CHORUS – providing a scalable solution for public access to scholarly research

CHORUS (Clearinghouse for the Open Research of the United States) offers an open technology platform in response to the public access requirements of US federal funding agencies, researchers, institutions and the public. It is focused on five principal sets of functions: *identification*, *preservation*, *discovery*, *access*, and *compliance*. CHORUS facilitates public access to peer-reviewed publications, after a determined embargo period (where applicable), for each discipline and agency. By leveraging existing tools such as CrossRef, FundRef and ORCID, CHORUS allows a greater proportion of funding to remain focused on research. CHORUS identifies articles that report on federally funded research and enables a reader to access the 'best available version' free of charge, via the publisher. It is a scalable solution that offers maximum efficiency for all parties by automating as much of the process as is possible. CHORUS launched in pilot phase in September 2013, and the production phase will begin in early 2014.

The road to CHORUS

Over the last five years, the United States government has evolved new approaches for improving public access to the results of publicly funded research[1]. In 2009, the White House sponsored the Scholarly Publishing Roundtable, drawing together US agencies, institutions and publishers into productive discussions; this helped to inform the America COMPETES Act of 2010, which aims to drive and maximize investment in innovation and research in the US.[2] In March 2012, the US White House Office for Science and Technology Policy (OSTP) issued 'Interagency Public Access Coordination', which reported on 'progress toward the coordination of policies' related to the 'dissemination and long-term stewardship of the results of Federally funded scientific research' required by America COMPETES[3] In response to these events, CrossRef[4], publishers, and funding agencies joined together to start the FundRef[5] pilot program. FundRef, officially launched in 2013, offers standard methods for identifying the funding sources of published articles, allowing funding agencies to track the results of their efforts.

The OSTP then proceeded to issue a memorandum in February 2013, stating that each federal agency with over $100m of expenditure on research and development had to develop a plan to support public access to this research.[6] This memo gave the agencies six months to create these plans to provide access to 'results published in peer-reviewed scholarly publications'. Particular emphasis was placed on the establishment of public-private partnerships to facilitate access and on leveraging existing infrastructure, rather than creating new, expensive ventures.

In response to the OSTP memo, it was proposed that the National Institutes of Health's PubMed Central serve as a universal solution for all funding agencies[7], while a coalition of publishers proposed CHORUS[8] and a group of university library organizations proposed SHARE[9]. While most agencies submitted their draft plans to OSTP by the end of August 2013 as required, these draft plans have not, at the time of writing this article, been made public.

How CHORUS fits the bill

As the first service of the non-profit CHOR, Inc. organization, CHORUS (Clearinghouse for the Open Research of the United States) represents the publishing industry's effort to create the public-private partnership with funding agencies requested by the OSTP. CHORUS

HOWARD RATNER
Executive Director
CHORUS

H FREDERICK DYLLA
Executive Director
and CEO
American Institute of
Physics

DAVID CROTTY
Senior Editor
Oxford University Press

CHORUS

leverages existing tools such as CrossRef, FundRef, and ORCID[10] to facilitate public access to peer-reviewed publications resulting from public funding. Through the reuse of existing infrastructure, CHORUS expects to offer a cost-effective solution, allowing research funds to remain where they are most needed – in the hands of researchers, funding research. There is no significant cost for agency use or participation in CHORUS.

By focusing on open standards and an open architecture, CHORUS is a scalable solution that offers maximum efficiency for all parties, automating as much of the process as is possible. This saves researcher time and effort and minimizes the costs to authors, their research institutions, funding agencies and publishers. The open standards create new opportunities for the creation of new discovery tools and economic development.

> "…open standards create new opportunities …"

Using FundRef protocols, CHORUS identifies those articles that are reporting on federally funded research and enables the reader to access the 'best available version' (BAV) free of charge, via the publisher's website. The BAV could be either the accepted manuscript (AM) or the version of record (VoR, the formally published version of the article[11]). Participating publishers will make either or both versions publicly accessible on their websites. CHORUS launched in pilot phase in September 2013; the production phase began in early 2014. The pilot version of CHORUS' initial services can be seen here: **http://chorusaccess.org/pilotservices/** .

Who are the stakeholders?

CHORUS has identified many key stakeholders and has been working closely with each group to best meet their needs:

- *funding agents* want to meet the OSTP guidelines, monitor grantee and agency compliance with OSTP requirements, measure the agency's impact (return on investment), and provide the widest possible access to articles reporting on the research funded by the agency

- *researchers* want to comply with their funding agency's requirements with minimal effort (in part to ensure they obtain funding for future research), know the sources of funding in their area of research, and have access to the best available version of content in their research area

- *librarians* want to have access to the best available version of content for their patrons, conduct text and data mining (TDM), have confidence that these articles will be readily available in perpetuity, help researchers comply with funding agency requirements, and build discovery tools for researchers

- *the public* wants to have access to the best available version of content (for example, to research a problem or drive economic development), see what the government is funding, learn the impact of specific agency grants, understand the latest developments in science, and have content connected to learning tools

- *publishers* want to help their authors and institutions comply with funder mandates, and retain traffic on journal websites to better demonstrate value to their customers.

> "The goal of CHORUS is to create a framework using already existing infrastructure …"

The goal of CHORUS is to create a framework using already existing infrastructure to see that these needs are met.

How does CHORUS work?

CHORUS enables: the *identification* of the appropriate article reporting on funded research; public *access* to the article; the *discovery* of the article on the publisher's site and *preservation* of the article in a long-term archive. CHORUS tracks each of these features and reports the level of *compliance* on its dashboards.

Identification

When a researcher submits a paper to a scholarly journal through a typical electronic submission system, they will receive a pull-down menu to *identify* the funding agencies that supported their research and an opportunity to fill in specific grant IDs (see Figure 1). That is all the effort required by the researcher; they have done their part and are now fully compliant with funder requirements. The importance of this cannot be stressed enough; other proposed systems require researchers to spend valuable time sorting through agency regulations, determining which version of a paper they need to submit, determining the proper embargo period and going through an archive submission process. Under CHORUS, this is all done automatically on the researcher's behalf, reducing time and effort and greatly improving *compliance* levels.

> "... all done automatically on the researcher's behalf ..."

The paper then goes through the publisher's regular peer-review process and, if accepted, the production cycle proceeds and the paper is published. The paper's digital object identifier (DOI) and associated metadata – including the grant information – is then registered at CrossRef, which in turn feeds a variety of metadata services including FundRef and CHORUS.
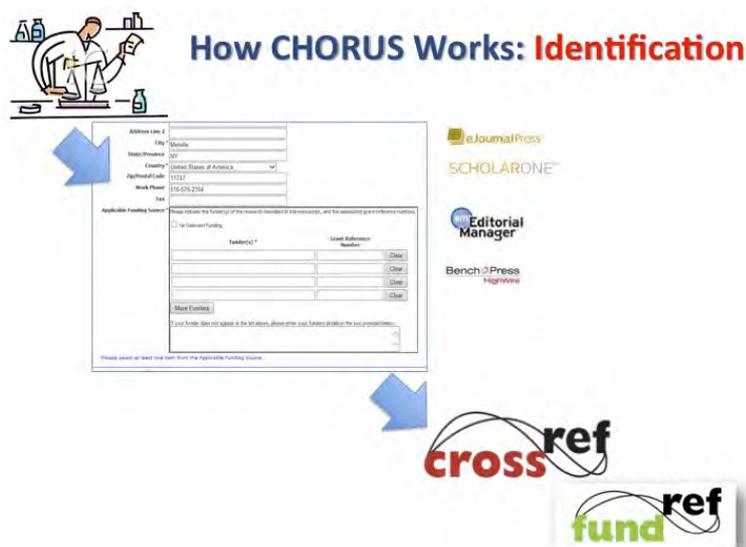

Figure 1. CHORUS identification

Access

Articles are then made publicly *accessible* by the publisher's host system, either after the funding agency embargo period expires or immediately, if the author has paid an article processing charge (APC) (see Figure 2). Article reuse terms are posted by the CHORUS-compliant publisher on their journal hosting websites and made transparent by the CHORUS services and application programming interface (API). Users will get access to the BAV – either the AM or the VoR. Access is always through the publisher's website for the publicly accessible version of the article. This ensures the reader is seeing a highly *discoverable* version of the paper in the context of the journal, where notifications of updates, corrections, or retractions will be available.

Figure 2. CHORUS access

Discovery

These publicly available articles will be indexed like any other journal article and can be *discovered* in a user's favorite search engine such as PubMed or Google. CHORUS provides an open API for these discovery tools to allow them to enhance their search results with information on the funding source and availability of the articles. (See Figure 3.) The open API also enables funding agencies to create their own discovery tools and build institutional portals if wished. New text- and data-mining services are also enabled via the API.

Figure 3. CHORUS discovery

Preservation

When a paper that has been publicly funded is published, the CHORUS member publisher also deposits a copy into one or more of what are called 'dark archives' for *preservation*. These dark archives include Portico, CLOCKSS and any other archive funding agencies require. The paper is permanently archived in these repositories, but not made available unless certain compliance trigger events occur.

Compliance

Trust is an important part of the CHORUS approach, so a great deal of work has been put into ensuring *compliance*. As noted above, CHORUS is built to drive maximum compliance with minimum effort. The CHORUS dashboard service (see Figure 4) allows stakeholders to constantly monitor that compliance, and provides automated mechanisms to make sure what has been promised is really happening. Currently, the dashboard keeps track of the number of articles identified for a funding agency, the number of articles preserved via archiving, the number of articles that are publicly accessible and the number of articles with agency accepted reuse licenses.

"… CHORUS is built to drive maximum compliance with minimum effort."

CHORUS checks to see whether papers that are supposed to be publicly available are indeed publicly available. If they are not, a trigger event occurs, and a notification is sent to the agency and the publisher. If the problem is not promptly fixed, the dark archived version of the paper is brought to light, and is substituted into the system in place of the publisher's version until public access is restored.

Over time, the dashboard will evolve and can certainly be customized for specific needs and reporting requirements. The open dashboard data can also be pulled into an agency's, institution's, or publisher's own system via the CHORUS API.
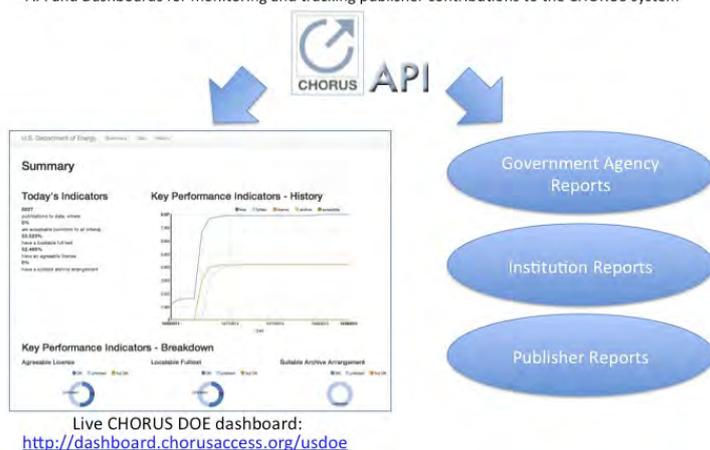
Figure 4. CHORUS compliance

Proof of concept

CHORUS launched its pilot system on 30 September 2013 with seven publishers (American Chemical Society, American Physical Society, AIP Publishing, Elsevier, IEEE, John Wiley & Sons and Oxford University Press) and the US Department of Energy (DOE). Pilot publishers submitted tens of thousands of articles reporting on research funded by the top-level US agencies. More than 4,000 records were rapidly harvested via the CHORUS system into the proposed DOE PAGES agency portal. This DOE discovery tool offers search results that point back to individual publisher websites via the DOIs of the papers reporting on funded research. This provides clear evidence that CHORUS is a workable solution for funding agencies.[12]

What about public access to scholarly data?

CHORUS is designed to address the first objective in the OSTP Public Access memorandum: 'Public Access to Scientific Publications' (the publications objective). However, CHOR, Inc. is often asked whether CHORUS can also play a role in the memo's second objective, 'Public Access to Scientific Data in Digital Formats' (the data objective). The publications and data objectives are listed separately in the OSTP memo, each with its own specific set of guidelines and goals. It is unclear whether there is one unified mechanism that can achieve both objectives. However, there is great value in bringing the results of both objectives together, linking data to the papers it supports, and simplifying the procedures for researcher compliance and funding agency monitoring.

The approaches and technologies behind CHORUS can be used to create those connections. However, while building CHORUS into a robust, fully-fledged solution to the publications objective has been a rapid process – because CHORUS relies entirely on systems and services that are already in place – the data objective is not so readily achieved. As explained, the OSTP memo repeatedly calls for leveraging existing archives and not duplicating existing mechanisms. Scholarly journals represent a widely used and accepted mechanism for the broad dissemination of research publications, enabling the CHORUS framework to bring established, dedicated services together into an easy-to-use system for ensuring access, compliance and discovery. In fulfilling the publications objective, therefore, CHORUS does not create new infrastructure; rather, it uses existing infrastructure to provide tools focused on public access.

"CHORUS is a workable solution for funding agencies."

Conversely, systems for archiving, organizing and providing access to enormous quantities of raw data are not yet as robust as those in place for publications. CHORUS' component

services and principles can nonetheless offer similar utility to fulfilling the data objective of the OSTP memo. As funding agencies and others plan and create data infrastructure, tools like CrossRef, for identifying digital data, and FundRef, for connecting that data to the appropriate funding agency, should prove highly valuable.

If the data infrastructure is built with these sorts of tools, then a common language will exist between the publications and data services, allowing for interaction or even possible integration. CHORUS is fully committed to openly sharing information about its technological systems and stands ready to offer strategic advice to those building research data systems.

The ultimate goal of CHORUS is to meet the needs of the research community and funding agencies with as robust a mechanism as possible, focusing on maximizing utility while minimizing the costs and efforts required by all parties. CHORUS is built on an open, expandable architecture and established standards, using commonly available tools. This strategy is deliberately designed to enable collaboration and integration with other services as they come into existence.

> "CHORUS … stands ready to offer strategic advice to those building research data systems."

Does CHORUS scale internationally?

CHORUS was designed to meet the specific needs of the US OSTP memorandum, but was deliberately built on open standards and designed for interoperability. CHORUS does not prevent or interfere with the existence or development of any other repository or archiving strategy, and the open API can be used to enhance efforts made in these areas. CHORUS works equally well with both 'gold' and 'green' open access, and the open standards used offer tremendous potential for further development.

While CHORUS is just in its pilot phase, international scalability is an important future direction. In fact, the very name of the not-for-profit organization behind CHORUS, CHOR, Inc., was deliberately chosen to recognize the potential for this framework to reach beyond the initial launch of CHORUS in the US.

Conclusion

CHORUS is a rapidly growing non-profit solution for public access to scholarly content. It is an open technology solution focused on the *identification*, *access*, *discovery*, *preservation*, and *compliance* of scholarly content reporting on federally funded research and is internationally scalable. Importantly, it keeps researcher funding in the hands of researchers and is as inclusive as possible. CHORUS addresses the public access needs of funding agencies, researchers, institutions, publishers and the public. To achieve maximum effectiveness, it needs maximum participation from the publishing community. Participating is easy – find more information at www.chorusaccess.org

References

1. Dylla, HF, Negotiating with governments: In search of pragmatic public access policy, *2013 ALPSP International Conference*, Belfry, UK, 12 September 2013:
   http://www.alpsp.org/Ebusiness/Libraries/1309AIC_Thurs/Fred_Dylla_Plenary_3_1309AIC_web.sflb.ashx?download=true (accessed 10 December 2013).

2. America COMPETES Act of 2010:
   https://www.govtrack.us/congress/bills/111/hr5116 (accessed 6 January 2014).

3. US White House Office for Science and Technology Policy Report: *Interagency Public Access Coordination*:
   http://www.whitehouse.gov/sites/default/files/microsites/ostp/public_access-final.pdf (accessed 9 January 2014).

4. CrossRef:
   http://www.crossref.org (accessed 6 January 2014).

5. FundRef:
   http://www.crossref.org/fundref (accessed 20 December 2013).

6. US Office for Science and Technology Policy (OSTP) February 22, 2013 memorandum:
   http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf (accessed 9 September 2013).

7. PubMed Central:
   http://www.ncbi.nlm.nih.gov/pmc (accessed 9 January 2014).

8. CHORUS:
   http://www.chorusaccess.org (accessed 6 January 2014).

9. SHARE:
   http://www.arl.org/news/arl-news/2773-shared-access-research-ecosystem-proposed-by-aau-aplu-arl (accessed 9 January 2014).

10. ORCID:
    http://www.orcid.org (accessed 6 January 2014).

11. NISO, *Journal Article Versions (JAV)*. Recommendations of the NISO/ALPSP JAV Technical Working Group (RP-8-2008), 2008, Baltimore, National Information Standards Organization.

12. Warnick, W, *How CrossRef has accelerated science and its promise for the future: A federal perspective*, CrossRef Annual Members Meeting, 13 November 2014:
    http://www.osti.gov/home/speeches_slides/FY2014-CrossRef (accessed 29 January 2014).

Corresponding author: Howard Ratner, Executive Director
CHORUS
ORCID ID: http://orcid.org/0000-0002-2123-6317
E-mail: hratner@chorusaccess.org │ Voice: 917-251-9235 │ Web: www.chorusaccess.org