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 typying 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

<table>
<thead>
<tr>
<th>Source</th>
<th>PubMed</th>
<th>Web</th>
<th>DVD</th>
<th>SCOPUS</th>
<th>CrossRef</th>
<th>RERO catalog</th>
<th>Library of Congress</th>
<th>OKDIR</th>
<th>Scientific Commons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Export ref. software</td>
<td>✅</td>
<td>✅</td>
<td>🟢</td>
<td>✅</td>
<td>✅</td>
<td>✅</td>
<td>✅</td>
<td>✅</td>
<td>✅</td>
</tr>
<tr>
<td>Crosswalks</td>
<td>✅</td>
<td>✅</td>
<td>🟢</td>
<td>✅</td>
<td>✅</td>
<td>✅</td>
<td>✅</td>
<td>✅</td>
<td>✅</td>
</tr>
<tr>
<td>Metadata support</td>
<td>✅</td>
<td>✅</td>
<td>🟢</td>
<td>✅</td>
<td>✅</td>
<td>✅</td>
<td>✅</td>
<td>✅</td>
<td>✅</td>
</tr>
</tbody>
</table>

- free 🟢 = subscription required
- (OAI6) (same format as the reference service) (same domain)
- (SERVeur Académique Lausannois) (OAI6)
- (OAI6) (PubMed + CrossRef.XML)
- (OAI6) (PubMed + CrossRef.XML)
- (OAI6) (PubMed + CrossRef.XML)

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

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 (MARCMX) 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 license): http://code.google.com/p/bib708/
- SCOPUS (Editeur Académique Lausannois): http://serval.unil.ch
- Archive Ouvverte UNIGE: http://archive-ouverte.unige.ch
- Infosciences EPFL: http://infoscience.epfl.ch
- RERO (Róbert Richard): http://www.rero.ch
- CrossRef: http://www.crossref.org
- GI Web of Knowledge: http://link. knowledge.com