



národní
úložiště
šedé
literatury

From XML to MARC

Nicolas, Yann
2018

Dostupný z <http://www.nusl.cz/ntk/nusl-380926>

Dílo je chráněno podle autorského zákona č. 121/2000 Sb.

Licence Creative Commons Uveďte původ-Neužívejte komerčně-Nezpracovávejte 4.0

Tento dokument byl stažen z Národního úložiště šedé literatury (NUŠL).

Datum stažení: 17.04.2024

Další dokumenty můžete najít prostřednictvím vyhledávacího rozhraní nusl.cz .

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



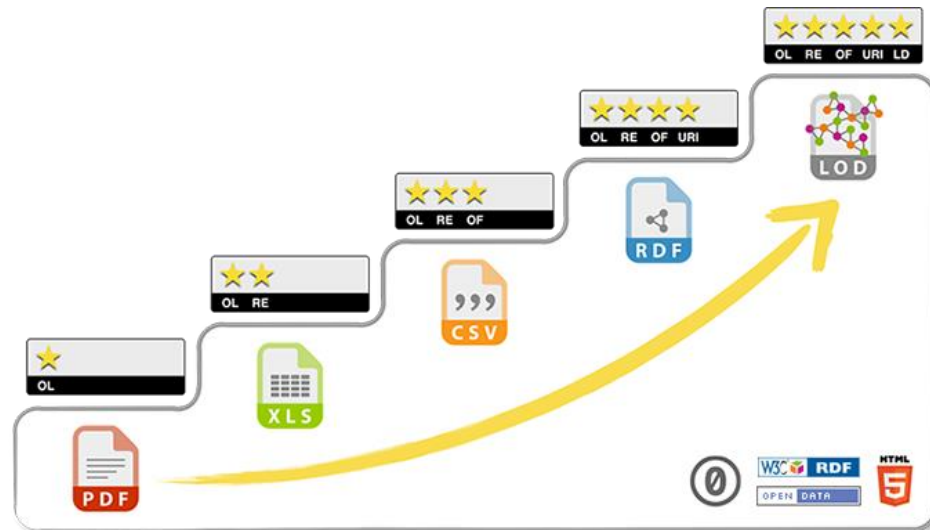
www.theses.fr

Dissertations (completed or in progress)



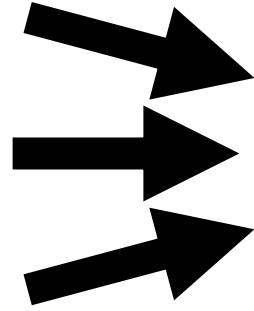
www.calames.abes.fr

Archives and manuscripts

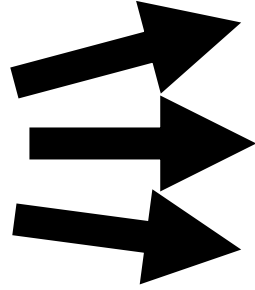


Publisher metadata

A
B
C



RDF



X
Y
Z

marc



marc



marc

Catalogues,
LOD, disco tools ...

Catalogues

COLLECTING

NORMALIZING

SHARING

XML

- ONIX
- JATS
- Etc.

KBART

Etc.



MARC

sudoc

LOD

Not yet #grrr

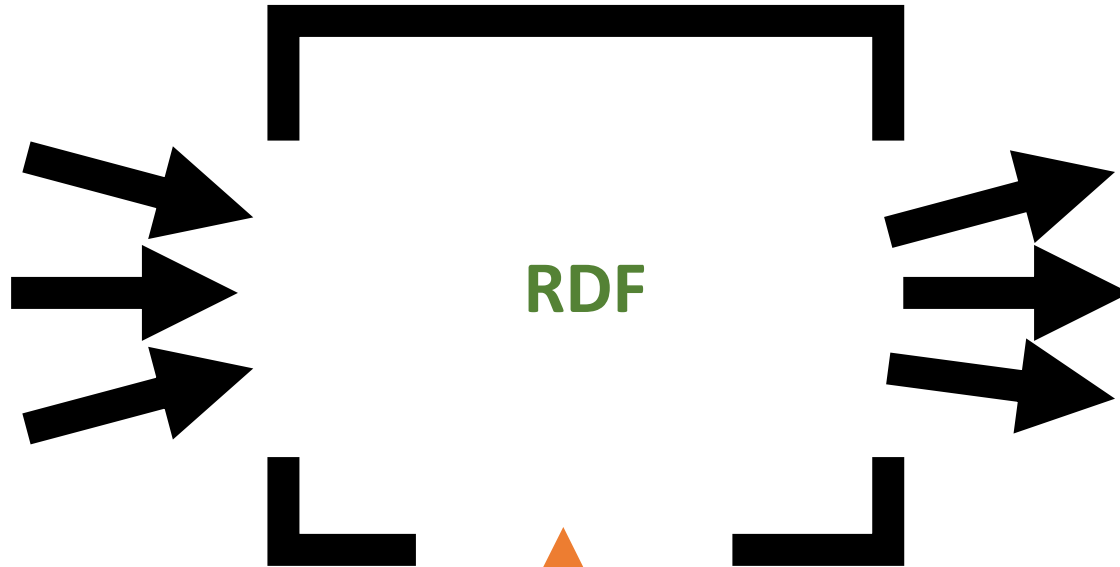
Etc.

XML

- ONIX
- JATS
- Etc.

KBART

Etc.



RDF

MARC

LOD

Etc.

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

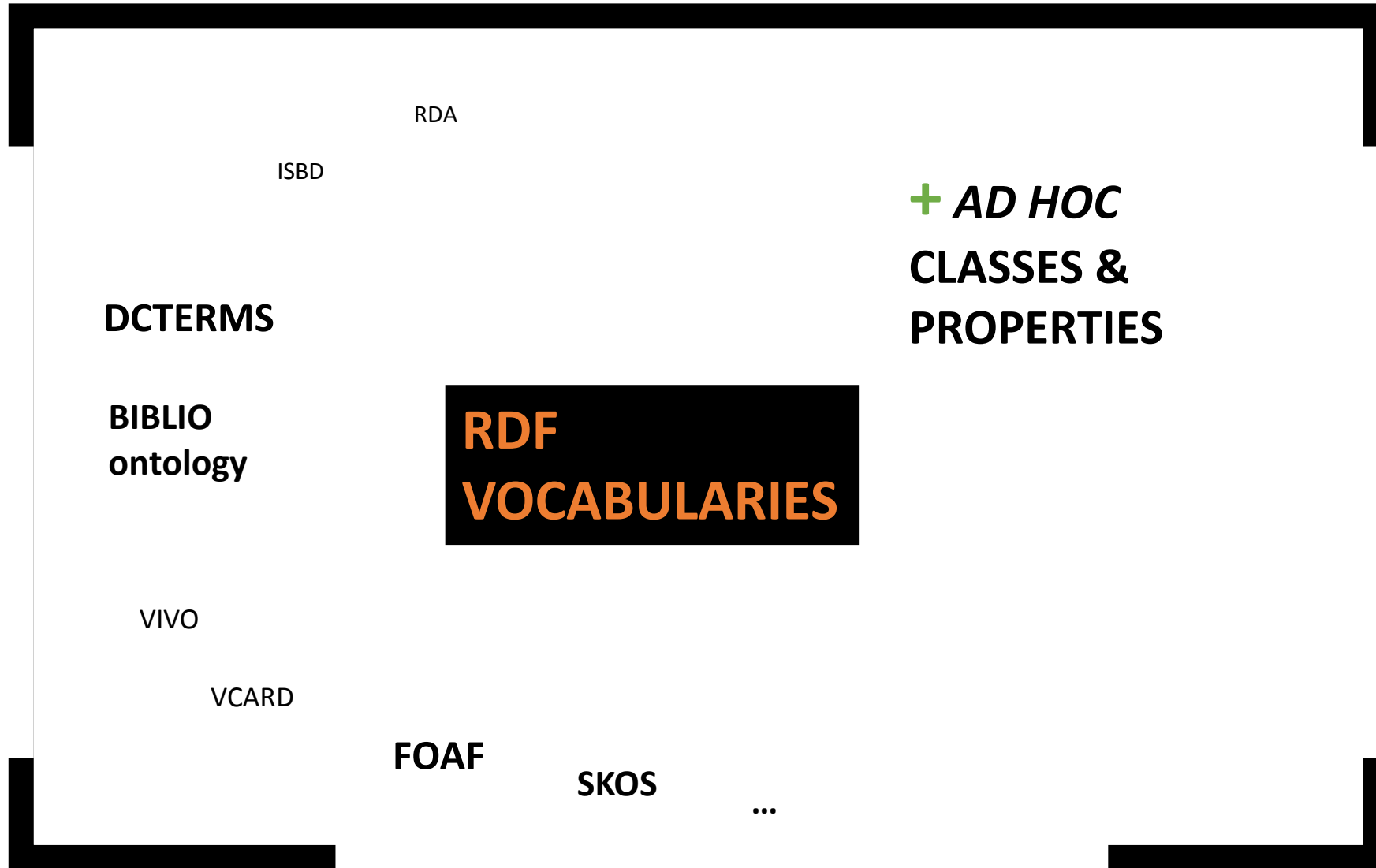
France related metadata sets
+ What is not (yet) in the LOD

We need to be flexible.

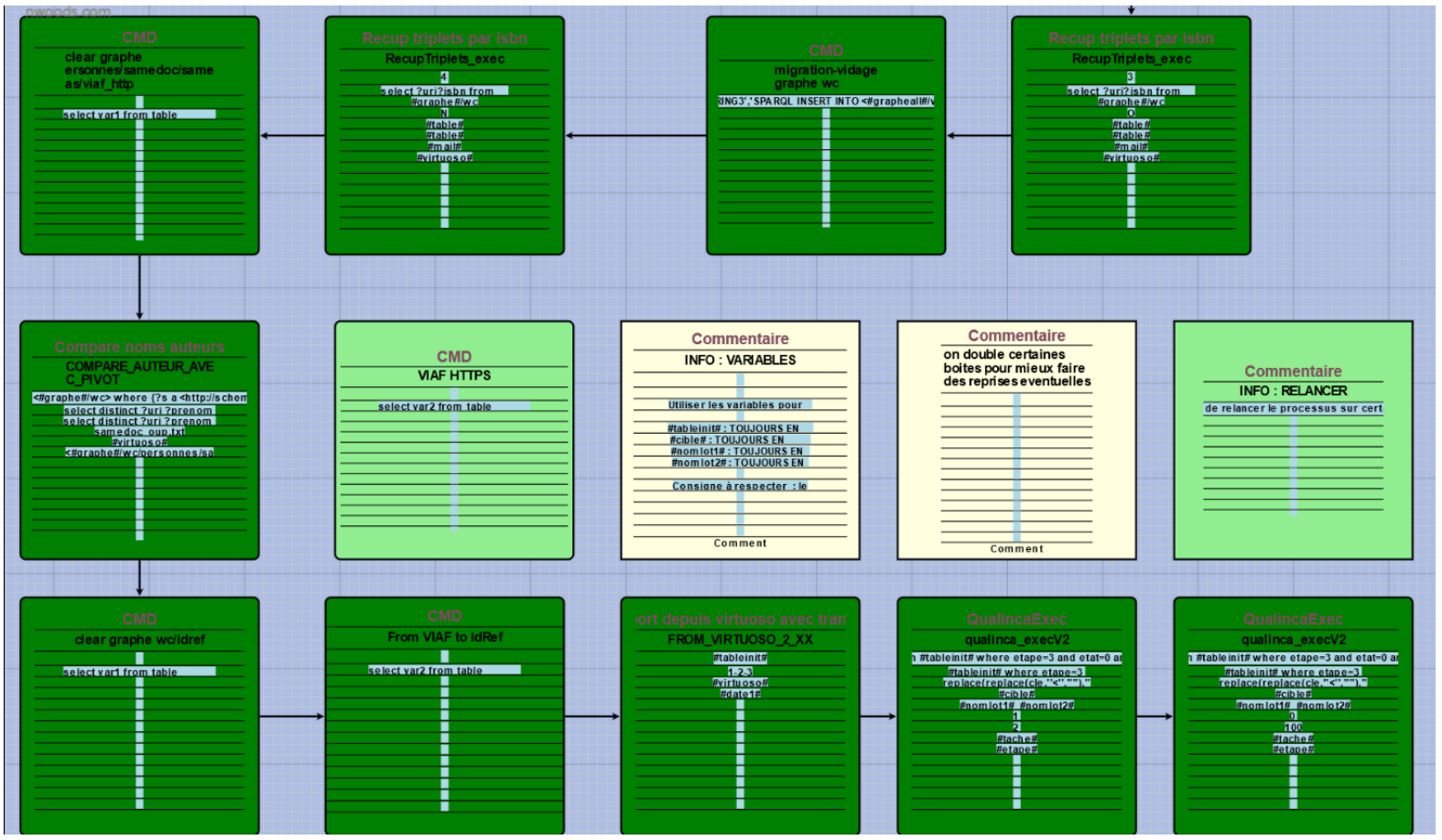
Data model

Workflows

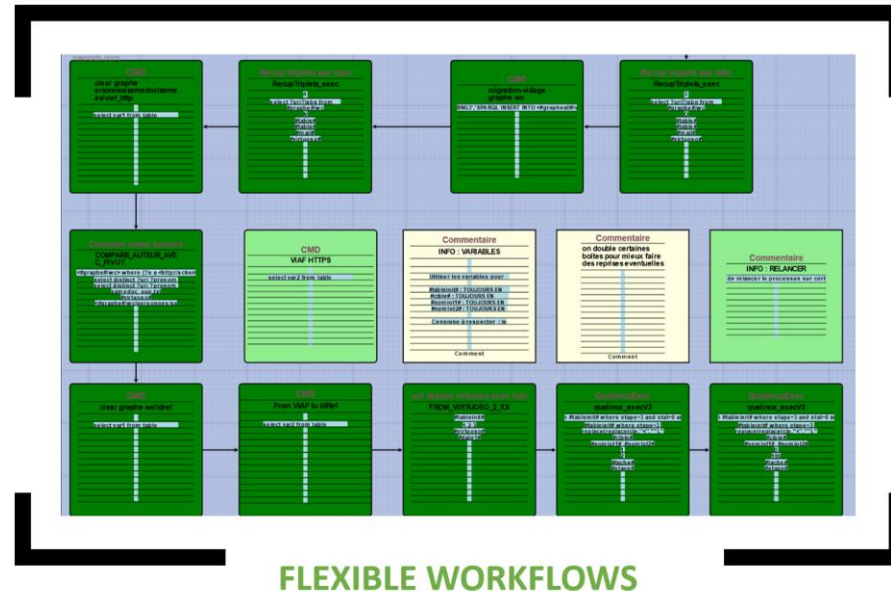
Technological choices



EXTENSIBLE DATA MODEL



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

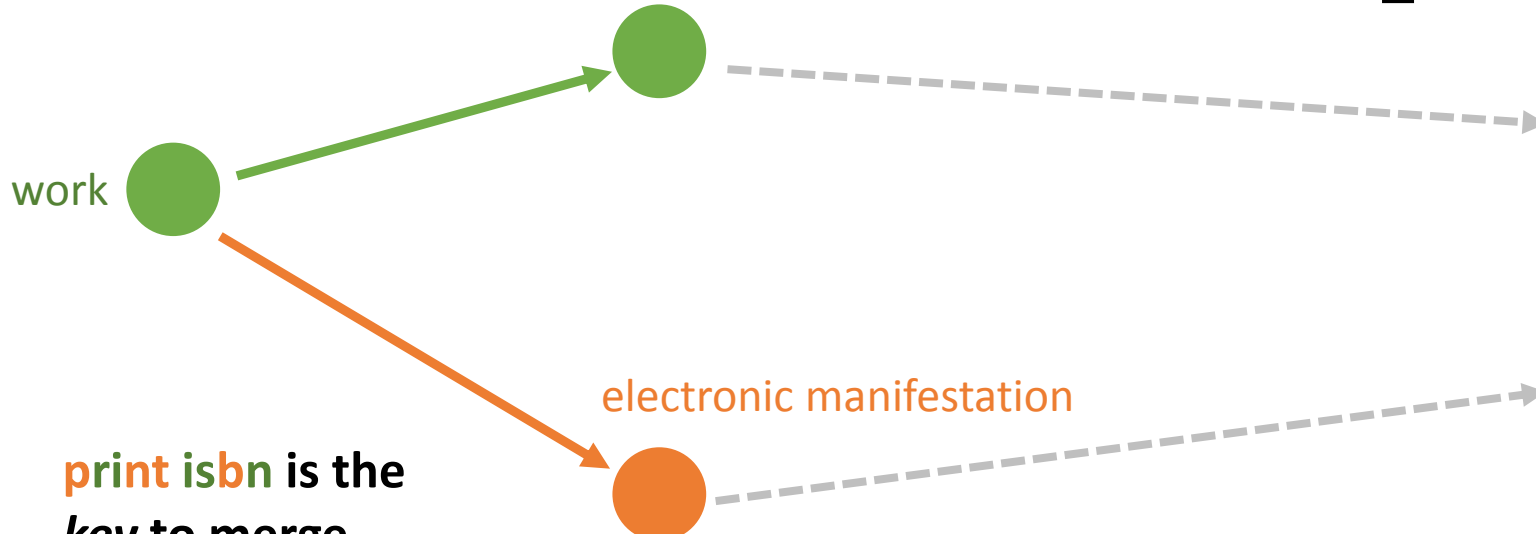
XML

RDF

MARC

<OUP_ONIX/>
about each
print book

work



MARC
for
print

MARC
for
ebook

print isbn is the
key to merge
the two graphs

From SQL to RDF

print isbn	elec. isbn	title	date	url

ORACLE

KBART about ebooks

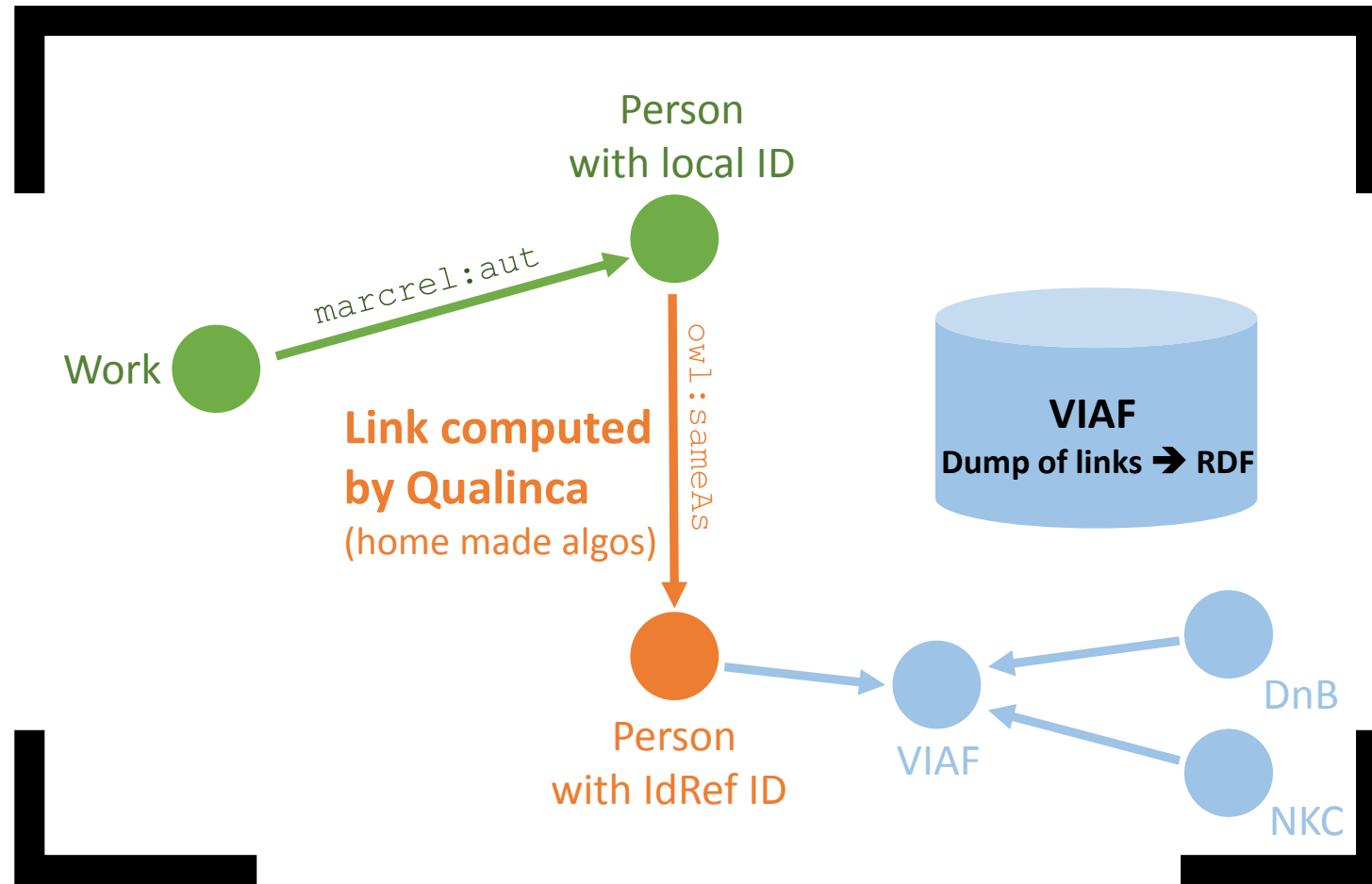
www.bacon.abes.fr



XML

RDF

<XML/>
about each
document



Link computed
by Qualinca
(home made algos)

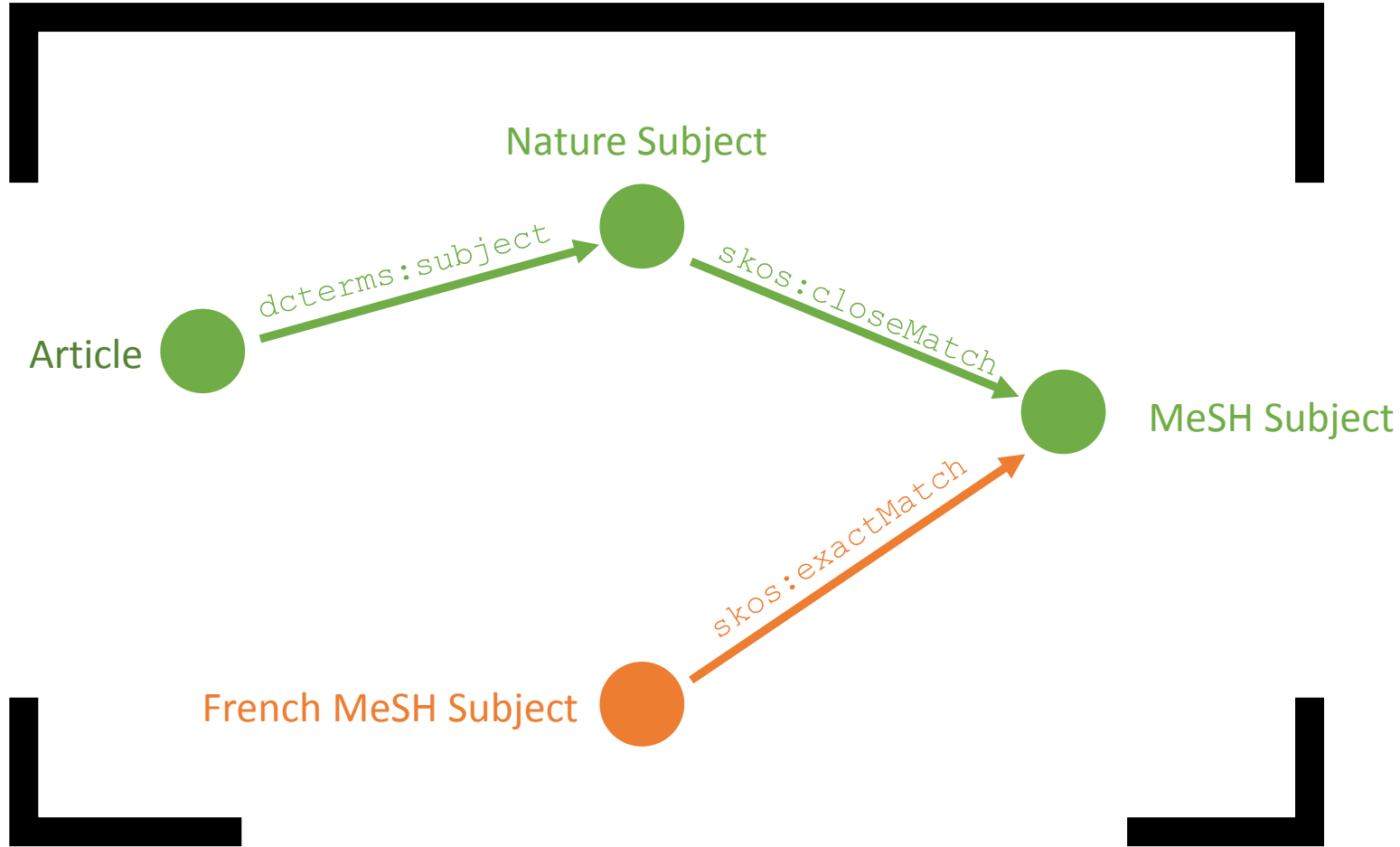
IdRef =
authority
database for
French HE



XML

RDF

<XML/>
about each
Nature article



F-MeSH is a
subset of
IdRef

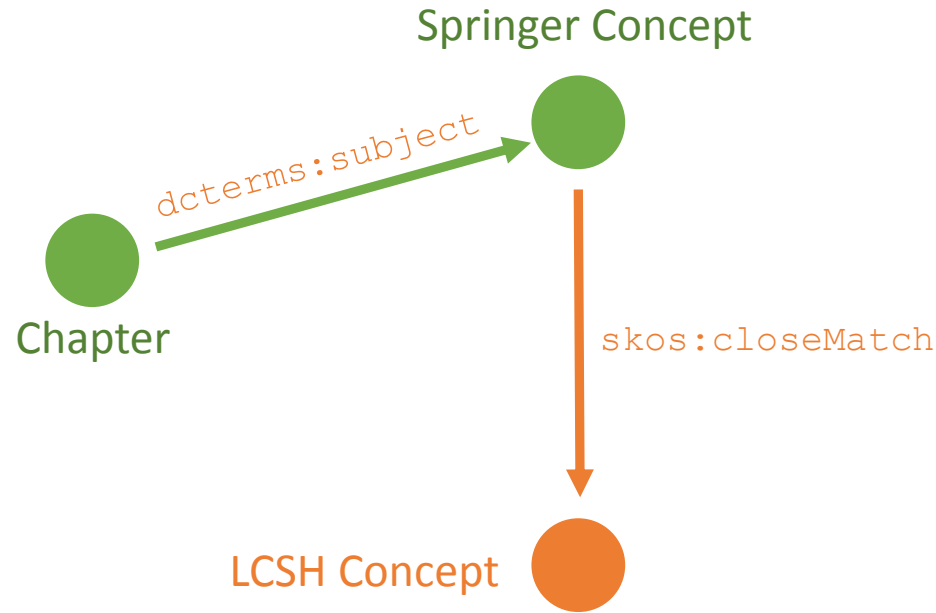


www.idref.fr

XML

RDF

<SPRINGER_XML/>
about each
chapter



Via Open Refine RDF export

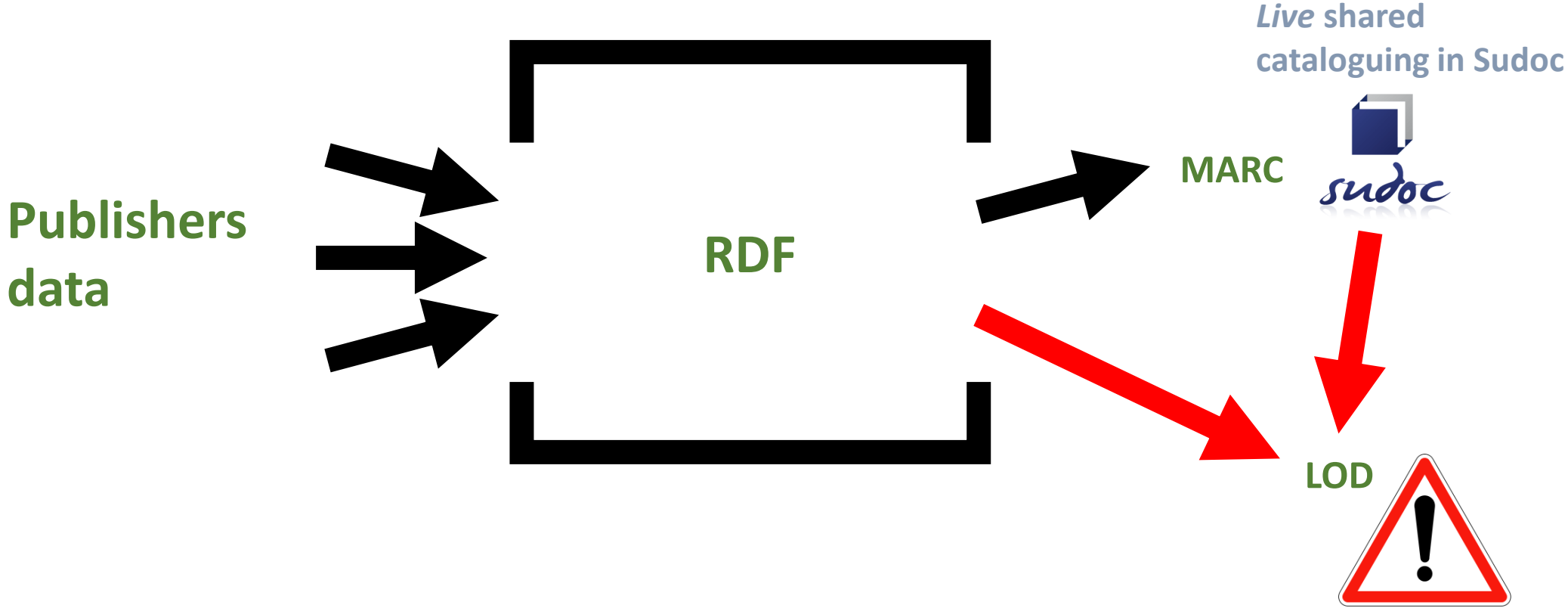
Springer Concept ID
is the key to merge
the two graphs

Springer Concept	LCSH Concept
Mapping	

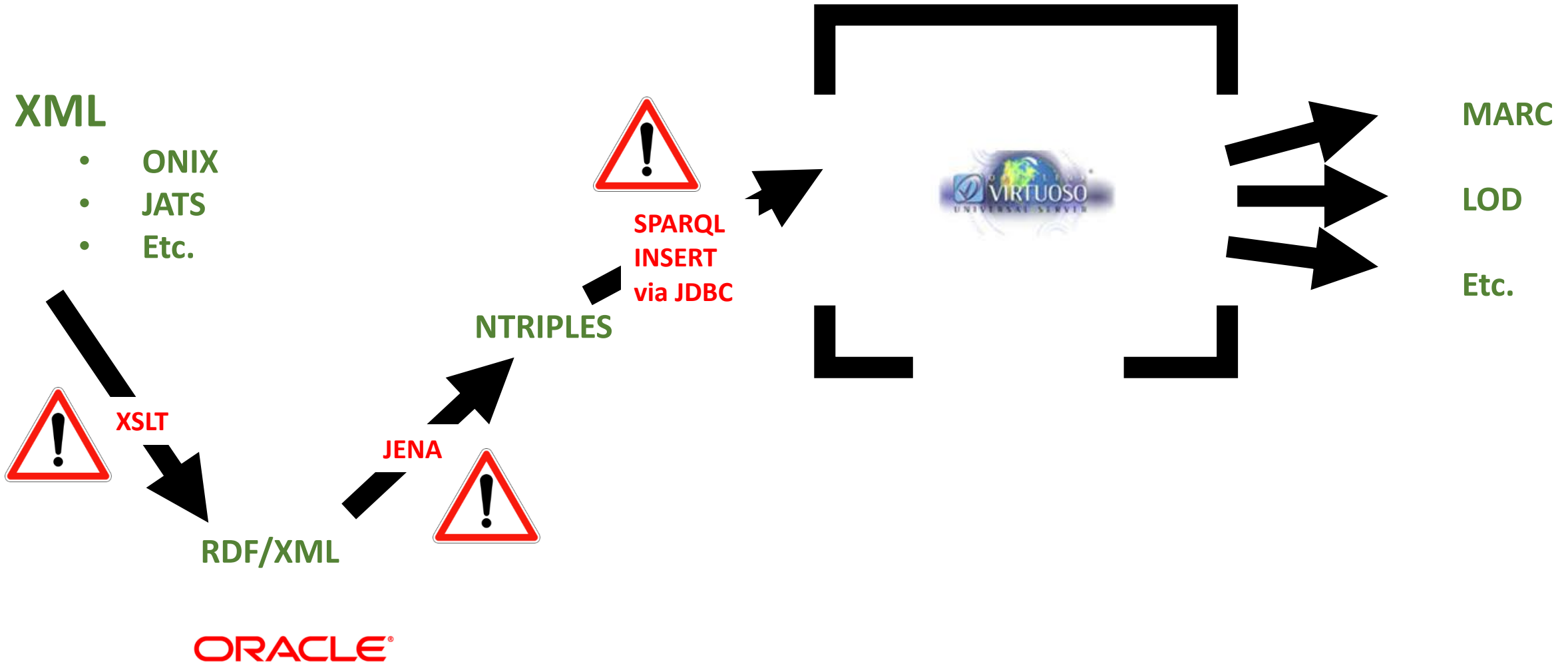


Some concerns...

AS CATALOGERS ARE CATALOGING, SHOULD WE FEED BACK THE RDF BLACK BOX ?

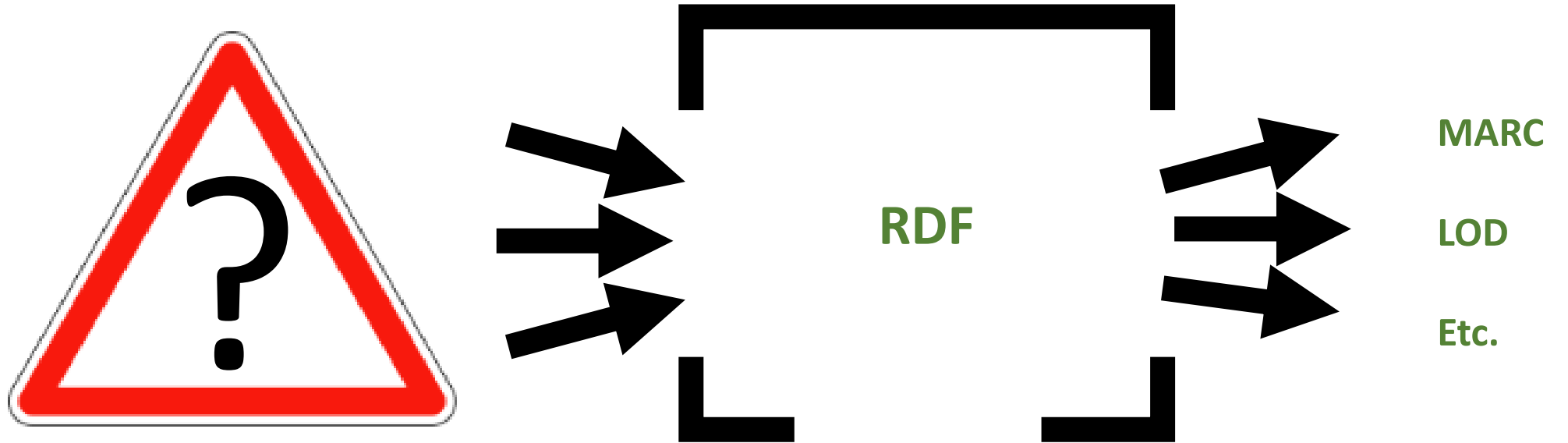


HUMAN and MACHINE BOTTLENECKS



OK...

BUT WHICH PUBLISHER DATA PROCESSING ?



NEED FOR A GLOBAL COLLABORATION !

→ NEED FOR A GLOBAL @TODO LIST !

GOK_b
Global Open Knowledgebase



Na zdraví !

Questions ?

