

Introduction à la notion d'interopérabilité des systèmes et des données

Stéphane Pouyllau,

directeur de Corpus-IR / blog.stephanepouyllau.org /
@spouyllau

Aix*Marseille
université



Interopérabilités : contextes

- Le *digital turn*
- Mais une diffusion des données fondée sur des principes hérités du XIXe siècle
- Catalogues « en ligne »
- IHM de recherche complexes et multiples



Photo : S. Pouyllau – cc-by-sa

Interopérabilités : contextes

Résultats ?

- Hétérogénéité des vocabulaires de description des données
- Multiplication des codages des données
- Multiplication des vocabulaires d'indexation
- Les données structurées, les référentiels (thésaurii, vocabulaires contrôlés) passent de mode...

Interopérabilités : contextes



Recherche Google

J'ai de la chance

Interopérabilités : contextes

- Les données (et les métadonnées) basculent dans le web
- On déporte dans « le moteur de recherche » l'ensemble des actions de recherche
- On imagine « les méta-moteurs »
- Des grandes masses de données ont été « déposées » sur le web

Interopérabilités : contextes



Veillez décrire votre requête

Titre



Auteur (nom)



Nom du laboratoire



Année



Trier les résultats suivant

date de publication, écriture

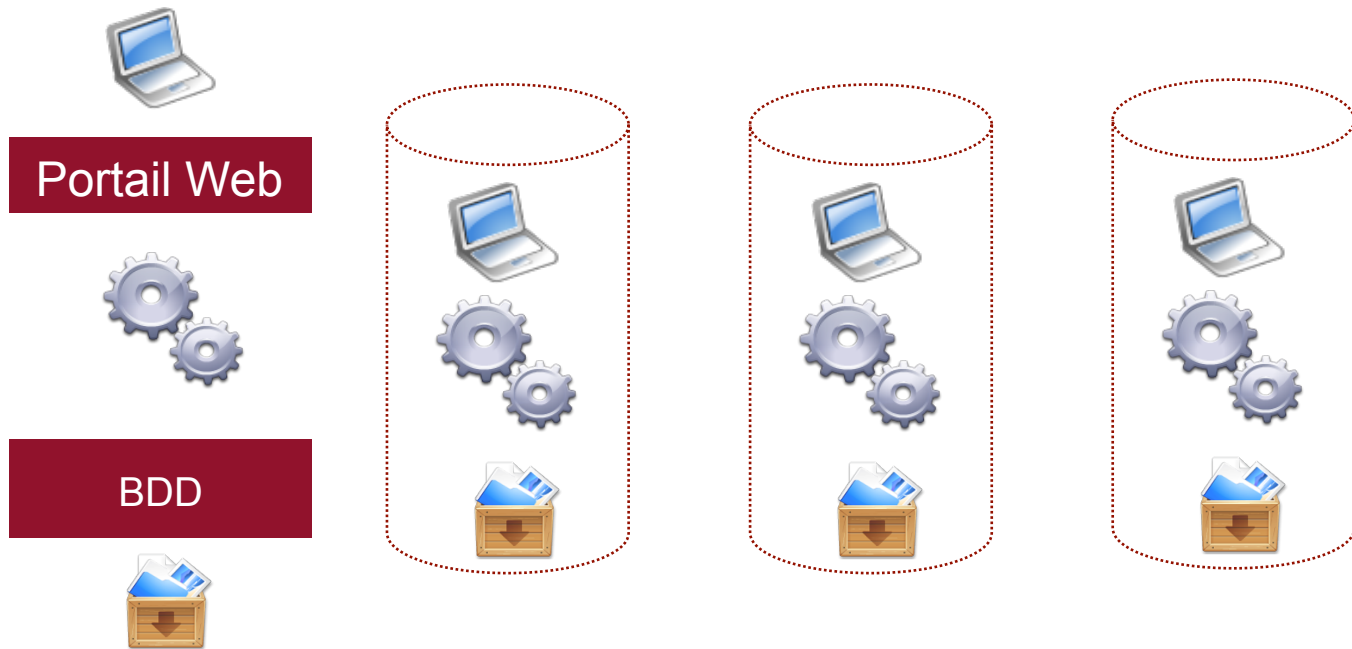
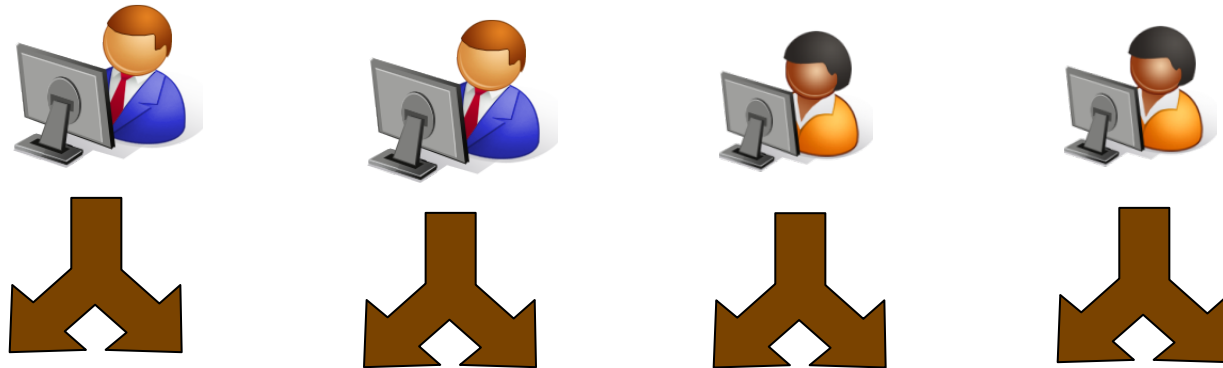
Interopérabilités : contextes

- Le web « invisible » empêche des systèmes automatiques (moteurs de recherche) de venir indexer, rechercher dans des données distantes
- Logique des bdd « en silos »



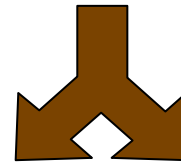
Photo : S. Pouyllau – cc-by-sa

Interopérabilités : contextes



Interopérabilités : définition(s)

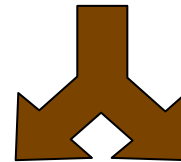
- Le phantasme :
« *pouvoir avec ses propres représentations interroger de multiples bases de données situés dans des lieux différents et ayant des structures différentes* »



Interopérabilités : définition(s)

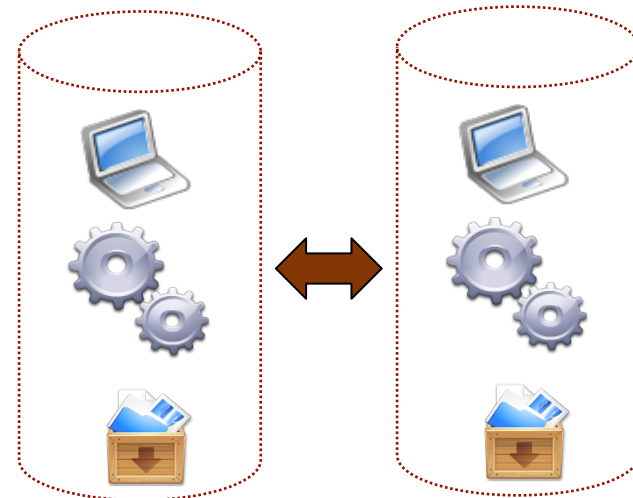
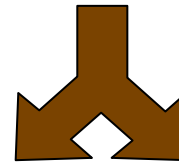
Définition(s):

- Prévoir des modalités d'échanges (métadonnées, données) entre différentes initiatives



Interopérabilités : définition(s)

- Interopérabilités des systèmes ?
 - ODBC
 - Z3950
 - SRU/SRW
 - APIs...
- Limites :
obsolescence rapide,
fragilité informatique,
rigidité...



Interopérabilités : définition(s)



The screenshot shows a web browser window displaying the Open Archives Initiative website. The browser's address bar shows the URL www.openarchives.org. The website features a navigation menu with links for Home, Projects, Specifications, Community, and About OAI. The main content area is divided into several sections:

- Standards for Web Content Interoperability**: A paragraph explaining the OAI's mission to develop interoperability standards for efficient content dissemination.
- Current OAI Projects**: A list of projects including:
 - Protocol for Metadata Harvesting (OAI-PMH), which includes:
 - OAI-PMH Version 2 Specification
 - OAI-PMH Implementation Guidelines
 - OAI-PMH Tools
 - Object Reuse and Exchange (OAI-ORE), which includes:
 - OAI-ORE Specifications and User Guides
 - OAI-ORE Press Release
 - OAI-ORE Production Release Press Release
- Contact us**: A link to openarchives@openarchives.org.
- News from the OAI Community**: A list of recent news items:
 - OAI-ORE Production Release**: Announces the availability of production versions of OAI-ORE specifications and implementation documents, with a link to <http://www.openarchives.org/ore/toc>.
 - April 4 OAI-ORE Roll-out at OR2008**: Reports on a meeting at the University of Southampton to roll-out the third alpha release of the OAI-ORE specifications.
 - March 3 OAI-ORE Roll-out at Johns Hopkins**: Reports on a meeting at Johns Hopkins University to roll-out the second alpha release of the OAI-ORE specifications.

The browser's status bar at the bottom shows a connection to oaisrv.nsd.cornell.edu and the Zotero logo.

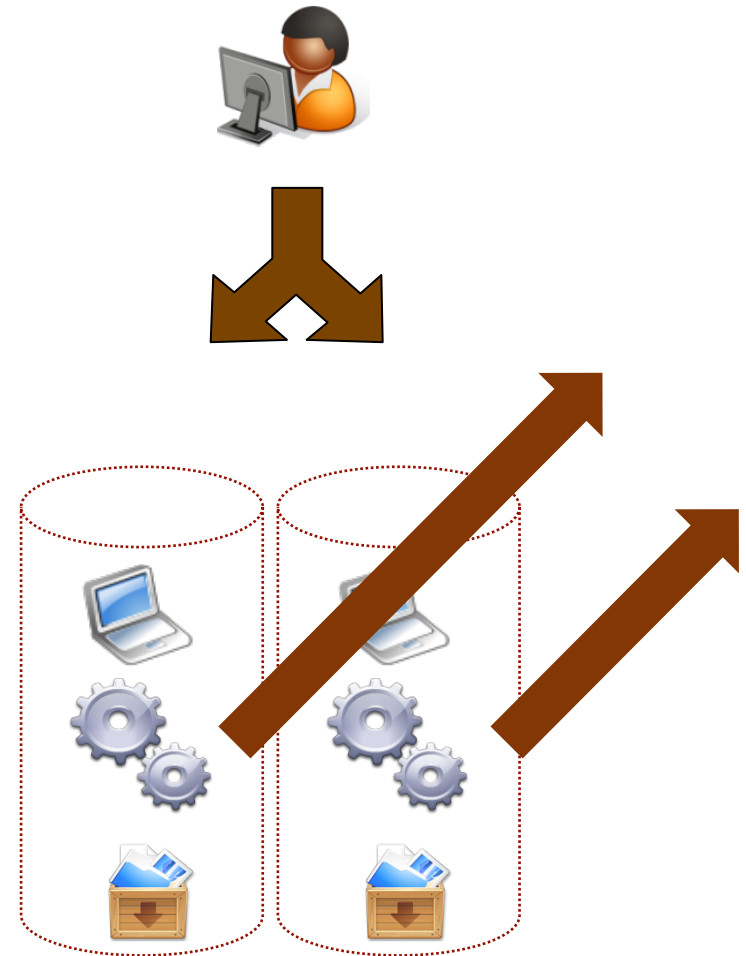
Interopérabilités : définition(s)

Le « dieu » OAI-PMH :

- Ajoute une « couche » (un protocole) au protocole du web HTTP (HyperText Transfer Protocol)
- Pour échanger uniquement des métadonnées (on se rapproche des données)
- Mais des possibilités de requêtes limitées (6 verbes OAI-PMH)

Interopérabilités : définition(s)

- On exprime, via des notices métadonnées, des données
- On propose les métadonnées sous la forme de flux XML
- On standardise les requêtes



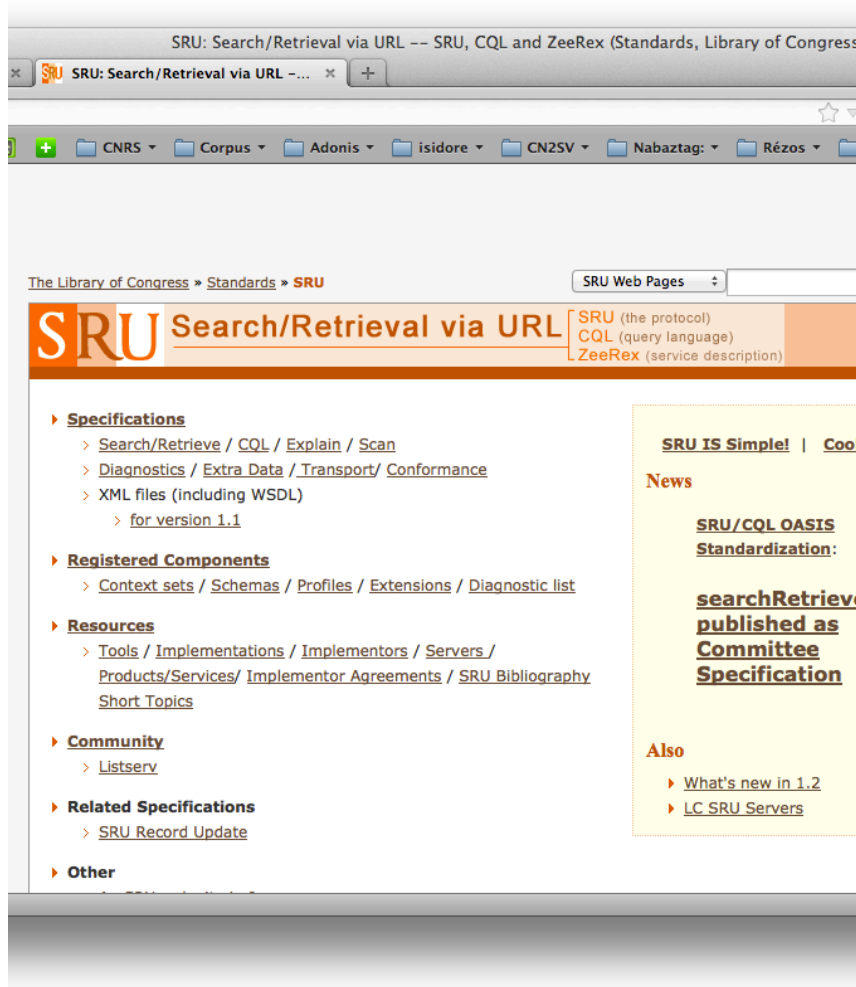
Interopérabilités : définition(s)



Aucune information de style ne semble associée à ce fichier XML. L'arbre du document est affiché ci-dessous.

```
--<OAI-PMH xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/ http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2012-10-12T04:52:56Z</responseDate>
  <request verb="ListRecords" metadataPrefix="oai_dc">http://www.arch.cn2sv.cnrs.fr/oai/oai2.php</request>
  -<ListRecords>
    -<record>
      -<header>
        -<identifier>oai:www.arch.cn2sv.cnrs.fr:hstl-000001</identifier>
        <datestamp>2011-06-06T10:00:00Z</datestamp>
        <setSpec>manuscript</setSpec>
        <setSpec>lamarck</setSpec>
      </header>
      -<metadata>
        --<oai_dc:dc xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/ http://www.openarchives.org/OAI/2.0/oai_dc.xsd">
          -<dc:identifier>
            http://www.lamarck.cnrs.fr/ice/ice_book_detail-fr-text-lamarck-notes_lamarck-3-300.html
          </dc:identifier>
          <dc:identifier>CN2SV:hstl-000001</dc:identifier>
          -<dc:title xml:lang="fr">
            Animaux sans vertèbres. Notes prises aux leçons de Lamarck au Muséum d'histoire naturelle du 29 floréal au 25 thermidor an X
          </dc:title>
          <dc:creator>Jacques-François BISSY</dc:creator>
          -<dc:description xml:lang="fr">
            Quatre cahiers de notes prises aux leçons de Lamarck, datées du 29 floréal an X au 25 thermidor an X (19 mai - 13 août 1802). Le document a été numérisé en mode image et transcrit par R. Bange.
          </dc:description>
          <dc:publisher>Pietro CORSI</dc:publisher>
        --
      </oai_dc:dc>
    </metadata>
  </record>
</ListRecords>
</OAI-PMH>
```


Interopérabilités, point d'étape : systèmes / données ?



Exemple SRU/SRW :

- Interrogation en temps réel entre des machines
- Réponses sérialisées en XML
- Réponse sur la capacité des serveurs « répondants »

Interopérabilités, point d'étape : systèmes / données ?



Open Archi

Home

Projects

Sp

Open Archives Initiative -> Home

Standards for Web Content Interoperability

The Open Archives Initiative develops and promotes interoperability standards that aim to facilitate the efficient dissemination of content. OAI has its roots in the open access and institutional repository movements. Continued support of this work remains a cornerstone of the Open Archives program. Over time, however, the work of OAI has expanded to promote broad access to digital resources for eScholarship, eLearning, and eScience.

Current OAI Projects

- [Protocol for Metadata Harvesting \(OAI-PMH\)](#)
 - [OAI-PMH Version 2 Specification](#)
 - [OAI-PMH Implementation Guidelines](#)
 - [OAI-PMH Tools](#)
- [Object Reuse and Exchange \(OAI-ORE\)](#)
 - [OAI-ORE Specifications and User Guides](#)
 - [OAI ORE Press Release](#)
 - [OAI-ORE Production Release Press Release](#)

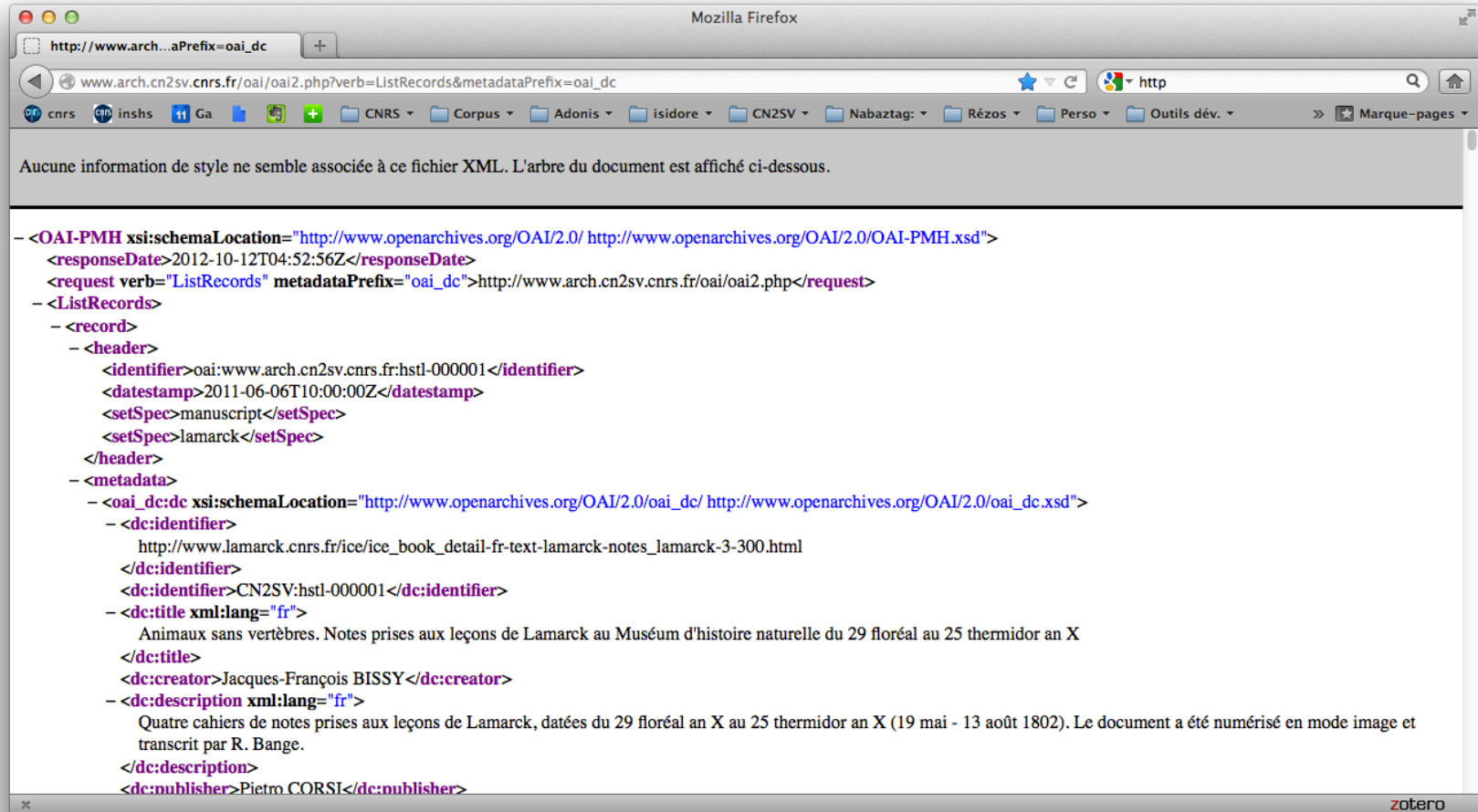
Contact us

- openarchives@openarchives.org

OAI-PMH :

- Moissonnage (on récupère les métadonnées de l'autres en local)
- Protocole normalisé (requêtes, structure XML)
- Possibilité de plusieurs vocabulaires pour décrire les métadonnées

DC element set



```
--<OAI-PMH xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/ http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2012-10-12T04:52:56Z</responseDate>
  <request verb="ListRecords" metadataPrefix="oai_dc">http://www.arch.cn2sv.cnrs.fr/oai/oai2.php</request>
--<ListRecords>
  --<record>
    --<header>
      --<identifier>oai:www.arch.cn2sv.cnrs.fr:hstl-000001</identifier>
      <datestamp>2011-06-06T10:00:00Z</datestamp>
      <setSpec>manuscript</setSpec>
      <setSpec>lamarck</setSpec>
    </header>
    --<metadata>
      --<oai_dc:dc xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/ http://www.openarchives.org/OAI/2.0/oai_dc.xsd">
        --<dc:identifier>
          http://www.lamarck.cnrs.fr/ice/ice_book_detail-fr-text-lamarck-notes_lamarck-3-300.html
        </dc:identifier>
        <dc:identifier>CN2SV:hstl-000001</dc:identifier>
        --<dc:title xml:lang="fr">
          Animaux sans vertèbres. Notes prises aux leçons de Lamarck au Muséum d'histoire naturelle du 29 floréal au 25 thermidor an X
        </dc:title>
        <dc:creator>Jacques-François BISSY</dc:creator>
        --<dc:description xml:lang="fr">
          Quatre cahiers de notes prises aux leçons de Lamarck, datées du 29 floréal an X au 25 thermidor an X (19 mai - 13 août 1802). Le document a été numérisé en mode image et transcrit par R. Bange.
        </dc:description>
        <dc:publisher>Pietro CORSI</dc:publisher>
      </oai_dc:dc>
    </metadata>
  </record>
--</ListRecords>
```

DC Terms

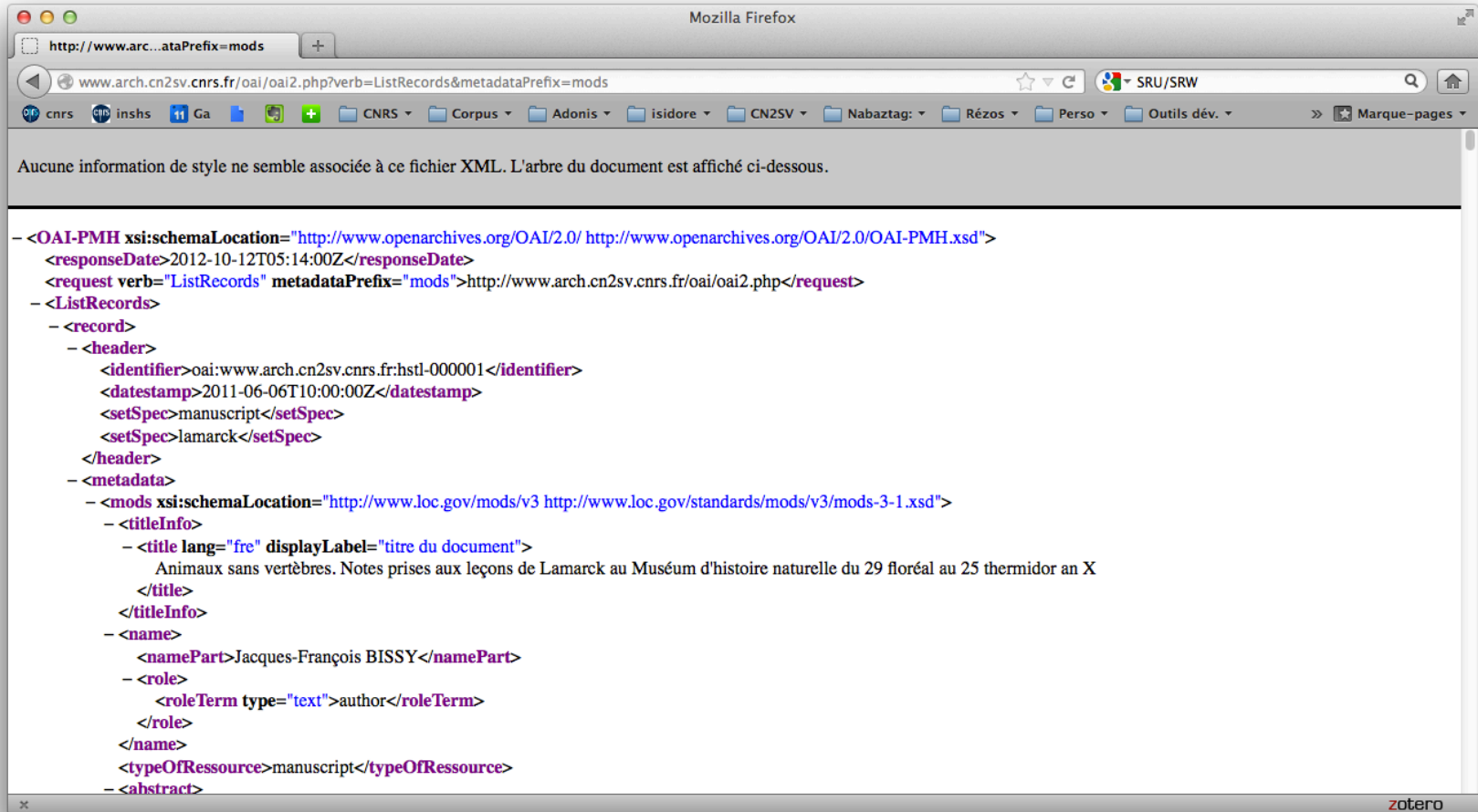


```

--<OAI-PMH xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/ http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2012-10-12T05:12:56Z</responseDate>
  <request verb="ListRecords" metadataPrefix="oai_dcterms">http://www.arch.cn2sv.cnrs.fr/oai/oai2.php</request>
  -<ListRecords>
    -<record>
      -<header>
        -<identifier>oai:www.arch.cn2sv.cnrs.fr:hstl-000001</identifier>
        <datestamp>2011-06-06T10:00:00Z</datestamp>
        <setSpec>manuscript</setSpec>
        <setSpec>lamarck</setSpec>
      </header>
      -<metadata>
        --<oai_dcterms:dcterms xsi:schemaLocation="http://dublincore.org/documents/dcmi-terms/ http://dublincore.org/schemas/xmls/qdc/dcterms.xsd">
          -<dcterms:identifier>
            http://www.lamarck.cnrs.fr/ice/ice_book_detail-fr-text-lamarck-notes_lamarck-3-300.html
          </dcterms:identifier>
          <dcterms:identifier>CN2SV:hstl-000001</dcterms:identifier>
          -<dcterms:title xml:lang="fre">
            Animaux sans vertèbres. Notes prises aux leçons de Lamarck au Muséum d'histoire naturelle du 29 floréal au 25 thermidor an X
          </dcterms:title>
          <dcterms:creator>Jacques-François BISSY</dcterms:creator>
          -<dcterms:abstract xml:lang="fre">
            Quatre cahiers de notes prises aux leçons de Lamarck, datées du 29 floréal an X au 25 thermidor an X (19 mai - 13 août 1802). Le document a été numérisé en mode image et transcrit par R. Bange.
          </dcterms:abstract>
          <dcterms:publisher>Pietro CORSI</dcterms:publisher>
        </oai_dcterms:dcterms>
      </metadata>
    </record>
  </ListRecords>
</OAI-PMH>

```

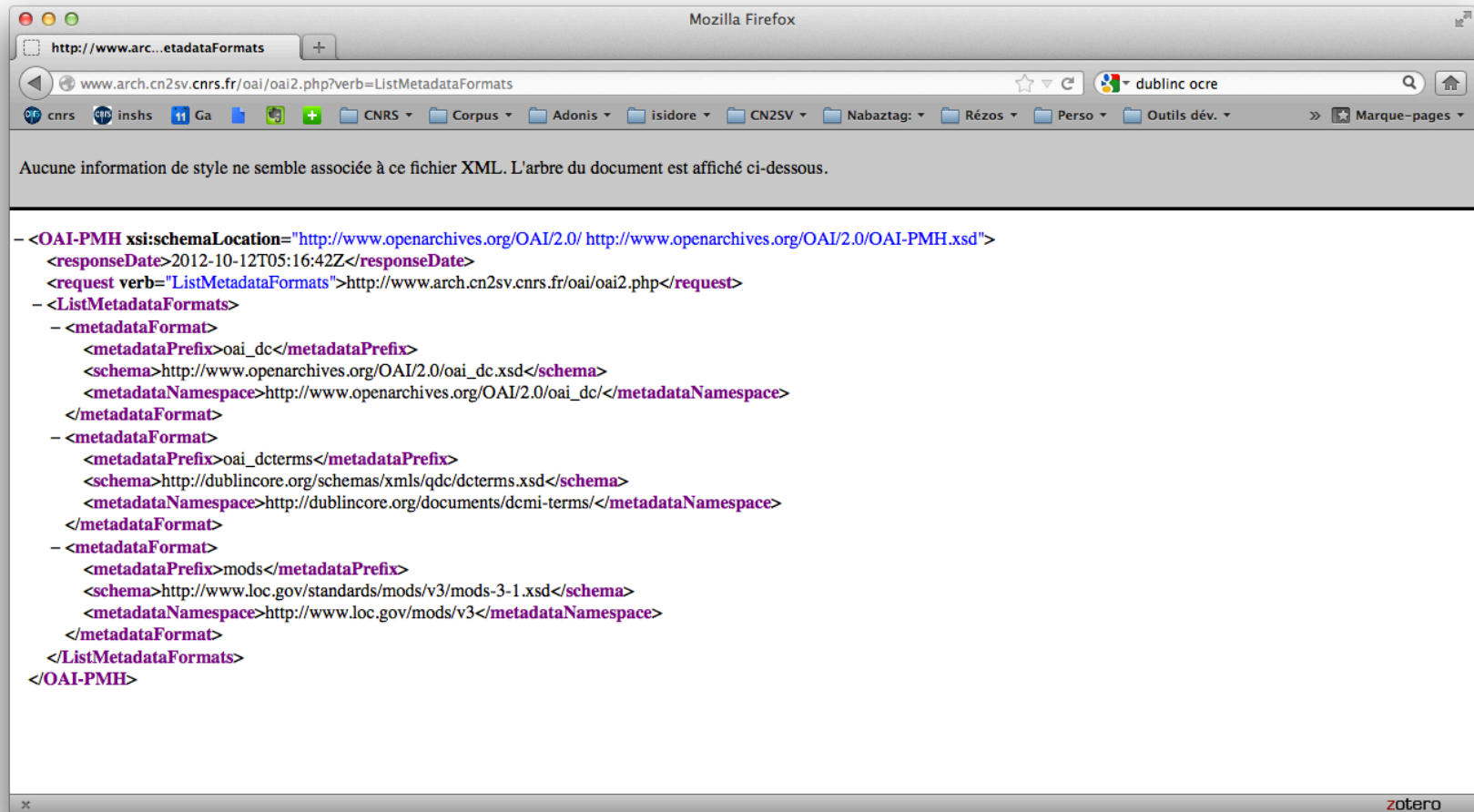
MODS...



Aucune information de style ne semble associée à ce fichier XML. L'arbre du document est affiché ci-dessous.

```
- <OAI-PMH xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/ http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2012-10-12T05:14:00Z</responseDate>
  <request verb="ListRecords" metadataPrefix="mods">http://www.arch.cn2sv.cnrs.fr/oai/oai2.php</request>
- <ListRecords>
  - <record>
    - <header>
      - <identifier>oai:www.arch.cn2sv.cnrs.fr:hstl-000001</identifier>
      <datestamp>2011-06-06T10:00:00Z</datestamp>
      <setSpec>manuscript</setSpec>
      <setSpec>lamarck</setSpec>
    </header>
  - <metadata>
    - <mods xsi:schemaLocation="http://www.loc.gov/mods/v3 http://www.loc.gov/standards/mods/v3/mods-3-1.xsd">
      - <titleInfo>
        - <title lang="fre" displayLabel="titre du document">
          Animaux sans vertèbres. Notes prises aux leçons de Lamarck au Muséum d'histoire naturelle du 29 floréal au 25 thermidor an X
        </title>
      </titleInfo>
      - <name>
        <namePart>Jacques-François BISSY</namePart>
        - <role>
          <roleTerm type="text">author</roleTerm>
        </role>
      </name>
      <typeOfResource>manuscript</typeOfResource>
    - <abstract>
```

Comment je structure...



Aucune information de style ne semble associée à ce fichier XML. L'arbre du document est affiché ci-dessous.

```
- <OAI-PMH xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/ http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2012-10-12T05:16:42Z</responseDate>
  <request verb="ListMetadataFormats">http://www.arch.cn2sv.cnrs.fr/oai/oai2.php</request>
  - <ListMetadataFormats>
    - <metadataFormat>
      <metadataPrefix>oai_dc</metadataPrefix>
      <schema>http://www.openarchives.org/OAI/2.0/oai_dc.xsd</schema>
      <metadataNamespace>http://www.openarchives.org/OAI/2.0/oai_dc/</metadataNamespace>
    </metadataFormat>
    - <metadataFormat>
      <metadataPrefix>oai_dcterms</metadataPrefix>
      <schema>http://dublincore.org/schemas/xmls/qdc/dcterms.xsd</schema>
      <metadataNamespace>http://dublincore.org/documents/dcmi-terms/</metadataNamespace>
    </metadataFormat>
    - <metadataFormat>
      <metadataPrefix>mods</metadataPrefix>
      <schema>http://www.loc.gov/standards/mods/v3/mods-3-1.xsd</schema>
      <metadataNamespace>http://www.loc.gov/mods/v3</metadataNamespace>
    </metadataFormat>
  </ListMetadataFormats>
</OAI-PMH>
```

Métadonnées aux données...

- OAI-PMH est la première « marche » vers une interopérabilité des données « portée » par les données elles-mêmes
- Utiliser les protocoles et standards du web (http, www, url, ...) pour structurer, diffuser, traiter des données
- Web de données, *linked data*

Un entre deux ?

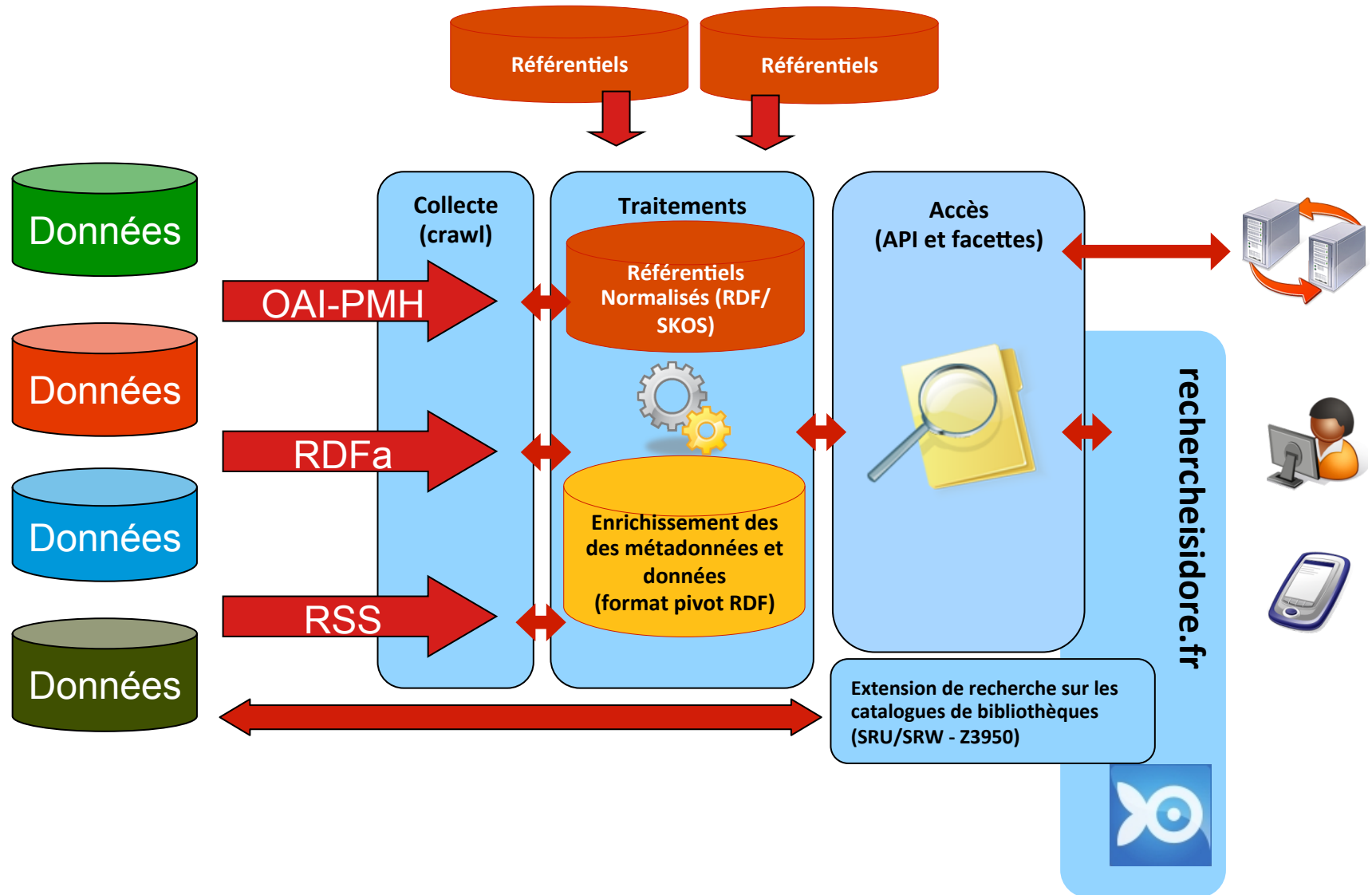
The screenshot shows the ISIDORE website interface. At the top, the browser title is "ISIDORE - Accès aux données et services numériques de SHS" and the address bar shows "www.rechercheisidore.fr". The page features a search bar with the text "Votre recherche" and an "OK" button. Below the search bar, there are two main navigation categories: "Sciences" and "Humaines et Sociales".

The page is divided into several sections:

- NAVIGUER**: A list of navigation options including "Par types de ressources", "Par catégories", "Par périodes historiques", "Par disciplines", "Par nature des sources", and "Par collections et organisations".
- ZOOM SUR...**: A featured section for "Musicologie" with a decorative image of a lute. Below it, there is a sub-section for "Traités de musique, [par Louis CHAVENEAU]" with a description: "Chaveneau, Louis (janv. 1601) Ce manuscrit est de la même main que le précédent ; il provient aussi de Louis Machon, et porte en tête (page I) la même note : « Ex bibliotheca Ludovici Machon, archidiaconi de Portu, canonici Tullensis et capellani".
- CHIFFRES**: A statistics section showing "Collections 65", "Sources 1 753", and "Ressources 1 818 617".
- CONTRIBUER**: A section for "Bibliothèques, centres de documentation, laboratoires et projets de recherche, vous pouvez contribuer à rechercheisidore.fr avec vos données."
- S'INFORMER**: A section for "Séminaire de recherche pluridisciplinaire semestriel 2012-2013 - La prosopographie : objets et méthodes. LARHRA UMR Laboratoire de Recherche".

The footer of the page includes the Zotero logo.


Principe... OAI-PMH, RSS, RDF > RDF / Web



ISIDORE : suivre les besoins

ISIDORE - Accès aux données et services numériques

chercheisidore.fr



Vous êtes ici : Accueil

Votre recherche

Votre recherche OK


Sciences
Humaines et Sociales

> NAVIGUER

- > Par types de ressources
- > Par catégories
- > Par périodes historiques
- > Par disciplines
- > Par nature des sources
- > Par collections et organisations

> SOURCES

> ZOOM SUR...



Traité de musique, [par Louis CHAVENEAU, Louis (janv. 1601)

Ce manuscrit est de la même main que le *Chant de Roland* qui provient aussi de Louis Machon, et porte le n° 1) la même note : « Ex bibliotheca Ludovici archidiaconi de Portu, canonici Tullensis etc »

ISIDORE - Accès aux données et services numériques

chercheisidore.fr/sparql



Vous êtes ici : Accueil > SPARQL endpoint

> SPARQL endpoint

Entrez votre requête SPARQL :

```
SELECT DISTINCT ?Concept WHERE {  
  [] a ?Concept.  
}
```

Format de sortie des résultats :
Tableau HTML

Ok RAZ

> Qu'est ce qu'un SPARQL endpoint ?

Un *SPARQL endpoint* est une interface web qui permet d'interroger de l'information numérique structurée en RDF à l'aide du langage *SPARQL*, langage normalisé et ouvert, développé et maintenu par le *W3C*.

> Avec le SPARQL endpoint d'ISIDORE, à quoi a-t-on accès ?

Le *SPARQL endpoint* d'ISIDORE permet d'interroger l'ensemble des métadonnées moissonnées, structurées et enrichies, qui est accessible au travers du site web rechercheisidore.fr. Les métadonnées moissonnées sont enrichies par des thésaurus, listes d'autorités, vocabulaires). Actuellement, l'ensemble des enrichissements est accessible via le S

Web de données et linked data

- Vers la fin du web invisible ?
- Il ne s'agit plus de « web-ifier » des bases de données
- Ni d'inventer des protocoles « métiers » (OAI-PMH) pour échanger des métadonnées
- Il s'agit d'utiliser naturellement le web pour stocker, diffuser des données