CHORUS – A solution for public access

Fred Dylla

Executive Director and CEO, American Institute of Physics, College Park, MD 20740, USA

The Clearinghouse for the Open Research of the United States (CHORUS) is a not-for-profit public-private partnership to provide public access to the peer-reviewed publications that report on federally funded research. CHORUS also provides enhanced search capabilities and links the public to articles directly on publisher platforms, where the articles can be read and preserved in their scholarly context. CrossRef provides the infrastructure for linking to the article or other document and collecting and retrieving the funding information, and CHORUS takes care of the rest. The CHORUS pilot program was launched in Fall 2013. It quickly garnered the support of the scholarly communications community, gaining endorsements from more than 100 publishers and organizations in less than six months.

This is a report on the presentation that I gave on CHORUS at the APE conference in January 2014. Since that time, the CHORUS pilot concluded and moved into full production on July 31, 2014. (See: http://chorusaccess.org/.)

Let me briefly reiterate the chronology of public access policy in the United States, which is expected to culminate as the year progresses in a federal open access policy for all large research funders. Next, I sketch three solutions for providing open access – as proposed by stakeholders. I then clarify the goals of CHORUS, the solution proposed by the publishing community. Subsequently, I turn in more detail to how CHORUS works. Finally, I look at the road ahead.

1. The chronology of public access policy in the US

The common ground among stakeholders is that scholarly publications are too important to allow a disruptive and unsustainable transition. At the same time, economic pressures are stressing all sectors of the community. Consequently, the US government has catalyzed multi-stakeholder discussion and collaboration to develop a solution.

Current public access policy in the United States is guided by the America COMPETES Act of 2010, which directed the Office of Science and Technology to coordinate study of the issues involved, culminating in the 2012 'Interagency Public Access Coordination' report. In 2013, the OSTP Public Access Memo 'Increasing Access to the Results of Federally Funded Scientific Research'[1] required agencies with $100M or more in annual R&D expenditures to develop public access plans. The memo stipulates that free public access to a full-text version of publications and data resulting from publicly funded research must be granted after a suitable embargo. Twelve months was provided as a guideline for this embargo. Beyond defining a national open access policy, the memo also described the means whereby funding agencies should pursue implementation:

---

[1]http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf.
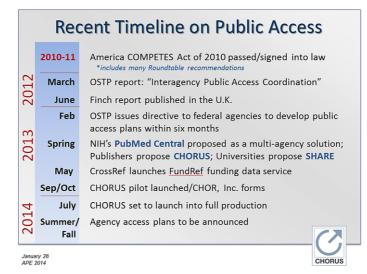
Fig. 1. Timeline on public access. (Colors are visible in the online version of the article; http://dx.doi.org/10.3233/ISU-140740.)

- Devise a strategy for leveraging existing resources while fostering public–private partnership;
- Optimize search and dissemination features to encourage innovation while also providing for long-term stewardship;
- Identify resources within existing agency budgets to complete implementation.

The following timeline (Fig. 1) summarizes development on public access and contextualizes the CHORUS solution.

2. Proposed solutions for implementing open access

One of the options proposed was extending the infrastructure of the National Institutes of Health's PubMed Central (PMC) repository to other disciplines. NIH, with an annual budget exceeding \$30 billion, has been a leader in open access policy, having mandated its public access policy in 2007. Under this paradigm, other research funding agencies would be invited to invest in PMC's expansion to satisfy their public access requirements. NIH or the funding agencies would both own and manage the repository system.

The Association of American Universities (AAU), the Association of Public and Land-Grant Universities (APLU) and the Association of Research Libraries (ARL) proposed a second option – SHARE (SHared Access to Research Ecosystem), scaling up the existing institutional repository infrastructure to host all articles based on federally funded research. As the principal investigators of federal grants and contracts are typically located at universities, the university would handle the article ingest process. It would be necessary to integrate diverse repositories into a single, searchable system.

Publishers, however, have the majority of the infrastructure already in place to deliver public access successfully without much additional investment, and with an advantage in their ability to ensure compliance given the close communications between authors and their publishers. In 2012, publishers started working with several funding agencies and CrossRef to craft a solution that identifies articles resulting from agency funding. Through CrossRef's FundRef funding data service, publishers collect or assign standardized funding information from a taxonomy of funders and add it to metadata they

already provide to CrossRef. This process, combined with the search and discovery protocols that have been developed over decades by commercial search engines, the publishing industry and others, allows articles to be identified and accessed through CrossRef's Digital Object Identifier (DOI) linking and other infrastructure services.

From this notion evolved the Clearinghouse for the Open Research of the United States (CHORUS) as an approach to address the OSTP requirements in partnership with federal agencies. A not-for-profit entity – CHOR Inc. – was incorporated on 1 October 2013, with startup funding from the Professional/Scholarly Publishing Division of the Association of American Publishers (PSP-AAP); it quickly garnered support from more than 100 signatory publishers and organizations, as well as significant sponsorship pledges.

As a not-for-profit and collaborative venture with authors and funding agencies, CHORUS is meant to steer clear of politics as it provides essential services in line with federal open access policies.

3. The goals of CHORUS

The main goals of CHORUS are:

1. Provide public access
2. Leverage existing infrastructure
3. Ensure compliance and archival presentation
4. Preserve agency funds
5. Provide for international scalability.

These goals are in line with the OSTP Public Access Memo, providing free public access to the article metadata immediately and to a full-text version after an appropriate embargo period, which in some cases will be zero. Moreover, it addresses the objectives of its key stakeholders – funding agencies, authors, researchers, institutions, librarians, and the public – because in addition to easy access, CHORUS maximizes chances of compliance, minimizes version control issues and provides for long-term preservation. Compliance rates for existing governmental open access policies have been improving, but require significant effort to achieve sub-optimal results: even long-standing policies only get 2/3 to 3/4 of the articles they expect. Building in identification of research funder(s) at the point of article submission simplifies the process. If publishers steward ingestion, this ensures not only that the correct version is used (i.e. publishers' version of record or author's final peer-reviewed draft submitted for publication), but also that potential subsequent addenda and errata can be connected to the article posted on the publisher's site. In addition, through CHORUS' integration with the publishers' preservation strategies (e.g. PORTICO, CLOCKSS), long-term access is guaranteed. Combined, this will create universal compliance for CHORUS-participating journals.

As an essential service provider for CHORUS, CrossRef, through its DOI infrastructure, directs readers to the identified article, where access to a freely available version is provided on the publisher's platform. The combination of CrossRef's and the member publishers' infrastructure provides a significant fraction of what is required to enable CHORUS; ongoing investments by publishers will be minimal. CrossRef also serves as an excellent role model for CHORUS for handling strategic partnerships and stakeholder interactions.

4. How CHORUS works

CHORUS manages identification and ingest. The identification of publications subject to the public access policy is built into the author's submission process to CHORUS member publishers' journals. The author inputs the funder name(s) from a publicly available taxonomy (CrossRef's FundRef Registry) from the manuscript submission software screen. Hitherto funding agencies could not easily identify articles, because funding information was not part of the standard metadata collected by publishers or other stakeholders, and the nomenclature used by authors was inconsistent. The following chart (Fig. 2) summarizes the development and deployment of CHORUS.

The FundRef registry originated from the SciVal database, which Elsevier donated to CrossRef for the project, and included more than 2000 agency names; it now exceeds 7000. FundRef was piloted successfully with the Department of Energy, National Science Foundation, NASA, and the Wellcome Trust. The FundRef taxonomy is now being implemented in all major editorial software products used by publishing companies. CrossRef also has supports publishers using their own editorial software. In sum, collecting, storing, and retrieving funding information is becoming as universal as collecting other metadata.

CHORUS manages free access and the embargo period. On the Internet, anyone using a CHORUS-powered search portal will be able to find any public access article and its digital object identifier. The user selects the article they want to view, and is directed to the article on the publisher's site. The publisher will provide free access to the article – either upon publication if the article is open access or after a determined embargo period. With articles released after an embargo period, publishers may decide to direct users to the version of record or the author's final accepted manuscript. Regardless, it will be presented in context so readers can find related content. CHORUS provides the infrastructure to enable these access functions.

CHORUS supports discovery by any existing search engine. CrossRef enhances discoverability of funding data through the many third parties that receive feeds from CrossRef Metadata Services to incorporate into their own discovery tools. Funding data is also available through CrossRef's search ser-
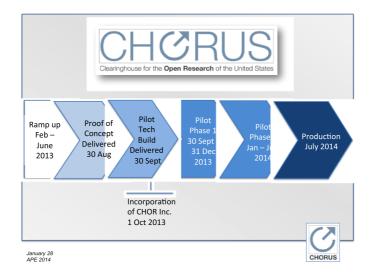

Fig. 2. CHORUS build up. (Colors are visible in the online version of the article; http://dx.doi.org/10.3233/ISU-140740.)

vices, such as FundRef Search, web query systems, CrossRef Metadata Search and the popular CrossRef application programming interfaces (APIs).

CHORUS has used a CrossRef API to build its live system at search.chorusaccess.org. More than ten publishers participated in its pilot, inputting articles funded by agencies. The service demonstrates how easy compliance, access and discovery are. CHORUS plans to extend the CrossRef APIs to improve its functionality and usefulness.

CHORUS will enable the monitoring of compliance with funder policies. It provides dashboard services to funders, and reports are available to all participating stakeholders. APIs enable funders and libraries to interact with CHORUS services. For example, the Department of Energy, the largest funder of physical sciences in the United States, has developed the portal PAGES to display DoE-funded articles. PAGES builds seamlessly on the services built by CHORUS, and also allows the DoE, at a glance, to understand where DoE-funded researchers publish.

CHORUS supports long-term access and preservation by making use of existing multiparty preservation strategies such as PORTICO and CLOCKSS. Also envisioned is the option for use of a government or third-party maintained archive.

5. Outlook: The road ahead

CHORUS aspires to be inclusive and to interoperate with scholarly repositories and other systems providing access to scholarly articles such as SHARE and PMC, for example through a joint initiative with SHARE to work on persistent identifiers and metrics. Since the time that this talk was given, CHORUS and SHARE have further agreed to work together on a research event notification system.

The infrastructure built by CHORUS could also be used to tackle the other major public access issue: the data resulting from publicly funded research. There is great value in initially solving the problem of dealing with the data that is linked to the articles that describe the relevant research. This would simplify procedures for research compliance regarding data deposit and the monitoring by funding agencies. CHORUS can easily link to data repositories when available.

While several approaches have been proposed for the infrastructure for public access, it is important to remember the common ground. On the whole, the important role of government as essential partner in the research enterprise should continue. The US Government has catalyzed solutions of public access and multi-stakeholder implementation. During this period and continuing to the near term, economic pressures are stressing all sectors of the research community, thus encouraging all stakeholders to identify and implement pragmatic and mutually beneficial solutions. For all, scholarly publications are too important to allow a disruptive transition that undermines the quality, stability, or sustainability of the scientific record.