Publishers and universities respond to the OSTP mandate

Denise Troll Covey, Carnegie Mellon University
Publishers and universities respond to the OSTP mandate

OSTP Mandate – February 22, 2013

The Office of Science and Technology Policy (OSTP) directive requires federal agencies spending > $100 million on R&D to draft policies mandating public access to metadata, peer-reviewed publications, and digital data arising from federal funding.

- Format and license must ensure interoperability, encourage reuse, and prohibit unauthorized mass distribution.
- Agencies are free to select a repository and encouraged to leverage existing repositories.
- Repositories must provide open access, interoperability, and long-term preservation.
- Metadata must be publicly accessible immediately upon publication.
- Publications (final peer-reviewed manuscripts) must be open access no later than 12 months after publication. Stakeholders may petition for a longer embargo based on evidence of need.
- Policies must comply with current intellectual property law.
- Agencies must coordinate policies.

Draft policies are due August 22, 2013.

IN A NUTSHELL

Publishers and universities have issued proposals for achieving the requirements of the OSTP mandate. The proposals leverage different existing infrastructures. Publishers want to use their existing systems, but build a portal of shared metadata that links to distributed journal websites for access to the full text; the version is ambiguous. Universities want to federate existing institutional and disciplinary repositories, distributing both metadata and access to peer-reviewed author manuscripts. The publishers’ proposal is not as detailed as the universities’ proposal, particularly in regard to providing the mandated functionality, the licensing required to allow it, and the formats required to enable it.

Publisher Response: CHORUS – June 4, 2013

The ClearingHouse for the Open Research of the United States (CHORUS) announced by the Association of American Publishers\(^1\) will integrate public access into the publishing system. CHORUS will create and support a portal (CHORUS.gov) of metadata for articles available open access. Discovery and search will be provided through CHORUS.gov, familiar search engines, or any integrated site. CHORUS will direct users to an open access copy – the final peer-reviewed manuscript or the published version – available on the journal website post-embargo. No mention of open formats or open data. No mention of licensing rights to agencies or end-users. One reference to text and data mining: CHORUS will leverage CrossRef’s Project Prospect to protect users and content.\(^2\) Publishers will join CHORUS via a standard Service Level Agreement committing them to compliance. CHORUS publishers will provide agencies and research institutions with lists of articles for accountability purposes.

CHORUS will (according to the publishers):

- Provide a comprehensive solution for agencies to comply with the OSTP mandate.
- Minimize costs and ensure preservation.

\(^1\) CHORUS is endorsed by more than 50 for-profit and non-profit publishers.

\(^2\) Prospect applies semantics to articles during the publishing workflow, provides new views, and adds value for subscribers. See R. Kidd, RSC Project Prospect: Introducing semantics into chemical science publishing, n.d.
- Protect limited research funding by eliminating the need to duplicate existing infrastructure.
- Streamline expansion of public access to articles – no burdensome deposit or prescribed format.
- Streamline author and agency compliance by integrating open access into the publishing system.
- Provide access to articles in the context of the journal, i.e., give researchers what they are familiar with and give subscribers what they pay for.
- Retain the traffic and customer relationships publishers need to survive, i.e., a win-win scenario for the U.S. economy and taxpayers.³
- Not give control of open access to publishers. Federal agencies will set the embargo periods.⁴
- Be governed by representatives of key stakeholder groups, including federal agencies, universities, and research libraries.

Timeline:
- June 14 – high-level system architecture
- July 26 – technical specifications
- August 30 – proof of concept

Criticisms of CHORUS:
- Incompatible with the spirit of open access.
- Monetizes open access.⁵
- Not a comprehensive solution: CHORUS participants do not publish all federally-funded research.
- Publishers will not provide easy access to the open access articles.
- Untrustworthy (flaky, selective, temporary, and late) open access to the published version will be far less useful than assured open access to the author’s peer-reviewed manuscript.
- Concern that open access to the published version will require payment of a fee.⁶
- Does not adequately address bulk downloading or computational analysis.
- Does not facilitate reuse of the full corpus of publicly funded research.
- Federal agencies do not fund publishers and therefore have no control over publishers.
- Allowing (participating) publishers to control the only point of access to publicly funded articles will not create a stable, sustainable long-term archive.
- The timetable and infrastructure for providing open access should belong to the research community. Public access requires public stewardship of publicly funded research.

University/Library Response: SHARE – June 7, 2013

SHared Access Research Ecosystem (SHARE) announced by the American Association of Universities, Association of Public and Land-grant Universities, and Association of Research Libraries will federate institutional, disciplinary, and federal repository infrastructure, tools, and services to preserve, provide

³ Publishers claim the NIH mandate and PubMed Central (PMC) are a win-lose scenario for the U.S. economy and taxpayers because they reduce publisher revenues by decreasing direct traffic.
⁴ Publishers see this as a major concession they are willing to make to keep users on their platform.
⁵ Fees paid by publishers to support CHORUS would be passed to authors and subscribers.
⁶ This is my personal concern given the AAP’s statement that CHORUS will provide access to the “final peer-reviewed manuscripts or final published documents.” See “Understanding CHORUS.”
public access, and support re-use of peer-reviewed publications and (eventually) data arising from federally funded research. Institutions without a repository will piggyback on existing repositories.

**SHARE repositories will (according to the universities):**
- Permit access without login, credentialing, or individual tracking.
- Make publications accessible to persons with disabilities consistent with Section 508 of the Rehabilitation Act of 1973. XML is the optimal accessible format.
- Integrate metrics and identifiers to provide information on access, use, and impact.
-Expose metadata (based on open standards, common elements and expressions) to search engines and other discovery tools immediately upon publication of the article. Metadata will include (a) links to where the full text and associated data will be available after the embargo period, (b) copyright licensing terms, including assignment to a designated SHARE repository and any subsequent preservation archive, and (c) preservation rights to the final published version in the event the publisher ceases business.
- Use formats and standards that support computational analysis and interoperability. Authors will submit final peer-reviewed manuscripts to the Principal Investigator’s designated repository; alternatively the publisher will submit the XML version of the final published article.
- Link to the publisher’s website during the embargo period and after it expires.
- Make the full text available and enable full-text, keyword, and metadata searches upon expiration of the embargo period.
- Allow and enable the public to read, download, and perform text mining of publications. Publications must be linked to their source data and allow for reuse and replication of results.
- Use automated methods to certify compliance with agency requirements by notifying the funding agency and the Principal Investigator’s institution that deposit has occurred. Existing protocols (e.g., SWORD) will enable repositories to designate additional recipients of either notice of deposit or a copy of the article, subject to embargo restrictions.
- Have an advisory board with representatives of all stakeholder groups, including federal agencies, to ensure interoperability and single point of contact for federal agencies.

**For the OSTP mandate to succeed, federal agencies must immediately require:**
- Manuscripts submitted for publication to include federal agency ID, award ID, Principal Investigator ID (e.g., ORCID), and repository designation ID.
- A non-exclusive copyright license that enables deposit in an open access repository, preservation, discovery, and sharing and machine-based services after the embargo expires. Bulk downloads of publications for research purposes must be allowed under terms and conditions of federal agency copyright licenses that encourage research while protecting the integrity of the work. **Licensing arrangements must ensure that no single entity or group secures exclusive rights to publications resulting from federal funding that would inhibit or prohibit access and use of services and tools developed in response to federal policies.**

**Timeline:**
SHARE will be minimally functional when federal policies go into effect. Four development phases are planned:
1. Phase One requirements (to be completed in 12-18 months) include publication ID, data set ID, reporting (feedback look from repository to federal agency and PI's research office), core usage statistics to authors (and agencies, if desired), e.g., downloads, use of SWORD protocol to lower barriers to deposit, and use of OpenURL tagged with metadata record to enhance connectivity between repository and article on the publisher’s website.

2. Phase Two (in parallel with Phase One, completed 6-12 months after Phase One) entails software development to streamline submission workflow, usage metrics that place articles in context, generating reports to funding agencies, incorporation of Open Archives Initiative Object Reuse and Exchange (OAI-ORE), certification that PIs have complied with agency mandates, and adoption of best practices.

3. Phase Three adds value from federating content: text and data mining (requires copyright licenses granting non-exclusive rights to funding agency and university), bulk harvesting, semantic data, APIs, real-time synchronization of web resources, and open annotation.

4. Phase Four focuses on development of infrastructure relationships to support data requirements: data curation and associated software, linked data, shared distributed resources in repositories.

**Criticisms of SHARE:**

- Does not provide access to the published version of record.
- Requires universities to change open access and intellectual property policies to require Principle Investigators to license permanent archiving, access, and reuse rights to their publications and data arising from federal funding.
- Will take a minimum of two years to be fully operational.
- Will not work – universities have a poor track record of getting Principal Investigators to comply with federal mandates.

**References**


