abes
agence bibliographique de l'enseignement supérieur

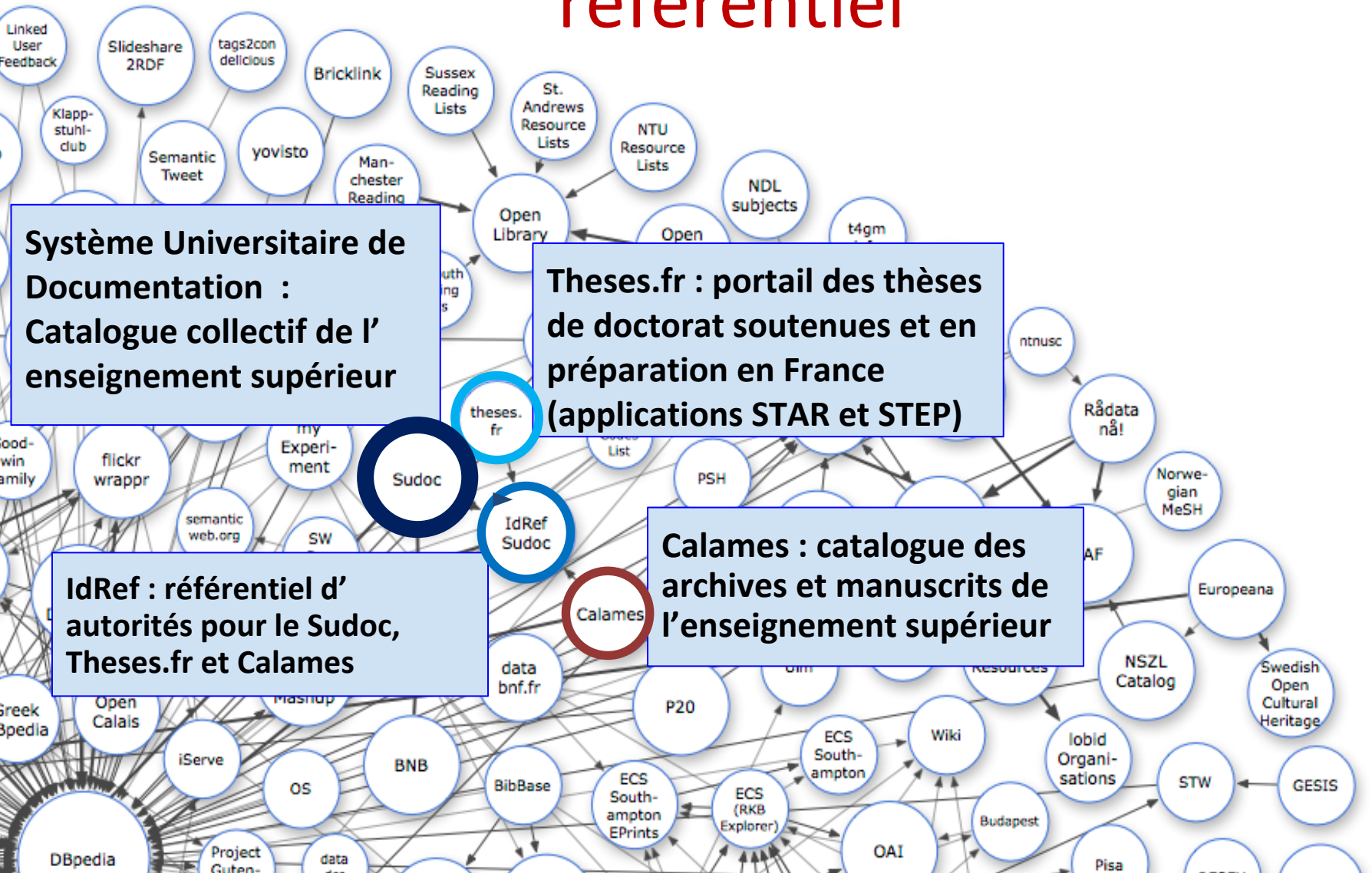
Du web sémantique à tous les étages ?

Yann Nicolas – Michael Jeulin

ABES

SemWeb.Pro 2014
Paris, 5/11/2014

Le Linked Data de l'ABES
Trois bases de données autour d'un référentiel



abes
agence bibliographique de l'enseignement supérieur

Exposer

L'ABES sur le web de données : pourquoi ?

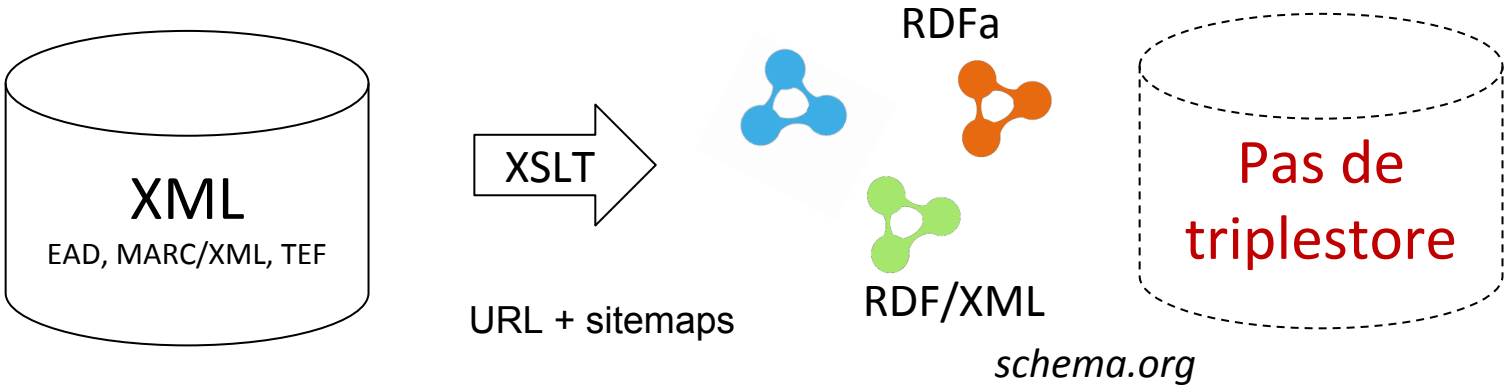
- Des données liées et structurées
  - pour les moteurs de recherche...
  - pour faciliter leur export et leur réutilisation
- Ouverture et mutualisation des données : une tradition dans les bibliothèques

  ISBD, MARC, catalogage partagé, Z39-50…
- Nouvelle étape : on ouvre plus, et à tout le monde

  OAI, webservices... et RDF

Méthode et principes

- Une entreprise au long cours
- Approche progressive, pragmatique et empirique
- Choix de standards du web



Conversion à la volée = pas de base RDF

Les chantiers RDF

Application	Format	Année	RDF	Dump	SPARQL ?	Qualité LOD
Calames	XML (EAD)	2008	RDFa	Non	Non	★★★★★
IdRef	MARC	2010	RDF/XML	Oui (mais pas public)	Non	★★★★★
Sudoc	MARC	2011	RDF/XML + schema.org	Oui (mais pas public)	Non (en cours)	★★★★★
www.theses.fr	XML (TEF)	2011	RDFa+RDF/XML	Non	Non	★★★★★

Linked open data :

★ non filtrées (presque)

★★ Structurées

★★★ Librement exploitables

★★★★ Identifiées (URL)

★★★★★ Données liées

Quels modèles de données ?

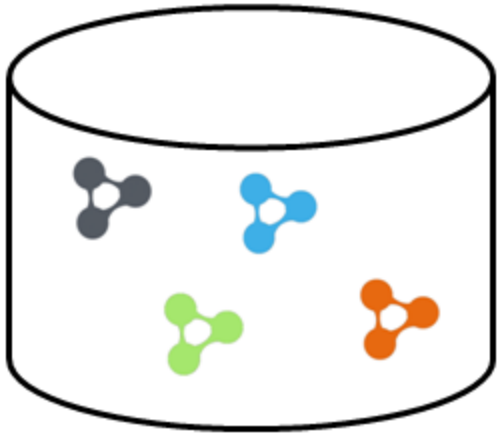
Vocabulaires déjà publiés et répandus : Dublin Core, Bibo, FOAF, bio, etc

=

Diffusion plus large, mais mal adapté aux données natives

vocabulaires « métiers » quand nécessaire : ISBD, RDA (Sudoc)

Proches des formats natifs, mais mal adaptés au web de données…



Vocabulaire ad hoc ?

Et jusqu'où raffiner ?

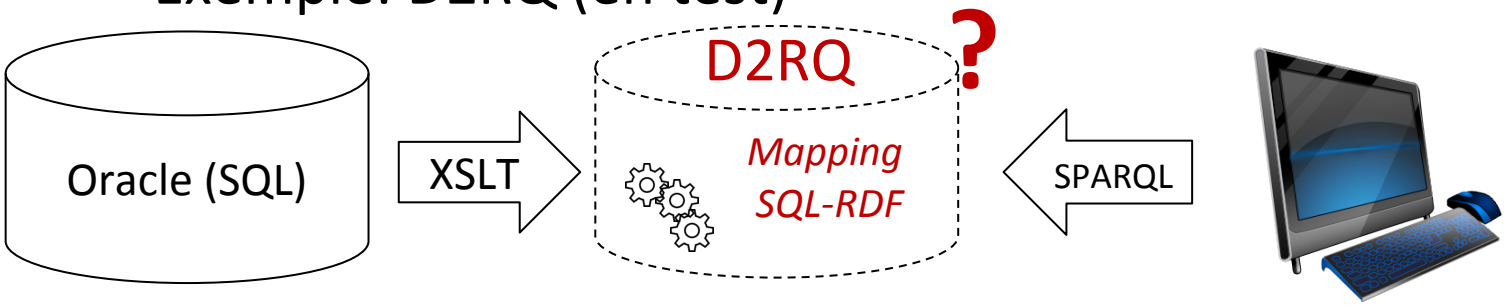
Interroger les données

Un Sparql endpoint pour le Sudoc et les autorités : on y travaille…

- Usages : interopérabilité Sudoc/Hub, BnF…
- Exigences : fraîcheur et exhaustivité
- L'écueil : la volumétrie

  Sudoc + IdRef = près d'un milliard de triplets
- Des alternatives au triplestore ?

  Exemple: D2RQ (en test)



Quel retour sur investissement ?

- Des exemples encore limités de réutilisations (connues)
  - Limités par l'absence d'un requêteur
  - Et de dumps vraiment exploitables
- Mais une montée en compétence réinvestie pour des usages internes

abes
agence bibliographique de l'enseignement supérieur

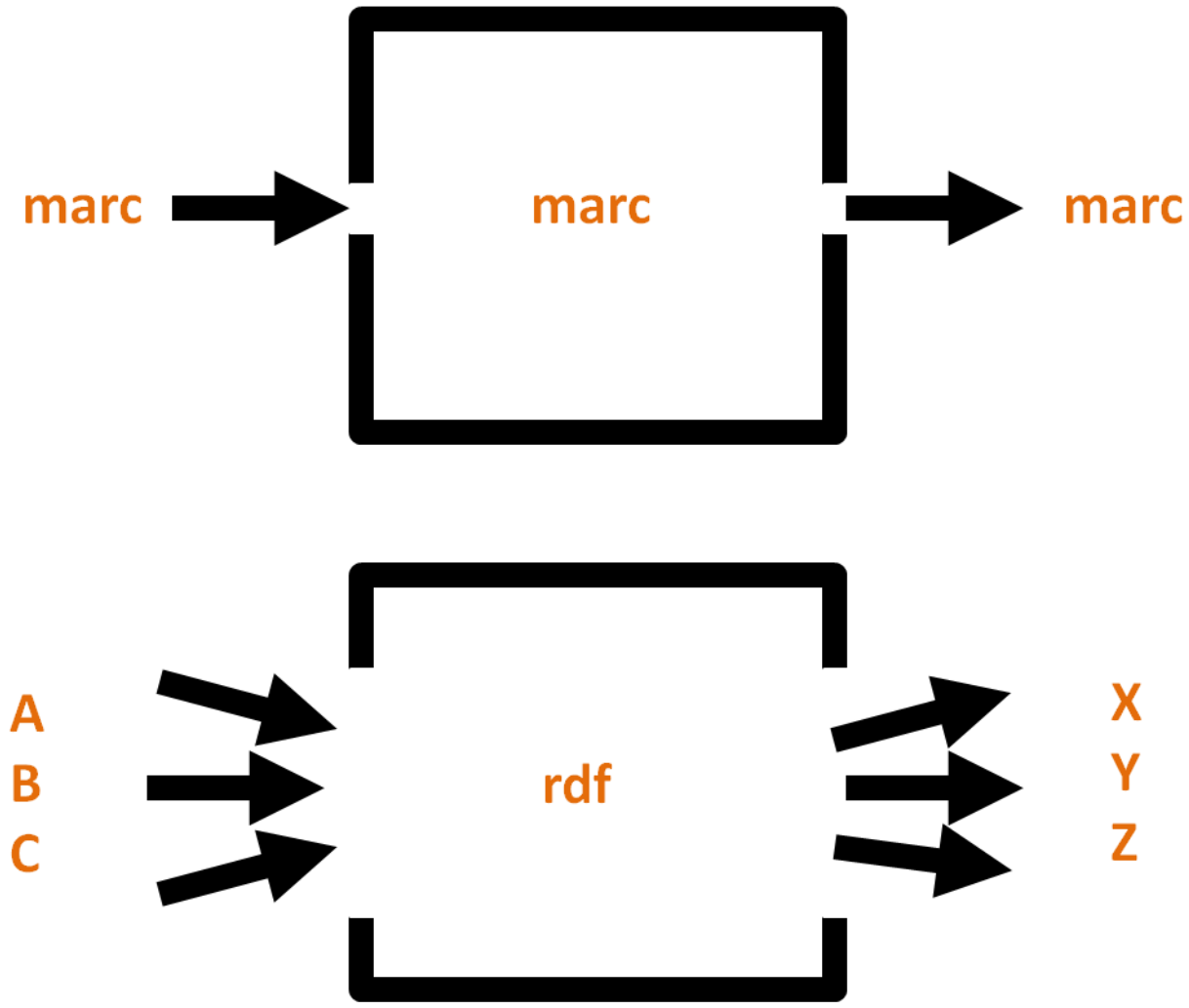
Gérer
en interne
des données hétérogènes

Le “hub de métadonnées” ABES

un hub de métadonnées

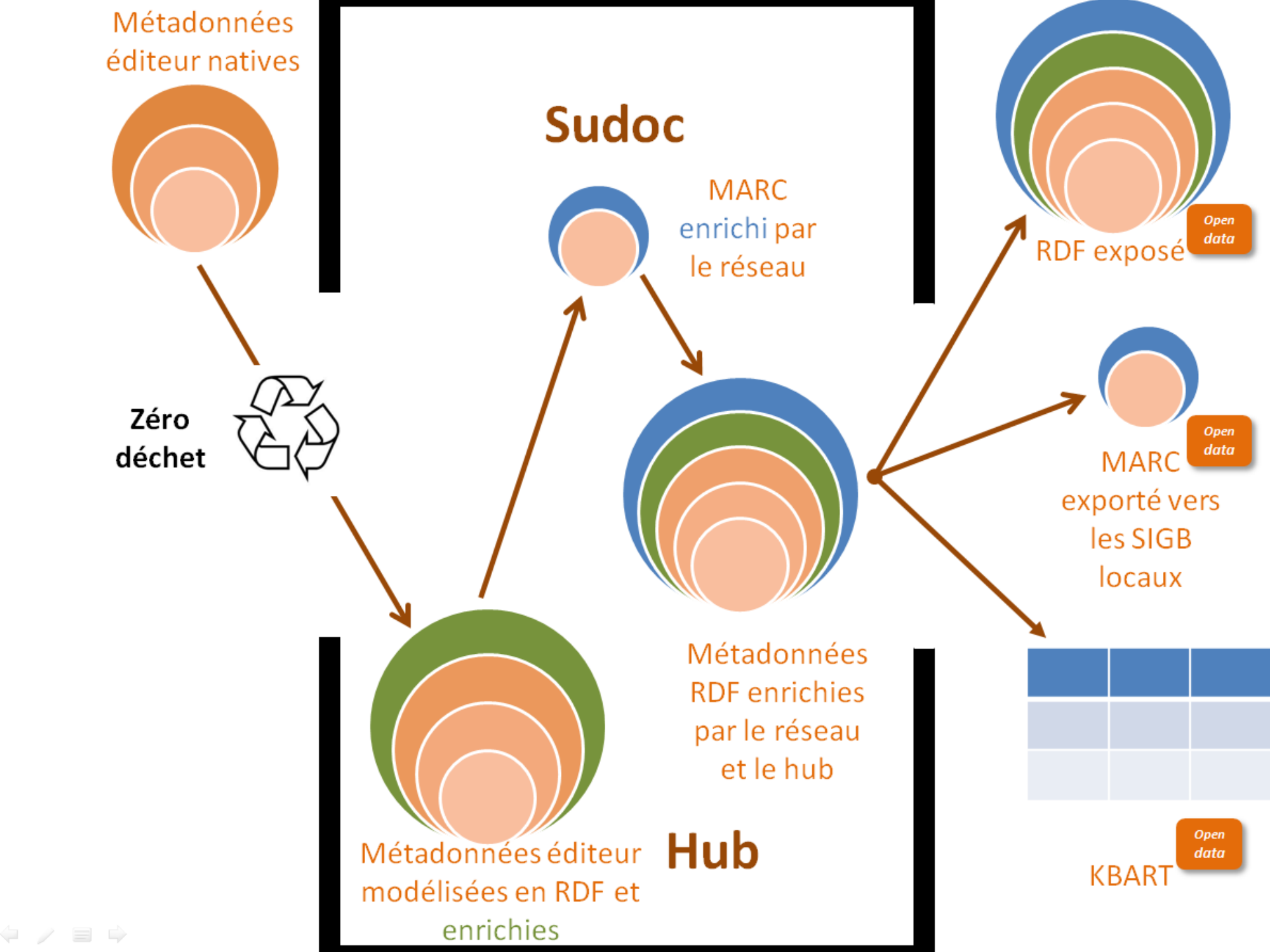
- ~~Une application~~
- ~~Une seule base de données~~
- Une approche
- Terrains d'application :
    - Aujourd'hui : les métadonnées fournies par les éditeurs internationaux dans le cadre du programme ISTEX (achat en masse de littérature scientifique online)

méta-
données
éditeur



catalogues

catalogues
+
discovery
tools
+
…
+
LOD


Métadonnées éditeur natives
Sudoc
MARC enrichi par le réseau
RDF exposé
Open data
Zéro déchet
MARC exporté vers les SIGB locaux
Open data
Métadonnées RDF enrichies par le réseau et le hub
Métadonnées éditeur modélisées en RDF et enrichies
Hub
KBART
Open data

Principes de conception

- **Modélisation et conversion zéro déchet**
  - ne rien perdre des données de départ
  - si nécessaire, forger classes et propriétés sans complexe

- **Corriger/Modifier dans la base RDF**
  - pas dans le format natif

- **Ré-exposer dans le LOD, sans le spammer** *#demain*
  - Si l'éditeur (ou un tiers) a déjà exposé les données, n' exposer que nos enrichissements
  - ***Mais*** quid des corrections/contradictions ?

Conclusions *avec des ?*

- **Résister à la tentation de mettre du semweb partout**
  #fétichisme

- **Cas d'usage** les plus pertinents :
  - Ouverture des données
  - Gestion de données hétérogènes *#hub*

- **Conséquences sur les priorités pratiques** :
  - Sparql OK, **mais** web services simples et efficaces avant
  - Produire les données en RDF ?
    - compliqué si données hétérogènes ?
    - inutile si données homogènes ?

Pour aller plus loin…

Calames

- http://calames.wordpress.com/2008/07/22/calames-yahoo-rdf/

IdRef

- http://punktokomo.abes.fr/2012/05/11/idref-dans-viaf-identifiants/
- http://punktokomo.abes.fr/2011/07/05/idref-des-pages-html-et-rdf-plus-riches/
- http://documentation.abes.fr/aideidref/developpeur/ch03s02.html

Thèses

- http://documentation.abes.fr/aidethesesfr/accueil/ch03.html
- http://punktokomo.abes.fr/2011/07/12/theses-fr-lapi-xml-des-theses/
- http://punktokomo.abes.fr/2011/07/12/theses-fr-lapi-xml-des-personnes/

Sudoc

- http://punktokomo.abes.fr/2011/07/04/le-sudoc-sur-le-web-de-donnees/
- http://documentation.abes.fr/sudoc/manuels/administration/sudoc_rdf/

Hub de métadonnées

- http://fil.abes.fr/2013/07/15/le-hub-de-metadonnees-rapport-final-et-plan-daction/
- http://fr.slideshare.net/abesweb/jabes14-yann-nicolasfocushub

SudocAD/Qualinca

- http://punktokomo.abes.fr/2012/02/02/sudocad-resume-du-projet/
- http://www.lirmm.fr/qualinca

Exemples de réutilisations externes (présentations Jabes 2014):

Julien Sicot, SCD Rennes 2

- http://fr.slideshare.net/abesweb/jabes14-julien-sicotutiliserwebservicesabes-35302040

Yves Tomic, SCD Université Paris Sud

- http://fr.slideshare.net/abesweb/jabes14-yves-tomicapipourquoifaire
- http://punktokomo.abes.fr/2014/02/18/domybiblio/