Exploiting and improving institutional repositories for bibliometrics

IRIARTE, Pablo, et al.

Abstract
Institutional repositories have a dual role, being both open archives, disseminating without barriers research results financed by public funds, and a directory of institutional scientific publications. Today, a new kind of functionality may be added to these repositories: they allow to evaluate University's scientific production through bibliometric analysis. Indeed, two special types of metadata are included into institutional repositories: bibliographic and administrative information (attribution of publications to each author or research unit from the institution for instance), making them particularly interesting for bibliometric use. It is common to make use of international and subject databases to feed and improve the content of the institutional repository. These mines of bibliographic information are currently used to import references into archives, individually or in batch. Using Web services and AJAX techniques, they are also exploited during data entry: with a simple identifier, bibliographic metadata is injected into a web form or proposed while typing. At the same time, bibliographic metadata is enriched [...]
Exploiting and improving institutional repositories for bibliometrics

Pablo Iriarte, Isabelle de Kaenel, Jan Krause, Nathalie Magnenat
pablo.iriarte@chuv.ch, isabelle.de-kaenel@chuv.ch, jan.krause@chuv.ch, nathalie.magnenat@unil.ch

Medical Library and Research Administration
Unit, Research and Education Department,
University Hospital Center and University of Lausanne, Switzerland

Objectives
Institutional repositories have a dual role, being both open archives, disseminating without barriers research results financed by public funds, and institutional scientific publications directory. Today, a new kind of functionality may be added to these: repositories as main source to evaluate University's scientific production through bibliometric analysis.

Connecting with external resources
Bibliographic metadata is systematically enriched with unique identifiers that connect the archive with external databases allowing metadata import and update. In bibliometrics, identifiers also play a major role: the ISSN makes the inclusion of journal impact indicators easier. Other unique identifiers can be used to import the citations count received for a publication, calculated by different databases and open archives: UT for the Web of Science, PMID for PubMed Central and DOI for Web of Science, SCOPUS and others.

Bibliometric resources explored in Lausanne

<table>
<thead>
<tr>
<th>Tool name</th>
<th>Accuracy</th>
<th>Confidence</th>
<th>Up to date</th>
<th>Unique identifiers</th>
<th>Crosslinking</th>
</tr>
</thead>
<tbody>
<tr>
<td>serval</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>SERVeurAcadémique Lausannois</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

Improving the archive together
The use of data from an institutional repository as a source of bibliometrics analysis has two advantages. On the one hand, this requires researchers, librarians and different units of the university to agree and work together to improve the archive’s metadata and keep it up to date. On the other hand, the effort invested in sorting, selecting and assigning publications to authors and research units may be exploited by the university evaluation services.

External resources

Results
Currently, the data from SERVAL is used to calculate bibliometric indicators for faculty of life sciences researchers and units:
- IF and RPU of journals in which researchers publish
- Total number of citation counts
- $h$-index

Those indicators are weighted by publication type and author position and calculated for the complete publication list and also for the temporal window of the last 6 years.

Conclusion
Managing this new kind of information requires to improve the current system and to make modifications at two levels: upstream, completing and correcting metadata from the archive; and downstream, processing and enriching data with citations information. Indeed, bibliometric indicators are changing rapidly, particularly the number of citations that needs to be updated frequently.

The aim of this project is to create a new flexible system that can be integrated with other institutional repositories. We believe that the use of these data as a source for bibliometrics is an opportunity to combine forces around the archive and fulfill even more important functions within the institution.

Bibliography
1. Hirsch JE. An index to quantify an individual’s scientific research output. PNAS. 2005 November 15, 2005;102(46):16569-72
5. Godel Spring S, De Kaenel I, Iriarte P. Web Services for Bibliometrics. EAHIL 2010 : 12th European Conference of Medical and Health - Lisboa, Portugal

Links
1. CERN workshop on Innovations in Scholarly Communication (OAI7) 22 - 24 June 2011 | University of Geneva

CERN workshop on Innovations in Scholarly Communication (OAI7) 22 - 24 June 2011 | University of Geneva