UKSG

# FundRef: connecting research funding to published outcomes

Funding sources for research are usually acknowledged by authors in their published papers, but until now there has been no practical way to collate this information from multiple publications. With funding bodies under increasing pressure to report on research outcomes and a growing number of national initiatives seeking to provide appropriate access to this research, a central solution has become necessary. The FundRef initiative from CrossRef meets this need by collecting standardized research funding data from published works and making it available for anyone to search and analyse.

*"To assist funders to effectively track the outputs of research to which they have contributed, either wholly or in part, it is important that investigators clearly acknowledge all relevant funders in research publications."*

Wellcome Trust[1]

KIRSTY MEDDINGS
Product Manager
CrossRef

With the awarding of a research grant come a number of obligations on the part of the researcher. In addition to reporting the research outcomes back to the funding organization, the researcher is expected to acknowledge the source of his/her funding in any publications that may result from the research. Funders often have guidelines to encourage and assist authors to comply with these obligations: UK research funders, for example, follow the recommendations issued by the Research Information Network in 2008[2], which offer a template for naming the funder(s) and associated grant numbers. While guidelines such as these offer a level of standardization, they are not always followed, and funding acknowledgements can take many forms, sometimes omitting grant numbers and even sometimes being omitted completely. Even when authors are fulfilling their obligations accurately, the information itself is usually at the end of the article, separate from the metadata, and soon fades from view; it is not collected, collated or shared beyond the article itself. Meanwhile, interest in this data is growing: funders are under greater pressure to account for their spending, and there are increasing moves to identify the published outcomes of publicly funded research with a view to making those outcomes available to the public.

There are a number of factors that contribute to the lack of visibility regarding funding acknowledgements in published content, and it is worth looking into each one in some detail in order to appreciate the problem.

If the assumption is that authors are complying with funders' requirements for them to acknowledge funding in publications, then there is a reasonable starting point: the information is available, and the reader is likely to be able to find it when browsing an article. The problem comes when someone wants to see all papers that had research funding from a particular source. This requires machine-readable information. To date, few publishers have been extracting funding data from papers and storing it as part of the article's structured metadata, which means it is difficult to do a fielded search on 'funder' as one might for 'author name' or 'publication title'. An attempt to look for it using a full-text search is likely to return results littered with other irrelevant instances of the search terms.

In cases where funding information is tagged up in a publication's XML there are still two substantial problems. The first is one of name ambiguity: 'NIH', 'N.I.H.', 'National Institutes

*"… funders are under greater pressure to account for their spending …"*

of Health', 'National Institute of Health' and 'US National Institutes of Health' are all very likely to refer to the same funding body. However, the variations, when mis-spellings are taken into account, are potentially very wide. Context is also important: there is a 'National Science Foundation' in more than one country, for example.

Secondly, the tags used to house the funding metadata may vary from publisher to publisher, making any kind of cross-platform search – unwieldy at the best of times – even more difficult. The National Information Standards Organization's (NISO's) Journal Article Tag Suite (JATS)[3], a recent successor to the NLM DTD, does include tags for funding information but is not yet universally used and does not extend to other content types such as books.

To make research funding information in publications accessible, it needs to be presented in a standard way and stored in a central location. This was the fundamental surmise of a presentation given by H Frederick Dylla, CEO of the American Institute of Physics, at the CrossRef Annual Meeting in November 2010. Following on from the recommendations in the report from US Scholarly Publishing Roundtable earlier that same year[4], Dylla's presentation 'Standardizing Funding Information in Scholarly Journal Articles'[5] proposed adding funding information to the standard CrossRef metadata fields so that publishers could deposit this data with CrossRef where it could be consolidated and made available to funders and other interested parties.

## FundRef pilot

After some further discussion amongst the CrossRef Board of Directors[6], it was agreed to pilot the FundRef project in early 2012. A working group was set up, comprising representatives not just from scholarly publishers, but also from funding bodies. This was a first for CrossRef: as a publisher membership organization, CrossRef's projects have always been collaborative in nature, but across publishers of different sizes, subjects and business models. The nature of the FundRef project required expert input from the funding bodies it was seeking to serve, and so volunteers from NASA, the Wellcome Trust, the US Department of Energy and the US National Science Foundation joined those from the American Institute of Physics, the American Psychological Association, Elsevier, the Institute of Electrical and Electronics Engineers, Nature Publishing Group, Oxford University Press and Wiley to work out the detail.

One of the key requirements the working group initially identified was the need for a controlled vocabulary and taxonomy of funding body names. If publishers simply submitted the information as supplied by authors in their manuscripts, the name ambiguity issue discussed above would render the system fairly ineffective unless a complex de-duplication and matching task was undertaken. The names of the funding bodies would have to be matched against an authoritative list before deposit with CrossRef, and for the pilot project Elsevier volunteered a funding body registry of over 4,000 names developed as part of its SciVal Funding product. The funders in the pilot group reviewed the list and provided feedback on the accuracy of how their organizations were listed, and the registry was updated accordingly.

"… the need for a controlled vocabulary and taxonomy of funding body names …"

With a taxonomy in place, the group mapped out the workflow that would capture and process the necessary funding information (see Figure 1). Using the registry of funder names, publishers would ask authors to select the relevant funding bodies and input associated grant numbers at the time of manuscript submission. This metadata would pass through publishers' production systems and upon publication be deposited with CrossRef using the CrossRef deposit schema. CrossRef would then make this information available for anyone to search or browse, and would also feed it back into publishers' websites via the CrossMark[7] service where appropriate.
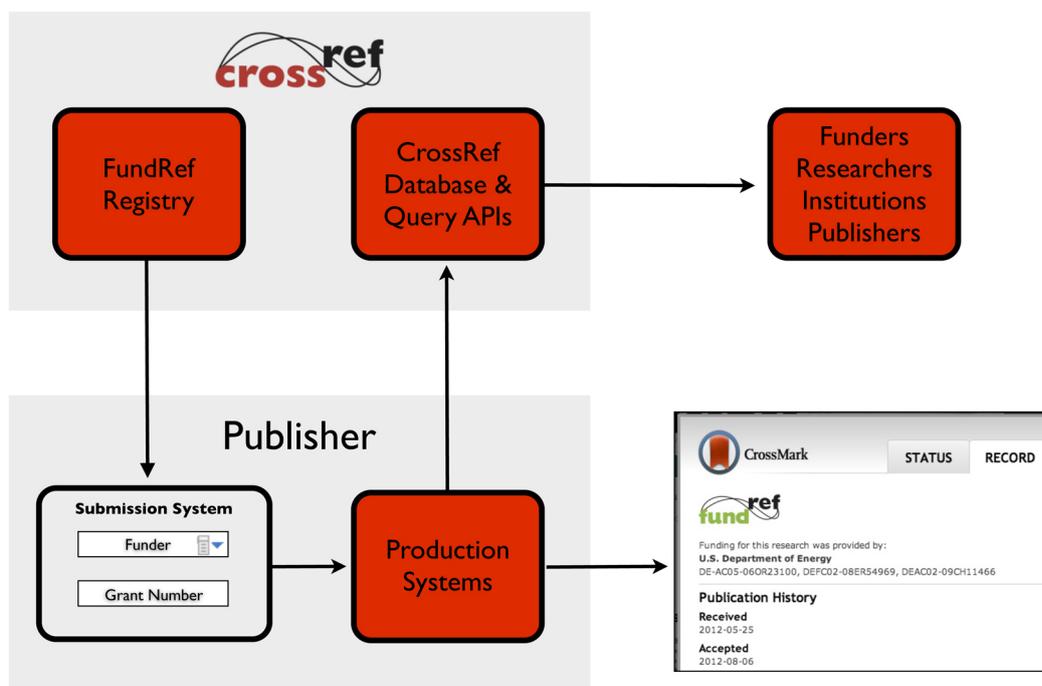
Figure 1. The FundRef workflow

This process was tested with a handful of journals from each of the participating publishers, and with the assistance of the manuscript tracking system vendors eJournalPress, Aries Systems and ScholarOne. All three successfully used the registry to normalize the names of funding bodies submitted by authors, and a sub-group of the FundRef working group went on to specify the process through which CrossRef would host and maintain this registry for use in the FundRef project. The pilot group assessed the idea of also verifying grant numbers, but cataloguing the many variations of numbers from so many funders was deemed too complicated for the first phase of the project, although it is something that may be revisited at a future date.

The FundRef pilot ran until early 2013 and in March 2013 the CrossRef Board reviewed the pilot project report[8] and approved the project to go into production. FundRef was officially launched on 28 May 2013 when it was opened up to receive deposits from CrossRef member publishers.

## The FundRef Registry

Based on the SciVal Funders list donated by Elsevier, the FundRef Registry[9] is hosted by CrossRef and freely available under a Creative Commons CC0 licence waiver for anybody to browse, download, or use as they wish. The Registry is presented as an RDF file containing funding body names, with a unique ID number for each organization in the form of a funder DOI, and some additional associated metadata such as country, funder type and alternative names or acronyms. The FundRef Registry also contains hierarchical relationships of funding bodies where they exist. The Registry will be updated and expanded in two ways: funders themselves are invited to send feedback to CrossRef on any missing funding organizations or changes that need to be incorporated[10]; additionally, whenever publishers deposit records containing funders that are not listed in the Registry, these names will be assessed, curated and added to the Registry on a monthly basis.

"… the FundRef Registry is … freely available …"

## Publisher workflow guidelines

The pilot tested integration of the FundRef Registry with manuscript submission systems, and the vendors of these systems have integrated or are in the process of integrating the

FundRef Registry so that authors are guided to choose the canonical funder name and to submit grant numbers. Since the launch, several publishers have indicated that they plan to extract the funding information automatically from manuscripts rather than ask the author to enter it into a form, with the authors validating the information at the time of acceptance. Either of these approaches is acceptable; the critical piece is that wherever a funding organization exists in the FundRef Registry, this name is used. Several publishers have already extracted funding data for back-file content and are in the process of matching this data against the FundRef Registry names for deposit with FundRef.

## Changes to the CrossRef deposit schema

The CrossRef metadata deposit schema[11] has been updated to include three new elements: 'funder_name', 'funder_identifier' and 'award_number'. Publishers will use these elements to deposit funding metadata with the rest of their CrossRef metadata at the time of publication. Further detail on how to deposit FundRef metadata can be found on the CrossRef Support site[12].

## FundRef Search

All FundRef funding metadata deposited by publishers is freely available for anyone to search, browse, export and analyse. It can be accessed through CrossRef's various search APIs (such as the OpenURL Query Interface used by many libraries), and also through the new FundRef Search form at **http://search.crossref.org/fundref**. FundRef Search is specifically for looking up a funder and retrieving a list of content that cites that organization as a funder of its research. When a user starts to type in the search, a list of matching funder names appears and they choose one from that list. They can then narrow or widen the results based on the funder heirarchies, and export search results. FundRef Search addresses the main use case for FundRef: that a funder (or any other interested party) will want to look up a funding body and retrieve a list of all the publications that have listed it in their acknowledgements.

However, there are myriad other use cases and CrossRef Metadata Search (**http://search.crossref.org**) allows for further exploration of the FundRef data. CrossRef Metadata Search searches all of the metadata fields in the CrossRef database. Entering an article title or DOI will return that article's metadata with funding information if it is available; entering an award number will return the content in which it is cited; entering an ORCID researcher ID[13] will return all of the publications that particular researcher has contributed to, and where funding information has been deposited it will be displayed for each paper. At the time of writing, the amount of FundRef data to have been deposited is very small but, as it grows throughout 2013 and beyond, there will be many ways in which many types of organizations and individuals can make use of FundRef Search and CrossRef Metadata Search for analysis, reporting and more.

"Entering an article title or DOI will return that article's metadata with funding information …"

## Next steps

For 2013, the focus is on publishers depositing funding metadata so that FundRef can quickly reach a critical mass and provide solid, useful data to funders, publishers, institutions and anyone with an interest in the outcomes of research funding.

FundRef is playing a central role in many initiatives concerned with providing public access to government-funded research, such as those currently addressing the issues confronting US Federal Agencies as a result of the Office for Science and Technology Policy Memo on public access.[14] With this in mind, a next step for FundRef will be to add the means for publishers to submit licensing information as part of the metadata indicating a publication's OA status or embargo period, and allowing funders to check that content is available in accordance with their mandates. A possible future enhancement, previously mentioned, is

a means by which to verify grant numbers to ensure that authors are correctly citing their grants, and we have had several suggestions for additional funder metadata that could be added to the FundRef Registry. All of this demonstrates that there is plenty of scope to develop FundRef; we think of this as FundRef+.

The FundRef pilot and launch are also excellent examples of how collaboration between different stakeholders can result in a solution that benefits all. This collaboration will continue, with FundRef's future development being guided by an Advisory Group made up of publishers and funders. This group will guide future service enhancements but will also ensure that FundRef retains its focus, as the success of the pilot was in a large part due to careful specification and limiting the initial release to the core features required to create a uniquely useful service for funders, publishers and institutions.

"… collaboration between different stakeholders can result in a solution that benefits all."

References and notes

1. Wellcome Trust:
   http://www.wellcome.ac.uk/managing-a-grant/end-of-a-grant/wtd037950.htm (accessed 24 July 2013).

2. Research Information Network:
   http://www.rin.ac.uk/our-work/research-funding-policy-and-guidance/acknowledgement-funders-journal-articles (accessed 24 July 2013).

3. NISO:
   http://www.niso.org/workrooms/journalmarkup (accessed 24 July 2013).

4. Association of American Universities:
   http://www.aau.edu/policy/scholarly_publishing_roundtable.aspx?id=6894 (accessed 24 July 2013).

5. SlideShare:
   http://www.slideshare.net/CrossRef/dylla-cross-refannual-general-mtg-nov2010 (accessed 24 July 2013).

6. CrossRef:
   http://www.crossref.org/01company/05board.html (accessed 24 July 2013).

7. CrossRef:
   http://www.crossref.org/crossmark/ (accessed 24 July 2013).

8. CrossRef:
   http://www.crossref.org/fundref/FundRef_Pilot_Project_Report.pdf (accessed 24 July 2013).

9. FundRef Registry:
   http://dx.doi.org/10.13039/fundref_registry (RDF file).

10. Feedback on the FundRef Registry can be sent to
    fundref@crossref.org

11. CrossRef:
    http://help.crossref.org/#deposit_schema (accessed 30 July 2013).

12. CrossRef Support:
    http://help.crossref.org/#fundref (accessed 30 July 2013).

13. ORCID:
    http://www.orcid.org (accessed 3 September 2013).

14. Office of Science and Technology Policy:
    http://www.whitehouse.gov/blog/2013/02/22/expanding-public-access-results-federally-funded-research (accessed 24 July 2013).

**Article © Kirsty Meddings**

Kirsty Meddings, Product Manager
CrossRef
Tel: +44 (0)1865 722644 | E-mail: kmeddings@crossref.org | skype: kirstymeddings web: www.crossref.org |
Twitter: @CrossRefNews