Criteria for the selection of a semantic repository for managing SKOS data

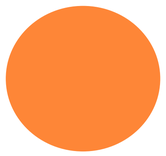
Convegno AIB CILW 2016

Ricardo Eito-Brun
Universidad Carlos III de Madrid

Moving to SKOS…

- Two sides of Linked Open Data projects:
  - Moving online, convert to RDF, our data.
  - Give other people the possibility of using your data (we mean "other software applications".

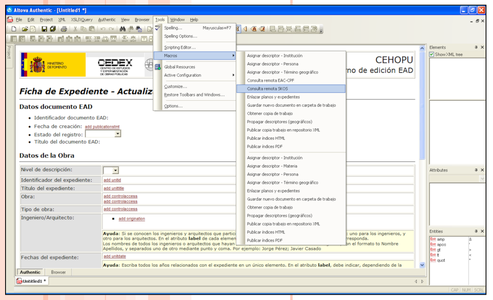
- Today, most organizations focus on data conversion and publishing
  - Convert Thesauri, Classification schemas or KOS to SKOS.
  - The results of this conversion are published…., as HTML data…, not reusable by machines.
  - Set up SPARQL end-points, that are hard to use
    - language syntax and
    - Complex, technical integration of "consumer applications" with SPARQL end points in specific context (e.g. non-Java desktop apps)
  - The last two points are done with the help of RDF repository/DB tools.



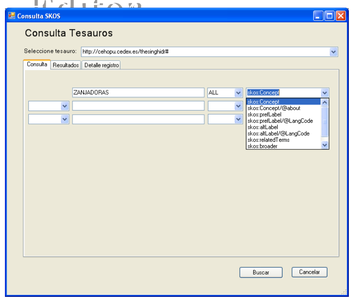
Tools for managing SW-data

- Functions related to data load / entry:
  - Bulk Import of RDF data.
- Functions related to data extraction / search
  - SPARQL end-points.
  - Full text indexing
  - Proprietary APIs
- Main problems:
  - Lack of Integration with the tools we use for editing RDF/SKOS data.
  - Data access restricted to complex SPARQL
  - Not all the data we must manage are RDF-based (e.g. EAC authority records)
- This scenario led to the selection of an XML database tool for managing SKOS data.

Integration of SKOS and SRU in a distributed collaboration environment for EAD description



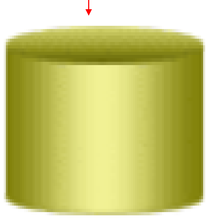
XML Metadata Editor



http://www.uc3m.es/bibdoc/sruSrvr/skos/processRequest.php?
version=1.2
&operation=searchRetrieve
&query=CQL_SEARCH
&maximumRecords=100
&recordSchema=skos_summary

SKOS files are generated by a conversion and uploaded in an Oracle XML DB database.

Web Server

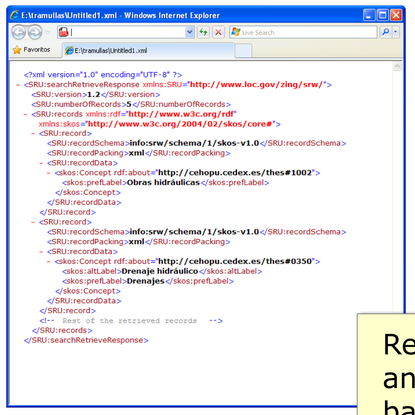


SKOS DB

1 Cataloguer proceeds to assing descriptor

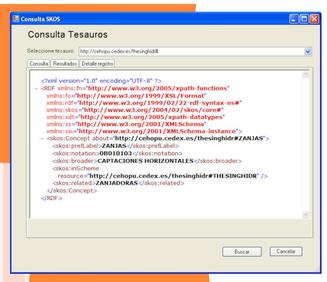
2 Cataloguer selects SKOS DB and search criteria.

3 SRU search requests is directed to remote web server.



4 Request is processed and response sent back to metadata editor.

5 Response data are shown to end-user, who can assign descriptor and see details.



1. Wider visibility to existing thesauri & classification systems
2. Leverage investments in the development of indexing languages.
3. Open, standard-based access to indexing languages.
4. Integration and reuse from metadata creation tools is easier.
5. New semantic integration / distributed search capabilities between repositories based on controlled vocabularies.

Ricardo Eito-Brun (UC3M)