Using Crossref metadata to track Open Access (and more)

OpenAIRE Workshop, Oslo, Feb 2017

Rachael Lammey
@CrossrefOrg
We are Crossref, a not-for-profit membership organization for scholarly publishing, working to make content easy to find, cite, link and assess. We do it in five ways: rallying the community; tagging metadata; running a shared infrastructure; playing with new technology; and making tools and services to improve research communications.

It’s as simple—and as complicated—as that.
Crossref Overview

- 5500+ publisher members, 1 member 1 vote
- Metadata store of over 85 million scholarly content items (including preprints)
- Persistent citation linking
- Funder identifiers
- Report and display corrections & retractions
- Check manuscripts for similarities
- Data about content activity e.g. social, dataset links
- Open REST API & Search
infrastructure |ˈɪnfrəstrʌktʃə|

noun

the basic physical and organizational structures and facilities (e.g. buildings, roads, power supplies) needed for the operation of a society or enterprise. The social and economic infrastructure of a country.
<metadata>
Why do publishers assign DOIs?

Show people where your content is located, and update that if/when the content moves

- Helps indexing
- Linking from other sources is persistent
- Full-text links can be used for text-mining and indexing

More metadata means that people have more ways to find publications
What are we trying to do?

- Institutions want to be able to track publications from their researchers and how much content is being published under OA licenses
- Funders need a way to track outputs from thousands of publishers
- Publishers will need a standard and efficient way to demonstrate conformance mandates
- All the stakeholders in the process (funders, publishers, institutions and researchers) will span disciplines, institutions, geographies and jurisdictions
How can we do it?

- Crossref has extended its metadata schemas and Application Programming Interfaces (APIs) to enable institutions, funders and publishers to use Crossref metadata to:
  - track research that is subject to these mandates
  - ensure that said research is being disseminated according to the requirements of the mandates
Metadata can be used to identify

- Publications relating to research supported by particular funders.
- Publications from particular researchers identified by their ORCID ID
- Bibliographic metadata
- License information (including embargoes)
- The location of the full text for reading and text & data mining application
- Long-term preservation arrangements
- Clinical trial information
- Updates (errata, corrigenda, retractions)
Funding and license information

- Standardisation of funder names using the Open Funder Registry
- 14,200 international funders
- Publisher deposit of funding metadata which they collect from authors upon submission (or add it afterwards)
- Facilitates large-scale analysis and reporting to funders
Made available in bulk or via an API

http://api.crossref.org
Search 14,200 funders connected to 1,252,198 published works with funding data

http://search.crossref.org/funding
<table>
<thead>
<tr>
<th>Title</th>
<th>Journal Article</th>
<th>Published Date</th>
<th>Volume</th>
<th>Research Funded By</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beyond the Medial Regions of Prefrontal Cortex in the Regulation of Fear and Anxiety</td>
<td>Frontiers in Systems Neuroscience</td>
<td>22 Feb 2016</td>
<td>10</td>
<td>Medical Research Council</td>
<td>Yoshiro Shiba, Andrea M. Santangelo, Angela C. Roberts</td>
</tr>
<tr>
<td>Sepsis Patients with First and Second-Hit Infections Show Different Outcomes Depending on the Causative Organism</td>
<td>Frontiers in Microbiology</td>
<td>26 Feb 2016</td>
<td>7</td>
<td>Medical Research Council</td>
<td>-</td>
</tr>
</tbody>
</table>
License information

- Information available as an overview or at the work level

```json
{
    status: "ok",
    message-type: "work-list",
    message-version: "1.0.0",
    message: {
        query: {
            search-terms: null,
            start-index: 0
        },
        items-per-page: 0,
        items: [],
        total-results: 216394,
        facets: {}
    }
}
```
At the work level
License information

- Can convey embargo & archiving information:

```xml
</fr:assertion>
</fr:program>
<ai:program name="AccessIndicators">
  <ai:license_ref applies_to="vor" start_date="2014-05-02">
    http://publisher.com/licenses/license_v1.html
  </ai:license_ref>
  <ai:license_ref applies_to="am">
    http://publisher.com/licenses/open_license_v1.html
  </ai:license_ref>
  <ai:license_ref applies_to="tdm" start_date="2014-12-02">
    http://publisher.com/licenses/tdm_license_v1.html
  </ai:license_ref>
</ai:program>
<archive_locations>
  <archive name="CLOCKSS"/>
</archive_locations>
```
ORCID iDs

• Nearly 0.6 million ORCID iDs deposited with Crossref from nearly 850 publishers
• Enables auto-update workflow
• Enables accurate mapping between authors and publications
Measuring impact

• Citations
  - Publishers can make these openly available via Crossref

• Crossref Event Data
  - Our data is open, auditable, transferrable, and portable
  - Regular updates from data contributors and reliable, secure storage
  - Raw data, free of charge with the addition of an SLA provided at favorable rates
  - A community-led initiative and data from a neutral source
  - https://www.crossref.org/services/event-data/
Crossref Event Data

Every 'thing that happens' is recorded as an individual Event. We gather Events from a wide range of sources, but examples include:

- an article was linked from DataCite dataset via its Crossref DOI
- an article was referenced in Wikipedia using its Crossref DOI
- an article was mentioned on Twitter using its Article Landing Page URL
- an article has been liked on Facebook 55 times as of June 15th
Also:

- Free to read tag
- Work on organisational identifiers (the OI project): [http://blog.crossref.org/2016/10/the-oi-project.html](http://blog.crossref.org/2016/10/the-oi-project.html)
The Crossref extended APIs, of course, will only work if publishers supply the appropriate metadata and follow the specified metadata deposit workflows.
Thank you

@CrossrefOrg
rlammey@crossref.org
https://crossref.org