

Les dieux sont aussi dans la
cuisine

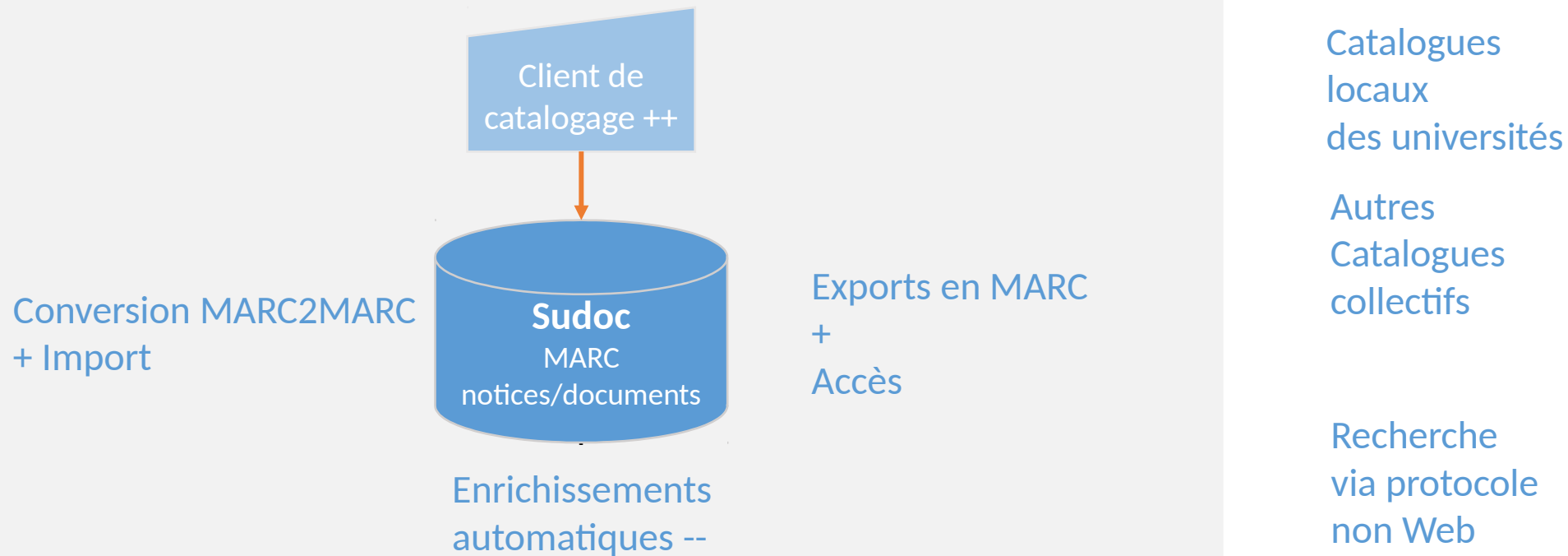
Héraclite

Bien (?) recevoir le web sémantique dans sa cuisine interne

Yann NICOLAS



2000



in

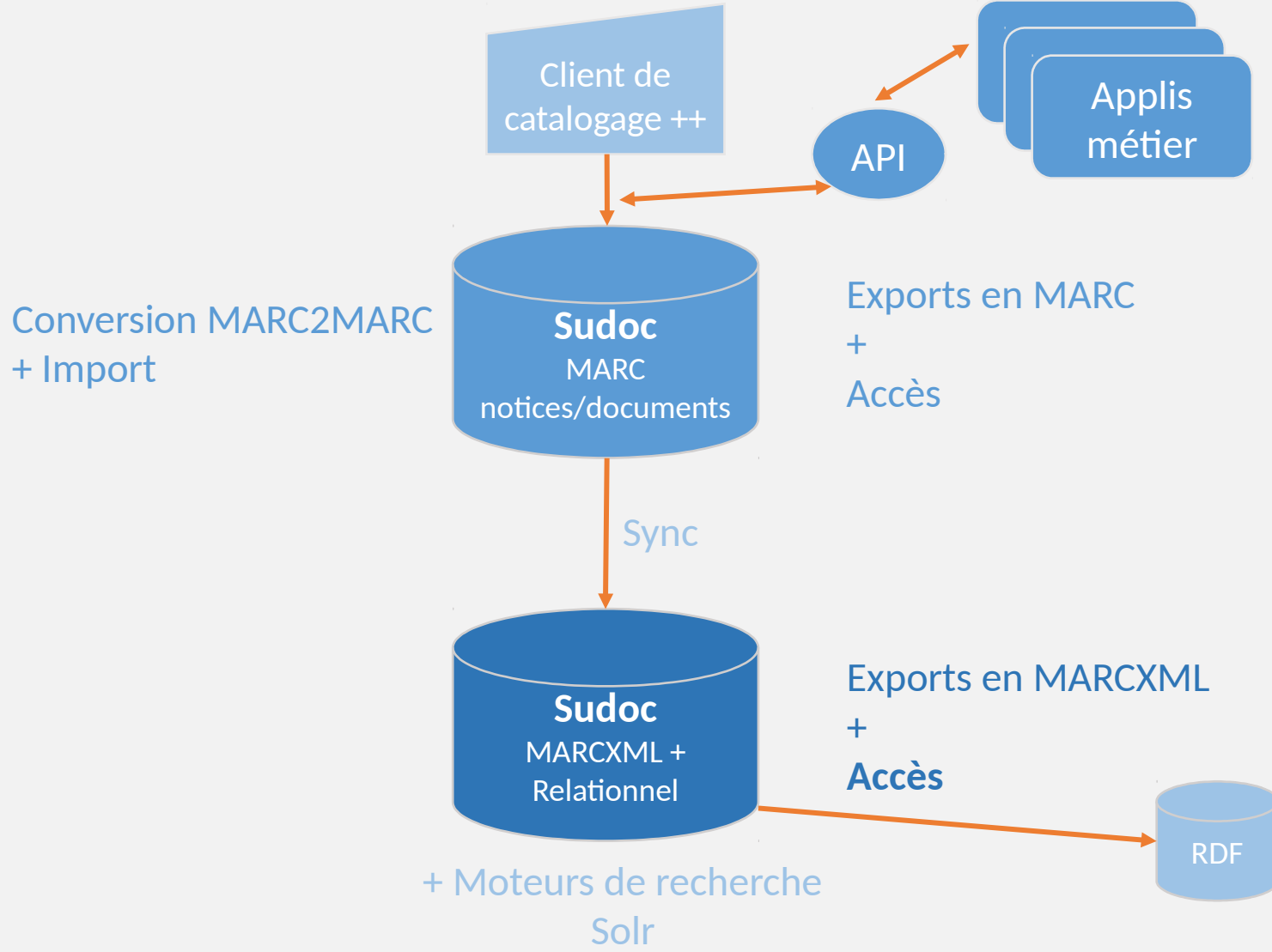
out

Problème

Progiciels + protocoles + formats propriétaires

- Dépendance par rapport au fournisseur (*vendor lock-in*)
- Manque d'autonomie opérationnelle (RH)
 - Ex : création d'index, nouveaux exports, etc.
- Technologies non standards (RH)
- Manque de souplesse

2010



Linked Open Data *****

MicroWebServices

SPARQL endpoint
data.idref.fr
(personnes, concepts,
lieux, etc.)
(2018)

Problème / Solution choisie

Solution propriétaire	Base XML + Solr (en complément)
Dépendance par rapport au fournisseur (<i>vendor lock-in</i>)	... euh ... ORACLE
Manque d'autonomie opérationnelle (RH) Technologies non standards (RH)	Autonomie
Manque de souplesse	La structure de la base évolue en fonction des opportunités ou besoins nouveaux ... au risque de d'une complexité croissante

RDF aurait pu faire l'affaire, non ?

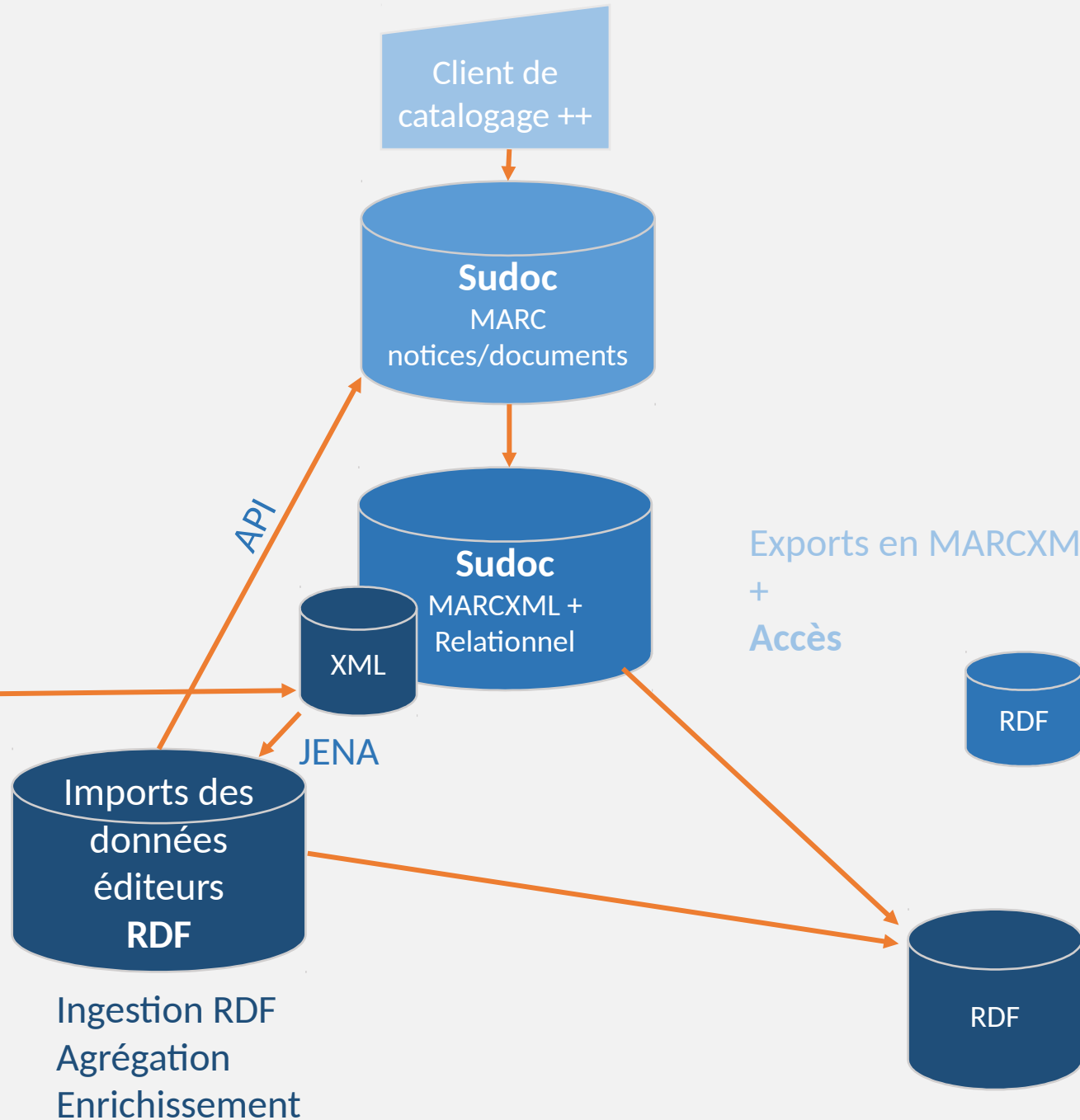
Pourquoi une solution XML en 2010 ?

Pourquoi pas une solution RDF ?

- Moins de familiarité avec les technologies RDF (RH)
- Doutes sur la scalabilité d'une base RDF
- Difficulté de tout modéliser en RDF (80/20)
 - Malgré une expérimentation « RDF concon »
- Performance et *usabilité* des web services : on voulait des web services simples à utiliser et comprendre, performants
 - vs SPARQL
- Performance et richesse du moteur de recherche
 - vs SPARQL

2017

Métadonnées
des éditeurs
scientifiques
(Variété + Volume)



Linked Open Data *****

MicroWebServices

Exports en MARCXML
+
Accès

SPARQL endpoint
data.idref.fr
(personnes, concepts,
lieux, etc.)
(2018)

SPARQL endpoint
(2019)

Ingestion RDF
Agrégation
Enrichissement

RDF

RDF

XML

JENA

Sudoc

MARC
notices/documents

Sudoc

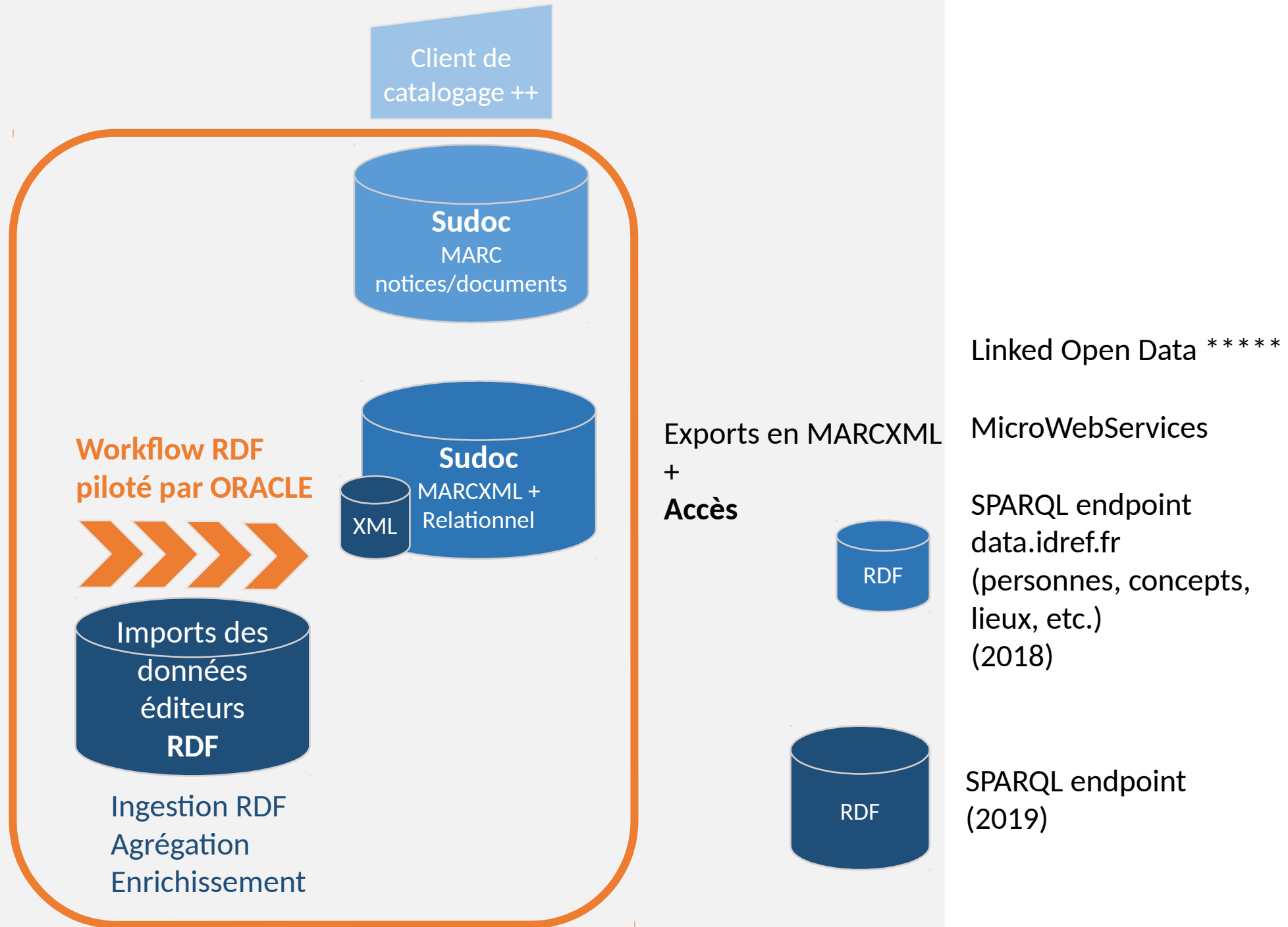
MARCXML +
Relationnel

Imports des
données
éditeurs
RDF

Client de
catalogage ++

Imports et enrichissements

Solution propriétaire en MARC	Base XML	Base RDF
Format MARC restrictif et complexe	DTD différentes selon les éditeurs	Modèle souple et extensible
		Mécanismes natif pour connecter/fusionner les données via les identifiants URI
Technologies ad hoc	Technologies standard	Technologies standard
Technologies ad hoc	Compétences standard et répandues	Compétences standard mais rares



Workflow RDF piloté par ORACLE



Composants du workflow :

- Variables globales
- Requête SPARQL
- Requête SQL
- Procédures stockées

Interface Web pour configurer et exécuter chaque workflow

Autonomie intellectuelle et opérationnelle
des experts données (un peu spéciaux, certes)

Conclusion provisoire

- On vide petit à petit le module propriétaire du SI
 - On déplace des fonctionnalités vers de nouveaux modules
 - On ajoute des logiciels, on complexifie ! □
- Mais :
 - On prépare la suite :
 - On se familiarise avec de nouvelles solutions
 - On étoffe la palette des compétences (RH)
 - On y voit plus clair, peut-être :
 - Chaque composant est spécialisé (plus ou moins)
 - Le workflow est en surplomb des modules □ Vision globale
- On rend les experts données autonomes
 - En travaillant près des données
 - En leur donnant outils et interfaces
 - Nos informaticiens n'ont pas eu à se *convertir* au Web sémantique !