

How to apply CrossMark and FundRef via CrossRef extensible markup language

Rachael Lammey

CrossRef, Oxford, United Kingdom

Abstract

CrossRef is an association of scholarly publishers that develops shared infrastructure to support more effective scholarly communications. In the 14 years since CrossRef launched, it has developed a number of services that hinge around the Digital Object Identifier (DOI) and the publication metadata submitted to CrossRef by publishers registering DOIs. CrossMark and FundRef are two additional services from CrossRef that publishers can choose to participate in, and both require publishers to provide additional information with the bibliographic metadata provided within their CrossRef deposits. This article will aim to show what this additional metadata is, how it should be formatted and why, so that publishers can use it to enhance their content and participate in additional CrossRef services that benefit the scholarly communications industry.

Keywords

Bibliographic metadata; CrossMark; CrossRef; Digital Object Identifier; FundRef

Introduction

It is important to explain the function of the CrossMark and FundRef services from CrossRef, before looking at the extensible markup language (XML) deposit details. CrossMark [1] was launched in April 2012 and sends a signal to researchers that publishers are committed to maintaining their scholarly content. It gives scholars the information they need to verify that they are using the most recent and reliable versions of a document. Readers simply click on the CrossMark logos on portable document format (PDF) or hypertext markup language (HTML) documents, and a status box tells them if the document is current or if updates are available. The CrossMark dialogue box can also be used to convey additional publication record information on things like the peer review process, copyright and licensing information and funder information by way of FundRef.

FundRef provides a standard way to report funding sources for published scholarly research. Publishers deposit funding information from articles and other content using a standard taxonomy of funder names [2]. This funding data is then made publicly available through CrossRef's

Received: May 23, 2014

Accepted: July 29, 2014

Correspondence to Rachael Lammey

rlammey@crossref.org

ORCID

Rachael Lammey

<http://orcid.org/0000-0001-5800-1434>

This article was presented at the workshop entitled 'Application of advanced information technology to scholarly journal publishing' during the 12th EASE General Assembly and Conference, Split, Croatia, July 12-13, 2014.

search interfaces and application programming interfaces (APIs) for funders and other interested parties to analyse. FundRef metadata can be deposited as part of the CrossMark metadata (which is preferred) or separately for publishers who have not yet implemented CrossMark.

Adding CrossMark Metadata to CrossRef XML

CrossMark metadata can be deposited as part of a regular CrossRef Digital Object Identifier (DOI) deposit for new deposits. For backfiles, publishers can provide a resource-only deposit which means that publishers only have to deposit the CrossMark metadata plus the DOI it relates to in order to populate the record for those deposits. This is often preferable for publishers as it means that they do not have to redeposit the complete record and associated metadata and it puts less weight on the CrossRef system to process the files in an efficient manner.

The CrossMark-specific section of the CrossRef XML needs to contain, at a minimum, the DOI of the content the CrossMark is being applied to, the DOI for the publisher's CrossMark Policy Page and the DOI of any piece of content that is being updated. An example of this section of the CrossRef metadata with the minimum information necessary is as follows:

```
<crossmark>
<crossmark_version>1</crossmark_version>
<crossmark_policy>10.5555/crossmark_policy</crossmark_policy>
<crossmark_domains>
<crossmark_domain>
<domain>psychoceramics.labs.crossref.org</domain>
</crossmark_domain>
</crossmark_domains>
<crossmark_domain_exclusive>true</crossmark_domain_exclusive>
</crossmark>
```

This section is taken from the Annotated Example Site [3] that CrossRef has created to show publishers how to apply CrossMark metadata in various scenarios. In standard, full DOI deposits, the CrossMark section sits above the <doi_data> section.

The parts of the deposit can be described as follows. The CrossMark version should be kept as '1' as standard. This is not tied to any article versioning that the publisher may implement but is the version of the CrossMark metadata being used. The CrossMark policy section should be the DOI of the CrossMark policy page that the publisher has created. When a publisher is implementing CrossMark, they must create a

CrossMark policy page on their own website. This page should be assigned a DOI and registered with CrossRef.

A publisher can choose to have one CrossMark policy page for all of their titles, or separate policy pages for each title if your policies and guidelines vary between publications. What the page should do is describe what CrossMark is and link to the publisher's correction and retraction policies (Fig. 1). Sample wording is available on the CrossMark support site at: <http://crossmarksupport.crossref.org/crossmark-policy-page/> and publishers are free to use this to help decide on their own wording.

CrossMark policy pages should be assigned a DOI so that they can be linked to permanently from the CrossMark dialogue box. Assigning a DOI to the page means that if a publisher updates the URL of their policy page, they only have to update the DOI of the policy page and all CrossMark dialogue boxes will update to point to that. CrossMark Policy pages should be deposited as datasets with a "database" called "PublisherName CrossMark Policy Statement". If a publisher has multiple policy pages (for example different policy pages for different journals) they should include them in the database deposit as multiple datasets. An example deposit file for the CrossMark policy page is available here: http://crossmarksupport.crossref.org/wp-content/uploads/2013/04/CrossMark_policy_page.xml [4].

The final section shown in the snippet of CrossMark XML is the CrossMark domain. A CrossMark domain is a domain name/website address (URL) where the content associated with the CrossMark metadata is hosted and maintained by the publisher. The CrossMark domain is an optional field in CrossMark metadata only to be supplied if a publisher wishes their content to be "domain exclusive".

When a user clicks on the CrossMark logo, the system will check whether the content is located on one of the specified CrossMark domains, and it will display the appropriate status message. If the content is on the publisher's site or a third party site that the publisher has committed to updating (a CrossMark domain), the known CrossMark status will appear in the status window in the CrossMark dialogue box. If the content is on a non-CrossMark domain the CrossMark status message will have a statement that the particular copy of the content is not being maintained by the publisher and may not be up to date, and will advise that the user should follow the CrossRef DOI link to the publisher-maintained copy. If a publisher does not supply any CrossMark domains, a researcher will see the latest status in the CrossMark dialogue box regardless of where they have located the content.

Publishers do not have to supply CrossMark domains. If a publisher's content is likely to legitimately appear in many places on different websites then they can choose to be "non

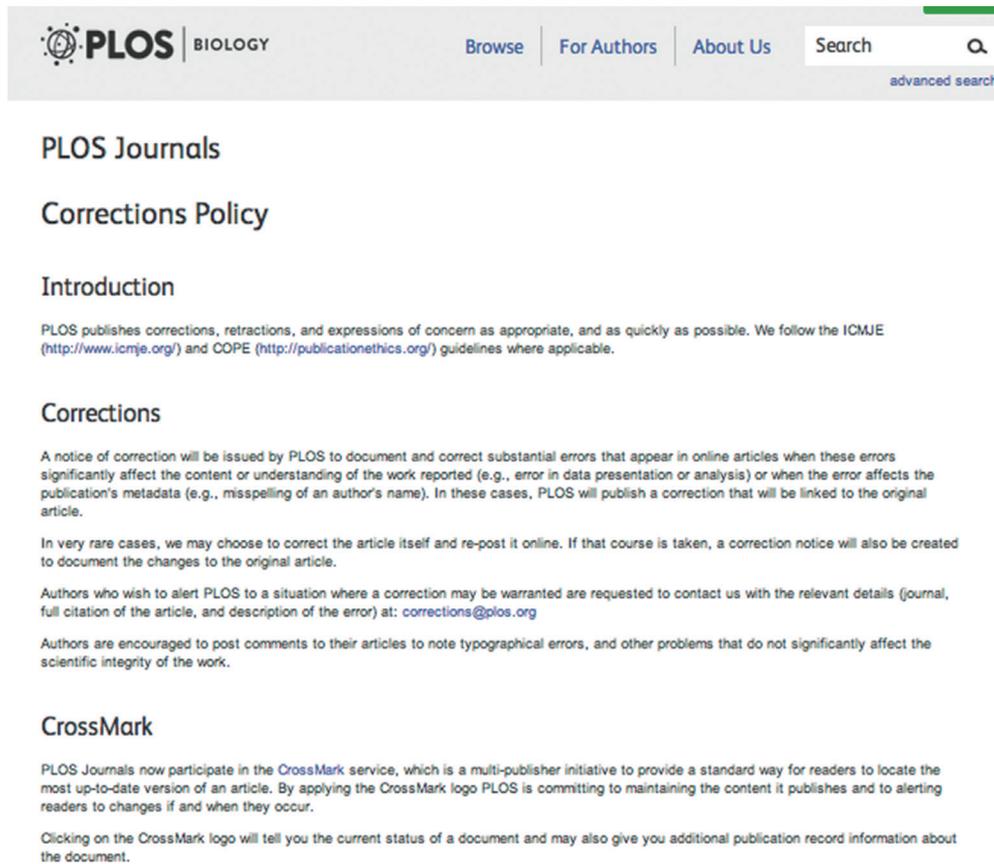


Fig. 1. The *PLOS Biology* CrossMark policy page (Available from: http://dx.doi.org/10.1371/journal.pbio.corrections_policy).

domain-exclusive” and not register any CrossMark domains. However, in the deposit schema, the domain is a required element. The reason CrossRef requires at least one CrossMark domain (even if the publisher marks it as not being domain exclusive i.e., `<crossmark_domain_exclusive> false </crossmark_domain_exclusive>`) is so that search engines can prioritise search results and point to the originating domain of the document.

Publishers entering multiple domains can enter them as follows:

```
<crossmark_domains >
<crossmark_domain >
<domain > domain 1 </domain >
</crossmark_domain >
<crossmark_domain >
<domain > domain 2 </domain >
</crossmark_domain >
<crossmark_domain >
<domain > domain 3 </domain >
</crossmark_domain >
</crossmark_domains >
```

The final element necessary for a CrossMark deposit is only required if the deposit is correcting or updating another piece of content.

An update section can be added as follows:

```
<crossmark >
<crossmark_policy > 10.5555/crossmark_policy </crossmark_policy >
<crossmark_domains >
<crossmark_domain >
<domain > psychoceramics.labs.crossref.org </domain >
</crossmark_domain >
</crossmark_domains >
<crossmark_domain_exclusive > true </crossmark_domain_exclusive >
<updates >
<update type = “retraction” label = “Retraction” date =
“2009-09-14” > 10.5555/12345678 </update >
</updates >
</crossmark >
```

The section highlighted shows what type of update is being applied (a retraction), the date that retraction was issued, and

the DOI of the piece of content is being updated by the retraction.

There are many “types” of updates. “Corrections”, “clarifications”, “retractions” and “withdrawals” are just a few of the better-known types. For these common types, CrossRef recommends that publishers use the values “correction,” “clarification,” “retraction” and “withdrawal” respectively.

The Terms & Conditions [5] for CrossMark stipulate that “updates” should only be deposited for changes that are likely to effect “the interpretation or crediting of the work.” In other words, updates should only be deposited for editorially significant changes. Updates should not be deposited for minor changes such as spelling corrections, formatting changes, etc.

When a correction is made in situ (i.e., replaces the earlier version completely), then the DOI of the corrected content will be the same as the DOI for the original CrossRef deposit. Examples of in situ corrections, and other more complex updates such as a correction that updates more than one DOI are illustrated, along with their XML on the annotated example site.

Along with the basic CrossMark metadata, publishers can also provide additional publication metadata that, if deposited, will appear in the CrossMark record tab as shown in Fig. 2.

Publishers are encouraged to provide any non-bibliographical metadata that they feel might help the researcher evaluate and make better use of the content that the CrossMark record refers to. For example, publishers might want to provide funding information, clinical trial numbers, information about the peer-review process or a summary of the publication history of the document. This particular data from another of our pilot participants, the IUCr (International

Union of Crystallography), and they are sharing some really useful information on the copyright, review process and publication history.

To add this metadata, CrossMark participants should add a custom metadata snippet of XML below the CrossMark domain entry (but above the </crossmark> enclosing tag).

```
<custom_metadata >
<assertion name = “orcid” label = “ORCID” group_name =
“identifiers” group_label = “Identifiers” order = “0”
href = “http://orcid.org/0000-0002-1825-0097” > http://orcid.org/0000-0002-1825-0097 </assertion >
<assertion name = “received” label = “Received” group_name =
“publication_history” group_label = “Publication History” order = “0” > 2012-07-24 </assertion >
<assertion name = “accepted” label = “Accepted” group_name =
“publication_history” group_label = “Publication History” order = “s1” > 2012-08-29 </assertion >
<assertion name = “published” label = “Published” group_name =
“publication_history” group_label = “Publication History” order = “2” > 2012-09-10 </assertion >
</custom_metadata >
```

To explain the metadata displayed here, it begins with an assertion, which is a piece of custom, non-bibliographic metadata that the publisher is stating about the content to which the CrossMark refers.

The assertion attributes that can be used are as follows.

explanation: If the publisher wants to provide a further explanation of what the particular “assertion” means, they can link to such an explanation by providing an appropriate URL on the “explanation” attribute.

group_label: This is the human-readable form of the “group_name” attribute. This is what will be displayed in the group headings on the CrossMark metadata record dialog. For example, ‘Publication History’ in Fig. 2 is a group label.

group_name: Some assertions could be logically “grouped” together in the CrossMark dialog. For instance, if the publisher is recording several pieces of metadata related to publication timelines, then they may want to make sure that these assertions are grouped next to each-other in the CrossMark dialog. The group_name attribute is used for grouping such assertions.

label: This is the human-readable version of the name attribute which will be displayed in the CrossMark dialog. If this attribute is missing, then the value of the assertion will not be displayed in the dialog. Publishers may want to “hide” assertions this way in cases where the assertion value is too large or too complex to display in the dialog, but where the assertion is nonetheless valuable enough to include in API queries and metadata dumps (e.g., detailed licensing terms).

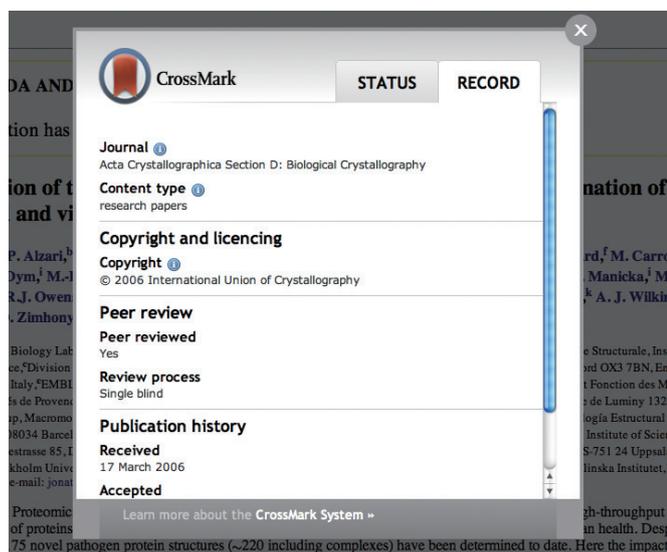


Fig. 2. The CrossMark record tab showing additional publication record information.

name: This is the machine-readable name of the assertion. Publishers can use the “label” attribute to provide a human-readable version of the name.

order: The publisher may want to control the order in which assertions are displayed to the user in the CrossMark dialog.

All assertions will be sorted by this element if it is present, but it is not required. As explained, this information is added at the publisher’s discretion and contain any information they deem relevant.

The CrossMark record tab can also be used to display funder information for the piece of content, by way of FundRef. If a publisher submits funding metadata as per the FundRef requirements, the information will display with the FundRef logo as seen in Fig. 3.

Adding FundRef Metadata to CrossRef extensible markup language

CrossRef supports the recording of funding information for a publication via the FundRef program. FundRef defines an open, standard registry of funder names and funder identifiers that can be used in order to increase the accuracy of the funding information recorded. FundRef benefits a host of stakeholders within the research and communications industries. More accurate funding information will aid funding organizations, which will be able to better track the results of their funding policies. It will help authors, in simplifying their sub-

mission process, as they will be able to choose from a list of funding sources and have their paper handled in accordance with funder mandates. Research institutions will be able to track the productivity of their employees and publishers will be able to better analyse the sources of funding for their published content. Finally, it will also benefit readers and the public, by providing greater transparency into the results of Research & Development funding.

CrossRef member publishers should deposit FundRef data as part of a CrossMark record if they already are (or are planning to become) a participant in CrossMark. There are two reasons for this: first, it ensures that the funder metadata is available both in a standard machine-readable format and via a standard interface for readers. Second, it ensures that the funder metadata is made maximally reusable via a CC Zero license waiver. Note that publishers do not need to have implemented CrossMark yet (i.e., displaying CrossMark logos on their HTML pages and on PDFs to deposit Funder metadata via CrossMark).

The example below shows FundRef metadata nested within the CrossMark metadata.

```
<crossmark>
  <crossmark_version> 1 </crossmark_version>
  <crossmark_policy> 10.555/cm_1 </crossmark_policy>
  <crossmark_domains>
    <crossmark_domain> <domain> www.crossref.org
    </domain> </crossmark_domain>
  </crossmark_domains>
  <crossmark_domain_exclusive> true </crossmark_domain_exclusive>
  <updates>
    <update type=“correction” label=“Correction” date=
    “2011-04-10”> 10.555/cm_test2.1 </update>
  </updates>
  <custom_metadata>
    <assertion name=“received” label=“Received” group_
    name=“publication_history” group_label=“publi-
    cation History”> 2011-04-12 </assertion>
    <fr:program name=“fundref”>
      <fr:assertion name=“funder_name”> IMA Funder
        <fr:assertion name=“funder_identifier”> http://
        dx.doi.org/10.13039/xxxxxxx </fr:assertion>
      </fr:assertion>
      <fr:assertion name=“award_number”> HPC-227 </
      fr:assertion>
    </fr:program>
  </custom_metadata>
</crossmark>
```

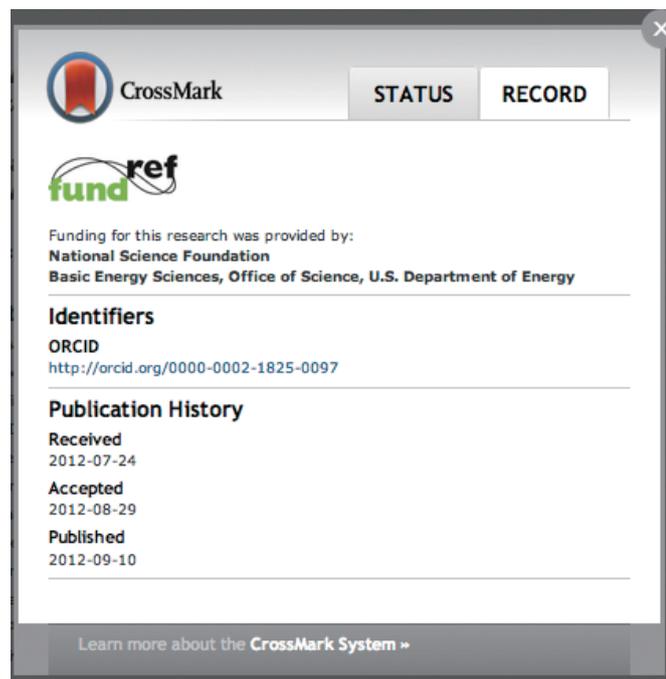


Fig. 3. The CrossMark record tab showing FundRef information.

Publishers can also deposit FundRef metadata as part of their

standard article metadata if they choose not to participate in CrossMark as below:

```

...
<publication_date media_type="print">
  <year>2011</year>
</publication_date>
<pages>
  <first_page>15</first_page>
</pages>
<fr:program name="fundref">
  <fr:assertion name="funder_name">ABC Inc.
  <fr:assertion name="funder_identifier">http://dx.doi.
    org/10.13039/xxxxxxxxxx</fr:assertion>
  </fr:assertion>
  <fr:assertion name="award_number">BXDFSDDS</
    fr:assertion>
</fr:program>
<doi_data>
  <doi>10.5555/cm_test_1.1</doi>
  <resource>http://www.crossref.org/crossmark/index.
    html</resource>
</doi_data>
</journal_article>

```

FundRef metadata must include the name of the funding organization and the funder identifier, where the funding organization is listed in the Registry, and may include an award/grant number assigned to the fund. Funder names should only be deposited without the accompanying ID when the funder is not found in the Registry. Although FundRef supports recording award numbers along with funder identifiers, FundRef does not define standards for recording award numbers as practice varies so greatly across funders.

The metadata attributes associated with FundRef are as follows.

fundgroup: This is used to group funding information for items with multiple funding sources. It is required for items with multiple *funder_name* or *funder_identifier* assertions.

funder_name: The name of the funding agency as it appears in the FundRef Registry. Funder names that do not match those in the registry will be accepted to cover instances where the funding organization is not listed.

funder_identifier: The funding agency identifier in the form of a DOI, must be nested within the “*funder_name*” assertion. The *funder_identifier* must be taken from the FundRef Registry.

award_number: The grant number or other fund identifier.

“*funder_name*” and “*funder_identifier*” must be present in a deposit where the funding body is listed in the FundRef Registry. Multiple “*funder_name*”, “*funder_identifier*”, and “*award_number*” assertions can be included.

When publishers are depositing FundRef metadata, they should add the additional element `<fr:program>` to the schema declaration at the top of their CrossRef XML deposits which is shown below:

```

<doi_batch xmlns="http://www.crossref.org/schema/4.3.1"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-in-
  stance"xmlns:fr="http://www.crossref.org/fundref.xsd"
  xsi:schemaLocation="http://www.crossref.org/schema/
  4.3.1 http://www.crossref.org/depositSchema/crossref
  4.3.1.xsd" version="4.3.1">

```

This basically means that the schema will support the FundRef program and therefore the language it adds to the CrossRef XML and can check it to make sure it's compliant before accepting the deposit. FundRef metadata is collected in this way to try to ensure the accuracy and consistency of the data.

From the previous examples, publishers should note that in the CrossRef XML, the relationship between “*funder_identifier*” and “*funder_name*” is established by nesting “*funder_identifier*” within “*funder_name*”. For example, the deposit excerpt shown below has the funder “National Science Foundation” with its corresponding funder identifier in the FundRef registry of “`http://dx.doi.org/10.13039.10000001`”:

```

<fr:assertion name="funder_name">National Science
  Foundation
  <fr:assertion name="funder_identifier">http://dx.doi.
    org/10.13039.10000001</fr:assertion>
</fr:assertion>

```

A “*funder_name*”, “*funder_identifier*”, and “*award_number*” should be included in FundRef deposits whenever possible. If the funder name cannot be matched in the registry, a publisher may submit *funder_name* only and the funding body will be reviewed and considered for addition to the official registry. Until it is added to the Registry the deposit will not appear in search results in FundRef search [6].

Obviously research is not always as simple as one funder for one paper, and the CrossRef schema has some flexibility that allows for this. If a publisher needs to add two grant numbers for one funding organisation, they can do so as follows within the CrossRef schema:

```

<fr:program name="fundref">
  <fr:assertion name="funder_name">National Science
    Foundation
  <fr:assertion name="funder_identifier">http://dx.doi.
    org/10.13039/100000001</fr:assertion>
</fr:assertion>
  <fr:assertion name="award_number">CBET-106</
    fr:assertion>
  <fr:assertion name="award_number">CBET-7259</

```

```
fr:assertion >
</fr:program >
```

Note that the `funder_identifier` is nested within the `funder_name` assertion, establishing “<http://dx.doi.org/10.13039.100000001>” as the FundRef identifier for funder name “National Science Foundation.” Two award numbers are present.

Multiple funding organisations and grant numbers may also be deposited as per the example shown below:

```
<fr:program name = “fundref” >
  <fr:assertion name = “fundgroup” >
    <fr:assertion name = “funder_name” > National Science
      Foundation
    <fr:assertion name = “funder_identifier” > http://dx.
        doi.org/10.13039/000000001 </fr:assertion >
    </fr:assertion >
    <fr:assertion name = “award_number” > CBET-106 </
      fr:assertion >
    <fr:assertion name = “award_number” > CBET-7259 </
      fr:assertion >
    </fr:assertion >
    <fr:assertion name = “fundgroup” >
    <fr:assertion name = “funder_name” > Basic Energy
      Sciences, Office of Science, U.S. Department of Energy
    <fr:assertion name = “funder_identifier” > http://
        dx.doi.org/10.13039/100006151 </fr:assertion >
    </fr:assertion >
    <fr:assertion name = “award_number” > 1245-ABDS </
      fr:assertion >
    <fr:assertion name = “award_number” > 98562-POIUB
      </fr:assertion >
    </fr:assertion >
  </fr:program >
```

This example contains two `funder_name/identifiers` and two `award_numbers` for each funder. Each funding organization is within its own “`fundgroup`” for clarity for the reader.

If a publisher has more complicated funder information to submit, they can contact support@crossref.org to get help on how to format their XML accordingly.

Conclusion

CrossMark and FundRef are services offered by CrossRef that provide important information to readers, funding bodies, librarians and other important stakeholders within the communications industry. This article has aimed to provide information for publishers and editors on how they should add CrossMark and FundRef data to their CrossRef XML in order to participate in both initiatives. Further information is available via CrossRef’s help documentation, and CrossRef staff are on-hand to help with any queries that may arise.

Conflict of Interest

No potential conflict of interest relevant to this article was reported.

References

1. CrossRef. CrossMark [Internet]. Lynnfield: CrossRef; 2014 [cited 2014 May 15]. Available from: <http://www.crossref.org/crossmark/>
2. CrossRef. Found at the FundRef section of the CrossRef website [Internet]. Lynnfield: CrossRef; 2014 [cited 2014 May 15]. Available from: http://www.crossref.org/fundref/fundref_registry.html/
3. CrossRef. CrossMark annotated example site [Internet]. Lynnfield: CrossRef; 2014 [cited 2014 May 16]. Available from: <http://crossmarksupport.crossref.org/annotated-example-site/>
4. CrossRef. CrossMark policy page [Internet]. Lynnfield: CrossRef; 2014 [cited 2014 May 16]. Available from: <http://crossmarksupport.crossref.org/crossmark-policy-page/>
5. CrossRef. CrossMark terms and conditions [Internet]. Lynnfield: CrossRef; 2014 [cited 2014 May 16]. Available from: <http://www.crossref.org/crossmark/AboutTermsConditions.htm/>
6. CrossRef. FundRef search [Internet]. Lynnfield: CrossRef; 2014 [cited 2014 May 17]. Available from: <http://search.crossref.org/fundref/>