solved the problem? Criteria of success may include quantitative metrics such economic impact and number of adopters; and qualitative changes in individual or organizational behavior, professional practice, and changing incentives.

This will be more of a continual-mitigation situation than a one-and-done problem-solving exercise, but a first-order marker of success would be widespread consensus, fostered by highly public and straightforward communication in the cultural heritage sector, on a limited number of areas where concrete action can be taken, leading to measurable movement along axes identified in that process. A mix of qualitative and quantitative metrics would be involved: some would have to do with attitudinal and curricular changes and connection of the community with governmental and NGO-based planners working on predictive models to inform the work—and others with steps toward increased robustness of infrastructure, relocation of collections, engagement of funders, allocation of financial and personnel resources to the problem, and so forth. I realize that I am proposing a “Grand Challenges” summit on this one grand challenge, but that seems necessary to taking on a problem this wicked.

MIT Libraries Grand Challenges Summit -- Problem Survey

Q6 Who is in the best position to address those challenges? Identify domains, stakeholders, institutions, funding agencies, particular individuals, and other parties as appropriate. In particular, who is addressing this grand challenge most effectively? How have they made a positive difference? Based on their experience, which components of the problem appear the most tractable and which the least? What other aspects of the challenge are most urgent to address?

Project ARCC (“archivists responding to climate change”) started conversations along these lines (though not narrowly focused on digital curation/preservation), but—as a small, volunteer collective—seemed to run out of steam. There have been a few small and disjointed conferences and colloquia on the issue, like the recent “Libraries and Archives in the Anthropocene.” A major problem seems to be that those interested in the subject are either scholars and practitioners far removed from positions of power or preservationist leaders/organizations concentrating on the more dire and immediate consequences of climate change and climate-fueled conflict on physical collections and historical sites. Sometimes organizations feel the need to expend energy on much more basic issues, passing resolutions (as did ALA at its 2017 Midwinter meeting) affirming that climate change actually exists and that information about it shouldn’t be suppressed. (Witness also recent DataRefuge, and DLF’s Endangered Data Week efforts focused on political threats to environmental data.) All this is to say that I can’t identify groups or individuals who have yet made a positive and focused effort to address the consequences of climate change on digital preservation. However, it’s clear that this problem calls for involvement by iSchools, professional and membership organizations, governmental agencies, private funders, lobbying groups, cultural heritage institutions, commercial entities, and local and civic data organizations.

Q7 If you like, please upload a document which answers the questions above or supplements your answers.

Respondent skipped this question

Q8 If we quote from your response, should we use your name for attribution?

Yes

#21

COMPLETE

First Name: Mary
Last Name: Minow
Custom Data: scholarly discovery

Page 4: Candidate Grand Challenge Problems

Q4 What is the most significant grand challenge problem in {{ contact.custom1 }}? Why it is significant and to whom? Identify the stakeholder groups in the problem and their relationship to it. Briefly identify which aspects of the problem are technical, social, political, economic, ethical, or institutional.

The most significant challenge is the scattering of rich resources in silos that are not easily found. Institutional repositories, openlibrary.org with its deep backlist of monographs, etc do not readily surface when doing a federated search. Proprietary databases are generally inaccessible to the independent scholar as well. While google scholar can be astonishingly good, its proprietary algorithms make it hard to know what is not surfacing (see Graham/Sengupta piece).

Stakeholders: Scholars including independent scholars and curious users of all kinds. That's it. Once we start focusing on backroom folks, we lose sight of the only stakeholders who really matter - the users. We need to constantly be circling back to users with focus groups, surveys etc to see what users need and what works for users.

Q5 How will we know we've solved the problem? Criteria of success may include quantitative metrics such economic impact and number of adopters; and qualitative changes in individual or organizational behavior, professional practice, and changing incentives.

Metrics, surveys etc.

Federated search tools that are increasingly broad in reach, yet powerful in user-defined limits (e.g. English language, item type, .edu etc etc).

Mainly a qualitative arena, however. Give users with a range of research sophistication a decent question to research and see what tools and results they get. Success measured by deep and relevant results.

Incentives needed in noncommercial world to compete with Google (or is this a lost cause) - with a public interest approach. See Algorithms of Oppression by Safiya Umoja Noble.

MIT Libraries Grand Challenges Summit -- Problem Survey

Q6 Who is in the best position to address those challenges? Identify domains, stakeholders, institutions, funding agencies, particular individuals, and other parties as appropriate. In particular, who is addressing this grand challenge most effectively? How have they made a positive difference? Based on their experience, which components of the problem appear the most tractable and which the least? What other aspects of the challenge are most urgent to address?

Funders: IMLS, Ford, Knight, Mellon, Macarthur

Institutions: DPLA, Internet Archive, LOC, Hathi Trust - have made a difference in aggregating scholarly info and making available (e.g. LOC/AAPB partnership has created online reading room of public broadcasting)

Biggest problem is sustainability. For example, DPLA is still run nearly 100% on grant funding. The Internet Archive was a semi-finalist in the Macarthur 100 and change, but did not win and its openlibrary project did not get the injection of funding it needs.

If The Grand Challenge could come up with models of sustainability that would help all of the visionary projects that libraries have.

Q7 If you like, please upload a document which answers the questions above or supplements your answers.

Respondent skipped this question

Q8 If we quote from your response, should we use your name for attribution?

Yes

#22

COMPLETE

First Name: Clifford
Last Name: Lynch
Custom Data: open scholarship

Page 4: Candidate Grand Challenge Problems

Q4 What is the most significant grand challenge problem in {{ contact.custom1 }}? Why it is significant and to whom? Identify the stakeholder groups in the problem and their relationship to it. Briefly identify which aspects of the problem are technical, social, political, economic, ethical, or institutional.

I think that the system has become overwhelming for researchers: there is too much published, publishing has become very complex due to various mandates, restrictions, and choices; the system is much too slow (except when there is an adoption of preprint culture in a discipline), the peer review system is both disrespectful of people's time and too demanding of time. And we don't offer good tools to help researchers and students identify the information that they need to be aware of. The rise of quantitative assesment for researchers rather than genuine deep peer review is a disaster and creating many perverse incentives.

Q5 How will we know we've solved the problem? Criteria of success may include quantitative metrics such economic impact and number of adopters; and qualitative changes in individual or organizational behavior, professional practice, and changing incentives.

Progress will be reflected in improved researcher productivity and improved quality of scholarship.

Q6 Who is in the best position to address those challenges? Identify domains, stakeholders, institutions, funding agencies, particular individuals, and other parties as appropriate. In particular, who is addressing this grand challenge most effectively? How have they made a positive difference? Based on their experience, which components of the problem appear the most tractable and which the least? What other aspects of the challenge are most urgent to address?

There are some very hard problems and some low-hanging fruit (for example, preprint servers). It's going to take a determined effort by the entire community over a long period of time to really make the progress we need to see however.

Q7 If you like, please upload a document which answers the questions above or supplements your answers.

Respondent skipped this question

Q8 If we quote from your response, should we use your name for attribution?

Yes

#23

COMPLETE

First Name: Safiya
Last Name: Noble
Custom Data: scholarly discovery

Page 4: Candidate Grand Challenge Problems

Q4 What is the most significant grand challenge problem in {{ contact.custom1 }}? Why it is significant and to whom? Identify the stakeholder groups in the problem and their relationship to it. Briefly identify which aspects of the problem are technical, social, political, economic, ethical, or institutional.

The public is largely unable to access scholarly information, and thus scholarly discovery is not available to them. I think we should think more deeply about what it means to keep research in formats and repositories that may seem inaccessible to activists, organizers, teachers, parents, community members, civic leaders, and so forth.

Q5 How will we know we've solved the problem? Criteria of success may include quantitative metrics such economic impact and number of adopters; and qualitative changes in individual or organizational behavior, professional practice, and changing incentives.

We will have a public interest, viable alternative to Google or large commercial platforms that can serve the information needs more broadly than advertising platforms. We will know we success when studies show the public using evidence-based research, vetted journalism and news media, and so forth in substantial ways for non-trivial information needs.

Q6 Who is in the best position to address those challenges? Identify domains, stakeholders, institutions, funding agencies, particular individuals, and other parties as appropriate. In particular, who is addressing this grand challenge most effectively? How have they made a positive difference? Based on their experience, which components of the problem appear the most tractable and which the least? What other aspects of the challenge are most urgent to address?

I think key domains of expertise include academic and public libraries, the Library of Congress, the National Archives and other Federal information agencies. I would include all of the major professional organizations that support the information, museum and library fields, and I would include DPLA, DLF, CLIR and so forth. I believe the key stakeholders and funding agencies are Congress, IMLS, NSF, NIH, and NEA, as well as State-funded public research universities. I don't think there is a fully coordinated effort or campaign yet, but I think there is interest among individuals in some of these organizations. I also think there is a low level of planning and intervention among key stakeholders about their role in addressing public trust in information, and I think this is urgent.

Q7 If you like, please upload a document which answers the questions above or supplements your answers. **Respondent skipped this question**

Q8 If we quote from your response, should we use your name for attribution? **Yes**

#24

COMPLETE

First Name: Patricia
Last Name: Hswe
Custom Data: digital curation and preservation

Page 4: Candidate Grand Challenge Problems

Q4 What is the most significant grand challenge problem in {{ contact.custom1 }}? Why it is significant and to whom? Identify the stakeholder groups in the problem and their relationship to it. Briefly identify which aspects of the problem are technical, social, political, economic, ethical, or institutional.

People - the shortage of people with expertise in these areas, which is both a professional problem and a problem with degree programs in LIS (the scale at which education and training should be happening isn't there). There aren't enough people with vision at leadership levels in higher ed institutions (not only at the library but also at the levels of deans, provosts, VPs for research, and university/college presidents) who are aware of digital curation & preservation problems and taking them seriously enough to invest in addressing them substantially. In addition to leaders across a higher ed institution, other stakeholders would include MLS students and faculty; librarians; publishers; technologists; and users writ large.

I think from the above it's probably clear which aspects of the problem are social and institutional. There are obviously technical and economic challenges to overcome (infrastructure and resources will always be needed), and ethical concerns are also at play (are there cultural issues that might cast into question whether certain materials should be curated? what are the unrecognized costs, like perhaps invisible labor, that can figure into digital curation?).

Q5 How will we know we've solved the problem? Criteria of success may include quantitative metrics such economic impact and number of adopters; and qualitative changes in individual or organizational behavior, professional practice, and changing incentives.

Is digital curation/preservation solvable or manageable over time, with incremental successes?

We may know we've solved part of the problem by achieving the necessary scale of people and thus practice through investments in LIS degree programs and commitments to hire specialists for this work (which may depend on qualitative changes in organizational behavior). Because I've identified people as the most significant grand challenge problem, changing incentives would definitely help pave a path toward solutions, or, at least, better management of the problem.

MIT Libraries Grand Challenges Summit -- Problem Survey

Q6 Who is in the best position to address those challenges? Identify domains, stakeholders, institutions, funding agencies, particular individuals, and other parties as appropriate. In particular, who is addressing this grand challenge most effectively? How have they made a positive difference? Based on their experience, which components of the problem appear the most tractable and which the least? What other aspects of the challenge are most urgent to address?

Funding agencies (e.g., IMLS, Sloan, Mellon, NEH) are addressing this grand challenge. Both Sloan and Mellon have funded data curation and software curation postdocs - representing an investment in people. IMLS funded the "Preserving digital Objects With Restricted Resources," or POWRR, which is an outreach and engagement program for digital preservation that is conducted via workshops - NEH's support has also helped here. In addition, both IMLS and the Library of Congress have partnered on the National Digital Stewardship Residence program. There are more academic libraries with digital preservation librarians than probably ever before. There isn't any one entity addressing this grand challenge most effectively, but a group of them (as noted in the foregoing) that are making headway. It's difficult to say which components of the problem appear the most tractable and which the least; I can't know without consulting the people who've been in the roles I've mentioned here, and I'm not sure there's been a study or investigation on the digital curation / digital preservation workforce in recent enough years that would help answer this question.

Q7 If you like, please upload a document which answers the questions above or supplements your answers.

Respondent skipped this question

Q8 If we quote from your response, should we use your name for attribution?

No

#25

COMPLETE

First Name: Jacob
Last Name: Foster
Custom Data: scholarly discovery

Page 4: Candidate Grand Challenge Problems

Q4 What is the most significant grand challenge problem in {{ contact.custom1 }}? Why it is significant and to whom? Identify the stakeholder groups in the problem and their relationship to it. Briefly identify which aspects of the problem are technical, social, political, economic, ethical, or institutional.

Why do we care about scholarly discovery? Discovery plays a critical role at many stages of scholarly workflows: it allows scholars to map out established knowledge; to identify pertinent tools, concepts, and perspectives; to shore up or revise their arguments in response to peer review; and (perhaps most critically) to uncover gaps, opportunities, and unexpected but promising recombinations. I am sure other participants will highlight issues like data discoverability; ease of data reuse; machine reading; knowledge mapping, organization, and ranking; task-specific search; etc. I want to focus on an underlying, sociological grand challenge. This challenge orients not (just) to the discovery of existing knowledge, but to the processes of knowledge production that determine the contents and validity of existing knowledge, and shape the ends of future scholarly discovery.

That was the wind-up; here's the pitch: to understand the complex ecology of scientific institutions shaping knowledge production just enough so that we can change them--whether in response to problems (replication crisis) or opportunities (machine science)--without wrecking the ecosystem. I'm primarily worried about unintended negative consequences (which often couple strongly to technologies and practices of scholarly discovery). Here follow a past, present, and future example.

Past: Tools for scholarly discovery affect what knowledge gets used; many increase the importance of citations, which both quantify past impact and increase the probability of future impact. But this spills over into knowledge production, with pathological consequences for the quality and distribution of knowledge: provocative claims with flimsy foundations being an example of the former, and fad-chasing an example of the latter. If we understood the ways that discovery, use, and recognition affect the research choices scientists make (mediated by a range of institutions) we could design tools for scholarly discovery that would better support the goals of scholarly communities and those that fund or host them.

Present: Replication crises in select disciplines have led to a bit of a moral panic, leading to proposed changes in scientific practice ranging from pre-registration to changing significance thresholds. On the one hand, better understanding of the relevant institutions (from pedagogy to publication to reward) would inform transformations of practice that improve the integrity of future knowledge production without unduly affecting substance (e.g., making new scholarship too conservative). On the other hand, understanding the institutional pressures and processes that led to the replication crisis in *specific* disciplinary contexts would allow discovery tools to identify problematic literatures and improve the integrity of existing knowledge bases, mitigating the generalized suspicion of current scientific knowledge that the replication crisis has provoked.

Future: Rapid advances in machine reading, knowledge representation, and knowledge synthesis suggest that within ten years we will see significant portions of scientific workflows executed algorithmically, with little human supervision. In some cases (e.g., analysis and combination of existing secondary data; analysis of streaming data sources; some bench science) everything from formulating the question to conducting the analyses will be done by machine. On the one hand, this future raises problems of machine explanation and understanding that are rightly in the spotlight. But it also creates a "management problem," and hence a problem in institutional design: how do we human scholars set up systems of "reward" that distribute these computational (and even robotic) resources across the knowledge-landscape to generate new knowledge that addresses key problems (scientific, technical, and social) while also pushing back knowledge frontiers? This is the outer limit of scholarly discovery, but (rather amusingly) this sci-fi scenario best underscores the importance of scientific institutions to the mission of discovery.

As should be clear, this grand challenge cuts across stakeholder groups: scholars, data producers, data consumers, corporate aggregators of information with scholarly relevance (publishers, social media, etc.), technologists, funders (foundations and federal government), and citizens. While the fundamental problem dimensions are social and institutional, these are inextricably entwined with technical questions of system design; with economic questions (e.g., the larger political economy of science); and with political questions (as the long-term aims of scholarly discovery are at stake).

Q5 How will we know we've solved the problem? Criteria of success may include quantitative metrics such economic impact and number of adopters; and qualitative changes in individual or organizational behavior, professional practice, and changing incentives.

It will be hard to "fully" solve this problem; in some sense, any change in scholarly institutions, practices, or technologies for discovery will create new, emergent sub-problems to be understood. That said, I think the replication crisis and the "machine science" case glossed above provide two settings where we can actually evaluate whether we've understood the relevant institutions. For the "replication crisis" setting: Do we have reliable tools for assessing the quality of the literature in various sub-disciplines; identifying problematic cases; and rejiggering discovery so that questionable findings are returned with an appropriately quantified grain of salt? Have changing practices (e.g., in response to new policies implemented by journals, funders, universities, etc.) led to a decrease in irreproducible research, without an unexpected, undesired reduction in the scope or pace of discovery? For the "machine science" setting: Are systems for automated discovery delivering the distribution of surprise and robustness that we desire? Are these systems asking questions we would "want" them to ask? Are we able to incorporate these systems seamlessly into existing institutions and professional practices, or are they largely curiosity factories, churning out findings that are irrelevant to disciplinary scholarship?

More generally, we will know we have succeeded when new tools and practices for scholarly discovery can be introduced with few or no negative unintended consequences: No alarms and no surprises, please. Or at least, only good surprises!

Q6 Who is in the best position to address those challenges? Identify domains, stakeholders, institutions, funding agencies, particular individuals, and other parties as appropriate. In particular, who is addressing this grand challenge most effectively? How have they made a positive difference? Based on their experience, which components of the problem appear the most tractable and which the least? What other aspects of the challenge are most urgent to address?

Because of the breadth of the challenge, it requires contributions from many disciplines. Scholars of knowledge production are certainly essential: sociologists, historians, anthropologists, psychologists, and philosophers of science. Expertise in organizational theory and mechanism design (coming from economics) would also be critical. This scholarship must be in conversation with the design of technical infrastructures for scholarly discovery. Corporate holders of relevant scholarly information are key stakeholders; the possibility of true machine science (or at least machine discovery) makes the monopoly problems already rampant in this space absolutely essential to solve. As knowledge relevant to this grand challenge develops, funding agencies, journals, universities, and the stewards of discovery infrastructure (whether libraries or corporations) will need to step in with bold experiments in institutional change: experiments informed by, and designed to add to, emerging ideas about institutional design and its intersections with scholarly discovery.

To me, the best example of experimental, reflexive institutional change to shape scholarly discovery is the Grand Challenges initiative pursued by the Gates Foundation and a range of partners. These programs have successfully stimulated more innovative science on a range of socially-important questions. In general, funders seem to be making the most headway with disciplined institutional experimentation. Journals are getting into the game now, too, supporting deep scholarship on the peer-review process.

By contrast, most universities seem to be ill-equipped to respond to (let alone to drive) changes in scholarly institutions or practices -- despite very good intentions! We need only cite the continued obstacles to interdisciplinary or collaborative scholarship thrown up by tenure and promotion processes. Similarly, while the creators of discovery infrastructure (e.g., GoogleScholar) *could* be critical participants in this process, to the best of my knowledge they have been minimally coupled to questions of institutional change.

Q7 If you like, please upload a document which answers the questions above or supplements your answers.

Respondent skipped this question

Q8 If we quote from your response, should we use your name for attribution?

Yes

#26

COMPLETE

First Name: Micah
Last Name: Altman
Custom Data: any area

Page 4: Candidate Grand Challenge Problems

Q4 What is the most significant grand challenge problem in {{ contact.custom1 }}? Why it is significant and to whom? Identify the stakeholder groups in the problem and their relationship to it. Briefly identify which aspects of the problem are technical, social, political, economic, ethical, or institutional.

MIT Libraries Grand Challenges Summit -- Problem Survey

Developing Systematic Value Models for Selection at Scale

It is neither desirable nor feasible to keep all research information forever – thus selection and appraisal are critical part of data curation.¹⁰⁴ However, 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.¹⁰⁵ Although assignment of intellectual property rights can address this issue to a limited extent, it is very challenging¹⁰⁶ – and hence markets for information goods are generally thin. And 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.¹⁰⁷ 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 sub-optimal.¹⁰⁸ Moreover, the future value of research information and its communication potential are notoriously difficult.

The development of economic models, methods, and empirical analysis that would lead to more rigorous, reliable, and systematic evaluation of the value of research information constitutes an important, but poorly understood, set of problems. Researchers and curators continually make implicit or explicit decisions regarding 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.

Correctly estimating the future value of a single specific information object or collection is often impossible or impractical—similar to trying to guess the future stock price of a single corporation. Estimating the value of portfolios, however, is standard practice in finance, and could become standard practice in digital curation.

Two promising areas to explore in this pilot are economic portfolio theory and information science threat taxonomies. Historically, selection criteria have been made locally, and in an ad-hoc manner, based on the history and local values of the institution selecting. In economics research, generally contingent valuation surveys¹⁰⁹ are a standard tool for measuring the value of non-market goods – yet this method has never, to our knowledge, been applied to valuing research data. Similarly, portfolio selection modeling¹¹⁰ is the primary tool used in economics to diversify across risky investments, but has never been applied to research data. Diversification is also an essential strategy for mitigating risks to future access. There is a well-identified taxonomy of potential single-points-of-failure (highly correlated risks), that at minimum, a trustworthy preservation system should mitigate. These risks include media failure, hardware failure software failure, communication errors, network failure, media and hardware obsolescence, software obsolescence, operator error, natural disaster, external attack, internal attack, economic failure, and organizational failure. ¹¹¹ Nonetheless, the reliability, design, and behavior of both centralized and distributed preservation networks are just beginning to be understood. A notable exception is Baker, et al.¹¹² which employs Monte-Carlo simulation to explore trade-offs in costs and reliability across bit-level replication technology choices. Designing technical diversification strategies for long-term access requires more extensive modeling along these lines.

102 K. De Souza, *Realizing the Promise of Big Data: Implementing Big Data Projects* (2014), Report. IBM Center for the Business of Government.□

103 Rosenthal, David SH. "Bit preservation: a solved problem?." *International Journal of Digital Curation* 5.1 (2010): 134-148.

104 Giaretta, D. (2011), *Advanced Digital Preservation*, Springer.

□105 Arrow, Kenneth J. "The value of and demand for information." *Decision and organisation*. Londres: North-Holland (1972).

106 Gans, J. S., & Stern, S. (2010). Is there a market for ideas?. *Industrial and Corporate Change*, 19(3), 805-837.□

107 Altman, M. (2012). "Mitigating Threats To Data Quality Throughout the Curation Lifecycle. In G. Marciano, C. Lee, & H. Bowden (Eds.), *Curating For Quality* (pp. 1–119). Retrieved from <http://datacuration.web.unc.edu/>□

108 Hess, C. & Ostrom, E. (2006). *Understanding Knowledge as a Commons: From Theory to Practice*. The MIT Press.

109 Mitchell, R. C., & Carson, R. T. (1989). *Using surveys to value public goods: the contingent valuation method*. Rff Press.□

110 Markowitz, H.M. (March 1952). "Portfolio Selection". *The Journal of Finance* 7 (1): 77-91□

111 Reich, V., Rosenthal, D. S. H., Robertson, T., Lipkis, T., & Morabito, S. (2005). *Requirements for Digital Preservation Systems: A Bottom-Up Approach*. *D-Lib Magazine*, 11(11).□

112 Baker, M., et al. (2006) A fresh look at the reliability of long-term digital storage, *Proceedings of the 1st ACM SIGOPS/EuroSys European Conference on Computer Systems*. ACM; 2006.; p. 221-234.

Q5 How will we know we've solved the problem? Criteria of success may include quantitative metrics such economic impact and number of adopters; and qualitative changes in individual or organizational behavior, professional practice, and changing incentives.

Portfolio models could be tested via computer simulation; retrospective evaluation (model fitting) against data; willingness-to-pay surveys; and experiments. Evidence of success would include consensus theoretical frameworks; best practice models; practice adoption in memory institutions; adoption in other sectors.

Q6 Who is in the best position to address those challenges? Identify domains, stakeholders, institutions, funding agencies, particular individuals, and other parties as appropriate. In particular, who is addressing this grand challenge most effectively? How have they made a positive difference? Based on their experience, which components of the problem appear the most tractable and which the least? What other aspects of the challenge are most urgent to address?

This requires multi-disciplinary research in social science, computer science, and information science. Archives

Q7 If you like, please upload a document which answers the questions above or supplements your answers.

2015NationalAgenda.pdf (416KB)

Q8 If we quote from your response, should we use your name for attribution? **Yes**
