Open Access discovery

ULiège experience with aggregators and discovery tools providers. Be proactive and apply best practices (if you can...)

Myriam Bastin, University of Liège Library
François Renaville, University of Liège Library

http://hdl.handle.net/2268/221340
• Let’s start with an assertion

Your repository must be visible
Your repository **CONTENT** must be visible

• Why?
  • Don’t expect many people going to your IR and search for items (except maybe your own researchers)...
  • They will use other channels (Google Scholar, MS Academic, OA harvester, discovery tool...)

• ULiège experience illustrated with ORBi ([https://orbi.uliege.be](https://orbi.uliege.be)) and PoPuPS ([https://popups.uliege.be](https://popups.uliege.be))
Agenda

• ULiège experience with:
  • Google Scholar
  • Primo Central Index
  • Summon
  • Ebsco Discovery Service
  • CORE
  • BASE
  • PubMed LinkOut

• Some tips and tools
Agenda

• ULiège experience with:
  • Google Scholar
  • Primo Central Index
  • Summon
  • Ebsco Discovery Service
  • CORE
  • BASE
  • PubMed LinkOut

• Some tips and tools
Google Scholar: Inclusion Guidelines

• **Content**: Two basic criteria

  1. **Scholarly articles**
     - The content hosted on your website must consist primarily of scholarly articles - journal papers, conference papers, technical reports, or their drafts, dissertations, pre-prints, post-prints, or abstracts. Content such as news or magazine articles, book reviews, and editorials is not appropriate for Google Scholar. Documents larger than 5MB, such as books and long dissertations, should be uploaded to Google Book Search; Google Scholar automatically includes scholarly works from Google Book Search.

  2. **Showing abstracts**
     - Users click through to your website to read your articles. To be included, your website must make either the full text of the articles or their complete author-written abstracts freely available and easy to see when users click on your URLs in Google search results. Your website must not require users (or search robots) to sign in, install special software, accept disclaimers, dismiss popup or interstitial advertisements, click on links or buttons, or scroll down the page before they can read the entire abstract of the paper. Sites that show login pages, error pages, or bare bibliographic data without abstracts will not be considered for inclusion and may be removed from Google Scholar.

Google Scholar: Inclusion Guidelines

• **Indexing:**
  
  • **Configuring the meta-tags**
    
    • If you're using repository or journal management software, such as Eprints, DSpace, Digital Commons or OJS, please *configure it to export bibliographic data in HTML "<meta>" tags.*
    
    • Google Scholar supports
      
      • Highwire Press tags (e.g., citation_title)
      • Eprints tags (e.g., eprints.title)
      • BE Press tags (e.g., bepress_citation_title)
      • PRISM tags (e.g., prism.title)
  
  • Other recommendations at
Google Scholar: ORBi

A message from Google Scholar our IT manager is proud of... 😊

Sujet : Re: Interstitial page for orbi.ulg.ac.be full-text PDFs and Google metas citation implementation
Date : Tue, 16 Sep 2014 10:19:44 -0700
De : Darcy Dapra @google.com>
Pour : Fabian Smagghe <f.smagghe@ulg.ac.be>
Copie à : ORBi <orbi@misc.ulg.ac.be>, Paul Thirion <paul.thirion@ulg.ac.be>

Hello Fabian,

Many thanks for your kind note! Regarding the text in the htm1map applied via "display:none" with CSS, you’re right, this can often be an issue, but here should be okay (so no need to change this).

Also, thanks for sending along the example journal article for review--your metatags look fantastic! Yours is the best implementation of machine-readable bibliographic information we have seen in DSpace repositories thus far (and that’s a near direct quote from the Scholar engineers). :-) 

I hope that you don’t mind my using ORBi as an example to other DSpace repositories of how best to implement the metatags for improved indexing in Scholar?
Agenda

• ULiège experience with:
  • Google Scholar
  • **Primo Central Index**
  • Summon
  • Ebsco Discovery Service
  • CORE
  • BASE
  • PubMed LinkOut

• Some tips and tools
Primo Central Index

- Primo Central Index (PCI) = name for Ex Libris’ mega-index of “hundreds of millions of scholarly e-resources of global and regional importance. These include journal articles, e-books, reviews, legal documents and more that are harvested from primary and secondary publishers and aggregators, and from open-access repositories.”
  - records are mainly provided by publishers, who provide the metadata from their publication platforms, and by aggregators (Gale, ProQuest) and their databases.
  - In addition: large collections and archives of free scholarly materials: ArXiv.org, HAL (Hyper Article en Ligne), OAPEN: Open Access Publishing in European Networks, SwePub, Norwegian Open Research Archives.

- Early 2013: Ex Libris started to add institutional repositories into Primo Central Index, see Ex Libris Institutional Repository Program
  https://knowledge.exlibrisgroup.com/Primo_Central/Product_Documentation/Institutional_Repository_Program
Update frequency is not always respected...
Growth of IRs in PCI

Note: There are currently less than 77 IRs in the PCI ‘Institutional Repositories’ section (73). Some IRs have been removed or moved to another section without any notification.
But Ex Libris is still working on OA...

Open Access flags and facet

Customer Benefits

- Easier identification and promotion of open access material

First Half of 2018

- Open Access material will be flagged on the User Interface
- Open Access facets will enable users to filter by such material

PCI: ORBi

• ORBi was added to PCI in March 2013.
• Update Frequency: Weekly

• Too many records!
  • Not only OA records had been harvested and added to PCI, but also records with no full text or with a campus access
  • We asked Ex Libris to run a full reload on OA records only

• Nowadays:
  • More than 57,000 ORBi records with OA content are in PCI and can be activated by other libraries
  • Activated by more than 3,000 PCI customers
 PCI: ORBi in the top 10 of IRs for articles

<table>
<thead>
<tr>
<th>Institutional Repositories</th>
<th>articles</th>
<th>% of the harvested IR content</th>
</tr>
</thead>
<tbody>
<tr>
<td>Repository Utrecht University</td>
<td>38,717</td>
<td>57.9%</td>
</tr>
<tr>
<td>Infoscience (École Polytechnique Fédérale de Lausanne (EPFL))</td>
<td>30,762</td>
<td>100.0%</td>
</tr>
<tr>
<td>Diposit Digital de Documents de la UAB (Universitat Autonoma de Barcelona)</td>
<td>21,189</td>
<td>36.1%</td>
</tr>
<tr>
<td>ORBi (Open Repository and Bibliography) (University of Liège)</td>
<td>19,665</td>
<td>49.2%</td>
</tr>
<tr>
<td>ZORA (University of Zurich)</td>
<td>19,152</td>
<td>70.6%</td>
</tr>
<tr>
<td>National Chung Hsing University Institutional Repository</td>
<td>17,499</td>
<td>26.0%</td>
</tr>
<tr>
<td>National Chung Hsing University, Taiwan</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VU-DARE (VU University Amsterdam)</td>
<td>16,271</td>
<td>66.8%</td>
</tr>
<tr>
<td>DiVA - Academic Archive Online (Uppsala University Library)</td>
<td>15,431</td>
<td>8.5%</td>
</tr>
<tr>
<td>Digital Access to Scholarship at Harvard (DASH) (Harvard University, Office for Scholarly Communication)</td>
<td>15,076</td>
<td>80.9%</td>
</tr>
<tr>
<td>Lirias (KU Leuven Association)</td>
<td>12,736</td>
<td>49.0%</td>
</tr>
</tbody>
</table>

*Figure 6: Top 10 Institutional Repositories in PCI - Articles (March 31, 2015)*

(Renaville, 2016)
OA Repositories and PCI

• 2015 Survey: How do Primo Libraries manage Green Open Access collections?

   Enhancement requests from the respondents:

1. Providing **only OA contents**: Too many false positives: some IRs are not 100% (no full text, embargo, login required) → bad experience and end-users are frustrated
2. Providing **more detailed information** about each OA repository
3. Adding **more OA contents** (more repositories)
4. Requiring OA content providers to provide **good metadata quality**
5. **Better collaboration** between Ex Libris and data providers
6. Allowing richer and more extended **record formats** for harvesting purposes

(Renaville, 2016)
PCI: PoPuPS

- PoPuPS was added to PCI in May 2012
- Update Frequency: Weekly

Nowadays:
- More than 5,600 PoPuPS journal articles

And Knowledge Bases?
- PoPuPS journals are also listed in KBs and journal finders:
  - A-to-Z (Ebsco)
  - Alma Community Zone (Ex Libris ProQuest)
  - SFX KB (Ex Libris ProQuest)
  - 360 KB (Ex Libris ProQuest)

(Pochet, 2017)
<table>
<thead>
<tr>
<th>Title</th>
<th>Authors</th>
<th>Journal</th>
<th>Year</th>
<th>Pages</th>
<th>DOI</th>
<th>Viewed</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agro-environmental scheme for segetal plant conservation in Wallonia (Belgium): an assessment in conventional and organic fields</td>
<td>Cyril Lemarche ; Emmanuel Senusse ; Gregory Melly ; Julien Piquemal</td>
<td>BASE, 18 January 2018, Vol.32(1) [Peer Reviewed Journal]</td>
<td>2018</td>
<td>1370-5229</td>
<td>10.11544/2018</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Agenda

• ULiège experience with:
  • Google Scholar
  • Primo Central Index
  • Summon
  • Ebsco Discovery Service
  • CORE
  • BASE
  • PubMed LinkOut

• Some tips and tools
Summon

- Summon = ProQuest’s discovery tool

**Summon and Open Access**

- Libraries can enhance resource discovery by making freely available Open Access content, outside of their own collection, more accessible and visible through a series of user interface enhancements.
- More than 200 million items identified as Open Access available through Summon
  - **icon** to indicate Open Access content appearing on the Summon results page
  - **filter** for Open Access content
Summon: ORBi

• ORBi was added to Summon in 2017.
• Update Frequency: ?

• Integrated, but not identified as being OA!
  • Summon team will be contacted by ULiège Library to fix this
Information Literacy in Students Entering Higher Education in the French Speaking Community of Belgium: lessons learned from an evaluation
by Thirion, Paul; Pochet, Bernard


Although universities are providing more and more information literacy training for their undergraduate students, the students' real level of information literacy at the beginning of their studies...

Journal Article: Full Text Online

Information Literacy in Students Entering Higher Education in the French Speaking Community of Belgium: lessons learned from an evaluation
by Thirion, Paul; Pochet, Bernard

2009

Although universities are providing more and more information literacy training for their undergraduate students, the students' real level of information literacy at the beginning of their studies...

Journal Article: Full Text Online
ORBi record has gone! Not considered as Open Access by Summon!
Another example with HAL (national French OA repository)
Open Access Content Sources Flagged in the Summon Index

- Content in the Summon index is considered Open Access if it meets the following general criteria:
  - An item is freely available and openly accessible without requiring authentication by the user
  - An item is identified by the provider/publisher as Open Access and has an indication in the metadata we index that signifies it is Open Access content
  - An item resides in a known Open Access repository or database or journal collection that we track in our Knowledgebase as being Open Access

We recognize that there are different flavors and models of Open Access that boil down to where the article is published, who shoulders the cost of publishing, reuse rights, embargo periods, etc. We do not distinguish or differentiate between any of these models. Instead, we identify Open Access content based on the broad criteria described above and methodologies described below.

We have identified more than 200 Million items in the Summon index as Open Access content across dozens of content types including journal articles, books, thesis, technical reports, patents, images, archival documents and more. In most cases, links for Open Access items in Summon will lead to the full text of an article, thesis, book (in the form of a PDF or HTML), to an image, patent or technical report, or to just descriptive metadata (such as a Table of Contents or an abstract) in the case of some open access repositories.

List of Sources with Open Access flags in the Summon index (spreadsheet updated April 28, 2017).

https://knowledge.exlibrisgroup.com/Summon/Product_Documentation/Summon%3A_Open_Access_Filter
Summon: PoPuPS

- PoPuPS was added to Summon in October 2017.
- Request via Idea Exchange Content forum
  https://ideas.exlibrisgroup.com/forums/574345-content/
Summon: PoPuPS

- PoPuPS is identified as being OA
Agenda

• ULiège experience with:
  • Google Scholar
  • Primo Central Index
  • Summon
  • Ebsco Discovery Service
  • CORE
  • BASE
  • PubMed LinkOut

• Some tips and tools
Ebsco Discovery Services

- Ebsco Discovery Services (EDS) = Ebsco’s discovery tool

- EBSCO supports harvesting IRs via OAI-PMH (Dublin Core, Dublin Core qualified, EAD, Junii2, METS, MODS) or FTP

- Institutional Repository Database Questionnaire
  https://help.ebsco.com/interfaces/EBSCO_Discovery_Service/EDS_Catalogs_IRs
EDS: ORBi

- ORBi was added to EDS in March 2016
- OA pdf full texts delivered for indexing purposes only

Nowadays:
- More than 57,000 ORBi records with OA content are in EDS and can be activated by EDS customers
- Activated by more than 2,700 customer (March 2017)
EDS: ORBi
Agenda

• ULiège experience with:
  • Google Scholar
  • Primo Central Index
  • Summon
  • Ebsco Discovery Service
  • CORE
  • BASE
  • PubMed LinkOut
• Some tips and tools
CORE

• Currently contains more than 125,747,000 open access articles, from over tens of thousands journals
• Collected from over 3,673 repositories around the world
• Becoming a data provider? [https://core.ac.uk/join](https://core.ac.uk/join)
Open Access discovery: ULiège experience with aggregators and discovery tools providers...
But searches on CORE may be very sloooowww... and results surprising...
CORE

• **CORE Recommender**: provides research article recommendations across the global network of repositories and journals (Knoth *et al*, 2017)

• Not only on CORE:
  • The plugin that can be installed in repositories and journal systems to suggest similar articles.
  • Purpose is to support users in finding articles relevant to what they read.
  • The current version of the plugin recommends full-text items in Open Access repositories that are related to:
    • a metadata record
    • a full-text item in pdf
    • any piece of text
    • any combination of the above

• Currently not on the roadmap for ORBi2
CORE

- CORE is on the roadmap to be added to Primo Central Index and Summon

CORE (https://core.ac.uk/) aims to aggregate all open access research outputs from repositories and journals worldwide and make these available to the public through a set of services including search, API and analytical tools. As of April 2017, CORE offers seamless access to 70,251,799 open access articles. CORE harvests, maintains, enriches and makes available metadata and full-text content (typically a PDF).

Developed by the Knowledge Management Institute at the Open University, CORE came top in a recent list of top 21 freely available online journal and research databases for academics; https://www.scribendi.com/advice/free_online_journal_and_research_databases_en.html

Galen Jones shared this idea · Apr 28, 2017

PLANNED · Dana Sharvit responded · Jun 4, 2017

After reviewing this idea, we decided to update our plans to include this suggested change. Thank you for your contribution

https://ideas.exlibrisgroup.com/forums/574345-content/suggestions/19079788-core
Agenda

• ULiège experience with:
  • Google Scholar
  • Primo Central Index
  • Summon
  • Ebsco Discovery Service
  • CORE
  • BASE
  • PubMed LinkOut
• Some tips and tools
BASE

- **BASE (Bielefeld Academic Search Engine)**
  - one of the leading global aggregator of open access content (journals, subject and institutional repositories) in the world.
  - claims to be the 2nd biggest academic search engine behind Google Scholar by size easily exceeding over 123 million records and indexing over 6,100 sources.
  - Boost option for OA records
  - Has been indexed in EDS since Dec. 2015 → if your IR contents are indexed properly in BASE, libraries using EDS who turn on BASE in EDS will expose your content to their users 😊

(Tay, 2018)
BASE

• Note that BASE is also requested by the Summon and Primo Central Index community

Content

BASE - Bielefeld Academic Search Engine

BASE is one of the world’s largest search engines especially for academic open access web resources. BASE is operated by Bielefeld University Library. It provides more than 100 million documents from more than 5,000 sources. About 60% of BASE’s contents are in Open Access. Only freely available materials should be indexed in Primo Central and Summon.

François Renaville shared this idea · Apr 24, 2017

UNDER REVIEW · Rael Elstein (Content Product Manager, Ex Libris) responded · Mar 4, 2018

Hello,

Thanks to community support, we have been able to renew discussions with BASE. We will provide an update when available.

Best,
Rael

https://ideas.exlibrisgroup.com/forums/574345-content/suggestions/19027240-base-bielefeld-academic-search-engine
BASE: ORBi

- ORBi was added to BASE in 2008
- Recent full reload of ORBi records
- https://www.base-search.net/Search/Results?q=dccoll:ftorbi&refid=dctableen
Agenda

• ULiège experience with:
  • Google Scholar
  • Primo Central Index
  • Summon
  • Ebsco Discovery Service
  • CORE
  • BASE
  • **PubMed LinkOut**
• Some tips and tools
PubMed LinkOut

- LinkOut is a service that allows you to link directly from PubMed and other NCBI databases to a wide range of information and services beyond the NCBI systems.
- LinkOut aims to facilitate access to relevant online resources in order to extend, clarify, and supplement information found in NCBI databases. Examples of LinkOut Resources include full-text publications, biological databases, consumer health information, research tools, and more.

- How do Institutional Repositories (IR) join LinkOut?

- Currently only 26 IRs participating in LinkOut
PubMed LinkOut

- ORBi was added to LinkOut in May 2017

Number of Hits

- May 2017: 1555
- Jun 2017: 6029
- Jul 2017: 5696
- Aug 2017: 6009
- Sept 2017: 6744
- Oct 2017: 7858
- Nov 2017: 7485
- Dec 2017: 5839
- Jan 2018: 6834
- Feb 2018: 5454
Cardiopulmonary exercise testing is a better outcome predictor than exercise echocardiography in asymptomatic aortic stenosis.

Domanski O¹, Richardson M², Caisne A³, Poëge AS², Mouton S², Godart P³, Edmé JL³, Matran R³, Lancelot F³, Montaigne D³.

Abstract

BACKGROUND: Objective assessment of maximal aerobic capacity using peak oxygen consumption (peak VO2) can be helpful in the management of patients with asymptomatic aortic stenosis (AS). The relationship between peak VO2 and AS severity criteria derived from rest and supine exercise echocardiography (SEE) has never been explored.

OBJECTIVES: We aimed to determine whether low peak VO2 (<85% of predicted value) is associated with severity parameters in SEE, and poor clinical outcome.

METHODS: Fifty one asymptomatic patients (mean age of 54±21 years) with moderate to severe aortic stenosis (Vmax>3m/s) and left ventricle ejection fraction>50% prospectively underwent resting and SEE and cardiopulmonary

Direct Access to the Open Access ORBi record
Agenda

• ULiège experience with:
  • Google Scholar
  • Primo Central Index
  • Summon
  • Ebsco Discovery Service
  • CORE
  • BASE
  • PubMed LinkOut

• Some tips and tools
Starting blocks

• Be **proactive** and **monitor** the harvesters, search engines and discovery tools
  • The big ones, but not only
  • The new ones (stay tuned!)
  • The one you use and do not use:
    • Objective: Pushing your OA content into the search environments of other researchers
    • Have a list of harvesters... and update it (ULiège Library has a 26 page document with admin data, repositories we are in contact with, update notes...)
  • Time consuming, but important!
• Follow **standards** (there are many) and be inspired (ODI...)
  • Some harvesters may have some specific requirements,
  • Be ready to adapt, but never forget the **KISS principle**!
• Have some harvesting **sets** ready to be used
  • All records vs records with full text vs OA records
  • By discipline, by document type?
Open Discovery Initiative (ODI)

- A technical recommendation for data exchange including data formats, method of delivery, usage reporting, frequency of updates and rights of use
- A way for libraries to assess content providers’ participation in discovery services
- A model by which content providers work with discovery service vendors via fair and unbiased indexing and linking


- Section 3.3.1: “Discovery service providers should make available to prospective and current customers sufficient information about the content of their repositories to ensure an adequate evaluation of that content against the customers’ needs.”
Open Discovery Initiative (ODI)

• Definition of Core metadata:
  • minimum metadata record that must be provided by Content Providers (CPs) to Discovery Service Providers (DSPs)
  • based on the KBART (NISO RP-9-2014) metadata encoding schema:
    • primarily for serial and monograph titles to improve A-Z lists, link resolution, and OpenURL, but
    • extended to capture elements for subject, abstract/description, content type, and content format

• Examples:
  • Open Access Designation
    • To comply with the NISO Open Access Metadata and Indicators (OAMI) group’s recommendations, if an item is open access, this status should be indicated with “free_to_read”.
  • Full Text Flag
    • A yes/no statement describing whether the content provider makes this item available in full text (or for non-print media, a full-length or high-resolution version) to the DSP for the purpose of indexing. It is expected that this will be disclosed by DSPs to libraries in future when describing indexing coverage for a title or collection.
Some best practices

- BASE has published **Golden Rules for Repository Managers**
  [https://www.base-search.net/about/en/faq_oai.php](https://www.base-search.net/about/en/faq_oai.php) focusing on structure, encoding, ISO standards...
  1. OAI interface working
  2. Comprehensive metadata
  3. Identifier (URLs) working
  4. Providing access information (Open Access)
  5. Providing information concerning re-use / licence (Creative Commons)
  6. Character encoding
  7. Publication date
  8. Document language
  9. Source / Suggested citation
  10. Items per page
  11. Contact person
  12. Changes / Updates
  13. Spread the word!

Open Access discovery: ULiège experience with aggregators and discovery tools providers...
Some best practices

- and an OAI-PMH validator OVAL http://oval.base-search.net/

**OVAL**

BASE OAI-PMH Validator

See also http://validator.oaipmh.com
References

- Ebsco. EDS - Catalogs and Institutional Repositories Guide [Link](https://help.ebsco.com/interfaces/EBSCO_Discovery_Service/EDS_Catalogs_IRs)
- Ex Libris. Primo Central Index [Link](http://www.exlibrisgroup.com/category/PrimoCentral)
- Ex Libris. Summon: Open Access Filter [Link](https://knowledge.exlibrisgroup.com/Summon/Product_Documentation/Summon%3A_Open_Access_Filter)
http://hdl.handle.net/2268/221340