



CHORUS Publisher Implementation Guide

Version 1.1 — 15 May 2014

This document was written by Evan Owens (AIP Publishing) in collaboration with the members of the CHORUS Technical Working Group.

TABLE OF CONTENTS

[INTRODUCTION](#)

[Simple Guide to Participating in CHORUS: What Do Publishers Need To Do?](#)

[IMPLEMENTATION CHOICE: AM OR VOR](#)

[IMPLEMENTATION COMPONENTS](#)

[1. FUNDING METADATA](#)

[2. CALCULATING ACCESS START DATES](#)

[3. LICENSE INFORMATION](#)

[4. CONTENT ACCESSIBILITY](#)

[5. URLS FOR PUBLIC ACCESS, AGENCY HARVESTING, AND TEXT MINING](#)

[6. CONTENT ARCHIVING](#)

[7. OTHER METADATA ELEMENTS](#)

[APPENDIX: Publisher Implementation Examples](#)

INTRODUCTION

Simple Guide to Participating in CHORUS: What Do Publishers Need To Do?

1. Collect funder information (via your manuscript tracking system or by extracting from your articles) and map it to entries in the FundRef Registry
2. Send the funder information to CrossRef's FundRef system
3. Link to a common license or post your proprietary license for content reuse on your site
4. Make at least one version of your articles reporting on funded research publicly accessible on your website, either Version of Record (VOR) or Accepted Manuscript (AM), with an appropriate embargo period if desired
5. Send to CrossRef the URL for the publicly accessible version paired with the URL for the reuse license with an appropriate start date commensurate with the chosen embargo period
6. Permit your publicly accessible versions to be harvested for indexing by agencies (agencies will link back to you by the DOI)
7. Archive the VOR at a recognized archive service

This document is designed to help publishers new to CHORUS plan and execute their implementation by providing background information, links to resources, and commentary and advice based on the CHORUS pilot publishers' experience. In the appendix are links to sample publisher pages that demonstrate CHORUS implementations.

IMPLEMENTATION CHOICE: AM OR VOR

CHORUS Policy: *Publisher member must make at least one version of their articles reporting on U.S. federally funded research publicly accessible through their website (either Version of Record [VOR] or Accepted Manuscript [AM]) with a stated embargo period if applicable.*

The decision as to whether to make the VOR or the AM available after embargo is an important business policy decision that impacts many of the CHORUS implementation components discussed in this document. It is primarily a business policy decision, but there may be technical issues depending on the publisher's current processes and systems.

For most hosting platforms, using the VOR to meet the mandate is likely to be the easier implementation: articles would just change from subscription access to public access at the end of the applicable embargo period. The entire VOR (PDF, HTML full text, etc.) would then be free to read in the same way that a Gold Open Access article would have been from date of publication. The implementation logistics depend on the mechanics of the access control system in the hosting platform.

Using the AM to meet the mandate is likely to require more hosting platform and production workflow changes. Implementation tasks could include adding the AM into the content on the hosting platform, adding a link to the AM from the article landing page, and any necessary changes to access control system so that the AM would be public access while the VOR is still subscription access.

Using the AM in this way also raises a workflow issue: when to load the AM into the hosting system? If it is loaded at the time of publication, a publisher might want to hide it from public view until the end of the applicable embargo period (or, equivalently, the access start date). Or the publisher might chose to wait and load the AM once the embargo period for the article ends. Another consideration is whether to suppress the AM for any authenticated subscribers who are entitled to the VOR. The timing of the loading of the AM has implications for the URL registration and agency harvesting requirements which are discussed below.

IMPLEMENTATION COMPONENTS

1. FUNDING METADATA

CHORUS Policy: *Participating publishers must collect and deposit funding metadata with CrossRef for articles that acknowledge funding from U.S. federal funding agency.*

The management of funding metadata is done using FundRef, a service of CrossRef that provides a standard way to report funding sources for published scholarly research. Publishers deposit funding information from articles and other content using the FundRef Registry, a taxonomy of international funder names, IDs, abbreviations, and alternate names. (The initial registry was donated and is maintained by Elsevier and is now freely available from CrossRef.) For more information about FundRef, see the CrossRef website: <http://www.crossref.org/fundref/>.

The FundRef funder registry is hierarchical and can thus represent parent-child relationships between agencies and agency programs. Ideally, the funding sources for an article should be identified at the finest grained level with the fallback to use sources higher up the hierarchy. The funding metadata for a given article consists of sets of funder name, funder identifier (omitted if the funder name is not in the registry), and award/grant/contract numbers (optional and repeatable). New funder names are reviewed by CrossRef as possible future additions to the registry. The FundRef registry is still evolving and growing rapidly; CrossRef is issuing monthly updates. One can sign up for notifications from CrossRef when an updated registry is available; sign up link is on bottom of the FundRef page: http://www.crossref.org/fundref/fundref_registry.html. Publisher implementation plans should include provision for keeping in sync with future updates to the registry.

Collecting Funding Metadata. There are two different approaches that publishers are taking

for collecting funding metadata for new articles: selected by the author in the peer review/manuscript tracking system UI (typically during manuscript submission) or extracted from acknowledgments during the production phase.

Multiple peer-review system vendors have added support for funding metadata collection into the user interface of their products. If you need to add an interface to your own software system, CrossRef has made available a reference UI widget that demonstrates best UI practices. See <http://labs.crossref.org/fundref-widget/> for details.

There are a number of operational challenges for publishers in getting funding information from authors:

- Authors don't always cite funders using the canonical registry funder names in their acknowledgments; e.g. acronyms versus full names
- Authors may not understand the hierarchy of the registry if they are asked to interact with it and/or may go for the parent agency rather than the sub-program
- Authors may input a variant of an existing funder name as a “new” funder name
- Authors may not include everything mentioned in acknowledgments section of their article in the FundRef data (or vice versa)

As a result of these challenges, some CHORUS pilot publishers that are relying on author input have built a formal review step into their process where staff compares the acknowledgments text in the submitted manuscript with the funding metadata entered by the author and then corrects the metadata and/or queries the author. Studies in the pilot period have found significant error rates, mostly omissions. Those error rates were reduced by improving the author instructions to emphasize that the funding information should match the acknowledgments text.

The other workflow solution is to make this a production operation by composition and typesetting vendors rather than author task: during the production process have the acknowledgments text read, mapped to FundRef registry, and the proposed FundRef metadata for the article returned to the author for verification at the proof stage. CrossRef Labs made available a tool to help with looking up funder names in the registry: <http://labs.crossref.org/fundref-reconciliation-service/>. There is also discussion in the industry about building this capability into copy editing tools.

Depositing Funding Metadata with CrossRef. In production mode (publishing new content), funding metadata would be part of the normal DOI registration process with CrossRef. So implementation would depend on a publisher’s current arrangements with CrossRef. For example, some publishers use their hosting platform vendors for DOI registration and thus implementation would depend on those vendors. FundRef metadata can be deposited with CrossRef either as part of article level metadata or as part of CrossMark metadata. If the publisher is participating in CrossMark, it is best practice to deposit the FundRef metadata within CrossMark so that it is visible in the displays of CrossMark data. According to CrossRef,

publishers can start depositing FundRef metadata in the CrossMark section of the deposit even before CrossMark has been enabled in publisher PDFs and web pages.

EXAMPLE 1. A simple example of funding information in CrossRef deposit format. See the CrossRef website for details: [http://help.crossref.org/#fundref\\$examples](http://help.crossref.org/#fundref$examples).

```
<fr:program name="fundref">
  <fr:assertion name="funder_name">U.S. Department of Energy
    <fr:assertion name="funder_identifier">
      http://dx.doi.org/10.13039/100000015
    </fr:assertion>
  </fr:assertion>
  <fr:assertion name="award_number">DE-FC26-07NT43098
  </fr:assertion>
</fr:program>
```

During the CHORUS pilot, publishers were adding FundRef metadata to articles already published. This can be done via the normal CrossRef update process using XML files; details are available at <http://help.crossref.org/>. This process may also be used by publishers who would otherwise be dependent on a vendor who has not yet implemented deposition of funding metadata. CrossRef also made it possible for publishers to supply just the FundRef metadata for articles already registered in CrossRef using a CSV file which requires less technical sophistication than the XML deposits.

Funding Metadata in XML Full Text. In designing a production process, a key decision will be where the FundRef metadata resides. One choice is to embed it into the article XML full text; examples of how to encode FundRef metadata in JATS follow below. (Note: JATS is being used here as an example. JATS is not a CHORUS requirement.) Another choice is store the information externally (e.g., as metadata in a content management system, in a database, or in an external file).

JATS XML for FundRef Metadata. The most recent versions of JATS (version 1.1d1 - <http://jats.nlm.nih.gov/archiving/1.1d1/>) include new features specifically to support FundRef funder identifiers, features that were not included in JATS version 1.0: <http://jats.nlm.nih.gov/archiving/1.0/>. The examples below show how to encode only the FundRef metadata elements in JATS; see the JATS documentation for examples of how to encode additional related information such as the principal award recipient, the principal investigator, and prose funding statements.

EXAMPLE 2: Funding metadata encoded using JATS 1.0. The attribute values on the funding group and named-content elements are suggestions only, not requirements; they match the assertion names in the CrossRef metadata deposit format.

```
<article-meta>
...

```

```

<funding-group specific-use="FundRef">
  <award-group>
    <funding-source>
      <named-content content-type="funder_name">U.S. Department of
        Energy</named-content>
      <named-content content-type="funder_identifier">
        http://dx.doi.org/10.13039/100000015
      </named-content>
    </funding-source>
    <award-id>
      DE-FC26-07NT43098
    </award-id>
  </award-group>
</funding-group>
...
</article-meta>

```

EXAMPLE 3: Funding metadata inserted into <article-meta> encoded using JATS 1.1. The attribute value on <funding-group> is a suggestion only.

```

<article-meta>
...
<funding-group specific-use="FundRef">
  <award-group>
    <funding-source>
      <institution-wrap>
        <institution>
          U.S. Department of Energy</institution>
        <institution-id>
          http://dx.doi.org/10.13039/100000015
        </institution-id>
      </institution-wrap>
    </funding-source>
    <award-id>DE-FC26-07NT43098</award-id>
  </award-group>
</funding-group>
...
</article-meta>

```

Funding Metadata Corrections. In discussion with funding agencies, it has been clear that they understand that the accuracy of publisher collected funding data depends on the actions of the authors and that if the agency spots an error or omission for a specific article, they should contact the author to pass the correction along to the publisher. The publisher can then update the FundRef metadata using the normal CrossRef update procedures. An implementation plan should include provisions for making these types of corrections as the FundRef metadata is critical to support the implementation of the anticipated agency public access requirements (discussed below).

Funding Metadata Display on Site. The FundRef FAQ suggests that publishers should display FundRef data on the DOI response pages, ideally via CrossMark. This makes sense because the FundRef data as submitted by the publisher can be revised by CrossRef (e.g. a “new” funder being added to the registry and the name corrected). The CrossMark widget fetches the current data from CrossRef so it is always up to date. To display the FundRef metadata directly in the DOI response page HTML, the publisher would need to be sure that its version of the data is current and reflects the latest changes to the registry by CrossRef. For example, if the funder name is not in the registry at time of publication but added later, CrossRef would update its copy of the funding information to incorporate the new funder ID. CrossRef is developing an API for publishers to get changes made to their articles’ FundRef data back from CrossRef.

Other Ideas, Observations, and Open Issues regarding Funding Metadata:

- Some publishers have been discussing making funder information a formal section of the article, not buried in the acknowledgments, and use the FundRef system to collect that information. This is a publisher decision, not a requirement or best practice recommendation.
- As funder agencies develop guidelines as to author best practices in acknowledging funding, CHORUS will include links to that in its documentation so that CHORUS participants are aware and can communicate this to authors if useful.
- There is discussion underway way about funding edge cases: what kinds of acknowledgments should or should not be included in FundRef (e.g., grant funding, facilities funding, fellowships, etc.) and which are subject to the OSTP public access mandate and the anticipated agency requirements.

2. CALCULATING ACCESS START DATES

If an article is identified by FundRef metadata as the output of U.S. government funded research, the next step is to identify the applicable funding agency access mandates and calculate the public access start date. As of the date of this document, those mandates have not been finalized. It is possible that different agencies will have different embargo periods and that the embargo periods may be different for different disciplines within the same agency. But there might be consensus on standard embargos.

Once the funding information has been finalized and an article is ready to be published, the steps to calculate the access start date could include the following:

- If the article is publicly accessible as of date of publication (e.g., Gold Open Access or made permanently open due to publisher policy such as free errata or free editorials), use the date of publication as the access start date.

- If the article becomes publicly accessible after the date of publication (e.g. Gold Open Access paid for after publication), compare that start date to the funder mandates in the next steps.
- Take the funder ID(s) from the article, look up the applicable embargo period(s) and calculate an access start date (e.g. date of publication plus one year).
 - If the agency embargos are different by discipline, select the appropriate one for this article.
 - If more than one embargo is applicable, select the earliest access start date.
- Use the result to populate the license metadata (described in next section).

Depending on the nature of your hosting platform, you may also need to update the access control system.

A future issue: the NISO Open Access Indicators work groups proposal for “free to read” metadata has not yet been finalized and not yet implemented in CrossRef or JATS. Once that is finalized and implemented, that may become the recommended way to express the public access start date along with the license metadata.

CHORUS and CrossRef are currently discussing how to capture funder mandates in a machine-readable format to support automating the process described above. Possibilities include incorporating funder mandates directly into the Funder Registry, CHORUS maintaining a machine-read file of U.S. government agency mandates, or CHORUS providing an API that can be queried. Details remain to be worked out and will be documented here when known.

3. LICENSE INFORMATION

CHORUS Policy: *A publisher should post a license on the publisher’s site for content reuse of all articles that acknowledge agency funding.*

CrossRef supports associating one or more licenses with an article’s DOI. The metadata properties for each license consists of a license URL, an optional start date, and an optional indicator as to what the license applies to (AM, VOR, or TDM (text data mining)). This metadata can be included with the initial DOI registration or added later.

The choice of license is up to the publisher: it can be either a standard public license such as one of Creative Commons licenses or a proprietary publisher-specific license, perhaps based on STM’s newly available licenses. For standard public licenses, the URL should point to the appropriate public web page and not to a page on the publisher’s web site; e.g., to <http://creativecommons.org/licenses/by/4.0/>.

For proprietary publisher licenses, a public web page must exist that documents the license; that URL would then be part of the CrossRef metadata deposit. It is important that those URLs be persistent. CrossRef does not currently support DOIs for licenses but they are looking at a

generic DOI capability for web pages that could be used to ensure the persistence of these licenses URLs. As licenses can potentially change over time, it is important that publisher proprietary licenses include version information. That can be as easy as building a version number into the URL, as in the example Creative Commons license cited above. In lieu of DOIs, publishers should try to use stable URLs that are unlikely to change to avoid having to do large updates of metadata deposited with CrossRef.

CrossRef License Metadata. The CrossRef license metadata (called “access indicators”) supports multiple licenses for the same article. Here are some scenarios:

- One license for during the embargo (subscription access) period, another starting afterwards for public access
- A different license for each version (AM or VOR)
- A separate license for TDM privileges
- Or any combination of the above

Posting licenses for the non-public access period is not required by CHORUS; however, a publisher might choose to do that for completeness or to document TDM licensing for subscribers. For publisher proprietary licenses, CHORUS must be informed which license URLs are for public access and which are for subscription access so that the CHORUS dashboard can filter the statistics correctly. A procedure for this has not yet been implemented. For now, CHORUS is working with agency partners to identify which publisher licenses will meet agency public access requirements.

For technical information on license information deposit, see the CrossRef help site: <http://help.crossref.org/#depositing-access-indicators>.

EXAMPLE 4. License metadata in CrossRef format.

```
<ai:program name="AccessIndicators">
  <ai:license_ref applies_to="am" start_date="2015-07-01">
    http://creativecommons.org/licenses/by/4.0/.
  </ai:license_ref>
</ai:program>
```

JATS markup for licenses does not currently support start date as a property. The specific-use attribute could be used for applies-to metadata; that is consistent with the meaning of “specific-use in JATS.

EXAMPLE 5. License metadata in JATS format.

```
<license specific-use="vor"
xlink:href="http://publishername.org/licenses/1.0v1"/>
```

```
<license specific-use="am"
xlink:href="http://creativecommons.org/licenses/by/2.0/" />
```

4. CONTENT ACCESSIBILITY

CHORUS Policy: *Publisher member must make at least one version of their articles reporting on U.S. federally funded research publicly accessible through their website (either Version of Record [VOR] or Accepted Manuscript [AM]) with a stated embargo period if applicable.*

Content declared to be public access as of a start date documented in the license metadata deposited with CrossRef must actually be publicly accessible on the public platform as of that date. Coordination of the license metadata with the hosting platform access control system is critical to the success of CHORUS. Errors in this area would seriously damage the reputation of CHORUS so publishers are urged to implement this robustly.

There is currently no requirement regarding display of a CHORUS logo on publisher sites.

5. URLS FOR PUBLIC ACCESS, AGENCY HARVESTING, AND TEXT MINING

CHORUS Policy: *Publishers must send to CrossRef the URL for the publicly accessible version paired with the URL for the reuse license with an appropriate start date commensurate with the stated embargo period.*

The core of CrossRef metadata is a DOI value and the URL that the DOI resolves to, usually the landing page for the article on the publisher hosting site. CrossRef has expanded its metadata to now support defining additional resource URLs for special purposes, defined by three attributes: mime-type of the returned file, content version (e.g., AM or VOR), and functionality (text-mining, syndication, crawler-based, or any). This allows a publisher to direct traffic to different locations as needed. This new metadata syntax is still under discussion and may evolve.

The simplest set of metadata to meet CHORUS requirements would be to add a single <collection> element identifying the content version being made publicly accessible, either AM or VOR. This must correspond to the license metadata; the link between the two is which version is identified using the applies_to attribute. A simple example:

EXAMPLE 6. Simple Example of Resource URL Metadata

```
<doi_data>
  <doi>10.5555/12345678</doi>
  <timestamp>201405151609</timestamp>
  <resource>http://publisher.com/content/12345678.html</resource>
</doi_data>
<license_ref applies_to="am" start_date="2015-07-01">
```

```

    http://creativecommons.org/licenses/by/4.0/.
  </license_ref>
  <collection property="any">
    <resource content_version="am" mime_type="application/pdf">
      http://publisher.com/content/12345678-AM.pdf
    </resource>
  </collection>

```

If the publisher wants to direct text mining, crawling, or syndication to different files and/or locations, those categories can be specified using the property attribute of the collection element. An example:

EXAMPLE 7. Complex Example of Resource URL Metadata

```

<doi_data>
  <doi>10.5555/12345678</doi>
  <timestamp>201405151609</timestamp>
  <resource>http://publisher.com/content/12345678.html</resource>
  <collection property="text-mining">
    <resource content_version="vor" mime_type="application/pdf">
      http://textmine.publisher.com/content/12345678.pdf
    </resource>
  </collection>
  <collection property="syndication">
    <resource content_version="vor" mime_type="text/xml">
      http://distribution.publisher.com/content/12345678.xml
    </resource>
    <resource content_version="am" mime_type="application/pdf">
      http://distribution.publisher.com/content/12345678-AM.pdf
    </resource>
  </collection>
  <collection property="crawler-based">
    <item crawler="google">
      <resource content_version="vor" mime_type="application/pdf">
        http://crawlsite.publisher.com/content/12345678.pdf
      </resource>
    </item>
  </collection>
</doi_data>

```

Including these resource URLs in CrossRef is completely separate from access control; that remains a function and responsibility of the publisher's systems. Different applications (text mining, syndication, crawling) might have different access control solutions.

6. CONTENT ARCHIVING

CHORUS Policy: *Publishers must archive the VOR or AM in at least one recognized archive service.*

To implement this policy, publishers must have a contractual arrangement and their content must be actively being deposited with or collected by one of the approved archiving services. Information about archiving arrangements must be included in the CrossRef metadata submission per article so that the CHORUS dashboard can document the archiving arrangements. For publishers who already participate in any of these archiving arrangements, adding this to the CrossRef DOI registration metadata deposit is a modest change; the metadata is described below. Publishers who do not currently participate in an archiving arrangement should contact the archiving services directly for information about the agreement and the data transfer process (crawling or feeding). For publishers using hosting platform vendors, some of those service providers support archiving arrangements so they may be able to facilitate the implementation.

Publishers should document their archiving arrangements as facts in the CrossRef metadata, regardless of whether or not a given agency has approved a specific archive.

Archive metadata is a new feature of CrossRef metadata schema 4.3.4. Although version 4.3.4 is not yet formally “released”, CrossRef is already accepting archive metadata using this addition to the schema. <http://www.crossref.org/schema/documentation/4.3.4/4.3.4.html>

EXAMPLE 8. Archive Metadata listing archive(s) for a specific article

```
<journal_article>
...
  <archive_locations>
    <archive name="CLOCKSS"/>
    <archive name="Portico"/>
  </archive_locations>
...
</journal_article>
```

7. OTHER METADATA ELEMENTS

In addition to all the metadata types described above which are essential to CHORUS, there are some additional metadata elements that are optional but desirable for a CHORUS participating publisher.

Abstracts. While not mandatory, CHORUS Publishers should include abstracts for funded articles in the metadata. CrossRef has recently added support for including abstracts in the metadata; the supported format is NISO JATS XML. For publishers whose XML is JATS 1.0 or NLM 3.0, it is straight forward to include this in the metadata submission to CrossRef (JATS 1.0 is backward compatible with NLM 3.0.) For publishers using other DTDs or earlier version of NLM, conversion programming would be required. Abstracts included in the CrossRef metadata are then available in the same way as other CrossRef bibliographic information.

ORCID Identifiers. CrossRef also supports inclusion of ORCID IDs as a component of contribution information. For publishers who collect ORCID IDs from authors, adding this to the metadata is of significant value to funding agencies as it gives them a tool to connect researchers to publications.

A technical note: CrossRef's data model collects researcher ORCID IDs and funding records as separate properties of the article not linked to each other within the deposit schema, so in the case of a multi-author article there is no way to link specific authors with specific funding records. Some of the peer-review system implementations of FundRef and ORCID data collection do actually align authors and funding thereby collecting more metadata than CrossRef currently supports and therefore it requires some manipulation to collapse that metadata down to the CrossRef flat model.

CrossMark. Although not a requirement for CHORUS, publisher participation in CrossMark to record updates, errata, etc., is a good mechanism to document versions of articles (updates, corrections, retractions, etc.) and direct readers to the latest version. A side benefit is that the CrossMark user interface widget can then display the FundRef metadata without any additional implementation requirements for the publisher platform.

APPENDIX: Publisher Implementation Examples

As publisher implementations of CHORUS become functional, links to examples will be added to this section of this document.

Publisher License and Embargo Web pages:

- APS: <http://journals.aps.org/licenses>
- Elsevier: <http://www.elsevier.com/about/open-access/green-open-access>