Publisher Implementation Guide

v2.0 – August 17, 2015

This document was written by Evan Owens (Cenveo Publisher Services), co-chair of the CHORUS Technical Working Group.

Updates and additions in this version (2.0)

- CHORUS policies and recommendations updated
- Simple Guide text updated
- XML examples updated for JATS 1.1 and NISO ALI; moved to Appendix 2
- Funder Harvesting URLs and Access information added
- Funder-Specific Implementation Tasks added as Appendix 1

TABLE OF CONTENTS

Introduction

This document is designed to help publishers new to CHORUS plan and execute their implementation by providing background information, links to resources, commentary, and advice based on the CHORUS publishers' experience.

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 CrossRef Funder Registry
2. Send the funder information to CrossRef's FundRef system
3. 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
4. Choose a license for content reuse for each article reporting on funded research, either a common public license such as Creative Commons (CC) or a publisher proprietary license
5. Send to CrossRef license metadata identifying the publicly accessible version, the reuse license, and the start date for public access commensurate with the selected embargo period
6. Send to CrossRef a URL for a full text version and make it available to be harvested for indexing by funders from date of publication (funders will link back to you by the DOI and only harvest articles reporting on research that they funded)
7. Archive the AM or VOR at a recognized archive service (CLOCKSS or Portico)

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 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 article 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 may require system changes, depending on the current production process and hosting system. 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. Some CHORUS participants using the AM as a PDF for public access are adding watermarks and/or cover pages to the AM PDF so as to document for the reader that the PDF is the AM not the VOR and to point the reader back to the VOR on their website via DOI links.

Using the AM as the public access version 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 choose to wait and load the AM once the embargo period for the article ends. The timing of the loading of the AM has implications for the URL registration and funder harvesting requirements which are discussed below in sections 5 and 6. Another optional consideration is whether a publisher wants to suppress the visibility of the AM from any authenticated subscribers who are entitled to the VOR so users access the best version available to them as well as whether to prevent Google from indexing accepted manuscripts.

Implementation Components

In addition to the standard CHORUS implementation components documented in this section, there are also funder-specific implementation tasks that arise from CHORUS funder participation agreements. See Appendix 1 for the current list of funder-specific implementation items.

1. Funding Metadata

CHORUS Policy*: Participating publishers must collect and deposit funding metadata with CrossRef for articles that acknowledge funding from a signed CHORUS Funder participant.*

CHORUS Recommendation: *Publishers may wish to make their existing open access content visible in CHORUS by depositing retroactive funding metadata with CrossRef. Retroactive deposits are not a requirement of CHORUS membership; however, they would be a valuable contribution to the scholarly community.*

The management of funding metadata is conducted using the CrossRef FundRef service, a service 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 funders. A FundRef entry consists of funder name, funder ID, abbreviations, alternate names, and country. The FundRef registry is hierarchic and can thus represent parent-child relationships between funders and funder sub-organizations. The initial FundRef 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 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). From the CrossRef website FundRef information page (http://www.crossref.org/fundref/):

> Funder IDs from the FundRef Registry ***must*** be included for any funder that is listed in the Registry. If a funder is not listed in the Registry it is acceptable to deposit the name only with no ID, but these deposits will not appear in FundRef Search until a funder ID has been created and the record redeposited (see details of how to query for new funder IDs at http://help.crossref.org/updating-fundref).

Funder names with no funder ID will also not be visible in the CHORUS dashboards and search until a funder ID is assigned and the article metadata updated accordingly.

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 location is http://www.crossref.org/fundref/fundref_registry.html. Publisher implementation plans should include provision for keeping in sync with future updates to the registry and updating their deposited information based on future additions to the registry.

Scope of Funder Metadata Collection: All or Part?. The FundRef registry is very large and collecting high-quality metadata for 10K funders can be a challenge. Some publishers have chosen to focus on a subset of the registry and only collect metadata for that subset.

CHORUS Recommendation: *Publishers should capture as much funder metadata as possible so that they don't have to go back and do it later. But if data collection for a subset is a practical solution, priority should be in this order:*

- *Signed CHORUS Funder participants*
- *All US Government Funding Agencies*
- *All Non-government US Funders (Bill and Melinda Gates Foundation, World Bank, Sloan, etc.)*
- *All international Funders with OA Policies*
- *All US State-level Funding Agencies*

- *All Funders (the full FundRef Registry)*

Funding Metadata Collection Processes. 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.

Most of the 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; see http://labs.crossref.org/fundref-widget/ for details. If you need real-time software access to the registry, CrossRef has an API that can be used to look up funder names; see http://api.crossref.org for documentation. Using this API ensures that your system is accessing the most up-to-date version of the FundRef registry.

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 or funding program names rather than funding organization names
- As the same organization name may occur in multiple countries, it is possible to accidentally select the wrong funder
- Authors may not understand the hierarchy of the registry if they are asked to interact with it and/or may go for the parent funder rather than the sub-organization
- Authors may input a variant of an existing funder name as a “new” funder name
- Authors may not include all funders mentioned in acknowledgments section of their article in the FundRef data (or vice versa)

As a result of these challenges, some CHORUS member publishers that are relying on author input have built a formal review step into their process where the funding metadata entered by the author is compared with the acknowledgments text in the submitted manuscript, any problems or inconsistencies identified, corrected, and then verified with the author. That can be done during the submission process or later during production. Some publishers are including the funding metadata verification as author queries on the article proofs. Studies by already implemented member publishers have found significant error rates in author metadata collected via peer review system UI, primarily 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 rather than an author task: have the funding information extracted from the acknowledgments text and mapped to the FundRef registry, verified during copy editing, and then the proposed FundRef metadata

submitted to the author for confirmation at the proof stage. CrossRef Labs made available a tool for looking up funder names: http://labs.crossref.org/fundref-reconciliation-service/.

Funding Metadata Special Cases. In addition to author grant funding, there are other kinds of funding relationships that appear in acknowledgements and may or may not be part of the FundRef metadata. Some examples currently being discussed:

- **Facility Acknowledgements versus Facility Funding.** Author acknowledgements may include mentioning the use of a government-funded facility (e.g. a laboratory) in the research that was the basis of the article. However, using a government facility is not necessarily the same as a funding grant and therefore may not have a public-access mandate and should not be captured as funding metadata. But there is also laboratory-funded research which should be captured as funding metadata. The US Department of Energy has recently adjusted its FundRef hierarchy: deprecated laboratory names and redirected them to a new entry named “Laboratory Directed Research and Development (LDRD)”.

- **Intramural Funding.** FundRef is having discussions about how to capture intramural funding information in the FundRef metadata system. Intramural funding is defined as when authors of an article are employees of a funder. So they are not grant funded but their research is subject to government public access mandates. As of July 2015, the FundRef metadata specification for intramural funding has not yet been finalized and the peer review vendors have not yet implemented this so this is a future requirement. As this goes live, it will be documented on the FundRef website and updated here.

Storing Funding Metadata. In designing a production process, a key decision will be where the FundRef metadata resides long-term. The most common process is to embed it into the article XML full text; examples of how to encode FundRef metadata in JATS or NLM XML are in Appendix 2. Publishers who use hosting vendors to deposit CrossRef metadata will have to conform to the vendors procedure...which may require JATS XML. Another choice is to store the information externally (e.g., as metadata in a content management system, in a database, or in an external file) rather than in the full text.

CHORUS Policy: *JATS is not a specific CHORUS requirement or recommendation; only submission of metadata to FundRef is a requirement.*

Depositing Funding Metadata with CrossRef. In production mode (publishing new content), funding metadata would normally 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 their vendor’s procedures. FundRef metadata can be deposited with CrossRef either as part of article level metadata or as part of CrossMark metadata.

CHORUS POLICY: *CrossMark is not a specific CHORUS requirement or recommendation; only submission of metadata to FundRef is a requirement.*

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.

For current documentation on FundRef depositing see the CrossRef website pages:

- http://help.crossref.org/fundref
- http://help.crossref.org/deposit_schema
- http://help.crossref.org/fundref-deposists
- http://help.crossref.org/updating-fundref

Depositing Retroactive or Updated Funding Metadata. CrossRef supports adding or updating funding metadata to existing article DOI metadata records by full metadata redeposit or "stand-alone deposit" of FundRef metadata only. Details are on the deposit information pages listed above. CrossRef also has a non-XML deposit process that uses CSV format; that can be used to facilitate retroactive metadata deposit: http://help.crossref.org/csv-upload

Funding Metadata Corrections. In discussion with funders, 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 funder spots an error or omission for a specific article, they should contact the author to pass the correction along to the publisher. The publisher would then update the FundRef metadata using the CrossRef procedures. An implementation plan should include provisions for making these types of corrections, as the accuracy of the FundRef metadata is critical to support the implementation of the funder's public access requirements.

Funding Metadata Display on Site. The FundRef FAQ suggests that publishers should display FundRef data on the DOI response pages, ideally via CrossMark.

CHORUS Policy: *CrossMark or its display is not a specific CHORUS requirement or recommendation; only submission of metadata to FundRef is a requirement.*

However, 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, a funder name might be corrected in response to input from the funder so the same funder ID might now point to an updated funder name.

Funding Metadata in Articles. Some CHORUS publishers have been discussing making funder information a formal section of the article, not buried in the acknowledgments, and using the FundRef system to organize that information. Journals that have already had separate funding and acknowledgements sections have found that they have less difficulty in implementing FundRef metadata because their authors already make the logical separation of funding information from other acknowledgements**.**

CHORUS POLICY: *This is a publisher decision, not a CHORUS requirement.*

2. Calculating Access Start Dates

If an article is identified by FundRef metadata as reporting on the output of U.S. government funded research, the next step is to identify the applicable funder public access mandates and calculate the public access start date. Links to US agency plans: http://www.chorusaccess.org/resources/us-agency-public-access-plans/

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 funder 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.

The date of publication used to calculate the embargo period should be the date deposited in the CrossRef metadata. So for journals that publish continuously (individual articles prior to issues), if the date of publication that is deposited with CrossRef is the date of the article posting rather than the issue cover date, the embargo period calculation would be based on that registered date.

3. License Information

CHORUS Policy: *A publisher must identify a license for content reuse of all articles that acknowledge funding, register that information with CrossRef, and link to it on their content website. The registered metadata must include the AM and/or VOR indicator flag and license start date.*

CrossRef supports associating one or more licenses with an article DOI. License metadata can be included with the initial DOI registration or added later. The metadata properties for each license consists of a license URL, an optional start date, and an optional indicator as to what version the license applies to (AM, VOR, or TDM). TDM (text and data mining) is not relevant to CHORUS; license metadata marked as "TDM" will be ignored by CHORUS Services.

The choice of license is up to the publisher: it can be either a standard public license such as one of the Creative Commons licenses or a proprietary publisher-specific license. Best practice information and sample publisher licenses are available from STM at http://www.stm-assoc.org/copyright-legal-affairs/licensing/open-access-licensing/

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 publisher-specific 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. 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. 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 (Not related to CHORUS)
- Or any combination of the above

CHORUS Policy: *Submitting license metadata 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 via the CHORUS Publisher Admin system (coming July/August 2015) which license URLs are for public access and which are for subscription access so that the CHORUS dashboard can filter the statistics correctly.

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

EXAMPLE. License metadata in CrossRef format. (v4.3.4)

```
<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>
```

As with funding metadata, publishers systems may store license metadata in the full text XML or store it externally in content management systems or database. Examples of encoding license metadata in full text XML are in Appendix 2. Because access to an article can change over time due to business considerations (e.g., an author going for gold open access post-publication rather than pre-publication), some CHORUS participants choose to store changes and updates to license information outside the full text XML.

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.*

CHORUS Recommendation: *Display of a CHORUS Member logo on publisher sites is encouraged but not required.*

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.

Public Access means unimpeded (unfettered) access. Some publishers require users to register on their platform even for public access content. Some funders may object to required registration. For example, the US DOE does not consider this to be public access, so content accessible with this requirement would not be linked to by DOE and author manuscripts would be made accessible by DOE directly.

5. URLs to Allow Funder Harvesting for Indexing

CHORUS Policy: *Publishers must send to CrossRef a URL for either the publicly accessible version or VOR version for Funder harvesting and indexing using the CrossRef Collection and Resource metadata syntax.*

CAVEAT: The CHORUS TWG is currently reviewing an alternative to using the CrossRef syndication metadata described in this section. That solution would involve embedding URL information into the article landing page. Publishers would have the option of using either solution. This document will be updated when the details of this new alternative solution are finalized.

CHORUS participating funders want to harvest funded content for indexing from the date of publication, which would often be well before the public access start date. So publisher implementation must include funder access to some full text version during the embargo period. Possible scenarios:

1. Funder access to AM from day 1 for indexing; public access to AM post-embargo
2. Funder access to VOR from day 1 for indexing; public access to AM post-embargo
3. Funder access to VOR from day 1 for indexing; public access to VOR post-embargo

Some publishers using AM for public access have chosen scenario 2 because they do not post the AM on their platform until the end of the embargo period. Each of these solutions has a different set of metadata values to be deposited with CrossRef, see examples below.

The core of CrossRef metadata is a DOI 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 URLs in groups known as collections of resources so as to point directly to specific content resources and not just the landing page. Each resource entry can include metadata for mime-type and content version. This CrossRef collection and resource metadata supports many different functionalities such as multiple resolutions, text mining, crawling, etc. Only a subset of this metadata is relevant to and used by CHORUS.

To support funder harvesting, a URL for the full text must be provided using the collection metadata element in the “syndication” collection. If a different version is intended for funder harvesting and for public access, the harvesting version should be in the “syndication” collection and the public access version in “unspecified” collection. If there is only one version URL, it must be in the “syndication” collection. There are other collection property values in the CrossRef metadata that are not relevant to CHORUS and will be ignored by CHORUS and funders: “list-based" (for multiple DOI resolution), “country-based", "crawler-based", and “text-mining".

The three examples below show the three scenarios listed above (AM/AM, VOR/AM, VOR/VOR). In scenario 1 (AM/AM), the publisher would also have the VOR available to subscribers, but is not required to deposit resource and license metadata for the non-publicly accessible versions, only that which fulfills the public access mandate. In Scenario 2, because the VOR is for harvesting and the AM for public access, the license_ref for the VOR is optional because it isn't the public access copy and the resource metadata for the AM is optional because it isn't the harvesting copy. Optional means not required by CHORUS; it is compatible with CrossRef practices to supply this metadata.

EXAMPLE. **Three versions of Funder harvesting (CrossRef schema v4.3.6)**

Scenario 1 (Funder AM harvesting AM and public access AM)

```
<ai:program name="AccessIndicators">
  <ai:license_ref applies_to="am" start_date="2016-07-01">
    http://creativecommons.org/licenses/by/4.0/
  </ai:license_ref>
</ai:program>
<doi_data>
  <doi>10.5555/12345678</doi>
  <timestamp>201505151609</timestamp>
  <resource>http://publisher.com/content/12345678.html</resource>
  <collection property="syndication">
    <item>
      <resource content_version="am" mime_type="application/pdf">
        http://publisher.com/content/12345678-AM.pdf
      </resource>
    </item>
  </collection>
</doi_data>
```

Scenario 2 (Funder harvesting VOR and public access AM)

```
<ai:program name="AccessIndicators">
  <ai:license_ref applies_to="am" start_date="2016-07-01">
    http://creativecommons.org/licenses/by/4.0/
  </ai:license_ref>
  <ai:license_ref applies_to="vor" start_date="2015-07-01">
    http://publisher.com/licenses/1.0/
  </ai:license_ref>
</ai:program>
<doi_data>
  <doi>10.5555/12345678</doi>
  <timestamp>201505151609</timestamp>
  <resource>http://publisher.com/content/12345678.html</resource>
  <collection property="syndication">
```

```
    <item>
      <resource content_version="vor" mime_type="application/pdf">
        http://publisher.com/content/12345678.pdf
      </resource>
    </item>
  </collection>
  <collection property="unspecified">
    <item>
      <resource content_version="am" mime_type="application/pdf">
        http://publisher.com/content/12345678-AM.pdf
      </resource>
    </item>
  </collection>
</doi_data>
```

Scenario 3 (Funder harvesting VOR and public access VOR)

```
<ai:program name="AccessIndicators">
  <ai:license_ref applies_to="vor" start_date="2016-07-01">
    http://creativecommons.org/licenses/by/4.0/
  </ai:license_ref>
</ai:program>
<doi_data>
  <doi>10.5555/12345678</doi>
  <timestamp>201505151609</timestamp>
  <resource>http://publisher.com/content/12345678.html</resource>
  <collection property="syndication">
    <item>
      <resource content_version="vor" mime_type="application/pdf">
        http://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.

6. Funder Harvesting Access

CHORUS Policy: *Members must permit the publicly accessible AM or VOR, or a VOR behind a paywall, to be available for indexing from the date of publication by Participating Funders.*

CHORUS POLICY: *Participating funders agree to harvest only articles they funded and for which FundRef metadata has been deposited, not to try to crawl a publisher's entire site unless the funder has a separate agreement with the publisher.*

Because CHORUS participating funders will want to harvest articles for indexing soon after publication which often will be prior to public access mandate start date, there is an access control implementation requirement for CHORUS publishers. Publishers are only required to grant funder access to the articles based on research that they funded, not the entire journal. As discussed in the previous section, the harvesting can be of the AM or VOR (publisher decision). CHORUS is supporting two different solutions for funder access: IP-based access and token-based authentication.

IP Access. Funders that conduct harvesting will register the IP address of their crawler with CHORUS. Publishers can sign up for notifications of any changes to funder crawler IP addresses on their CHORUS Publisher Admin page. IP-based access can be implemented on the hosting platform in various ways. Some solutions: by registering the IP address on the hosting server as a crawler, by giving the funder IP-based subscription access to the journals, or by constructing a separate product that consists of the funded articles and giving the funder a subscription to that product. Given that the funder agrees to harvest only funded articles, giving the funder access to the entire journals might be a technically easier solution. The best solution will be based on the hosting platform requirements.

Token Authentication. CHORUS has implemented a token-based authentication scheme based on the CrossRef TDM implementation. The token authentication system issues a token to the funder that is then included in the content request to the publisher's site; the publisher's system can then check the token via a CHORUS API to authenticate the content request. Implementation details are in the document cited above.

7. Content Archiving

CHORUS Policy: *Publishers must archive the VOR or AM in at least one CHORUS recognized archive (CLOCKSS or Portico)*

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. 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.

Information about archiving arrangements may be included in the CrossRef metadata submission but this is not a requirement for CHORUS participants. 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 in the section on other metadata elements.

8. Other Metadata Elements

In additional 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. As CrossRef updates its metadata schema several times per year, for the latest specification be sure to check the CrossRef website: http://help.crossref.org.

Abstracts. While not mandatory, CHORUS Publishers should include abstracts for funded articles in the metadata. CrossRef supports including abstracts in the metadata; the supported format is currently NISO JATS XML 1.0. 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 might be required. Abstracts included in the CrossRef metadata are then available in the same way as other CrossRef bibliographic information.

Archiving. Archive metadata was ~~is~~ a new feature of CrossRef metadata schema version 4.3.4:

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

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

For publishers who want to update their CrossRef metadata for already published materials to reflect their archiving arrangements, they should contact CrossRef support who can help facilitate such an update. For example, rather than a publisher redepositing all of its CrossRef records to record that all those articles are in a specific archive, CrossRef staff can organize a batch update that just adds that information across the set of records. For new publications going forward, the archive metadata would be added to new CrossRef submissions.

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.

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 funders 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. This may change in the future as CrossRef implements support for intramural funding information in FundRef; details not finalized yet.

APPENDIX 1: Funder-Specific Implementation Tasks

As Funders become CHORUS participants, some funder-specific implementation tasks may be part of the participation agreement. This appendix is for documenting these tasks. If a task becomes standard for all of CHORUS it will be moved to the main text as a standard implementation task.

1. DOE Author Prompting

CHORUS Policy: *Publishers who make AMs publicly accessible and want DOE to link to the AM on their site from DOE PAGES must prompt authors of DOE funded research to submit the AM to DOE and must provide a screenshot or example demonstrating their prompting practice via their CHORUS Publisher Admin page.*

From the CHORUS-DOE Funder Participation Agreement:

> If Publisher Member does not make a VOR publicly accessible on its website, but does make an AM publicly accessible on its website, Funder will link in a prominent manner to the AM on the Publisher Member website for those Publisher Members who prompt authors to submit AMs to Funder.

Publishers are not required to submit the AM to DOE; only to prompt the author to do the submission at the appropriate point in the manuscript submission workflow (accepted manuscripts only). With the goal of integrating prompting into the manuscript submission workflow, interim steps could include measures such as a web page denoting author instructions, author communication during the submission or revision process, and/or other ways).

Because CHORUS has to report to DOE how publisher members are prompting, a feature has been added the Publisher Admin page in CHORUS to allow publishers to upload screen shots of their prompt implementation. The menu item is “Prompt Samples | Upload”. Depending on how the prompt is executed, the publisher should upload a sample screen shot (e.g. the author instructions web page) or a sample of the author communication (e.g. an acceptance email that includes prompting).

DOE had provided CHORUS with a URL for their manuscript submission systems for use in the prompts: https://www.osti.gov/elink/am-submission/ In the future, this URL will be available on a CHORUS web page and via the CHORUS API; details to be finalized.

APPENDIX 2: XML Markup for FundRef and License Metadata

This appendix consists of examples of ways to encode CrossRef metadata needed for CHORUS in a variety of XML full text DTDs such as JATS and NLM. This is for information only; the only CHORUS requirement is to deposit metadata with CrossRef so how publishers manage that metadata in their systems and in their content is their decision.

As CrossRef updates its deposit schema several times a year, be sure to check on the CrossRef website for the latest version and for full documentation.

1. FundRef Metadata XML examples

EXAMPLE 1. Funding information in CrossRef deposit format (schema v4.3.6)

```
<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>
```

JATS XML for FundRef Metadata. JATS version 1.1 is currently in the final stages of review and voting; it is not expected to be formally approved until October 2015. The final draft (version 1.1d3) available on the website (http://jats.nlm.nih.gov/) is what is being voted on for approval as the formal release version 1.1. This version include new features specifically to support FundRef funder identifiers, features that were not included in JATS version 1.0. The examples below show only how to encode the FundRef metadata elements in JATS 1.1d3; 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 1.1d3; 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>
```

EXAMPLE 3: 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>
```

NLM DTD XML. JATS 1.0 is fully backward compatible with NLM DTD version 3.0 so the JATS 1.0 example above would also be possible in NLM DTD 3.0. For publishers still using NLM DTD version 2.3 or earlier, there is no easy way to include FundRef information as the award-group element shown above was new in NLM 3.0. One solution would be to use the custom metadata markup:

EXAMPLE 4: Funding metadata encoded using NLM DTD 2.3.

```
<article-meta>
  ...
  <custom-meta-wrap>
    <custom-meta>
      <meta-name>fundref:funder_name</meta-name>
```

```
      <meta-value>U.S. Department of Energy</meta-value>
    </custom-meta>
    <custom-meta>
      <meta-name>fundref:funder_identifier</meta-name>
      <meta-value>
        http://dx.doi.org/10.13039/100000015
      </meta-value>
    </custom-meta>
    <custom-meta>
      <meta-name>fundref:award_number</meta-name>
      <meta-value>DE-FC26-07NT43098</meta-value>
    </custom-meta>
  </custom-meta-wrap>
  ...
</article-meta>
```

If there is more than one set of funding information (name/identifier/award), a separate custom-meta-wrap element could be used for each set.

2. License Metadata XML Examples

Example 5. License Metadata in CrossRef Schema (v4.3.4)

```
<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 version 1.1d3 added support for NISO Access License Indicator metadata, which is related to the CrossRef Access Indicators metadata but not exactly the same. As part of the change to support NISO ALI, the URL for the license moved from an attribute to an element content.

Example 6. License Metadata using JATS 1.1d3. There is no equivalent to the CrossRef applies_to attribute; specific_use is a suggestion as to how to encode that in JATS 1.1d3.

```
<license>
  <ali:license_ref xmlns:ali="http://www.niso.org/schemas/ali/1.0"
          specific-use="am" start_date="2015-07-01">
     http://creativecommons.org/licenses/by/4.0/
  </ali:license_ref>
  …
<license>
```

JATS 1.0 encoded licenses URLs as attributes and did not support somewhere to record the start date as a property. Using @specific-use to capture the CrossRef applies-to information is a suggestion.

EXAMPLE 7. License metadata in JATS 1.0

```
<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/"/>
```

The examples above do not include the <free_to_read> element in NISO ALI and CrossRef because using that element is not a CHORUS requirement. For more information about <free_to_read> see the websites of CrossRef, NISO ALI, and NISO JATS.