AMERICAN UNIVERSITY
WASHINGTON
COLLEGE OF LAW

creative commons

Workshop on Research and Resource Commons in Scientific Research: Final Report

American University, Washington College of Law
November 17-18, 2011

Michael Carroll, Professor of Law and Director, Program on Information Justice and Intellectual Property, Washington College of Law at American University, Washington, D.C.

In November of 2011, the Washington College of Law at American University convened and hosted a two-day workshop in collaboration with the Creative Commons to develop a strategy for promoting a commons or scientific research and related resources. The workshop brought together interested stakeholders from across the scientific research enterprise: scientists, administrators, librarians, publishers, societies, technologists, lawyers, policy makers, students, funders, and Open Science advocates, including both U.S. and international representatives. This diverse group discussed the current state of policy and technology as it relates to a scientific research commons, and identified key opportunities and challenges, as well as next steps, for the scientific community in general and Creative Commons in particular. These opportunities will inform the next phase of the Science program at Creative Commons and include legal and policy issues, education and technology efforts, and partnerships that will better leverage our efforts going forward.

NSF®

ALFRED P. SLOAN FOUNDATION

Table of Contents

Introduction ………………………………………………………………………. p. 2
Section I: Policy ……………………………………………………………………. p. 5
Section II: Infrastructure ……………………………………………………. p. 7
Section III: Organizations ……………………………………………………. p. 9
Conclusions ……………………………………………………………………….. p. 11
Acknowledgements and credits ………………………………………….. p. 14
Appendix I: Proposal ……………………………………………………………. p. 15
Appendix II: Agenda ……………………………………………………………. p. 18
Appendix III: Attendees ………………………………………………………. p. 19

Introduction

Rapid advances in information technology, and their uses in the inputs and outputs of scientific research, are far ahead of the legal and policy framework that supports science. A range of initiatives have emerged in the past decade to catch up, including changes in the grants policy at the National Institutes of Health to require public access through PubMed Central to peer-reviewed journal articles arising from NIH-supported research, university-based initiatives to improve open access to the scientific literature and to use institutional repositories as sites for data sharing, and the recent National Science Foundation requirements concerning grantees' data management plans. The time is ripe to review these and other initiatives to assess what lessons can be generalized. In particular, the rapid growth of digital scientific data, the complex status of these data under intellectual property law, and requirements that these data be managed responsibly, suggest that an open, commons-based approach could be particularly useful for addressing these phenomena.

A research or resource commons requires agreement among providers and participants about its legal structure, the technical requirements for its resources, and a shared understanding about how to sustain the commons. Legal issues, usually involving intellectual property or contract law, often arise as researchers, or research funders, seek to build commons or commons-based tools, such as Creative Commons[1] licenses. The objectives for the workshop were:

- To review lessons learned from those who have worked to build or to promote the use of commons structures to support scientific research from within the federal government and from the private sector, including the non-profit sector. This would include review of case studies from existing initiatives to provide open access to the scientific and scholarly literature, attempts to streamline and standardize the sharing of biological materials, and successful data-sharing projects, such as Sage Bionetworks and existing and proposed methods for sharing earth observation data.

- To identify the legal, technical, and cultural requirements for a successful commons, with a particular focus on scientific data. The key themes will be the respective roles of standardization and interoperability at the legal and technical levels necessary for resources to be shared in a commons, whether those resources are literature, data, physical inputs, or others.

[1] http://creativecommons.org/

- To discuss how the federal government, the university and non-profit sector, and industry can best work together to support existing successful resource commons in science and to create new commons or commons-based tools to improve the speed and efficiency of publicly funded scientific research. Attention will be given to how existing commons standards, such as legal and technical tools supplied by Creative Commons, are currently being used in the sciences and how these might be made more useful with respect to emergent forms of scientific communication.

A commons, in this context, is a standard set of rules by which people access the shared resources, including the infrastructure (standards, protocols, security methods, etc.) as well as the policies and terms of its use (e.g. methods of covering its costs). For the progress of science, we also promote commons that allows for maximal reusability and re-purposability of resources –i.e., the ability to combine large amounts of texts or data for new research purposes often unforeseen by the resource producers but having great potential benefit to science and the public.

The scientific research resources under consideration are primarily the research articles (analysis and conclusions drawn from research) and primary research data produced during the conduct of a research project. Also in scope for consideration are emerging forms of scholarly communication: websites, wikis, blogs, pre-prints, technical reports and white papers, databases, data visualizations, etc. All of these resources may be used in their entirety or in parts, e.g., the reference section of a research article, or a subset of a dataset, so different policies and infrastructure may pertain. Rearch outputs are normally the product of one or more researchers working for a grantee (university, research institution, government agency, etc.) often across international boundaries, so we are interested in a wide range of research materials, research stakeholders, and of international scope.

Historically, formal research publications are copyrighted to the publisher (society, non-profit or commercial), the author and their institution retain no rights, and funders don’t enforce what rights they could retain. This is changing in two ways. First, Open Access publications use different funding models and allow for free public access to articles, but often the publisher retains further rights (e.g. to “mine” the texts or build added-value commercial services from them). Second, a lot of scientific communication now occurs outside the formal research publications, on websites, wikis, blogs, and via informal publications. Many of these are unlicensed and the copyright owner is unclear, others are published under open licenses like CC-BY[2].

As background for the workshop, a survey of participants conducted before the event helped identify consensus around key questions:

1. Taking account of the ways in which scientific data vary in type and scale, what policies and practices would you say would constitute **successful data sharing** in science?

We need to clearly define what is meant by “sharing” (i.e. of what, in what manner and to whom) but in general this will require community buy-in as well as mandates (and enforcement of them), new financial and human resources, and recognition of effort (i.e. academic credit via citation, etc.). Successful sharing will enable reuse of data, implying that it is citable, documented and well-structured.

[2] http://creativecommons.org/licenses/by/3.0/

2. From your perspective, what are one or two of the biggest social, legal, or technical **barriers** to openly sharing scientific data?

The greatest perceived barriers were lack of incentives and demonstrations of the value and impact of open data sharing; concerns about data privacy, confidentiality or loss of personal use (i.e., competitive advantage); problems of data heterogeneity, quality and incompatible formats; and uncertainty about legal rights to reuse or re-share data; and the general lack of expertise or best practices to emulate.

3. In which disciplines can we find the biggest successes and failures to share scientific data (e.g. biology, economics, energy, astronomy) and what characteristics of the policy framework, institutional settings or disciplinary norms were most important to success or failure?

In general, scientific disciplines that have adopted a culture of sharing data are basic (as opposed to applied), have lower commercial potential, do not involve human subjects, and are not classified. Specific areas of success have been in genomics and other -omics, based on ethical and economic principles as well as strong incentives from funders, the geospatial and social sciences, and “big” science such as astronomy and high energy physics where there are large, shared infrastructure costs. Failures can be found in engineering and computer science, ecology and biodiversity, and chemistry (an example of an older discipline in which data was enclosed long ago).

4. What incentive structures most motivate scientists and/or their institutions to share, e.g., academic credit or easier access to data?

Recognition and credit by institutions and funders was uniformly seen as the key motivator for researchers. Additionally funder and publisher requirements or mandates are good motivators, as well as scientific community values and norms, including ethics and principles such as reproducibility. Finally, demonstrated value of and public demand for data sharing are good incentives.

5. Which resources for scientific research (e.g. research articles, research data, patents, materials, biobanks) are best managed as common resources, and how would you delineate who should have rights of access and reuse to these common resources?

Ideal candidates for a research commons are datasets that can be aggregated centrally to create a valuable resource (e.g. GenBank). Each discipline has to look at the costs versus benefits of sharing and it is not necessary to share everything that’s possible, but knowing the specific benefits is difficult and sometimes unexpected benefits appear and in unpredictable time frames. Additional pieces of the answer relate to where is it on the spectrum from “raw” or “processed and analyzed” (e.g., the more processed it is the more useful to others) and who its audience is beyond researchers, e.g. students or the public. Some resources with low research value may have high value as educational training data or to support public policy. Linking data to social benefit would improve public awareness of its value, and increase its likelihood of being shared.

In discussing the survey responses, it was clear that much work is still needed to document the cost/benefit ratio across different types of data in different disciplines. The case for open access to research data is already made in some areas (e.g. genomics) but still very difficult in others. Administrators of research institutions are worried about excessive costs, taxpayers want the benefits of research and access to its results but it is uncertain if they see this as justifying additional funding. Solutions to providing access to research resources must be efficient and affordable. Implementing stronger policy at funding agencies will require strong will and credible incentives for researchers and their institutions.

The workshop was organized are three themes: policy, technology infrastructure, and organizations. This report is similarly organized, with conclusions drawn from each area.

Section I: Policy

The policy goals of government research funders include the advancement of knowledge with high-quality scientific research (i.e. responsible and reproducible), and maximizing the impact of financial investments across the research enterprise, including researchers, grantees, and the public. Workshop participants started from the position that Open Science, including the existence of an openly accessible commons for research resources, is valued by many research stakeholders and should be supported by good policy and range of infrastructure and practices that support its growth and use over time.

Workshop participants reviewed and discussed current policies and activities across a range of U.S. federal agencies, international agencies and specific countries. This gave them a shared sense of current positions on providing access to research in the U.S. and elsewhere, and the types of activities that are influencing both policy and behavior across a range of stakeholders in the research enterprise. Providing a complete picture of current policies, attitudes and activities across such a complex domain is not possible in workshop format so the goal was instead to provide a "snapshot" of a moment in time. The group considered three white papers (provided as appendices to this report), presentations from two major U.S. science funding agencies (NSF and NIH) and various international efforts (CODATA, and initiatives in the EU and China), and had general discussions around new developments and activities in support of improved sharing of research results. These are summarized briefly in this report, along with initial findings and recommendations.

A white paper on U.S. federal policy with regard to research data, focused particularly on unclassified research arising from federal funding agencies (e.g. contracts, cooperative agreements, etc.) was presented and discussed. It examined both government-wide principles and detailed, specific policies for each agency or program that was discoverable by the public Web. The high-level conclusions were that:

- Congress has delegated data-related policy to the funding agencies (although the OMB has a few high-level rules like 'rights in data' including copyright, patent, and other rights;
- It appears that in cases where no rights accrue for research data (i.e. for factual data) there is no way for funding agencies to prevent grantees or others from adding contractual access restrictions to the data.
- The NIH has robust and aggressive rules (e.g. for all grants above $500,000) and the NSF is requiring a data management plan that includes policy disclosures;
- In general, individual programs vary from no policy to voluntary sharing to mandates and there is no consistency across agencies. Programs are serving as labs for policy experimentation within different disciplines.

Participants discussed the merits and challenges of this approach to research data policy, whether it was appropriate and at the right level. Many participants felt that data policy cannot be 'one-size-fits-all' in science, to allow for an evolving understanding of competing needs across the stakeholders. However it may be possible to create a 'default' policy that can be overridden in specific cases, to avoid the common situation of no policy decision due to conflicting interests. But any discussion of research access policies must address incentives for sharing, e.g., how data can be cited for credit in the scholarly record, and how metadata associated with data (citation or other types) will also be shared.

It was also clear that the challenges with data can be more complex than other types of research resources since there are often privacy, confidentiality or security concerns, and the balance of access and other concerns is delicate. Furthermore, both access and privacy requirements for data have deep technical implications for what researchers can actually do with the data once accessed. Policies that require data to be shared must also address *how* it is shared to achieve these goals.

Many federal funding agencies leave decisions about releasing data to the individual researchers and grantees, but there are a growing number of agencies that are beginning to set policy for how and when data should be shared, in what formats and under what terms and conditions. Across these agencies there exists a range of practice, from strict policy and mandates for open sharing of results within a defined time period to complete silence on the matter. This variability and inconsistency is itself a challenge, since researchers and grantees need to examine the policies and guidelines for each agency (and sometimes each program within an agency) to discover what it requires and how to comply. Universities are well-placed to help research agencies know what each other are doing, and to share the mechanisms and lessons learned from different approaches.

A noteworthy and troubling phenomenon is the common convention of releasing research data under a “Terms of Use” contract (often referred to as a “usage license” or “data user license agreement”). This practice has evolved in part because in the U.S. factual data is in the public domain so researchers and grantees need to impose a contract to control its use, e.g., to insure proper credit is given or to capture financial resources to support its ongoing availability. Because no standard “Terms of Use” contract exists, many data repositories use custom contracts that contradict each other, are difficult to enforce, do not work across international jurisdictions, overreach the original intent of the data producer, and have a chilling effect on researchers who might like to reuse or repurpose the data.

There was consensus that in the context of policy, we need clear definitions of “data” and levels of data (e.g., raw, processed, analyzed, or published). A shared understanding – perhaps a standard taxonomy – would facilitate discussions. Similarly we need to develop a shared understanding of “sharing” and “access” if we intend that a research resources commons will enable reuse of those resources in new contexts.

Most policy related to research resources, particularly data, is made at the private level: private contracts; community norms; licenses and waivers; or explicit agreements, e.g. with publishers. Public policies such as government funding agency mandates are difficult to set and the higher the level, across the spectrum from local to international, the harder the decisions are to make. Given this, when considering at what level a policy should be made it is worth considering how easy it would be to block it, and how close to the local level it could be made without leaving it to individual decision.

Some research data is particularly susceptible to forming a commons via policy, for example, clinical trial data that the U.S. FDA wants available but that data producers may resist sharing. A better course than starting with contentious areas is to choose a research subject that is already a commons, like Antarctica (for which there is already an international treaty to share data). Such a public commons would demonstrate both the value and the costs of aggregating and sharing data in the public domain.

As for software tools, the best investment will be in those that inspire researchers to share rather simply supporting policy mandates. Creating demonstration projects in a research area that includes multiple publically accessible datasets connected to a specific location seems promising, and could mix both professional and citizen science aspects. Such a demonstrator could then be generalized to other kinds of non-geographic "maps" (e.g., a genome or brain structure). Again, Antarctica seems to be a logical candidate for such a demonstrator.

Opportunity: to assist the government and its science agencies with establishing policy for sharing research results in all forms to the broadest possible audience, to gain the maximum benefit for science and the public. To further assist these agencies in identifying and establishing best practices for setting and implementing policy for a variety of types of research results (articles, data and gray literature, especially published online). These policies involve a wide range of stakeholders committed to open sharing of science (many represented at the workshop) and these consultative efforts will need a forum for discussion beyond Creative Commons.

Next Step: establish a science advisory group, convened and led by CC, to respond to requests for comment, information, and other opportunities to present the arguments for open sharing when possible.

Other areas that affect open science and have historically constrained the commons (e.g. patents or materials transfer agreements) are still of concern but less tractable to short-term action. The focus for the next five years should be on effect sharing of research outputs in the commons.

Section II: Infrastructure

The goals of technical infrastructure in support of open science include

- providing credit (and not just legally-required attribution)
- enabling large-scale interoperability of research data
- supporting scientific research reproducibility
- leveraging digital technology for linking research products (e.g. papers to data)

For certain shared infrastructure needs (e.g., unique identifiers for researchers, institutions, funding agencies, and data products) there is clear benefit in defining them as interdisciplinary and interagency standards. Other infrastructure will probably differ across disciplines but should be architected to support future interoperability and integration, to support interdisciplinary research and other innovation.

A high priority is to expand the scope of "data" in the international discussion of these issues to include related outputs such as software, provenance (or workflow) documentation, and metadata, to which a potentially different set of rights attach.

To support large-scale data discovery, use and integration will require automation and lower cost for infrastructure, which suggests using standards. The particular components of those standards, i.e., machine-processing, integration of heterogeneous data, globally unique identifiers, and inferencing capabilities (to extract information of the data structure) points to Semantic Web as an existing set of standards that meet these requirements.

Data requires documentation (e.g., metadata or paradata) to be discoverable, understandable, reusable and reproducible, and re-purposable for new research. To build trust that data is of good quality, some form of peer review or similar quality assurance process is necessary. The nuances of what level of data, from unprocessed to fully analyzed, also merits more discussion since the costs and benefits vary widely across that spectrum and vary by discipline, type of research and type of data. Given the high cost of good documentation for data, better standards are needed to leverage that investment.

For example, better mechanisms for machine-processability of data and metadata, for example, to achieve scientific goals such as credit, research reproducibility, or integration of multiple datasets. Look to semantic web standards for this, and possibly standardized ontologies to support ease of integration across disciplines. There has been growing recognition of and initiatives to address this need[3], but so far they have been at the discipline level, putting interdisciplinary research at risk and failing to achieve economies of scale across all of research.

Recognizing that putting research data in the public domain and making it openly accessible can be a very difficult proposal for many researchers, an interesting concept for data sharing in science is a "collaboration club" or "walled garden" in which members openly share data with each other but not outside the boundary of the local community. The locally-shared data has been sufficiently structured and documented to support its reuse, so it might be easier to broaden access after a defined time period, e.g. an embargo period, possibly automatically. Such an approach may hinder large-scale interoperability (as discussed previously) but offers a socially acceptable step towards sharing. Such an approach could also lend itself to standards and best practices for ad hoc collaborations, with sharing policies that can be quickly adopted without custom negotiation. Also built into this model is the concept of concentric circles of access (with necessary documentation) beginning with the data producer, broadening to close collaborators, a circle of peers, the larger discipline, the scientific community, and ultimately the general public. If a "commons kit" was developed to support this, and particularly if it included automatic decay of exclusivity, this would gradually lead to a research commons as a natural evolution.

Infrastructure and interoperability, like governance, require coordination and should address policy, stewardship, persistence, provenance, ownership, and licensing.

A key driver for a research commons is getting access to data, with the associated challenge being the axis of ease-of-production versus interoperability of that data. Discovering the existence of available research data is a challenge, as is tracking its ongoing use and impact. Mechanisms for citing and referencing data, learning its provenance, understanding its structure and semantics, and making use of it in new contexts are all known challenges for shared data. We lack tools to calculate the value of a dataset based on its quality (e.g. format, documentation and reusability). It would useful to embed better standards and policies into current tools that scientists know, rather than creating new tools and expecting scientists to change their procedures. On the other hand, reliance on popular infrastructure (e.g. Flickr for images) will be a big challenge if long-term preservation is a requirement for the data.

New tools seen as valuable for the commons include:

- A publishing platform to support "active publications" that change as the data evolves. This requires a modular object design that could link sections of research papers (e.g. the methodology and

[3] The "Investigation-Study-Assay" or "ISA Commons" http://www.isacommons.org/ is a prime example of a promising collaborative data integration framework within the bioscience community.

protocols) to related data; Additionally, tools that link datasets to research papers or other post-publication connections;
- Machine-readable data usage licenses;
- Open version of Google Scholar or Web of Science to advance credit mechanisms (COUNTER allows for metrics via publishers websites). Microsoft's Academic Search is an example of a large search engine that has an API to support the creation of new applications;
- Tools to visualize impact based on use of a particular dataset;
- ORCID for article claiming to link articles to authors to co-authors, to create citation graphs;
- Curation workflows tools;
- Registration for discovering active research projects via search systems;
- Metadata registries

APIs to data carry risk, since they inherently constrain what can be done with the data unless they just support bulk download. Also need to consider that the future may require the computing to come to the data rather than the other way around, given the size of many datasets and the problem of moving them across the network.
Demonstrator projects for large-scale data integration within particular disciplines, such as Antarctica, the atmosphere, geospatial data (climate), computer science data, possibly clinical trial data, and combined geographic visualizations (e.g., Cornell bird migration).

Opportunity: existing tools to manage data are not being taught or promoted (e.g., dropbox, if-this-then-that). Documenting existing tools that already support collaboration already (e.g. Microsoft list); educational campaign, e.g. an OER module for using the tools that goes into Connexions etc., drives demand for commons resources.

Section III: Cultural and Organizational Issues

The research enterprise involves multiple stakeholders whose interests must be balanced to achieve success and who respond differently to the prospect of a research commons. Stakeholders include individual researchers and students, their institutions (research universities and institutes), libraries and archives, funding agencies, scholarly societies and publishers, and the public as the ultimate beneficiaries of the research. The value of research resources differs across these stakeholders so each has different drivers and challenges for creating a research commons that must be simultaneously addressed.

Among the stakeholders' concerns about a research commons are credit, cost, quality, responsibility, accountability, mission-fit, and control. Understanding the costs and benefits to each stakeholder of contributing to and participating in the commons is an important prerequisite to understanding which actors will set the policies that will advance or preclude the commons. Ultimately, the policies and decision-making around the commons will be distributed so that no single stakeholder's interests dominate, so that the commons is less likely to rise or fall with the fate of that stakeholder over time.

A key component of a research resources commons is the infrastructure that supports it and which is provided by one or more of the organizational stakeholders. Large-scale infrastructure like a research commons must be a joint investment of several stakeholders, e.g., universities and funding agencies, as was the case with the Internet backbone. Also required of a successful commons infrastructure are standards: for data formats and transfer protocols, documentation and metadata, legal terms and conditions, software dependencies.

Discussion of science always comes back to the differences across disciplines. Furthermore, a robust and reliable digital research commons should be virtual/distributed, like the Internet, rather than centralized to mitigate the risk of sudden, total failure. Add these together and it might be more profitable to work towards creating a series of commonses across science rather than a single, unified commons for all science. A distributed, disciplinary-oriented approach may not be ideal for the needs of interdisciplinary research, but it would likely be easier to build, and if architected correctly could be federated in the future.

The group observed that there are natural roles for the various stakeholder organizations in creating a research commons that can be exploited:

- Funders can provide the "nudges" like the data management plan requirement, e.g., requesting dataset citations in addition to publications
- Societies (of some kinds) can play the gatekeeper role for research data quality and reproducibility
- Publishers can provide incentives to deposit data, software and related resources in trustworthy archives
- Universities can help provide the infrastructure and policies to support increased sharing
- Libraries need to help develop the collection policies for what should be archived, shared, how, and for how long (i.e., delineate the cost/benefit tradeoffs).

But top-down and government- or university-led policies won't succeed without simultaneously building understanding and support for the goals of open science and a commons of research resources. This will best be accomplished via outreach and education, especially of graduate students and young researchers who are already familiar with the affordances of the Web for many other activities.

Next steps: develop basic online educational modules on open science, open access, and open data, with associated policy and technology choices, for distribution to institutional, society, publisher, and other stakeholder partners. Consider collaboration platforms as a means of injecting sharing policies and data curation into the workflow early.

Opportunity: work with universities, research institutions, individual researchers and government agencies on specific aspects of data management plans that lack clear guidelines, such as the means for sharing data, under what license or waiver and in what time frame, to best serve the funding agencies' and the public's expectations.

Next Steps: identify institutions for partnership opportunities to work on integrating policy into data management plan processes and infrastructure[4].

Opportunity: establish a certification process (inspired by LEED-certification in the Architecture, Engineering and Construction industry) as an incentive for using good infrastructure and complying with best practices.

[4] The California Digital Library has a project underway to roll out the "DMP Tool" https://dmp.cdlib.org/ to help researchers create compliant plans for new grant proposals. Collaborating with this tool provider would allow a large number of researchers to immediately leverage open science tools like the CC licenses and waivers.

Next steps: review how LEED certification was introduced and structured, and consider stakeholders in science who could serve as a certification body.

Conclusions

There is increasing visibility of the Open Science movement world-wide[5] and that is an opportune time to accelerate progress towards a public commons of research resources. We offer the following observations and recommendations derived from the workshop:

- **High-level government policies are key, and good examples exist to develop best practices**
  - A very high leverage activity is connected to the National Science Board data policy recommendations, as these policies will inform policy at NSF and elsewhere in the federal government, and have knock-on effects in other countries;
  - Another high-level opportunity is working on Office of Science, Technology, and Policy's open policy recommendations, as this is affects multiple funding agencies;
  - Agencies that currently have policies related to openly sharing research results (primarily publications and data) include the NSF, NIH, National Statistics Bureau, Bureau of Justice Statistics, and the Census Bureau.
- **Policy for funder-mandated data management plans should require direction on open availability of results (publications and data) to the extent possible, clear terms of access and use, and timing of availability**.
  - Developing best practices should be a working document over time, incorporated experiences from the US, from Europe (Netherlands, Germany, etc.), from China, and elsewhere;
  - Implementing automatic "decay times" for intellectual property rights accruing to research results may address concerns of researchers and grantees.
- **In addition to government and funder policy, focusing on a middle ground or bottom up support is also critical and can be supported through education and outreach, especially to researchers and students**
  - Top-down policies alone won't work. As China's experience illustrates, top-down policies are good but getting researchers on board is equally necessary to insure success;
  - Education, outreach, and training are critical components to developing this bottom up support for the commons;
  - Simpler, shared licenses (or waivers) for data rights and clearer definitions around attribution can also support increased commons.
- **Improving incentives for data sharing, demonstrating its value and removing barriers are also very important**

[5] As evidence, during the preparation of this report there were several articles published in the popular press on the subject, including this article in the NY Times: http://www.nytimes.com/2012/01/17/science/open-science-challenges-journal-tradition-with-web-collaboration.html?_r=1

  - New incentives include providing better mechanisms for assigning credit (e.g., in tenure decisions), such as formal publication and attribution/citation tracking;
  - Providing funding for data sharing is a key to reducing barriers.
- **Data sharing policy should address multiple levels of data that are useful for a variety of audiences, including scientists but also students and the general public**
  - The lowest level of unprocessed data (Level 0) is not often immediately useful, but the next level (Level 1) of cleaned data is very useful for research. Making this data openly available involving costs for data center management;
  - Higher levels (Level 2-5) of data, incorporating more metadata, analysis, and tools, are even more useful to students and the broader public;
  - Data-focused efforts are generally distinct from publications, but the line is increasingly blurred so that policies need to address both;
  - While there is growing agreement that research articles should be openly shared, open access to 100% of data is unlikely for a variety of regulatory and practical reasons.
- **Usefully shared data requires specific information about methodology, semantics, and other appropriate metadata, in standardized terms.** Sharing metadata for research data supports its discovery, analysis and reuse, and standardization increases interdisciplinary research potential. As with the primary data, terms of access and use of metadata should be clear and open to the greatest extent possible.
- **Data management infrastructure will likely be characterized by a mix of centralized holding and distributed centers at universities**
  - Funding and incentive models to support decentralized storage should be pursued;
  - Creating certification for data management facilities (along the lines of LEED-certification for buildings) could be a practical incentive to create high-quality infrastructure, especially at universities.
- **Creating exemplars on specific scientific research topics that use a commons approach will demonstrate value,** e.g. Antarctica, (see discussion in Section II, under Policy)

Key Opportunities

Improved legal tools for data (licenses, waivers, contracts)

Shared licenses are a good direction, must be easy for researchers; look for agreement to simplify terms; data usage agreement that includes appropriate attribution; goal is simplicity for researchers to understand and use.

Policy advocacy

Collaborate with publishers and societies to develop joint data archiving policies[6], use of appropriate persistent identifiers and other metadata, and platforms and tools for research sharing.

[6] Joint Data Archiving Policy (JDAP) requires data deposit in an appropriate public archive. http://datadryad.org/jdap describes a requirement that supporting data be publicly available. This policy

Technology

Create demonstrators for scientific research commons in key subject areas, e.g. Antarctica, to include registries of relevant data, portals to aggregate data, tools to use data, and better User Interface tools.

Education, outreach, and training

Develop a cross-discplinary curriculum to express the vision and definition of a commons, its policy and regulatory landscape, normative practices in Open Science, and available legal tools and relevant technical tools. Target

- Individual instructors, e.g., to have students reproduce prior results as part of course work;
- Graduate Students, e.g., 3rd year students teaching 1st year students;
- Universities, libraries, and scholarly societies, e.g., for continuing education or professional training);
- Funders, e.g., to create young investigator awards for contributions to the commons;
- iSchools, to develop policy and commons-based modules for training data curators.

Creative Commons Activities

The scope of Creative Commons' particular efforts to create and support a research resource commons could include:

1. Continue work to document current Terms & Conditions contracts in use by the federal government research and funding agencies; identify common patterns and associated goals;
2. Identify societies and publishers to trial data archiving policies under open waivers or licenses, develop best practices for publisher/society policy on data sharing (following the example of JDAP);
3. Continued legal analysis of copyrightability of bibliographic data and the citing sentence (also fair use);
4. Defining best practices for standards (e.g. open, documented, IP-free)

Additional recommendations

- Creative Commons should limit its focus to digital research products (particularly publications and data) and not pursue open patents or materials transfer agreements, but should expand its 'data' focus to include software that supports data.
- It should also limit its involvement in developing technology standards, since those efforts can be confusing and contentious, drawing focus away from the legal and policy aspects of the commons.
- There is a multiplicity of technological tools that support the commons, but there is no clear role for Creative Commons in developing or promoting these.

was adopted in a joint and coordinated fashion by many top-tier journals in the field of evolution in 2011, and has since been adopted by additional journals across different disciplines.

Acknowledgements and credits

The **Workshop on Research and Resource Commons in Scientific Research** had the following major contributors:

Principal Investigator: Michael W. Carroll, Professor of Law and Director, Program on Information Justice and Intellectual Property, Washington College of Law at American University, Washington, D.C.

Co-Principal Investigators: Hal Abelson, Class of 1922 Professor of Computer Science and Engineering, Massachusetts Institute of Technology; John Wilbanks, Vice President for Science, Creative Commons, Inc

Co-Organizers: MacKenzie Smith, Research Director, MIT Libraries and Science Fellow, Creative Commons; Meredith Jacob, Pharmacy Fellow, Program on Information Justice and Intellectual Property, Washington College of Law at American University

Notetakers: Valerie Hovland, Redstone Strategy Group, LLC; David Kindler, DTKindler Communications

Funders: National Science Foundation; Alfred P. Sloan Foundation.

Appendix I: Proposal

Proposal for a Workshop on Research and Resource Commons in Scientific Research

Submitted by: American University Washington College of Law
Program on Information Justice and Intellectual Property

The Program on Information Justice and Intellectual Property at American University Washington College of Law proposes to hold a two-day workshop in Fall 2011 attended by approximately 15 - 20 experts to analyze and discuss opportunities for improved use of commons-based initiatives to improve resource and information sharing among scientific researchers.

I. Workshop Objectives

A research or resource commons requires agreement among providers or participants about its legal structure, the technical requirements for common resources, and a shared understanding about how to sustain the commons. Legal issues, usually involving intellectual property or contract law, often arise as researchers, or research funders, seek to build commons or commons-based tools, such as Creative Commons licenses.

The objectives for the workshop are:

- To review lessons learned from those who have worked to build or to promote the use of commons structures to support scientific research from within the federal government and from the private sector, including the non-profit sector. This would include review of case studies from existing initiatives to provide open access to the scientific and scholarly literature, attempts to streamline and standardize the sharing of biological materials, and successful data-sharing projects, such as Sage Bionetworks and existing and proposed methods for sharing earth observation data.

- To identify the legal, technical, and cultural requirements for a successful commons, with a particular focus on scientific data. The key themes will be the respective roles of standardization and interoperability at the legal and technical levels necessary for resources to be shared in a commons, whether those resources are literature, data, physical inputs, or others.

- To discuss how the federal government, the university and non-profit sector, and industry can best work together to support existing successful resource commons in science and to create new commons or commons-based tools to improve the speed and efficiency of publicly funded scientific research. Attention will be given to how existing commons standards, such as legal and technical tools supplied by Creative Commons, are currently being used in the sciences and how these might be made more useful with respect to emergent forms of scientific communication.

II. Statement of Need

Rapid advances in information technology, and their uses in the inputs and outputs of scientific research, are far ahead of the legal and policy framework that supports scientific research. A range of initiatives have emerged in the past decade to catch up, including changes in the grants policy at the National Institutes of Health to require public access through PubMed Central to peer reviewed journal articles arising from NIH-supported research, university-based initiatives to improve open access to the scientific literature and to use institutional repositories as sites for data sharing, and the new National Science Foundation requirements concerning grantees' data management plans. The time is ripe to review these and other initiatives to assess what lessons can be generalized. In particular, the rapid growth of digital scientific data, the complex status of these data under intellectual property law, and requirements that these data be managed responsibly, suggest that a commons-based approach could be particularly useful for addressing these phenomena.

III. Related Meetings

We are not aware of any other meetings that have specifically addressed this topic, although we are aware that the National Research Council's Board on Research Data and Information (on which Professor Carroll sits) intends to propose a workshop specifically on the issue of the intellectual property rights in scientific data. This meeting differs from that proposal because this meeting will focus on commons-based strategies for sharing a range of scientific resources, including data, and this meeting will address both legal and technical interoperability requirements for effective resource sharing.

IV. Organizing Committee

Chairperson: Michael W. Carroll, Professor of Law and Director, Program on Information Justice and Intellectual Property, American University

Committee Members: Hal Abelson, Class of 1922 Professor of Computer Science and Engineering, Massachusetts Institute of Technology; John Wilbanks, Vice President for Science, Creative Commons, Inc.

V. Workshop Details

The meeting will be held at American University Washington College of Law in late September or early October 2011. Experts will be invited based on their relevant expertise and experience. The meeting will be cross-cutting, and participants will be recruited from a wide range of scientific disciplines, in order to better identify the general legal and technical requirements for an effective research commons, while also identifying those issues, such as privacy, which may have more domain-specific or discipline-specific application. A substantial portion of the meeting will be open to the public, however there will be at least one closed working session.

The meeting will open with a review of lessons learned from existing and attempted commons-based initiatives for sharing scientific resources, including the work of Science Commons. Participants will discuss existing policy, legal and technical barriers to effective sharing of information and other resources among researchers and possible commons-based strategies that might be deployed or developed to overcome these.

The briefing materials will include case studies that will have been researched and documented prior to the meeting. At a minimum, these case studies and a report from the meeting will be published at least on the Internet through the Washington College of Law's Digital Commons portal, which will ensure that these materials will be properly preserved and will remain publicly accessible for the foreseeable future. We do not envision producing a print-based report.

The meeting will contribute to the enhancement and improvement of scientific and engineering research by identifying means to make more efficient and effective use of the outcomes of scientific research, particularly digital scientific data arising from NSF-funded research, through development and deployment of commons, whether centralized or distributed.

The organizing committee is committed to seeking out and inviting members of underrepresented groups in science and engineering as both speakers and attendees. The public sessions will be widely advertised through the Program on Information Justice and Intellectual Property's existing listservs (with about 3,000 members), through web-based advertising and announcements on relevant listservs, and throughout the university.

Appendix II: Agenda

Workshop on Research and Resource Commons in Scientific Research

Washington, D.C., November 17-18, 2011

Agenda

Thursday, November 17: Examining the Landscape

09:00 – 9:15 Welcome and introductions
09:15 -- 9:45 Attendee introductions
09:45 -- 10:45 White paper: Funding Agency Data Policies
10:45 – 11:00 Coffee break
11:00 – 12:00 U.S. government scientific data policy activities
12:00 – 13:00 Lunch break
13:00 – 14:00 International scientific data policy activities
14:00 – 15:00 White paper: Technical Interoperability as Implicit Policy
15:00 – 15:30 Coffee break
15:30 – 16:30 White paper: Lessons Learned from the Science Commons
16:30 – 17:15 Responses to survey questions
17:15 Adjourn
18:30 – 21:00 Workshop dinner

Friday, November 18: Opportunities and Challenges Ahead

09:00 – 9:30 Review of Thursday and setup for day two breakout groups (policy, technology, organizations)
09:30 – 10:30 Breakout group #1: What should drive the Research Resource Commons and what are its most pressing challenges?
10:30 – 11:00 Coffee break
11:00 – 12:00 Breakout group #2: What new tools are necessary to achieve a Research Resource Commons, e.g., standards, protocols, licenses, usage agreements, exemplar institutional policies, funder mandates?
12:00 – 13:00 Lunch break
13:00 – 13:45 Reports from breakout groups
13:45 – 14:30 Strategy session #1: Creating a 5-year plan. How can a Research Resource Commons be advanced over the next few years? What are the key principles for advancing this Commons?
14:30 -- 14:45 Coffee break
14:45 -- 15:30 Strategy session #2: Creating a map of the landscape. What is the best role for the non-profit and NGO sector, e.g., Creative Commons? Higher Education? Research funders? What other partnerships or collaborations are needed? What are you or your organization planning to do in this area next year?
15:30 Adjourn

Appendix III: List of Attendees

Name	Title	Organization
Josh Greenberg	Program Officer	Sloan Foundation
Chris Mentzel	Program Officer, Science Program	Moore Foundation
Sylvia Spengler	Program Director	National Science Foundation
Jorge Contreras	Professor	Washington College of Law
Michael R. Nelson	Research Associate	CSC Leading Edge Forum
Jerry Sheehan	Assistant Director for Policy Development	National Library of Medecine - National Institutes of Health
Paul Uhlir	Director, Board on Research Data and Information	National Academies of Science
Sayeed Choudhury	Associate Dean for Library Digital Programs & Hodson Director of the Digital Research and Curation Center	Johns Hopkins University
Clifford Lynch	Executive Director	Coalition for Networked Information
Peter Hirtle	Senior Policy Advisor and Fellow, Society of American Archivists	Cornell University Library
Christine Borgman	Professor & Presidential Chair in Information Studies	University of California, Los Angeles
Arfon Smith	Galaxy Zoo Technical Lead	University of Oxford
Hans Pfeiffenberger	Head, AWI-IT-infrastructure	Alfred Wegener Institute
Heather Piwowar	Postdoctoral Research Associate	NESCent
Debbie Crawford	Senior Vice Provost, Research	Drexel University
Liu Chuang	Director of Global Change Information and Research Center, Institute of Geography and Natural Resource	Chinese Academy of Sciences
Iain Hrynaszkiewicz	Associate Journal Publisher	BioMed Central
Elizabeth Nolan	Chief Publishing Officer	Optical Society of America
Dan Kulp	Editorial Director	American Physical Society
Mike Carroll	Professor & Director, Program on Information Justice and Intellectual Property	Washington College of Law
Hal Abelson	Professor	Massachusetts Institute of Technology
John Wilbanks	Senior Advisor	Creative Commons
Eric Saltzman	Creative Commons Board Member	Creative Commons
MacKenzie Smith	Science Fellow & Research Director, MIT Libraries	Creative Commons & Massachusetts Institute of Technology

Name	Title	Organization
Jonathan Rees	Principal Scientist	Creative Commons
Meredith Jacob	Assistant Director, Program on Information Justice and Intellectual Property	Washington College of Law
Cathy Casserly	CEO	Creative Commons
David Kindler	Communications Consultant	Creative Commons
Alan Ruttenberg	Principal Scientist	Creative Commons
Val Hovland	Consultant	Redstone Strategy Group, LLC