FROM XML TO MARC

RDF

BEHIND THE SCENES
ABES = French national agency for HE libraries

www.sudoc.abes.fr  General union catalog
www.idref.fr  Authority data
www.theses.fr  Dissertations (completed or in progress)
www.calames.abes.fr  Archives and manuscripts
Publisher metadata

RDF

Catalogues, LOD, disco tools ...
ENRICHING

Other bib. Data + reference data (authority data...)
WHAT WE DO
Collecting, Normalizing, Enriching, Sharing

WHAT FOR
Contributing to the **web of open and linked and high quality data** of scientific content

ON WHICH STUFF
France related metadata sets
+ What is not (yet) in the LOD
We need to be flexible.

Data model
Workflows
Technological choices
RDF VOCABULARIES

EXTENSIBLE DATA MODEL

AD HOC CLASSES & PROPERTIES

ISBD
DCTERMS
BIBLIO ontology
VIVO
VCARD
FOAF
SKOS
...
FLEXIBLE WORKFLOWS
Basic operations on data > SQL & SPARQL Queries + Stored procedures
Workflow engine > Stored procedures in a relational database
Administration interface for data experts >
TECHNOLOGICAL CHOICES
WHY ORACLE?

• Need to program close to the data (for performance reasons)
  • PL SQL
  • Java inside Oracle

• Our dependencies
  • Not being dependent on a specific RDF DB ➔ no Virtuoso PL SQL
  • We are already dependent on ORACLE... (and have the skills)

• Joining with more data already in Oracle
  • Mix of SPARQL and SQL (or XQUERY)

• Management of errors ➔ logs in tables, easy to query
4 use cases
<OUP_ONIX/> about each print book

print isbn is the key to merge the two graphs

From SQL to RDF

<table>
<thead>
<tr>
<th>print isbn</th>
<th>elec. isbn</th>
<th>title</th>
<th>date</th>
<th>url</th>
</tr>
</thead>
</table>

KBART about ebooks

www.bacon.abes.fr
<XML/>

about each document

IdRef = authority database for French HE

VIAF Dump of links → RDF

Link computed by Qualinca
(home made algos)

Person with IdRef ID

Person with local ID

Work

marcrel:aut

VIAF

DnB

NKC
<XML/>
about each Nature article

F-MeSH is a subset of IdRef

www.idref.fr
Springer Concept ID is the key to merge the two graphs.

Via Open Refine RDF export

**Mapping**

<table>
<thead>
<tr>
<th>Springer Concept</th>
<th>LCSH Concept</th>
</tr>
</thead>
</table>

Chapter

**dcterms:subject**

Springer Concept

**skos:closeMatch**

LCSH Concept

about each chapter

```SPRINGER_XML/```
Some concerns...
AS CATALOGERS ARE CATALOGING, SHOULD WE FEED BACK THE RDF BLACK BOX?
HUMAN and MACHINE BOTTLENECKS

XML
- ONIX
- JATS
- Etc.

XSLT

RDF/XML

NTRIPLES

SPARQL INSERT via JDBC

MARYC
LOD
Etc.
OK…

BUT WHICH PUBLISHER DATA PROCESSING?

NEED FOR A GLOBAL COLLABORATION!

⇒ NEED FOR A GLOBAL @TODO LIST!
Na zdraví!

Questions?