

# Efficacy and benefits of web services for metadata acquisition

## an overview based on Swiss institutional repositories

Isabelle de Kaenel, Pablo Iriarte  
 Medical Library, University Hospital  
 CHUV BH-08, CH-1011 Lausanne, Switzerland  
 isabelle.de-kaenel@chuv.ch, pablo.iriarte@chuv.ch

### Objectives

One of the challenges of institutional repositories is to assist end-users in the creation of metadata in order to help the submission process and to ensure a high level of data quality in reducing typing errors or multiple key strokes, such as the tedious « copy/paste » combination.

Clean, authoritative and accurate datasets are now available from various sources : bibliographic databases, library catalogs, controlled lists and repertories. The information providers facilitate data transfer and integration into local applications through many channels :

- Export functions to personal reference software that support various formats (RIS, BibTeX, etc.) and can generate files for upload into repositories
- OpenURLs for single record creation in the repository
- Web Services for single record creation and completion and for batch input routines

### Sources and techniques to assist metadata population

	PubMed	WoS	OVID	SCOPUS	CrossRef	RERO catalog	Library of Congress	OAlster	Scientific Commons
Export to ref. software	✓	✓	✓	✓	✗	✓ <sup>4</sup>	✓ <sup>4</sup>	✓ <sup>4</sup>	✓
OpenURL	✓	✓	✓ <sup>2</sup>	✓	✗	✓	✗	✗	✗
Web Service	✓	✓ <sup>1</sup>	✗	✓ <sup>3</sup>	✓	✓	✓	✓	✗

✓ = free      ✓ = subscription required

<sup>1</sup> Agreement from WoS required for the use of web service

<sup>2</sup> OVID allows the export of other fields not included in the OpenURL standard (abstract, complete list of authors, keywords...)

<sup>3</sup> The SCOPUS Web Services are open to non subscribers in a limited way (the abstract and the complete list of authors are reserved)

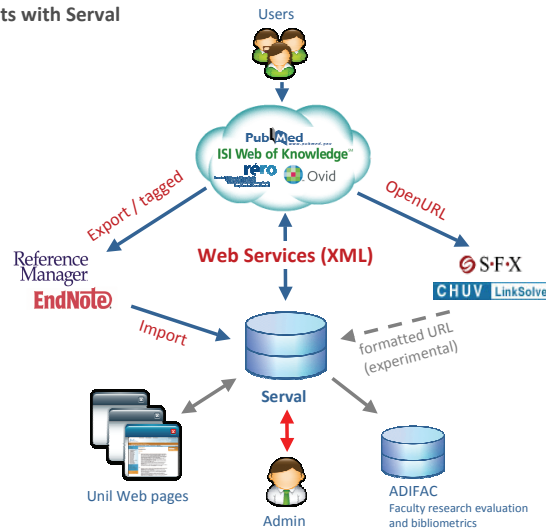
<sup>4</sup> Only for ZOTERO users

### Design and Settings

The potentialities of these technologies have been investigated and tested in Swiss institutional repositories such as **Serval** at *Lausanne University*, **Archive Ouverte UNIGE** at *Geneva University* and **Infoscience** at *Ecole Polytechnique Fédérale de Lausanne*.

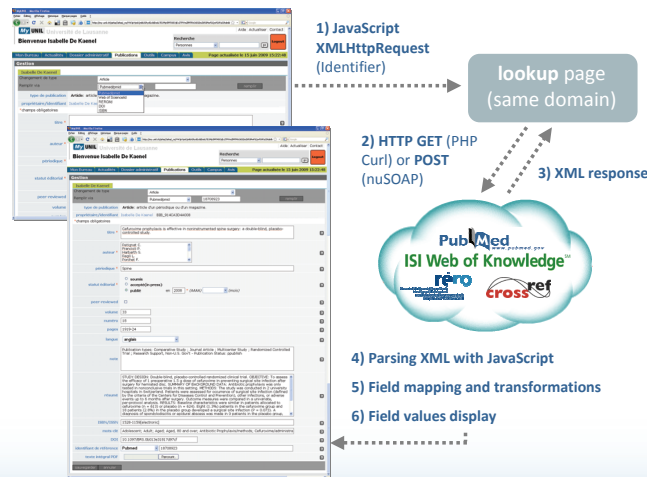


### Tests with Serval



### Implementation

An AJAX solution has now been implemented in Serval to search/retrieve records using DOI, ISBN and database unique identifiers (PubMed, Web of Science and RERO), parsing XML response, transforming and mapping metadata fields prior to introduction into the repository record entry form.

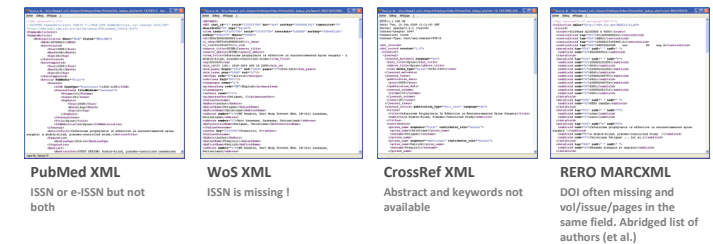


### Results

Web services are particularly efficient to fill in the web entry forms with external metadata. Their benefits are promoted to the end users as well as the repository content managers.

Different databases imply different kinds and quality of metadata. PubMed excepted, the API of the bibliographic Web Services are still poor. The XML format provided notably varies from one source to another. Only library catalogs (MARCXML) and repositories (OAI-PMH) use standards.

### Same document but different metadata and XML



### Conclusion

Higher standardization of access to the data providers (OAI-PMH or SRU/SRW) would increase efficiency of Web services. Currently, rich and accurate metadata can only be retrieved using unique identifiers.

The AJAX model, allowing seamless integration of metadata directly into the repository ingest mechanisms, improves the usability of data entry tools and assists the process of capturing content from external sources. The repository administrators can also apply the same parsing and mapping methods to build automated workflows.

### Future work

Further implementations are planned, the spectrum is large. The aim is to interact with other types of data providers : repertories of administrative metadata, authors names and identifiers, lists of ISSN, sources of rights management metadata like **SHERPA/RoMEo**, etc. Another ambition is to merge metadata from multiple sources (PubMed + Web of Science reference data using a DOI, RERO + Amazon book data using an ISBN, etc.)

### Links

- Source code (GNU licence) : <http://code.google.com/p/bibliofill/>
- SERVAL (SERVeur Académique Lausannois) : <http://serval.unil.ch>
- Archive Ouverte UniGe : <http://archive-ouverte.unige.ch>
- Infoscience EPFL : <http://infoscience.epfl.ch>
- RERO (REseau ROmand) : <http://www.rero.ch>
- PubMed e-utilities : [http://eutils.ncbi.nlm.nih.gov/corehtml/query/static/eutils\\_help.html](http://eutils.ncbi.nlm.nih.gov/corehtml/query/static/eutils_help.html)
- CrossRef : <http://www.crossref.org>
- ISI Web of Knowledge : <http://isiknowledge.com>



CERN workshop on Innovations in Scholarly Communication (OAI6)

17 - 19 June 2009 | University of Geneva

