

# Bibliothèques numériques, XML et métadonnées



---

Formation URFIST Bordeaux, 8 décembre 2009  
Catherine Morel-Pair, INIST-CNRS

[catherine.morel@inist.fr](mailto:catherine.morel@inist.fr)



partageable sous [Licence Creative Commons](#)

# Un monde complexe !!!

---



**Pour** ... trouver  
choisir

échanger – 'interopérabilité' -  
conserver

- ... **Nécessité de références communes**
  - Modèles partagés % objectifs ... ex : FRBR
  - ... Formalisés et validés : **standards et normes**
    - ... projets et outils



# Normes et standards (fre) ou 'Standards' (eng)

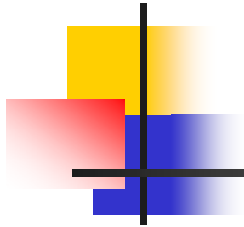
---

## ■ Norme

- Un ensemble de règles de conformité édictées par un organisme de normalisation institutionnelle, comme l'ISO au niveau international, l'AFNOR au niveau français...
- **ISBD, ISO 2709, Dublin Core simple**

## ■ Standard

- un ensemble de recommandations développées et préconisées par un groupe représentatif d'utilisateurs réunis au sein de forums, comme l'IETF, le W3C, le DCMI ...
- **HTML, XML, Dublin Core '+', MODS ...**



- Le standard XML, pourquoi ... comment ?
  - Le fichier XML et ses modèles,
  - Un environnement stratégique et technique avancé
- Métadonnées, généralités, quelques standards et exemples d'utilisations
  - Dublin Core, un standard "universel"
  - Autres standards utilisés en bibliothèques : MODS, LOM ...
  - XMP pour les images
- XML et structuration de documents texte : TEI
- Des outils d'implémentation ?
- ... RDF et le Web sémantique

**Bonus ?**

# eXtensible Mark-up Language

- **Structuration logique** du document ...
  - **Hiéarchique** : livre, chapitres, titres, section niveau 1,2,3, paragraphes ; citations ; notes ...
  - **Sémantique** : marquage d'éléments spécifiques
- ... via des "éléments" représentés par des **<balises>**
  
- **Standard ouvert, pérenne, interopérable** ...
  - Mode texte, Unicode et tous encodages
  - "Métalangage", base de différents modèles et applications
    - Y compris industrie, images, structuration multimédia
- **Standard "efficace" - comprend**
  - Le document XML et ses modèles (DTD, schémas)
  - Des **modules de traitement**
    - XPath, **XSL**, XQuery ... → **transformations et réutilisations**



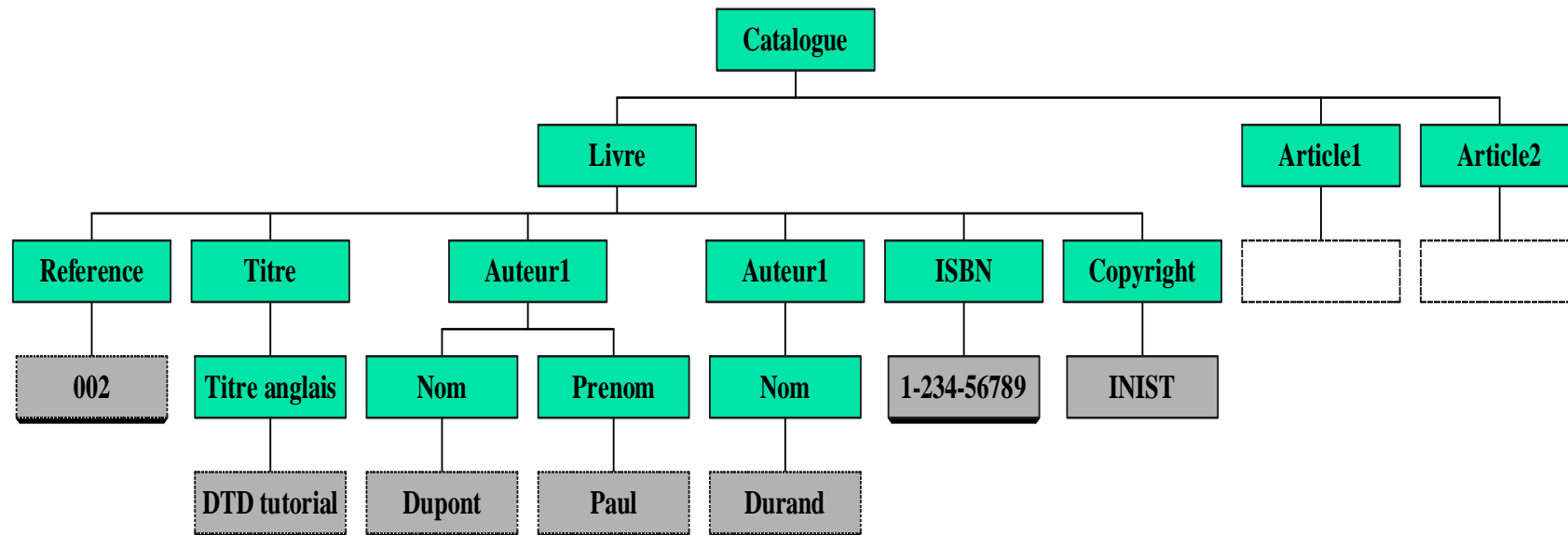
# La famille "Mark-up Language"

---

- **SGML** : représentation des données selon une structure logique avec des **<balises>**, norme ISO 8879:1986
  - Dissocie le contenu de la présentation
  - Très utilisé dans les documents de l'industrie (aéronautique)
  - **COMPLEXE** pour une utilisation large
- **HTML**
  - Application simple de la norme SGML
  - Documents simples pour le Web, liens hypertexte
    - ... **Présentation plus que structuration - vers XHTML**
- **XML**, e**X**tensible **M**arkup **L**anguage, 1998
  - Simplification de SGML, structuration – interopérabilité
  - **Documents, modèles + modules de traitement**

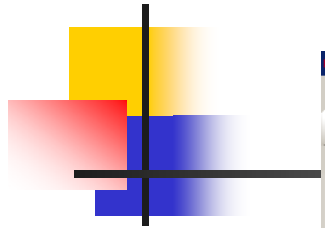
# XML

## Structure arborescente des données



Représentation graphique

# Le document XML



19/02/2





# XML, première démo

---

- Editeurs XML
  - Libres : XMLPad, EditiX, Saxe, XMLmind, JEdit ...
  - Oxygen, XMetal, **XMLSpy (Altova)**
    - cf [Altova](#) : toute une gamme de produits selon besoins
- Le fichier `axe.xml`, ouvert dans XMLSpy
  - Beaucoup de fonctions, de cadres contextuels ...
  - On modifie le fichier
    - Modifier des **valeurs** d'éléments
    - Ajouter des éléments : des enfants à Fille ... puis à Fils
      - **On utiliser l'interactivité du logiciel !**
  - On peut vérifier que le fichier reste 'correct' à chaque fois
    - 'V' : bien formé
    - 'V' : valide par rapport au modèle

# XML, les modèles de données

- Document XML

- **Bien formé** selon la logique XML

- Racine – balises ouvrantes et fermantes

- **Valide** : **conforme à un modèle** décrivant la structure de l'information



- Deux types de modèles

- DTD, Document Type Description

- Ici : `axe.dtd`

- DTD métiers : **BiblioML, TEI, EAD ...**

- **Schéma XML**, plus récent

- Ici : `axe.xsd` - équivalent de `axe.dtd`

- Schémas métiers : **Dublin Core, MODS, TEI ... EAD**



# Modèle de données XML la DTD

- Le modèle de axe.xml, axe.dtd

```
<!ELEMENT ArriereGrandParent (nom, GrandOncle,
GrandParent)>
<!ELEMENT nom (#PCDATA)>
<!ELEMENT GrandOncle (nom, Oncle , Tante)>
<!ELEMENT GrandParent (nom, (Oncle | Tante)*, Parent)>
<!ELEMENT Tante (nom)>
<!ELEMENT Parent (nom, (Fille | Fils)+)>
<!ELEMENT Fils (nom, PetitFille, PetitFils)>
<!ELEMENT Fille (nom , (PetitFils | PetitFille)*)>
<!ATTLIST Fille
    position (cadette | ainée) #REQUIRED>
<!ELEMENT PetitFils (nom)> (...)
```

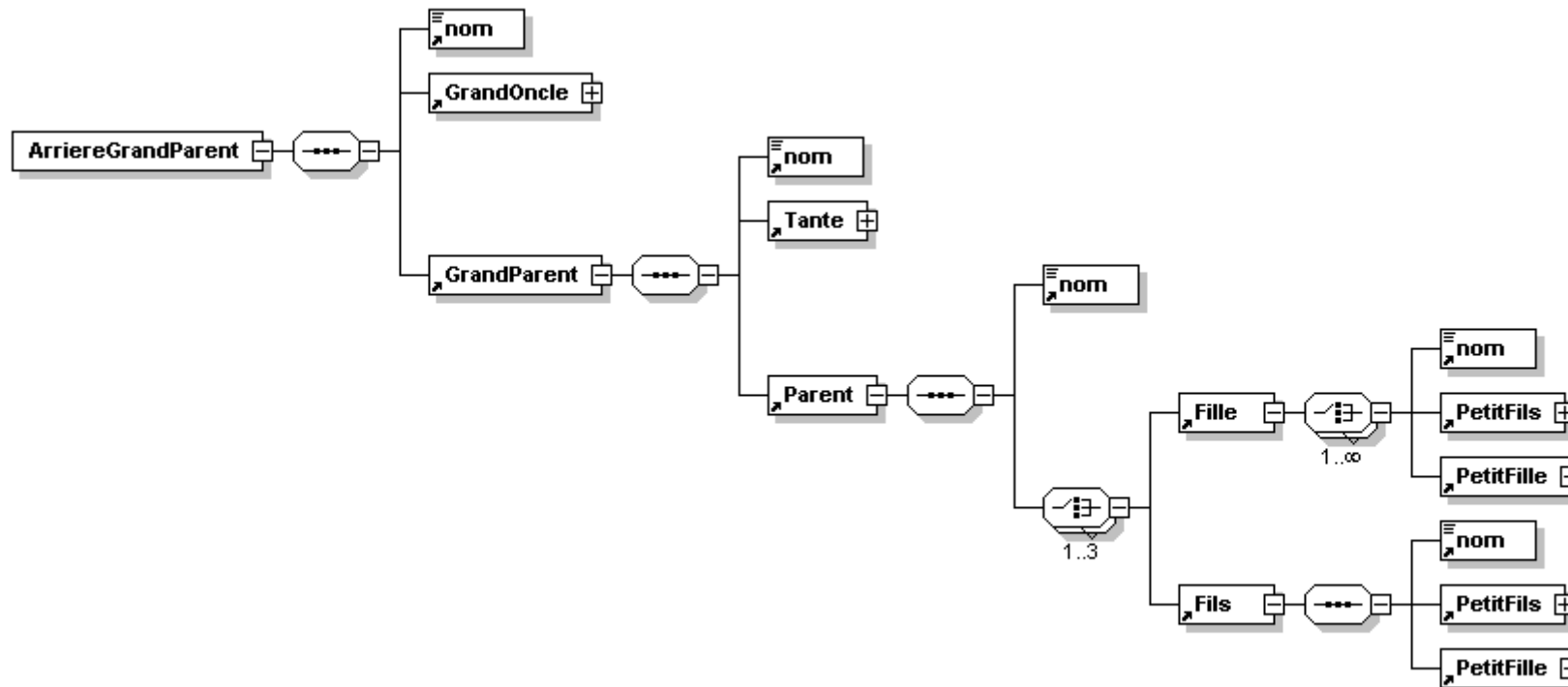


# Modèle de données XML

## le schéma W3C

- **Avantages sur la DTD :**
  - C'est un document XML → mêmes manipulations
  - Déclarer plus de contraintes que les DTD
  - Faire référence à des "espaces de noms"
    - ensemble fermé d'éléments définis, accessible par un URI/URL : métadonnées, vocabulaires ...
      - *Réutilisations, homogénéisation, interopérabilité !*
  - Imports d'autres schémas - modulaire
- **Inconvénient** : lourd !
  - Mais ++ : éditeurs XML → représentations graphiques
- **Autres modèles** pour les documents XML
  - Schémas RelaxNG, Schematron

# Axe.xsd, représentation graphique




Generated with XMLSpy Schema Editor [www.xmlspy.com](http://www.xmlspy.com)



# XML, environnement

---

- Soutenu par de gros consortiums industriels
  - OASIS : Organization for the Advancement of Structured Information Standards, organisme de normalisation industriel international,
    - plus de 600 organismes adhérents : Airbus, General Motors, Dell, Nokia, Oracle, Adobe, Reuters ...
    - Liste des centaines de standards / normes de documents XML + diffuse les spécifications techniques du W3C
  - Standard *effectif* pour l'échange de documents
- Bien implanté dans l'environnement informatique
  - Bureautique, BD, tous systèmes d'exploitation, Web2, fichiers de paramétrage de logiciels variés ...
  - Éditeurs XML performants



Hosted by OASIS

Online resource for markup language technologies

---

**SEARCH**    Advanced Search

Advanced Search

**ABOUT**  
Site Map  
CP RSS Channel  
Contact Us  
Sponsoring CP  
About Our Sponsors

**NEWS**  
Cover Stories  
Articles & Papers  
Press Releases

**CORE STANDARDS**  
XML  
SGML  
Schemas  
XSL/XSLT/XPath  
XLink  
XML Query  
CSS  
SVG

**TECHNOLOGY REPORTS**  
XML Applications  
General Apps  
Government Apps  
Academic Apps

**EVENTS**  
Introductions  
FAQs  
Bibliography  
Technology and Society

**LIBRARY**  
Introductions  
FAQs  
Bibliography  
Technology and Society  
Semantics  
Tech Topics  
Software  
Related Standards  
Historic

**ABOUT**


## XML Applications and Initiatives

**Contents Listing for XML Applications and Industry Initiatives**






Many markup language applications in industry, government, and academia which began in the 1985-1998 timeframe as SGML applications have subsequently migrated to XML using XML DTDs, W3C XML Schemas, or other schema languages. Projects listed below are principally those which began as XML applications in 1998 or later. These applications and initiatives are not (here) categorized under fixed labels, but in many cases one will find related or similar topics grouped together. Should you fail to find a relevant TOC entry here, you may use the [website search facilities](#) or execute a Web browser search in the text of the [Site Index document](#) (load the document, then *Edit -> Find* or *Search -> Find*).

- [W3C Specifications Documentation](#)
- [IEEE Standard DTD](#)
- [OASIS Specification DTDs and Tools](#)
- [Using XML for RFCs](#)
- [Text Encoding Initiative \(TEI\)](#)
- [Markup Language for Complex Documents \(Bergen MLCD Project\)](#)
- [Manuscript Access through Standards for Electronic Records \(MASTER\)](#)
- [XCES: Corpus Encoding Standard for XML](#)
- [Global Document Annotation Initiative \(GDA\)](#)
- [Electronic Metadata for Endangered Languages Data \(EMELD\)](#)
- [Encoding and Markup for Texts of the Ancient Near East](#)
- [Electronic Text Corpus of Sumerian Literature \(ETCSL\)](#)
- [XML System for Textual and Archaeological Research \(XSTAR\)](#)
- [Perseus Project](#)
- [The Making of America II Project](#)
- [Metadata Encoding and Transmission Standard \(METS\)](#)
- [Harvard University E-Journal Archive Project](#)
- [Image Metadata Aggregation for Enhanced Searching \(IMAGES\)](#)
- [Encoded Archival Description \(EAD\)](#)
- [Encoded Archival Context Initiative \(EAC\)](#)
- [Linking and Exploring Authority Files \(LEAF\)](#)
- [STEP/EXPRESS and XML](#)
- [STEPml XML Specifications](#)
- [Atom Publishing Format and Protocol](#)
- [Channel Definition Format, CDF \(Based on XML\)](#)
- [RDF Rich Site Summary \(RSS\)](#)
- [Open Content Syndication \(OCS\)](#)
- [Web Modeling Language \(WebML\)](#)
- [Portable Site Information \(PSI\)](#)

HOSTED BY



SPONSORED BY

XML DAILY NEWSLINK

Receive [daily news](#) updates from Managing Editor, Robin Cover.



# XML et bibliothèques ?

---

- Des standards de métadonnées pour **décrire** les ressources
  - Généralités
  - Standards
    - Dublin Core –MAR-XML, MODS – LOM – METS, EAD ...
  - Usages :
    - OAI-PMH, portails, flux de données XML ...
- Des standards pour **structurer** les documents
  - TEI, [DocBook, Erudit]
    - ... et EAD: un inventaire est "aussi" un document en texte intégral



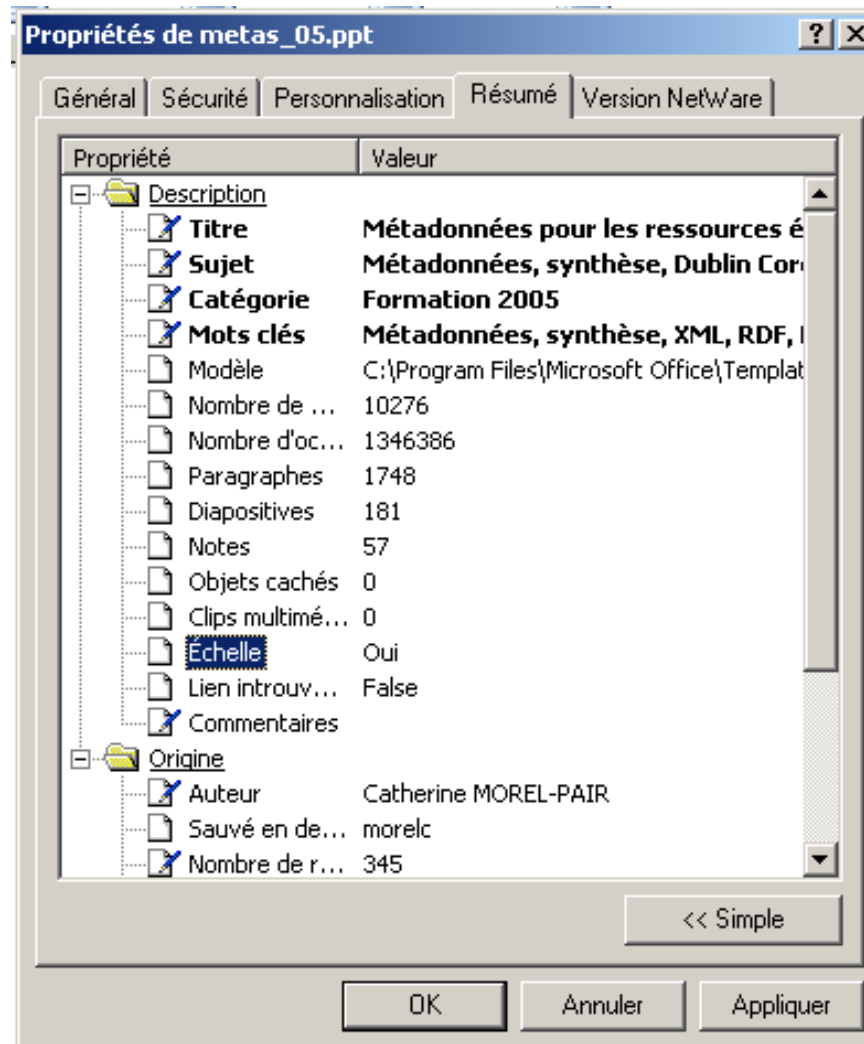


## II - Métadonnées ?

---

- **Métadonnée** : une donnée à propos d'une autre donnée
- **Un jeu de métadonnées** : un ensemble structuré d'information pour "**représenter**" des ressources
  - Ensemble de propriétés/éléments : titre, auteur, type ...
  - Décrit, explicite, localise la ressource
  - Pour en faciliter la recherche, l'usage, la gestion
- **Web et tous ensembles de ressources électroniques**
  - Ou physiques ...
  - Terme ancien dans certains domaines (BD, images, ressources géospatiales), plus récent dans les métiers bib-doc

# Métadonnées courantes Windows



19/02/2013

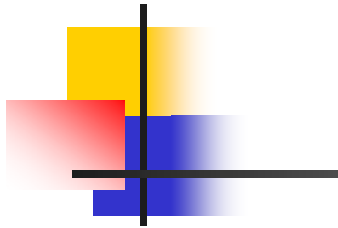


Photo Studio - [eglise\_de\_Ballan-Mire.jpg]

File Browse View Edit Image Tools Window Help

Photo Studio - [eglise\_de\_Ballan-Mire.jpg]

File Browse View Edit Image Tools Window Help

Image Markers Exit Info Thumbnail Camera Info File Info

Field	Value
Time Taken	13:34:43 30/08/2005
Flash	no
Exposure Time	1/200
FNumber	8
ISO	100
Metering Mode	multi-segment
Focal Length	22
Aperture Value	6
Shutter Speed Value	7.64
Brightness Value	8.64
Sensitivity Value	5
Exposure Value	13.64
Make	Canon
Model	Canon EOS 300D DIGITAL
Compression Ratio	3
Exposure Bias	0
Max Aperture	3.61

00000

000001.jpg eglise\_de\_Ball...

Mangaz...

Mangazachi.jpg

uments\lo

19/02/20 Ready

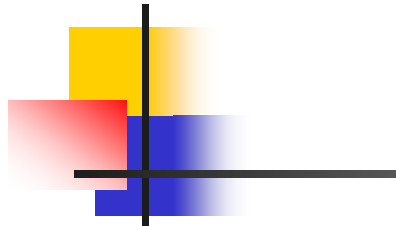
uments\logiciels\Images\Pictures

Déma Ready

NUM

# Photothèque collaborative

## Flick'r



**flickr** Vous n'êtes pas connecté Connexion Aide

Accueil La visite Inscription Explorer Rechercher

**Notes:**  
Title from the Detroit Publishing Co., Catalogue J--foreign section. Detroit, Mich. : Detroit Publishing Company, 1905.  
Print no. 7107.  
Forms part of: Landscape and marine views of Norway in the Photochrom print collection.

**Subjects:**  
Norway--Trondheim.

**Format:** Photochrom prints--Color--1890-1900.

**Rights Info:** No known restrictions on publication.

**Repository:** Library of Congress, Prints and Photographs Division, Washington, D.C. 20540 USA, [hdl.loc.gov/loc/pp/print](http://hdl.loc.gov/loc/pp/print)

**Part Of:** Landscape and marine views of Norway (DLC) 2001699563

More information about the Photochrom Print Collection is available at [hdl.loc.gov/loc/pp/ppz](http://hdl.loc.gov/loc/pp/ppz)

**Persistent URL:** [hdl.loc.gov/loc/pp/ppmsc.06249](http://hdl.loc.gov/loc/pp/ppmsc.06249)

**Call Number:** LOT 13432, no. 147 [item]  
Des notes sont associées à cette photo. Déplacez le curseur sur la photo pour afficher les notes.

**Commentaires**

**Christian Sesseng** [pro](#) dit :  
View from the west-side of Trondheim, Byåsen.  
Posté il y a 2 mois. ([permalien](#))

**dyben** [pro](#) dit :  
I belieaf the title should be. Trondheim (correct spelling) seen from west.  
--  
[Seen in your Photochrom Travel Views set. \(?\)](#)  
Posté il y a 2 mois. ([permalien](#))

**Christian Sesseng** [pro](#) dit :  
These titles are most likely the original ones. And at the time this photo was taken, the correct spelling was Trondhjem.  
Posté il y a 2 mois. ([permalien](#))

**Souhaitez-vous faire un commentaire ?**

[Créez votre compte gratuit](#) ou [connectez-vous](#) (si vous êtes déjà membre).

Byåsen  
Utsikten  
Trøndelag  
photochrom  
color  
colorized  
Nidarø  
Sør-Trøndelag  
Norge  
Noreg  
Nidelv  
Nidelven  
Nidelva

Masquer les tags de programmation (4)

dc:identifier=http://hdl.loc.gov/loc/pp/ppmsc.06249  
xmlns:dc=http://purl.org/dc/elements/1.1/  
geo:lat=63.42118  
geo:lon=10.37071

Informations supplémentaires

Aucune restriction de copyright connue [?]  
 Tout le monde peut voir cette photo

Prise le 1890  
 12 personnes l'ont ajouté à leurs favoris  
 Vue 4 691 fois

19/02/2013

[Accueil](#)

[Supméca](#) ▶ [Accueil](#) ▶ [Plan du site](#)

[Rechercher](#)

[Évènements](#)

[Recrutements personnels](#)

[Appel à fournitures](#)

[Rechercher](#)

[Plan du site](#)



**Supmeca**

**Accueil**

[Évènements](#)

[TIE 2005 à Paris](#)

[Recrutements personnels](#)

[IATOS Enseignants](#)

[Appel à fournitures](#)

[Rechercher](#)

[Plan du site](#)

**Enseignement**

[Admissions](#)

[Rentrée](#)

[Cycle Ingénieur Supméca](#)

[1ère Année](#)  
[2ème Année](#)  
[3ème Année](#)  
[MP](#)  
[SAPI](#)  
[SEM](#)  
[Options Partenaires](#)

[Départements d'enseignement](#)

[MP](#)  
[SAPI](#)  
[SEM](#)  
[SESH](#)  
[Langues](#)

[Débouchés](#)

**Relations Industrielles**

[Taxe d'apprentissage](#)

**Relations Internationales**

[Recherche](#)

[Formation](#)

[Nos partenaires Programmes](#)

[Entreprises](#)

**Recherche**

[Actualité](#)

[Archives 2004-05](#)  
[Archives 2003-04](#)

[Thématiques](#)

[Logistique](#)  
[MDEM](#)  
[Physique des Matériaux](#)  
[Structures](#)  
[Tribologie](#)  
[Vibro-acoustique](#)

[Formations](#)

[Formation Tribologie](#)

[Personnel](#)

[Annuaire Tremblay](#)

### 3. Technologies de l'Information et de la Communication et la santé (Les) - [2003]

Par : M. Majed B - M. Marqué G - [ Site éditeur : ORS Nord - Pas-de-Calais Observatoire Régional de la Santé ]  
résumé, contexte sanitaire des TIC (système de santé, principales orientations politiques en France, aspects  
juridiques), recours aux TIC (population et internet, professionnels de santé), offre des TIC dans la santé, expériences  
TIC dans la région Nord & Pas-de-Calais, analyse, conclusion et perspectives ; bibliographie, 92 pages [France ]

mots-clés : ➔ \*dossier médical informatisé; \*informatique médicale; \*Internet; \*Internet/Utilisation; \*réseaux coordonnés; \*sciences  
information; \*système de santé; \*télémedecine;

types : \*rapport technique; \*recommandation pour la politique de santé;

accès : <http://www.orsnpsc.org/etudes/etude03-4.htm>

### 4. IPM 2002 - 3ème Séminaire International Inter-Universitaire Internet et Pédagogie Médicale, 28-29 novembre 2002 à Lille [2002]

[ Site éditeur : Université du Droit et de la Santé de Lille 2, Faculté de Médecine Henri-Warembourg ]  
présentation générale du Comité et du programme, vidéos des interventions [France Lille]

mots-clés : ➔ \*enseignement et éducation; \*Internet; \*pédagogie;

types : \*actes de congrès; \*matériel d'enseignement audio-visuel;

accès : <http://w3med.univ-lille2.fr/ipm2002/>

### 5. Formation et conseil en soins palliatifs : réflexions pédagogiques - [2001]

[ Site éditeur : Centre de Ressources National François-Xavier Bagnoud ]  
démonstration de l'importance de la formation, de l'enseignement en soins palliatifs [France ]

mots-clés : ➔ \*pédagogie; \*soins aux mourants;

types : \*texte;

accès : <http://www.cdrnfb.org/content/view/119/>

### 6. 13èmes Journées Universitaires Francophones de Pédagogie Médicale - l'enseignant de médecine : rôles et devenir [1999]

[ Site éditeur : CIDMEF Conférence Internationale des Doyens des Facultés de Médecine d'Expression Française ]

Faculté de Médecine, Université de Nantes, du 13 au 16 avril 1999 ; présentation, conférences inaugurales, enjeux des  
industries du savoir, F.M.C., communications libres orales, communications affichées NTIC, communications affichées  
F.M.C., rôle des spécialistes dans la formation initiale [France ]

mots-clés : ➔ \*enseignement et éducation; \*pédagogie;

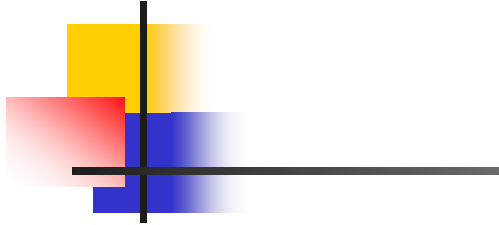
types : \*actes de congrès;

accès : <http://www.cidmef.u-bordeaux2.fr/wnantes/>

### 7. Recommandations pour la formation pédagogique des enseignants de médecine - Guide du conseil pédagogique de la CIDMEF Conférence Internationale des Doyens des Facultés de Médecine d'Expression Française [1998]

[ Site éditeur : CIDMEF - Conférence Internationale des Doyens des Facultés de Médecine d'Expression Française ]

# Un monde sans métadonnées



Source : <http://www.collectionscanada.gc.ca/obj/s37/f2/s37-4016.1-f.pdf>



# → des challenges à relever pour le monde numérique

---

- Comment donner accès à l'information pertinente ?
  - Dans une masse exponentielle, distribuée et hétérogène, pour un arc-en-ciel d'utilisateurs
    - Recherche, tri, évaluation - Interprétation, contextualisation
- Comment garantir l'accessibilité technique et la traçabilité, à moyen et long terme ?
  - Ressources très dépendantes de la technologie
  - Contenu plastique : évolutif, réutilisations ...
- Comment gérer les questions "administratives" ?
  - Propriété intellectuelle, Droits d'accès/usage
- Comment faciliter la mutualisation, les échanges, les réutilisations des ressources ?
  - Interopérabilité → standardisation des formats et descriptions





# Curation digitale et métadonnées

---

- *“Digital curation”*
  - Activités nécessaires pour rendre les documents et collections plus pérennes et interopérables\*
    - Et constituer de vrais entrepôts et infrastructures numériques
  - Digital Curation Center, consortium d’universités, Royaume Uni
  
- **Les métadonnées sont la colonne vertébrale de la curation digitale”**
  - Donnent un **sens structurel et cognitif** à l’information
  - Participent à sa gestion à court et long terme
    - Importantes modèles de préservation long terme
  - Aident à mutualiser : interopérabilité
  - Partageables par tous
    - **Internationales ... multidisciplinaires**

# Un monde de métadonnées ...

“Pour gérer  
la diversité,  
rien ne vaut  
la diversité  
des standards”  
(W3C ?)



# Métadonnées, fonctions une typologie (1a)

- **Métadonnées descriptives du contenu intellectuel**
  - Titre, mots-clés, résumé, classement, type, relations ...
  - **Recherche d'information, labellisation, contextualisation**
- **Métadonnées de structure**
  - Organisation / appartenance – “collection”
  - Structuration d'un document : sections, titres ...
- **Métadonnées techniques, “instancielles”**
  - Identifiant unique, local ou standard
  - Format, dates, “comportement” ...
  - **Gestion et accessibilité techniques**
- **Métadonnées sur les droits**
  - Propriété intellectuelle, responsabilité administrative
  - Droits d'usage // **métadonnées sur utilisateurs**



Métadonnées  
“administratives”



# Métadonnées, fonctions Typologie (1b)

- Métadonnées de préservation
  - Conservation - accessibilité à long terme
  - Descriptives contenu, droits, structure
  - Techniques + + + : accessibilité, traçabilité
  
- ... Métadonnées sur des "acteurs"
  - Utilisateurs
    - Identification → droits d'usage, personnalisation d'applications
    - Comportements
    - Web multimodal : dispositifs de lecture
  - Créateurs de ressources : contexte
  - ... et métadonnées sur des événements, des lieux ...
  
- Métadonnées sur les métadonnées : contexte de production



# A fonctions diverses standards et pratiques divers ...

---

- **Nombreux standards complémentaires**, parfois concurrents
  - Objectifs et objets décrits très divers
  - **Acteurs** divers : bib-doc-archives, NTIC, commerce
  - Complexité variable
- **Éléments de contenu** varié
  - Valeurs texte, contrôlées ou modélisées / texte libre
  - Valeurs complexes : formules, vignettes ...
- **Créées** manuellement / automatiquement / par transformation
- Lisibles par **l'homme et la machine**

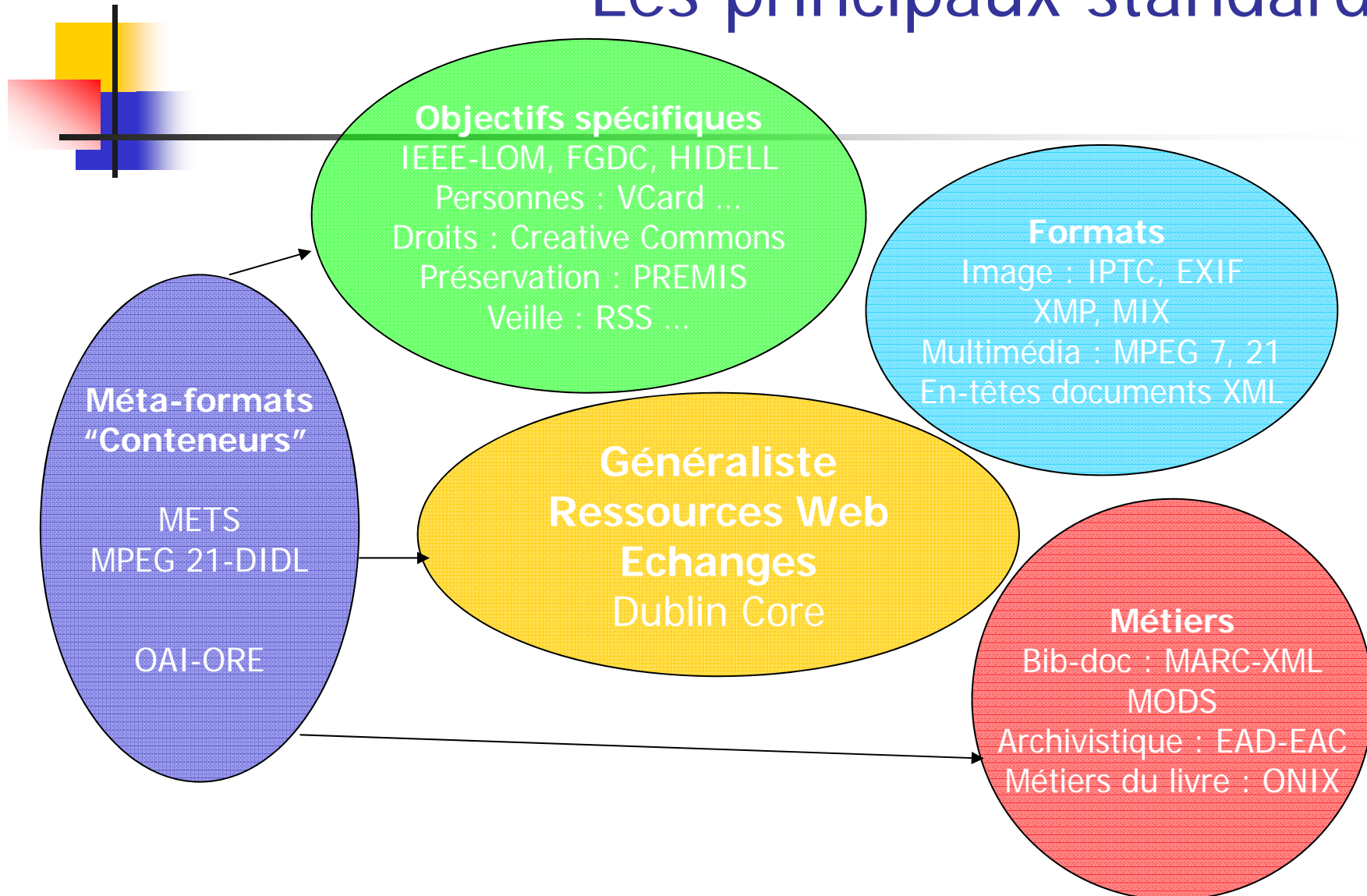


# Embarquées ou externes ?

---

- “Embarquées” : fichier commun ressource-métadonnées
  - Transportées avec la ressource
  - Internes à une section
    - en-tête (header) des documents **HTML**
    - Métadonnées **images** et **multimédia** natives
  - Ou “embedded” dans l’ensemble du fichier : EAD, TEI
- Externes, dans un autre fichier
  - Indispensable pour rechercher, gérer, échanger
  - Catalogues, annuaires, notices bibliographiques
  - Archives ouvertes, protocole OAI-PMH et **entrepôt de métadonnées**
  - Métadonnées image ou multimedia **exportées** → BD

# Les principaux standards





# Ressources générales

- *Métadonnées et XML : des standards efficaces de l'environnement numérique*, C. Morel-Pair, 2007, <http://www.enssib.fr/bibliotheque-numerique/notice-1842>
- *Des métadonnées pour bien utiliser les ressources électroniques*, Journée d'information AFNOR/CG46, 7 juin 2005, Bibliothèque nationale de France, <http://www.bnf.fr/pages/infopro/journeespro/no-Afnor2005.htm>
- *Understanding Metadata*, NISO, 2004 (+ 2007), <http://www.niso.org/standards/resources/UnderstandingMetadata.pdf>
- *Métadonnées : mutations et perspectives*, Séminaire INRIA, 29 sept-3 oct 2008, Dijon, <http://www.inria.fr/actualites/colloques/2008/ist08/programme.fr.html>





## I Ib – Quelques standards ...

---

- *Un mot sur les métadonnées HTML*
- Jeu d'éléments **Dublin Core**
  - éléments, écriture, outils
- Autres jeux XML et bibliothèques
  - MARC-XML, MODS, LOM
  - METS, EAD ...
- **Des applications**
  - OAI-PMH
  - ORI-OAI
  - Thèses électroniques françaises
  - Métadonnées ePub, standard dominant des eBooks
  - Et les logiciels métier ?



# 1 - Dublin Core

---

- Dublin Core Metadata Initiative, 1995,
  - objectif : “Pallier les insuffisances des métadonnées HTML” pour donner une sémantique au Web [...] et améliorer la “resource discovery”
  - Par un standard généraliste
    - “description bibliographique simplifiée”
    - ... pour tout type de ressource, de producteur
  - Fondation active
    - Groupes de travail thématiques évolutifs, workshop annuel
  - <http://dublincore.org>



# Jeu Dublin Core, modules

---

- **DC simple**

- 15 éléments **caractérisés** par 10 propriétés
- **Norme ISO 15836-2003**, RFC 5013, ANSI-NISO 39.85

- **DC qualifié**

- + “qualificatifs” (DCTERMS, “refinements, qualifiers, extensions”) **caractérisés**

- **DC étendu** :

- + éléments issus **d'autres standards** (Learning Object Metadata, IEEE-LOM)
- + **éléments récents** : gestion de collections ...

- **“Encoding schemes”**, standardisation des **valeurs**

- Vocabulaires contrôlés, classifications reconnus, codes (ISO)
- Modèles de valeurs



# Dublin Core, des éléments bien définis !

---

- **Chacun des 15 éléments de base est défini par 10 propriétés** (norme ISO/IEC 11179)
  - **Name (identifiant invariable) et Label (traductions possibles)**
  - **Définition et Commentaire**
  - Version, autorité de gestion (DCMI)
  - Langue du Name (anglais), type du contenu (tous = texte)
  - Obligation et nombre d'occurrences de l'élément
    - tous sont facultatifs et répétables

<http://www-rocq.inria.fr/~vercoust/METADATA/DC-fr.1.1.html>
- **5 propriétés pour chaque qualificatif**

# Dublin Core simple

Contenu intellectuel

<title>  
<subject>  
<description>  
<type>  
<relation>  
<coverage>  
<language>

Droits

<creator>  
<publisher>  
<contributor>  
<rights>  
<source>

Techniques

<date>  
<format>  
<identifier>

# DC, éléments et qualificatifs descriptifs du contenu (1)

Title	Nom donné à la ressource, titre du document	<b>alternative</b> : autre titre (traduit, abrégé ...)
Subject	Mots-clés, codes de classement, sujet. Utilisation de vocabulaires ou classifications contrôlés recommandée (LCC, UDC,DDC LCSH, Mesh)	
Description	Description du document: résumé, table des matières, texte libre	<b>abstract</b> <b>tableOfContents</b>

## DC, éléments et qualificatifs descriptifs du contenu (2)

<b>Coverage</b>	Portée du document : domaine géographique, laps de temps, ou juridiction administrative. Recommandation : utiliser des représentations normalisées de ces types de données.
<b><i>qualificatifs</i></b>	<b>spatial</b> (couverture spatiale). On peut utiliser les codages Point (point géographique), ISO3166 (codes de pays à deux lettres), Box (régions géographiques), ou TGN (dictionnaire de noms de lieux) <b>temporal</b> (couverture temporelle). On peut utiliser les codages Period (intervalle de temps) ou W3CDTF (dates).

## DC, éléments et qualificatifs descriptifs du contenu (3)

<b>Type</b>	<p>Nature ou genre du contenu (grandes catégories). Il est recommandé d'utiliser des termes clairement définis au sein de son organisation. Le Dublin Core définit 12 types dans le vocabulaire DCMITypes ( Text, Event, Image ...).</p> <p>[ Cf aussi CRGI, Canada : <a href="http://www.tbs-sct.gc.ca/im-gi/mwg-gtm/typ-&lt;br/&gt;typ/docs/2003/schem/schem-fra.asp">http://www.tbs-sct.gc.ca/im-gi/mwg-gtm/typ- typ/docs/2003/schem/schem-fra.asp</a> ]</p>	
<b>Language</b>	<p>Langue du document Recommandation : code de langue conforme au format RFC 3066 : ISO 639 (2 ou 3 lettres), subtag optionnel code pays (ISO 3166)</p>	



# DC, éléments et qualificatifs sur les droits (1)

Element	Définition	Q.
<b>Creator</b>	Auteur(s) du document, personne, organisation ou service	
<b>Contributor</b>	Personne, organisation ou service qui contribue ou a contribué à l'élaboration du document	
<b>Publisher</b>	Personne, organisation ou service à l'origine de la publication du document	



# DC, éléments et qualificatifs sur les droits (2)

<b>Source</b>	<p>Ressource dont dérive le document, en totalité ou en partie.</p> <p>Recommandation : dénomination formelle des ressources (URI)</p>
<b>Rights</b> <i>accessRights</i> <i>licence</i> <b>RightsHolder</b>	<p>Droits relatifs à la ressource : copyright ou autre licence, droits d'accès, lien vers le détenteur des droits.</p> <p>L'absence de ces éléments ne présume pas que le document est libre de droits.</p>



# DC, éléments et qualificatifs élément "relation"

<b>Relation</b>	<p>Lien vers une ressource liée. Il est recommandé d'utiliser une dénomination formelle des ressources (URI)</p>
<b><i>Qualificatifs</i></b>	<p>isFormatOf / hasFormat isVersionOf / hasVersion isReplacedBy / replaces isPartOf / hasPart isReferencedBy / references isRequiredBy / requires conformsTo</p>

# DC, éléments et qualificatifs techniques (1)

<b>Format</b>	<p>Format physique ou électronique du document : type de média ou dimensions (taille, durée), matériel et logiciel nécessaires pour accéder au document.</p> <p>Recommandation: termes clairement définis ( types MIME ou IMT )</p>	<p><b>extent</b> : taille ou durée</p> <p><b>medium</b> : support physique ou informatique (format de fichier)</p>
<b>Identifier</b>	<p>Identificateur non ambigu.</p> <p>Recommandation : système de référencement précis ( URI, numéro ISBN...)</p>	<p><i>BibliographicCitation</i></p>

# DC, éléments et qualificatifs

## élément "date"

<b>Date</b>	Date d'un évènement dans le cycle de vie du document. Recommandation : format W3CDTF (AAAA-MM-JJ), norme ISO et Period (DCMI).
<i>Qualificatifs</i>	<b>created</b> : date de création <b>valid</b> : date ou période de validité <b>available</b> : date/période de mise à disposition <b>issued</b> : date de publication <b>modified</b> : date de modification <b><i>dateAccepted</i></b> : <i>date d'acceptation (thèse, article...)</i> <b><i>dateCopyrighted</i></b> : <i>date du copyright</i> <b><i>dateSubmitted</i></b> : <i>date de soumission (article...)</i>

# Dublin Core, éléments et qualificatifs plus récents

<b>audience</b> <b>mediator</b> <b>educationLevel</b> <b>Instructional method</b>	Groupe de personnes à qui le document est destiné.  Méthode d'apprentissage
<b>provenance</b>	Changements signifiants pour authenticité, intégrité, interprétation depuis création : propriétaire, dépôt ...
<b>accrualMedium</b> <b>accrualPeriodicity</b> <b>accrualPolicy</b>	Accroissement collections : Méthode d'ajouts de nouveaux items Fréquence des ajouts Politique d'accroissement



# Dublin Core, intérêt et limites

---

- Consensus international et interprofessionnel
- Sémantique “commune” interdisciplinaire ; “multilinguisme”
- Simplicité – 2 niveaux d’usage
- Extensibilité et flexibilité
- Normalisation ET évolutivité

→ **largement adopté aujourd’hui**

- Description généraliste pour la recherche d’informations
- Ressource plutôt électronique et isolée
- Métadonnées techniques et administratives limitées
- Architecture peu contraignante

→ **... en association avec d’autres jeux**



# Dublin Core, ressources

---

- Métadonnées et Dublin Core, OpenWeb group  
[http://www.openweb.eu.org/articles/dublin\\_core/](http://www.openweb.eu.org/articles/dublin_core/)
- Liste de discussion francophone Dublin Core  
<http://listserv.inist.fr/wwwsympa.fcgi/info/dcmi-fr>
- Site officiel DCMI, <http://dublincore.org>
  - Guide d'usage "Using Dublin Core" ,  
<http://dublincore.org/documents/usageguide/>
    - Généralités - Syntaxes – liens vers documents spécifiques
  - Registre :  
<http://www.dublincore.org/documents/dcmi-terms/>
- Bonnes traductions de documents du DCMI  
<http://www.yoyodesign.org/doc/dcmi/>





# TP: éléments Dublin Core

---

- **Objectif** : découvrir les éléments Dublin Core en décrivant une ressource
- A partir de la notice [crozat\\_notice.htm](#) décrivant une thèse
- ... renseigner la "notice" [crozat\\_DC.xls](#) (papier)
  - 2 onglets : Dublin Core simple ou qualifié
  - Tous les éléments sont facultatifs et répétables
  - **Préciser la valeur** avec un "**attribut**" ???
    - Disponibles : langue ... type
- **Conclusions ...**

## Formats d'écriture interopérables

- **HTML, XHTML** : métadonnées dans la ressource
  - Un exemple : site INIST, et bien d'autres
- **XML** : métadonnées dans un autre fichier
- **RDF** : les 2 possibilités
  
- **Flexibilité** : éléments et qualificatifs facultatifs et répétables, formats divers
- **Extensivité** : association possible avec des métadonnées d'autres origines
  
- **NB** : "modèle abstrait" 2008 → évolutions en cours
  - Un seul espace de nom, des classes, une syntaxe différente




# Créer des descriptions Dublin Core - des outils

---

- Voir <http://dublincore.org>, Tools and Software (sous Firefox)
- Une classification valable pour tout projet !
  - 1 - des outils d'aide à la découverte/création
    - Aide-mémoire DC-assist
    - add-on Firefox d'affichage des métadonnées DC
      - "Firefox Dublin Core Viewer Extension"

page Dernières nouvelles ... Google Grand dictionnaire t... Serveur Miriad Spurl! Spurl bar

CSS Forms Images Information Miscellaneous Outline Resize Validation View Source Options



**Dublin Core Metadata Initiative®**

ABOUT THE INITIATIVE DOCUMENTS GROUPS RESOURCES  
DCMI NEWS TOOLS AND SOFTWARE PROJECTS AskDCMI

> [Tools](#) >

Enter keyword  Search

## Tools and Software

es  
[Finding Metadata \(updates\)](#)  
[Tools for the Creation/Change of Plates](#)  
[Automatic Collection/Gathering of data](#)  
[Automatic Production of data](#)  
[Conversion Between Metadata Formats](#)  
[Integrated \(Tool\) Comments](#)  
[Commercially Available Software](#)  
[Software Tool or Software](#)

### Utilities

**Mozilla Firefox Dublin Core Viewer**  
 This extension adds a button to the list of Dublin Core Metadata embedded in web pages.

**DC-assist**  
 A small, flexible help utility for metadata within existing software.

### Creating Metadata (Templates)

**AHDS; Newcastle DC creator**

**DC-dot (UKOLN)**  
<http://www.ukoln.ac.uk/metadata/>

**DC-dot Java (UKOLN)**, requires  
<ftp://ftp.ukoln.ac.uk/metadata/>

**Dublin Core Metadata**

The following DC metadata was found:

Element	Content	Scheme
title	Dublin Core Metadata Initiative Tools and Software	
description	A listing of tools and software to assist in the creat...	
date	2005-05-10	
format	text/html	
language	en	
publisher	Dublin Core Metadata Initiative	

Details

OK

rechercher :  moteurs Occurrence suivante Occurrence précédente Surligner Respecter la casse

19/02/2013



# DC, outils

## 2 - Convertisseurs

---

- ❖ Exemple : **DC-Dot**, Dublin Core Metadata Editor
  - De métadonnées (X)HTML
    - ... vers des métadonnées DC en (X)HTML, XML, RDF
    - Et vers d'autres jeux d'éléments : MARC, TEI ...
    - Convertisseur + formulaire de saisie + « *extracteur* »
  - Démo
    - Entrer une URL, ex. <http://www.loc.gov/standards/>
    - Résultat généré
    - Un formulaire pour compléter les métadonnées
      - soumettre à nouveau
  - Utilisation des métadonnées DC générées
    - En HTML : coller dans l'en-tête
    - Autres : "notice DC XML" pour catalogue, portail, BD ...



# Dublin Core, outils (3)

## Autres

---

- 3 - Des “templates”, formulaires de saisie
  - DC-Dot
  - Interface des systèmes de gestion de bases de données
- 4 - Des **extracteurs**
  - Externalisent les métadonnées embarquées
- 5 - Des “**outils intégrés**” à des applications
  - Outils **OAI-PMH** et **archives ouvertes**
  - **Plateformes éditoriales** : Lodel, Dspace ... ..
  - Logiciels documentaires et archivistiques
    - → interopérabilité avec le protocole OAI-PMH
  - Logiciels images : embarquées, exportées ... pour **DC-XML**

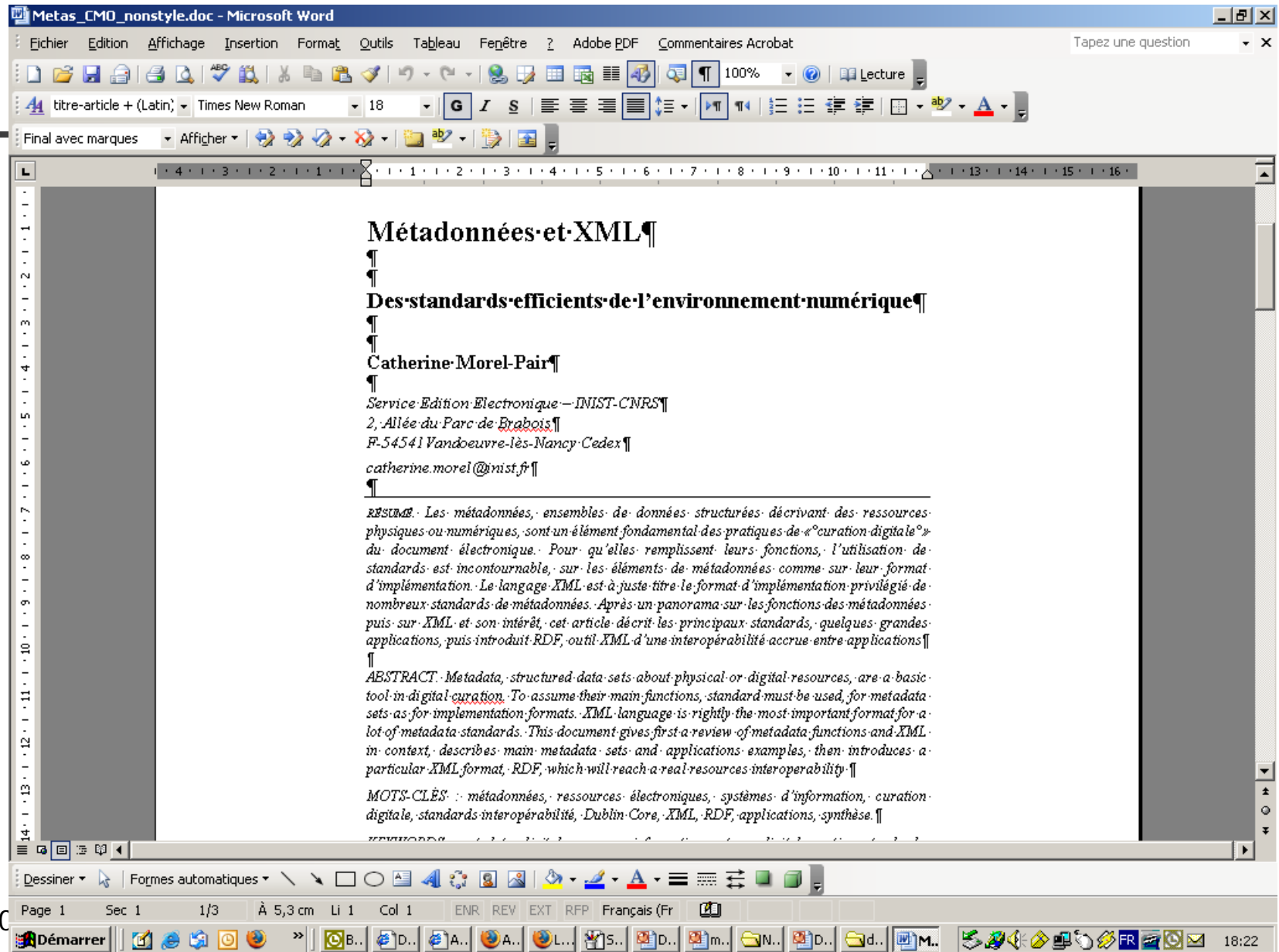


# Lodel et Dublin Core

---

- **Lodel** : plate-forme éditoriale développée par le CLEO
  - Revues SHS
- **Présentation paramétrable**
  - Voir [www.revues.org](http://www.revues.org)
  - et <http://lodel.documents.inist.fr>
- **Organisation arborescente**
  - revue, année, numéro ... par exemple
- **Processus : d'un document Word ou Open Office à la mise en ligne en format XHTML**
  - Document et métadonnées
  - Contrainte : documents **stylés** selon le modèle Lodel
    - Préparation : stylage/marquage des éléments du .doc
    - Soumission, publication

# D'un article publié ailleurs à la publication avec Lodel ...



Metas\_CMO\_nonstyle.doc - Microsoft Word

Fichier Edition Affichage Insertion Format Outils Tableau Fenêtre ? Adobe PDF Commentaires Acrobat Tapez une question

titre-article + (Latin) Times New Roman 18

Final avec marques Afficher

## Métadonnées et XML

### Des standards efficaces de l'environnement numérique

Catherine Morel-Pair

Service Edition Electronique - INIST-CNRS  
2, Allée du Parc de Brabois  
F-54541 Vandœuvre-lès-Nancy Cedex  
catherine.morel@nist.fr

*RESUME* : Les métadonnées, ensembles de données structurées décrivant des ressources physiques ou numériques, sont un élément fondamental des pratiques de « curation digitale » du document électronique. Pour qu'elles remplissent leurs fonctions, l'utilisation de standards est incontournable, sur les éléments de métadonnées comme sur leur format d'implémentation. Le langage XML est à juste titre le format d'implémentation privilégié de nombreux standards de métadonnées. Après un panorama sur les fonctions des métadonnées puis sur XML et son intérêt, cet article décrit les principaux standards, quelques grandes applications, puis introduit RDF, outil XML d'une interopérabilité accrue entre applications.

*ABSTRACT* : Metadata, structured data sets about physical or digital resources, are a basic tool in digital curation. To assume their main functions, standard must be used, for metadata sets as for implementation formats. XML language is rightly the most important format for a lot of metadata standards. This document gives first a review of metadata functions and XML in context, describes main metadata sets and applications examples, then introduces a particular XML format, RDF, which will reach a real resources interoperability.

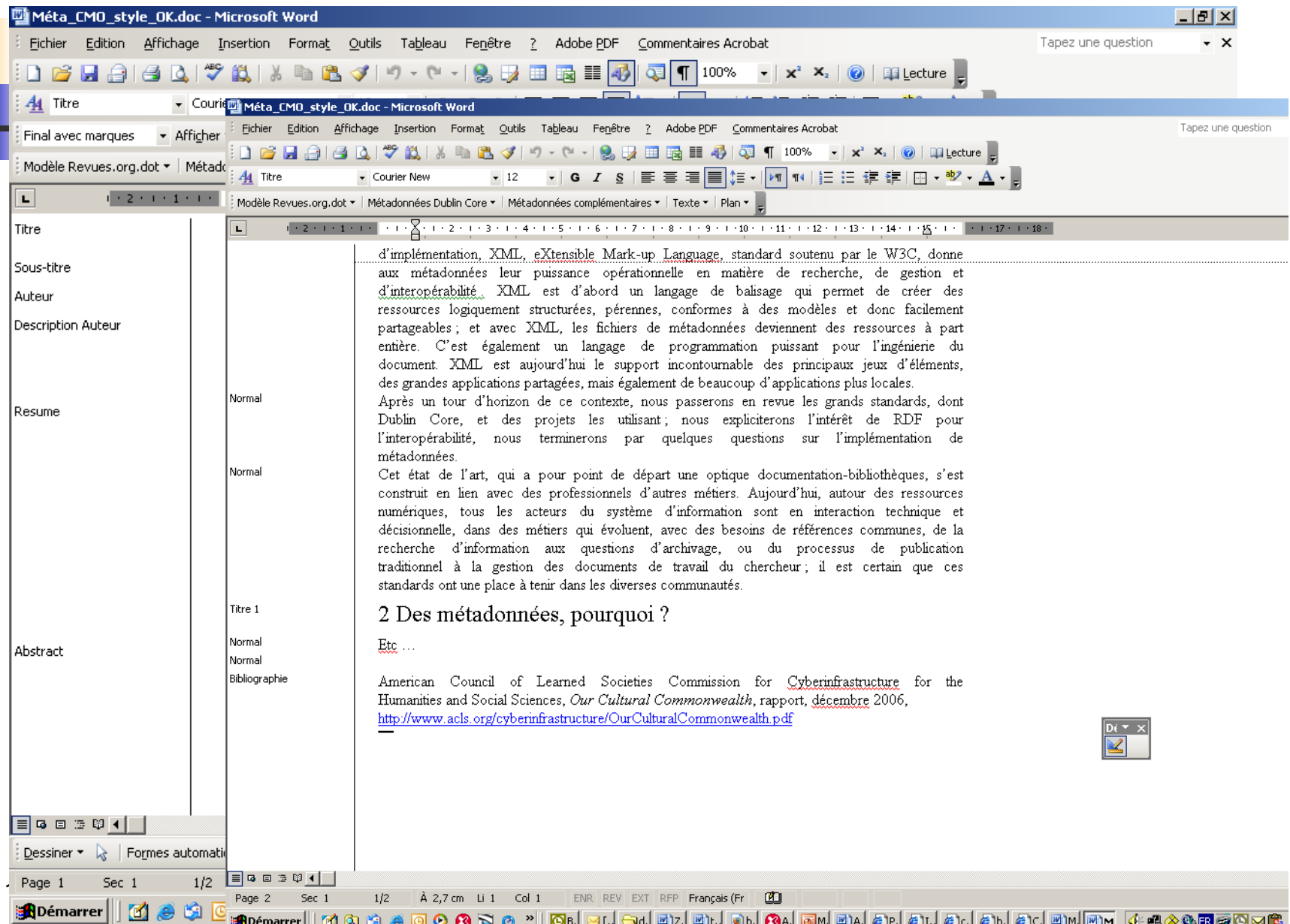
*MOTS-CLÉS* : métadonnées, ressources électroniques, systèmes d'information, curation digitale, standards interopérabilité, Dublin Core, XML, RDF, applications, synthèse.

Page 1 Sec 1 1/3 À 5,3 cm Li 1 Col 1 ENR REV EXT RFP Français (Fr)

Démarrer B... D... A... A... L... S... D... m... N... D... d... M... 18:22



# Article stylé selon le modèle Lodel



Méta\_CMD\_style\_OK.doc - Microsoft Word

Eichier Edition Affichage Insertion Format Outils Tableau Fenêtre ? Adobe PDF Commentaires Acrobat

Titre

Final avec marques

Modèle Revues.org.dot

Titre

Courier New

12

Modèle Revues.org.dot

Métadonnées Dublin Core

Métadonnées complémentaires

Texte

Plan

Titre

Sous-titre

Auteur

Description Auteur

Resume

Abstract

Normal

Normal

Titre 1

Normal

Normal

Bibliographie

d'implémentation, XML, eXtensible Mark-up Language, standard soutenu par le W3C, donne aux métadonnées leur puissance opérationnelle en matière de recherche, de gestion et d'interopérabilité. XML est d'abord un langage de balisage qui permet de créer des ressources logiquement structurées, pérennes, conformes à des modèles et donc facilement partageables ; et avec XML, les fichiers de métadonnées deviennent des ressources à part entière. C'est également un langage de programmation puissant pour l'ingénierie du document. XML est aujourd'hui le support incontournable des principaux jeux d'éléments, des grandes applications partagées, mais également de beaucoup d'applications plus locales. Après un tour d'horizon de ce contexte, nous passerons en revue les grands standards, dont Dublin Core, et des projets les utilisant ; nous expliciterons l'intérêt de RDF pour l'interopérabilité, nous terminerons par quelques questions sur l'implémentation de métadonnées.

Cet état de l'art, qui a pour point de départ une optique documentation-bibliothèques, s'est construit en lien avec des professionnels d'autres métiers. Aujourd'hui, autour des ressources numériques, tous les acteurs du système d'information sont en interaction technique et décisionnelle, dans des métiers qui évoluent, avec des besoins de références communes, de la recherche d'information aux questions d'archivage, ou du processus de publication traditionnel à la gestion des documents de travail du chercheur ; il est certain que ces standards ont une place à tenir dans les diverses communautés.

## 2 Des métadonnées, pourquoi ?

Etc ...

American Council of Learned Societies Commission for Cyberinfrastructure for the Humanities and Social Sciences, *Our Cultural Commonwealth*, rapport, décembre 2006, <http://www.acls.org/cyberinfrastructure/OurCulturalCommonwealth.pdf>

Page 1 Sec 1 1/2

Page 2 Sec 1 1/2 À 2,7 cm Li 1 Col 1 ENR REV EXT RFP Français (Fr)

Importation avec le servOO - Microsoft Internet Explorer fourni par l'INIST

Fichier Edition Affichage Favoris Outils ?

← Précédente → Recherche Favoris Média

Adresse <http://lodel.demo.inist.fr/formation/lodel/edition/oochargement.php?idparent=56&idtype=2> OK Liens >>

Google yahoo Envoyer Mes favoris 25 bloqué(s) Orthographe Traduire Envoyer à Paramètres

**LODEL** logiciel d'édition électronique

Site	formation
Utilisateur	morel ( Administrateur )

Édition Tableau de bord Informations Site Administration

Parcourir Chercher Gérer les index Gérer les index de personnes Options du site Chercher Ok

Racine > Formation Métadonnées > Métadonnées démo

### Importation d'un fichier stylé Ajouter à partir de l'interface

**Sélectionnez le fichier à charger**

Au cours du processus de téléchargement, l'usage des flèches de navigation du navigateur est déconseillé.

Choisir un fichier sur votre disque dur

D:\morelc\mes documents\Metadonnees\Cour: Parcourir...

Choisir un fichier sur le serveur

La conversion est assurée avec l'aide de : **ServOO**

**Validation**

Télécharger Annuler

Terminé

Intranet local

Démarrer Boîte... Le ST... Imp... le bu... d:\m... Micro... Tr : ... Netvi... Alert... 10:39



Édit

Parce

Vérif

Il existe aujourd'hui quelques dizaines de standards de métadonnées, basés sur un modèle de données spécifique, et complémentaires dans leurs objectifs. Par rapport à d'autres formats d'implémentation, XML, eXtensible Mark-up Language, standard soutenu par le W3C, donne aux métadonnées leur puissance opérationnelle en matière de recherche, de gestion et d'interopérabilité. XML est d'abord un langage de balisage qui permet de créer des ressources logiquement structurées, pérennes, conformes à des modèles et donc facilement partageables ; et avec XML, les fichiers de métadonnées deviennent des ressources à part entière. C'est également un langage de programmation puissant pour l'ingénierie du document. XML est aujourd'hui le support incontournable des principaux jeux d'éléments, des grandes applications partagées, mais également de beaucoup d'applications plus locales.

Après un tour d'horizon de ce contexte, nous passerons en revue les grands standards, dont Dublin Core, et des projets les utilisant ; nous expliciterons l'intérêt de RDF pour l'interopérabilité, nous terminerons par quelques questions sur l'implémentation de métadonnées.

Cet état de l'art, qui a pour point de départ une optique documentation-bibliothèques, s'est construit en lien avec des professionnels d'autres métiers. Aujourd'hui, autour des ressources numériques, tous les acteurs du système d'information sont en interaction technique et décisionnelle, dans des métiers qui évoluent, avec des besoins de références communes, de la recherche d'information aux questions d'archivage, ou du processus de publication traditionnel à la gestion des documents de travail du chercheur ; il est certain que ces standards ont une place à tenir dans les diverses communautés.

## 2 Des métadonnées, pourquoi ?

Etc ...

<b>Bibliographie du document</b>	American Council of Learned Societies Commission for Cyberinfrastructure for the Humanities and Social Sciences, <i>Our Cultural Commonwealth</i> , rapport, décembre 2006, <a href="http://www.acls.org/cyberinfrastructure/OurCulturalCommonwealth.pdf">http://www.acls.org/cyberinfrastructure/OurCulturalCommonwealth.pdf</a>
<b>Notes de bas de page</b>	1 "maintaining and adding value to a trusted body of digital information for current and future use", Chris Rusbridge (DCC), in "The challenge of managing and preserving e-Research", <a href="http://www.apsr.edu.au/documents/rusbridge_NLA_talk.pdf">http://www.apsr.edu.au/documents/rusbridge_NLA_talk.pdf</a>

textes ↑

Validation

Continuer

Annuler

Importer à nouveau

Importer sans passer par le formulaire

Importer et visualiser

Métadonnées et XML - Microsoft Internet Explorer fourni par l'INIST

Fichier Edition Affichage Favoris Outils ?

← Précédente → × 🏠 🔍 Rechercher 📌 Favoris 📺 Média 📄 📄 📄 📄 📄 📄

Adresse http://lodel.demo.inist.fr/formation/index.php?id=57

Google Ispy pro free download Envoyer Mes favoris 25 bloquée(s) Orthographe Traduire Envoyer à Paramètres

professionnels d'autres métiers. Aujourd'hui, autour des ressources numériques, tous les acteurs du système d'information sont en interaction technique et décisionnelle, dans des métiers qui évoluent, avec des besoins de références communes, de la recherche d'information aux questions d'archivage, ou du processus de publication traditionnel à la gestion des documents de travail du chercheur ; il est certain que ces standards ont une place à tenir dans les diverses communautés.

## 2 Des métadonnées, pourquoi ?

5 Etc ...

### Bibliographie

American Council of Learned Societies Commission for Cyberinfrastructure for the Humanities and Social Sciences, *Our Cultural Commonwealth*, rapport, décembre 2006, <http://www.acls.org/cyberinfrastructure/OurCulturalCommonwealth.pdf>

### Notes

1 "maintaining and adding value to a trusted body of digital information for current and future use", Chris Rusbridge (DCC), in "The challenge of managing and preserving e-Research", [http://www.apsr.edu.au/documents/rusbridge\\_NLA\\_talk.pdf](http://www.apsr.edu.au/documents/rusbridge_NLA_talk.pdf)

### Pour citer ce document

Catherine MOREL-PAIR, «Métadonnées et XML», *Formations* [En ligne], Formation Métadonnées, Métadonnées démo, mis à jour le : 29/01/2008, URL : <http://172.16.128.66:50010/formation/index.php?id=57>.

### Quelques mots à propos de : Catherine Morel-Pair

Service Edition Electronique - INIST-CNRS  
2, Allée du Parc de Brabois  
F-54541 Vandoeuvre-lès-Nancy Cedex  
[catherine.morel@inist.fr](mailto:catherine.morel@inist.fr)

http://lodel.de Intranet local

Démarrer Démarrer Boîte d... xmlspy... d:\mor... Micros... Tr : Ne... MICHA... Métad... 11:44

## L'article dans Lodel, la source du document

[en-tête : métadonnées de type HTML puis **DC** : ]

```
<meta name="DC.Format" content="text/html" />
<meta name="DC.Language" content="fr" />
<meta name="DC.Identifiant"
content="http://lodel.demo.inist.fr/formation/document.php?id=202" />
<meta name="DC.Title" content="Métadonnées et XML Des
standards efficients de l'environnement numérique" />
<meta name="DC.Subject" content="ressources électroniques" /> [
autres DC.* ...]
```

[Puis corps du texte]

# Dublin Core, HTML et Internet

- Non pris en compte par les “grands” moteurs !

« la question que doivent se poser les webmestres à cet instant est : les « grands » moteurs de recherche comprennent-ils le Dublin Core ? Non. La question qui suit, logiquement est : et donc, par rapport aux meta-tags classiques, que m'apporte le Dublin Core ? Rien. »

## ➤ Signes d'espoir ...

- Moteurs spé – plug-in Firefox
- Sites professionnels + + +
- Usage / outils (DOAJ, Zotero)
- Intérêt des webmasters

« cependant, il est deux choses à rendre en compte :

- le Dublin Core est un standard et plus ce standard sera accepté, plus il sera employé, et par les webmestres et par les moteurs de recherche,
- dans le cas de sites intranet, il existe des outils qui prennent en compte les balises Dublin Core. Si les balises sont bien renseignées, ces outils sont d'une redoutable efficacité ! »

# Fichier Dublin Core en XML

## Déclaration des "espaces de nom"

```
<metadata [...] xmlns:dc="http://purl.org/dc/elements/1.1/"  
xmlns:dcterms="http://purl.org/dc/terms/...">
```

( fait par la machine si on a un schéma pour modèle )

## Dublin Core simple :

```
<dc:language>fr</dc:language>
```

```
<dc:subject>Recherche d'information</dc:subject>
```

## Dublin Core qualifié :

```
<dcterms:available>2002-06</dcterms:available>
```

```
<dcterms:alternative xml:lang="eng">Numeric... ..</dcterms:alternative>
```

- Plus "pro" : attributs, valeurs contrôlées

# Démo

## Dublin Core en XML

■ **Objectif** : compléter une notice Dublin Core en XML en restant conforme au schéma

- À partir de la notice de thèse **contenant les informations** ([crozat\\_notice.htm](#))

■ ... On renseigne le fichier XML avec un éditeur XML, ici XMLSpy

- Une "notice" en DC-XML
- Ajouter quelques éléments
  - Plusieurs possibilités car ordre indifférent
  - S'aider des propositions du logiciel ...
  - **Débutant : valider à chaque modification !!!**



## 2 – Standards bibliographiques Marc-XML

```
<record>  
<leader>01142cam 2200301 a 4500</leader>  
...  
<datafield tag="260" ind1="" ind2="">  
  <subfield code="a">San Diego.....</subfield>  
  <subfield code="b">Harcourt Jovanovich..</subfield>  
  <subfield code="c">c1993.</subfield>  
</datafield> ...  
</record>
```

# MODS

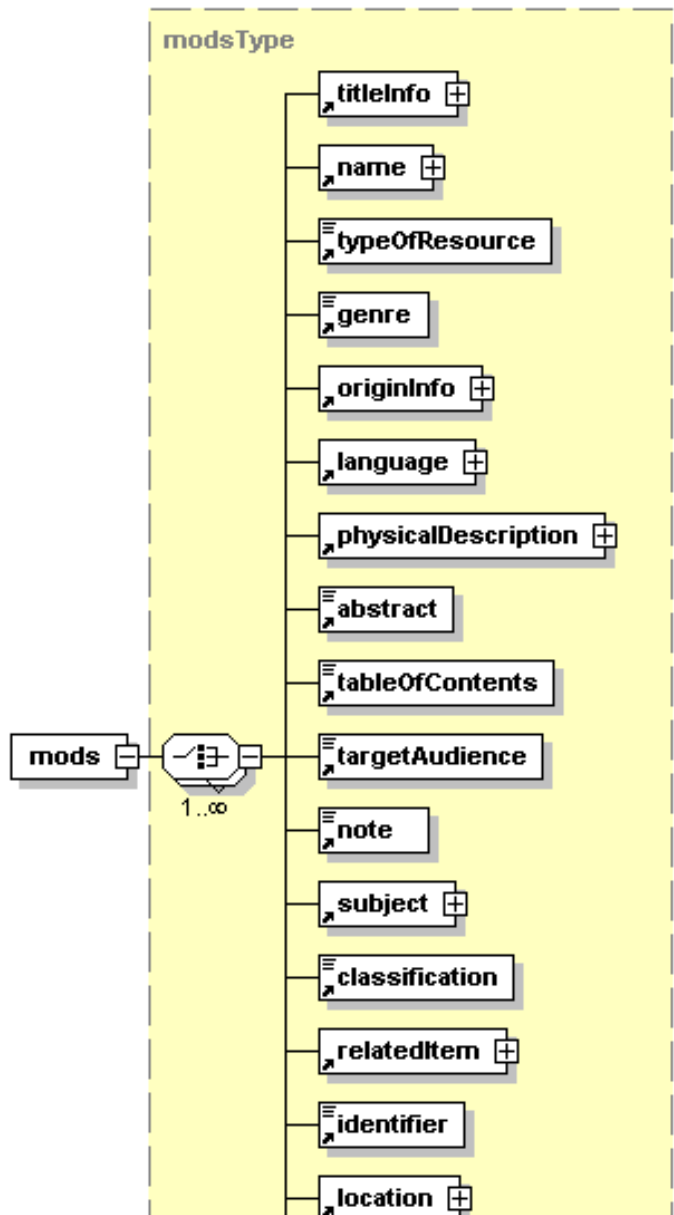
## Metadata Object Description Schema

Sous-ensemble d'éléments MARC verbalisés, "suffisant pour créer un enregistrement complet"

- + convivial, + simple et + orienté document électronique que MARC-XML
- + riche que Dublin Core, + proche habitudes bibliothèques que ONIX
- flexible, bon niveau de granularité
- Né et stabilisé en 2003 - <http://www.loc.gov/standards/mods/>

Applications variées : collections textes, images, AV, partitions, sites Web, bâtiments ...

- implémentées avec des ensembles d'outils variés
- France : base de données DAPHNE, <http://daphne.cnrs.fr>, CN2SV



orelc\mes documents\Metadonnees\Cours2007\2007\_09\_ENSSIB\MODS\modsbook.xml \*

ject XML DTD/Schema Schema design XSL Authentic Convert View Browser Tools Window Help



<b>xmlns:xlink</b>	http://www.w3.org/1999/xlink
<b>version</b>	3.0
<b>xmlns:xsi</b>	http://www.w3.org/2001/XMLSchema-instance
<b>xmlns</b>	http://www.loc.gov/mods/v3
<b>xsi:schemaLoca...</b>	http://www.loc.gov/mods/v3 http://www.loc.gov/standards/mods/v3/mods-3-0.xsd

<b>titleInfo</b>	
<b>title</b>	Sound and fury :
<b>subTitle</b>	the making of the punditocracy /

<b>name</b>	
<b>type</b>	personal
<b>namePart (2)</b>	
<b>type</b>	Abc Text
<b>1 family</b>	Alterman
<b>2 given</b>	Eric
<b>Comment</b>	séparation de nom-prénom et typage, facultatif
<b>role</b>	
<b>roleTerm</b>	
<b>type</b>	text
<b>Abc Text</b>	creator

**typeOfResource** text

**genre** authority=marc

<b>originInfo</b>	
<b>place (2)</b>	
<b>publisher</b>	Cornell University Press
<b>dateIssued (2)</b>	
<b>issuance</b>	monographic

**language**

**physicalDescription**

<b>note (2)</b>	
<b>type</b>	Abc Text
<b>1 statement of responsibility</b>	Eric Alterman.
<b>2</b>	Includes bibliographical

Elements

- titleInfo
- name
- typeOfResource
- genre
- originInfo
- language
- physicalDescription
- abstract
- tableOfContents
- targetAudience
- note

Append Insert A1 A2

Attributes

- ID
- xsi:type
- version

Append Insert A1 A2

Entities

- Ent amp &
- Ent apos '
- Ent gt >
- Ent lt <
- Ent quot "

# MODS pour DAPHNE

Recherche | DAPHNE - Mozilla Firefox

http://www.daphne.cnrs.fr/daphne/search.html

Le CNRS | Département SHS | BiblioSHS

DAPHNE - Données en Archéologie Préhistoire et Histoire sur le NET

Recherch

DAPHNE ?

Bases parten

Recherche

Mon espace

Administrati

Déconnexion

DAPHNE ?

Dases partenaires

Recherche

Mon espace

Administrati

Déconnexion

MAISONS DES SCIENCES DE L'HOMME RESEAU

MAISONS DES SCIENCES DE L'HOMME RESEAU

Résultat abrégé de la recherche

CSV MODS

Afficher la notice complète Retour à la liste des résultats

Titre du document / titre traduit :	Le Dénari des origines
Auteur (A) :	Hachimi Mallika
Titre de revue / nom d'ouvrage déposé :	Le Dénarien
Édition :	
Année millésime / date de publication :	2006
Tomaison/Volumaison :	178
Pagination :	p. 3-39
ISSN/ISSN :	
Type de document :	extra
Mots clés :	paléontologie humaine - paléontologie humaine - désert - animal - environnement - art rupestre - vie quotidienne - Sahara - Afrique du Nord - Afrique occidentale - Préhistoire - Préhistoire - Histoire
Base d'origine :	Fréridq

# BiblioML

■ Initiative du Ministère de la Culture, *Mission de la recherche et de la technologie*, 1999, en lien avec AJLSM

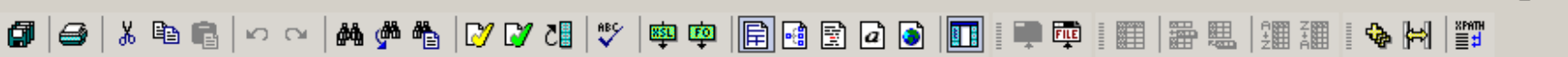
## ■ Représentation XML des formats biblio. et autorités UNIMARC

- modèle : DTD(s), Version 0.5/0.6
  - 5 grands blocs, 224 éléments, une arborescence profonde
- Voir <http://90plan.ovh.net/~adnx/biblioml/doku.php>

## ■ Bibliographie nationale française (2001)

- <http://bibliographienationale.bnf.fr>

## ■ Fonds anciens de la BU de Toulouse, doc numérisés



<b>Language</b>	fre
<b>id</b>	B590092101_00010.337_001
<b>Meta</b>	
CreationDate	Value=20020425
TransactionDate	Value=
Status	Value=c
RecordType	Value=a
BibliographicLevel	Value=m
Completeness	Value=0
Transliteration	Value=
Origin	Role=Original System=

<b>Description</b>	
LanguageInfo TitleScript= f	
<b>TitleAndResponsibility</b>	
Significant	True
<b>Work</b>	
<b>TitleGroup</b>	
<b>Title (2)</b>	
<b>Type</b>	Rbc Text
1	Exposition universelle de 1889 :
2	OtherInfo les institutions ouvrières et sociales du département du Nord /
<b>Responsibility</b>	
<b>Type</b>	Primary
Rbc Text	par Alfred Renouard,... et L. Moy,... ; [éd. par la]Commission de l'économie sociale.

<b>IntellectualResponsibility</b>			
<b>PersonalName (2)</b>			
<b>Role</b>	<b>FormOfName</b>	<b>Entry</b>	<b>OtherPart</b>
1 Primary	Surname	Renouard,	Alfred
2 Secondary	Surname	Moy	Léon-Charles-Mari

Aide en ligne Dictionnaire des Antiquités -- Bibliotheca-tholosana -- Horace, *Opera* - Bâle, 1555

**Recherche :**

Titre du titre  ->

Auteur  ->

Sujet  ->

Imprimeur  ->

Année  ->

[Sujets / par provenances](#)

## Liste alphabétique des titres

**1. Actions forenses de Maistre Simon d'Olive du Mesnil,... Divisées en quatre parties. Ensemble quelques lettres du mesme auteur**

par Olive Du Mesnil Simon d' (15...-1645)  
Imprimeur : Chez Pierre Camusat - A Tolose (1626)

Pertinence : (0.996008)

**2. Adresse pour trouver espoir en desespoir; et repos en adversité par Maistre Touss. Giboult...**

par Giboult Toussaint (....-1581)  
Imprimeur : Chez G. Boudeville - A Tolose (1559)

Pertinence : (0.996008)

**3. Altercacion en forme de dialogue, de l'empereur Adrian, & du philosophe Epictète, contenant soixante questions, & autant de réponses rendu de latin en françois par monsieur maître Jean de Coras...**

par Coras Jean de (1515-1572)

Le SICD a mis en place dans le courant de l'année 2003 une politique de numérisation de documents anciens. Les corpus sont établis selon les thèmes de recherche des universitaires et les ressources des fonds anciens. La numérisation est réalisée en interne. Au 19 février 2008, 335 ouvrages sont accessibles en ligne soit 97368 pages.

----- [Plus d'infos ...](#) -/-

[Flores manuscrites](#)  
[Catalogues de fonds particuliers](#) -/- [Fonds Edouard et Louis Lartet](#)  
-/- [Fonds hispaniques anciens des bibliothèques toulousaines \(ouvrages numérisés\)](#) -----





Aide en ligne Dictionnaire des Antiquités -- Bibliotheca-tholosana -- Horace, *Opera* - Bâle, 1555

Recherche :

mot du titre  
->  
Auteur  
->  
Sujet  
->  
Imprimeur  
->  
Lieu  
->

Sujets / par provenances



**Titre :** **Actions forenses de Maistre Simon d'Olive du Mesnil,... Divisées en quatre parties. Ensemble quelques lettres du mesme auteur**

**Auteur(s) :** Olive Du Mesnil Simon d' (15...-1645) (*Auteur*)

**Imprimeur(s) :** Colomiez. Veuve de Jacques. (15...-16...) imprimeur.  
Camusat. Pierre. (15...-16...)

**Provenance :** Pitteau. Fernand. (1865-1940?) collectionneur.

**Lieu d'éd., Editeur :** A Tolose, Chez Pierre Camusat

**Date de publication :** 1626

**Collation :** [24], 991, [1] p., In-8

**Particularité :** Reliure en basane, dos à nerfs, orné de fleurons en caissons  
Note manuscrite de Fernand Pitteau indiquant que l'ouvrage est sorti des presses de Jacques Colomiez

**Provenance (Note) :** Ancienne collection Pitteau

**Notes :** Lettre ornée, bandeau et marque au titre gravés sur bois. Imprimé par la veuve Colomiez d'après le matériel typographique  
Texte numérisé d'après l'exemplaire Resp Pf XVII-176, de la Bibliothèque universitaire de l'Arsenal (SCD Toulouse 1), consultable en ligne

**Lieu de publication :** Toulouse

**Mots-Clés (libres) :** Parlement de Toulouse (plaidoyer) ;

**Visualiser le document :** [Cliquer ICI](#)

**Document(s) associé(s) :**

Le SICD a mis en place dans le courant de l'année 2003 une politique de numérisation de documents anciens. Les corpus sont établis selon les thèmes de recherche des universitaires et les ressources des fonds anciens. La numérisation est réalisée en interne.  
Au 19 février 2008, 335 ouvrages sont accessibles en ligne soit 97368 pages.  
----- [Plus d'infos ...](#) -/-  
[Flores manuscrites](#)  
[Catalogues de fonds particuliers](#) -/- [Fonds Edouard et Louis Lartet](#)  
-/- [Fonds hispaniques anciens des bibliothèques toulousaines \(ouvrages numérisés\)](#) -----

# 3 - Métadonnées et pédagogie

■ <http://www.educnet.education.fr/dossier/metadata/default.htm>

■ Des normes sur ressources pédagogiques pour ...

- Accessibilité
- Interopérabilité - simplifier de nouveaux usages
- Pérennité et adaptabilité

■ Description de ressources éducatives

- IEEE-LOM : ressources électroniques ou physiques

■ Description de formations : CMD, CMD-fr

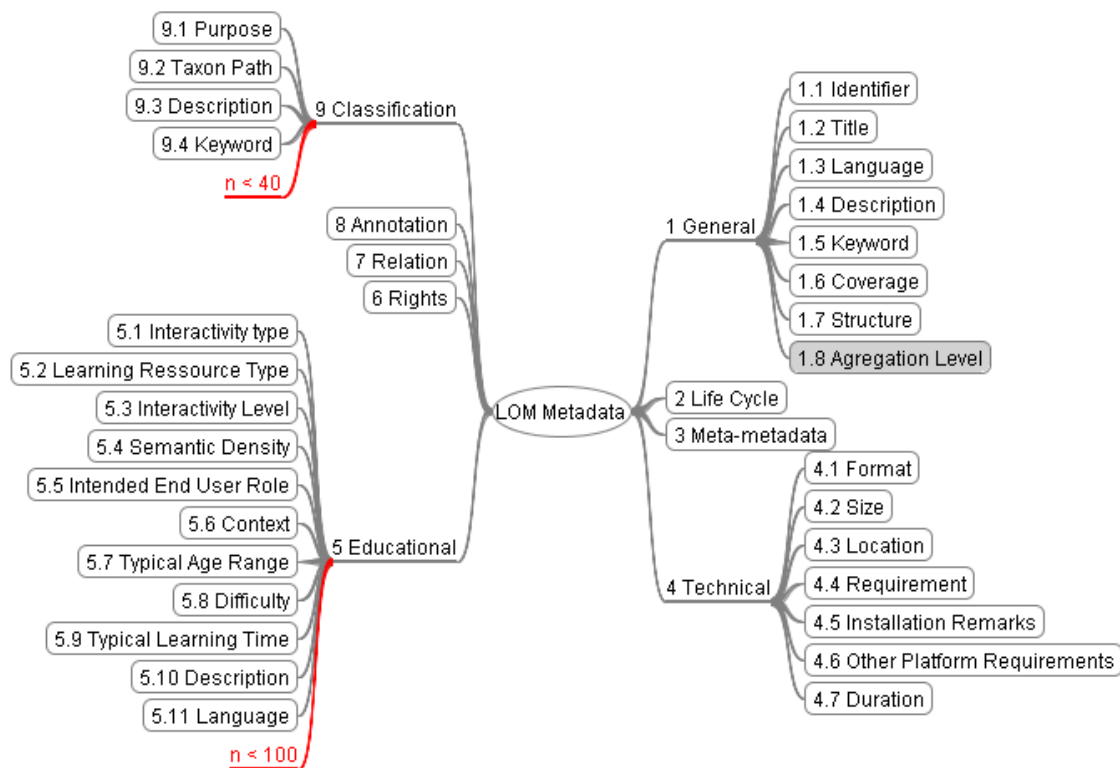
- Course Description Metadata - XML

# LOM, Learning Object Metadata

- Développé par l'IEEE
- Standard international examiné par l'ISO
  - Adaptation française, LOM-fr (2005-2006), AFNOR, CG46, CN357
- Utilisateurs
  - Editeurs commerciaux
  - Communautés d'enseignement, de formation professionnelle et continue
    - Projets (SCORM, Normetic, Université Virtuelle...)

# Métadonnées LOM

- Trois niveaux d'arborescence
  - 9 sections de niveau 1
    - 78 éléments, définis chacun par 8 propriétés
- Valeurs typées et valeurs contrôlées
- Nombreux "profils d'application"
- Pour en savoir plus
  - [http://fr.wikipedia.org/wiki/Learning\\_Object\\_Metadata](http://fr.wikipedia.org/wiki/Learning_Object_Metadata)
  - [Métadonnées : quelques cartes du LOM et de profils d'application](http://dicos.ens-lyon.fr/~dicos/metadata/), dicos.ens-lyon
  - Liste d'éléments et correspondance DC-LOM  
<http://www.ischool.washington.edu/sasutton/IEEE1484.html>



pages ... !

Bibliothèques numériques

protocole OAI-PMH  
Archives ouvertes ...

Profils  
d'applications

Mutualisation des services  
culturels et patrimoniaux

OpenURL

Cyberthèses  
TEF

Annuaire et portails

Corpus Recherche  
OLAC, IMDI ...

Informations publiques  
INSPIRE, URBADOC

Web 2

Archivistes  
outils pédagogiques  
et livres

Flux de production



# 1 - Accès libre, archives ouvertes protocole OAI-PMH

- Un principe : le libre accès à l'information scientifique et technique
- Intérêt :
  - Accessibilité, visibilité, impact
  - Modèle économique et social alternatif
- Des pionniers, 1990-93 : ArXiv, WoPEc, ADS astrophysique, CogPrints
- Conférences/conventions de Budapest, Bethesda, Berlin, Genève...:
  - Le libre accès aux résultats de la recherche devient un postulat
  - Différents pays et institutions adhèrent formellement à ce principe
  - Distinctions formelles
    - revue libre accès / archive ouverte
    - archive thématique / institutionnelle
- Voir <http://openaccess.inist.fr/>



# Interopérabilité des archives ouvertes

## Le protocole OAI-PMH

### ■ Extension rapide du mouvement

- BioMedCentral, serveurs du CCSD, PLoS...
- Près de 2000 **entrepôts** répertoriées, contenant plus de 20 millions de ressources
  - <http://gita.grainger.uiuc.edu/registry/> (2000)
  - <http://www.openoar.org/index.html> (1500)
  - <http://roar.eprints.org/> (1500 archives)

### ■ Protocole OAI-PMH, 2001

- Convention de Santa Fé et Open Archive Initiative, 1999 : définir des protocoles de structuration et d'interrogation communs pour l'interopérabilité ... Z39.50 ? Ou ...
- HTTP, XML, Dublin Core



# OAI-PMH du côté utilisateurs

Recherche dans un ensemble d'archives avec des moteurs spécialisés

- [ <http://oaister.umdl.umich.edu/o/oaister/>  
nov. 2005 : 477 archives, 5 277 000 enregistrements  
janvier 2008 : 14 673 000 enregistrements ]
- <http://citebase.eprints.org/cgi-bin/search> (facteur d'impact")
- <http://www.scientificcommons.org> (2007: 898 archives, 16M enr.)

Auto-archivage, exemple dans HAL, Hyper Articles on Line, CCSD-CNRS

<http://hal.archives-ouvertes.fr>



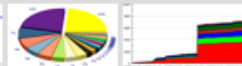
# Registry of Open Access Repositories (ROAR)

Search Content Search Register a Repository Help Login Register in ROARMAP

Country: Any System Software: Any Content Type: Sort by Name: Filter

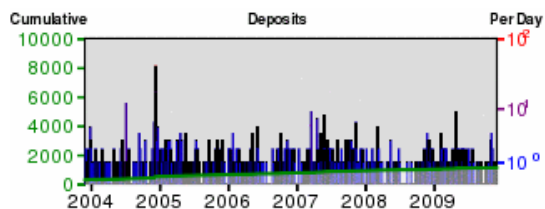
Registry of Open Access Repositories (ROAR) resulted in 1543 repositories

Export this result set as a chart:



Export this result set as:  
[Comma-Separated Format](#)  
[Google Earth Overlay \(Thumbnails / Bars\)](#)  
[... other formats](#)  
 Analyse the **content** of these repositories ([caveats](#))  
 Export repository records summary as: [Cumulative Graph](#) [Cumulative Table](#)  
 View repository contents as: [File Format Graph](#)

101 records)



Running **HAL**, based in **France** and is registered as **Research Cross-Institutional**  
 Registered on 2002-05-17  
 Cumulative deposits: 1101 total [\[table\]](#) [\[graph\]](#)  
 Daily deposits in last year: 73 days of 1-9, 0 days of 10-99, 0 days of 100+ [\[table\]](#)  
[\[graph \(PNG format\)\]](#) [\[interactive graph \(requires SVG format support\)\]](#)  
 OAI Interface: [Identify](#) [List Metadata Formats](#) [List Sets](#) [\[harvest status\]](#)

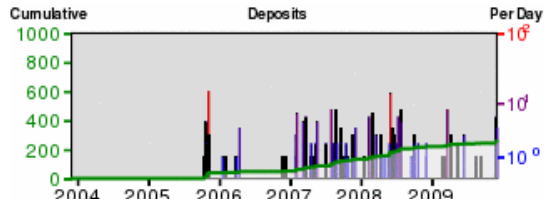
an Agriculture and Natural Resource Online



[OAI interface not registered](#)

Running **Other softwares (various)**, based in **Australia** and is registered as **Research Cross-Institutional**  
 Registered on 2009-09-09  
 No successful harvest yet.

University Research Archive: AURA (257 records)



Running **DSpace**, based in **United Kingdom** and is registered as **Research Institutional or Departmental**  
 Registered on 2006-01-27  
 Cumulative deposits: 257 total [\[table\]](#) [\[graph\]](#) [\[Preserv Profile\]](#)  
 Daily deposits in last year: 15 days of 1-9, 0 days of 10-99, 0 days of 100+ [\[table\]](#)  
[\[graph \(PNG format\)\]](#) [\[interactive graph \(requires SVG format support\)\]](#)  
 OAI Interface: [Identify](#) [List Metadata Formats](#) [List Sets](#) [\[harvest status\]](#)

... in the **Title** field for "**metadata**".

**records.**  
... [your search](#) to retrieve fewer records.  
... results, starting with **records 1 to 10 of 818.**

**Institution** Next 10 R

Sort by  Jump to Records: [1](#) | [11](#) | [21](#) ... [81](#) ... [161](#) ... [241](#) ... [321](#) ... [401](#) ... [481](#) ... [571](#) ... [651](#)

ledge  
(KT) EPrints

Online (AJOL)

ities Data

Archive

l Brasileira de  
DBComp)

l de Teses e  
(DTD), Instituto  
Formacao em  
logia (IBICT)

ce-Zentrum  
berg (BSZ BW),  
aller

**Record 1 of 818**

Title	Digital collection management software employed by libraries and museums: examination of <b>metadata</b> sen mapping functionality
Contributor	Park, Jung-ran
Contributor	Park, Sang Joon
Year	2005-09-06T20:37:49Z
Resource Type	Presentation
Resource Format	91252 bytes,application/pdf
Language	en_US
Note	A growing number of organizations are building digital collections using both commercial digital collection management software such as CONTENTdm, Encompass, etc., and open source software such as Greenstone D-Space. This rapidly growing number of distributed digital collections has brought to the fore the critical iss resource discovery and sharing across these collections. The goal of this project is to examine the functional <b>metadata</b> creation and mapping and the configuration of digital collection management software. This goal to the issue of semantic interoperability of concept representation across digital collections. For this, this pro aims at examining how digital collection management software provides a mechanism for semantic mapping between different <b>metadata</b> schemes such as Dublin Core (DC) and MARC or between cataloger-defined fi names and a given <b>metadata</b> scheme such as DC. As a first step, we will examine features related to <b>meti</b> semantic mapping of CONTENTdm software, which provides a feature that allows for catalogers to map cataloger-defined field names onto DC <b>metadata</b> elements.
Note	IST Research Day 2005 posters;no. 105F
Subject	<b>Metadata</b> ; Data conversion; Electronic libraries



ligatoires

Soumettre dans HAL

# HAL, soumettre

[laissez ce champ vide si vous considérez votre document comme récent; si vous le remplissez, il entrera dans la catégorie des "rétro-dépôts" qui ne sont pas envoyés vers arXiv et ne déclenchent pas les systèmes d'alerte automatique (abonnements)]

[pour une soumission qui doit être transmise sur arXiv, il est nécessaire de fournir un résumé en anglais, et de ne pas dépasser un total de 1400 caractères (15 lignes) environ]

[Large empty text area for abstract]

...gies de l'Information et de la Communication

## Liste des domaines

HAL :: DOMAINE - Microsoft Internet Explorer fourni par l'INIST

## Liste des domaines

- HAL : http://hal.ccsd.cnrs.fr/
  - Physique [physics]
  - Science non linéaire [nlin]
  - Mathématiques [math]
  - Informatique [cs]
  - Chimie

Anglais

[mots-clés séparés par le caractère ";"]

[Empty text area for keywords]

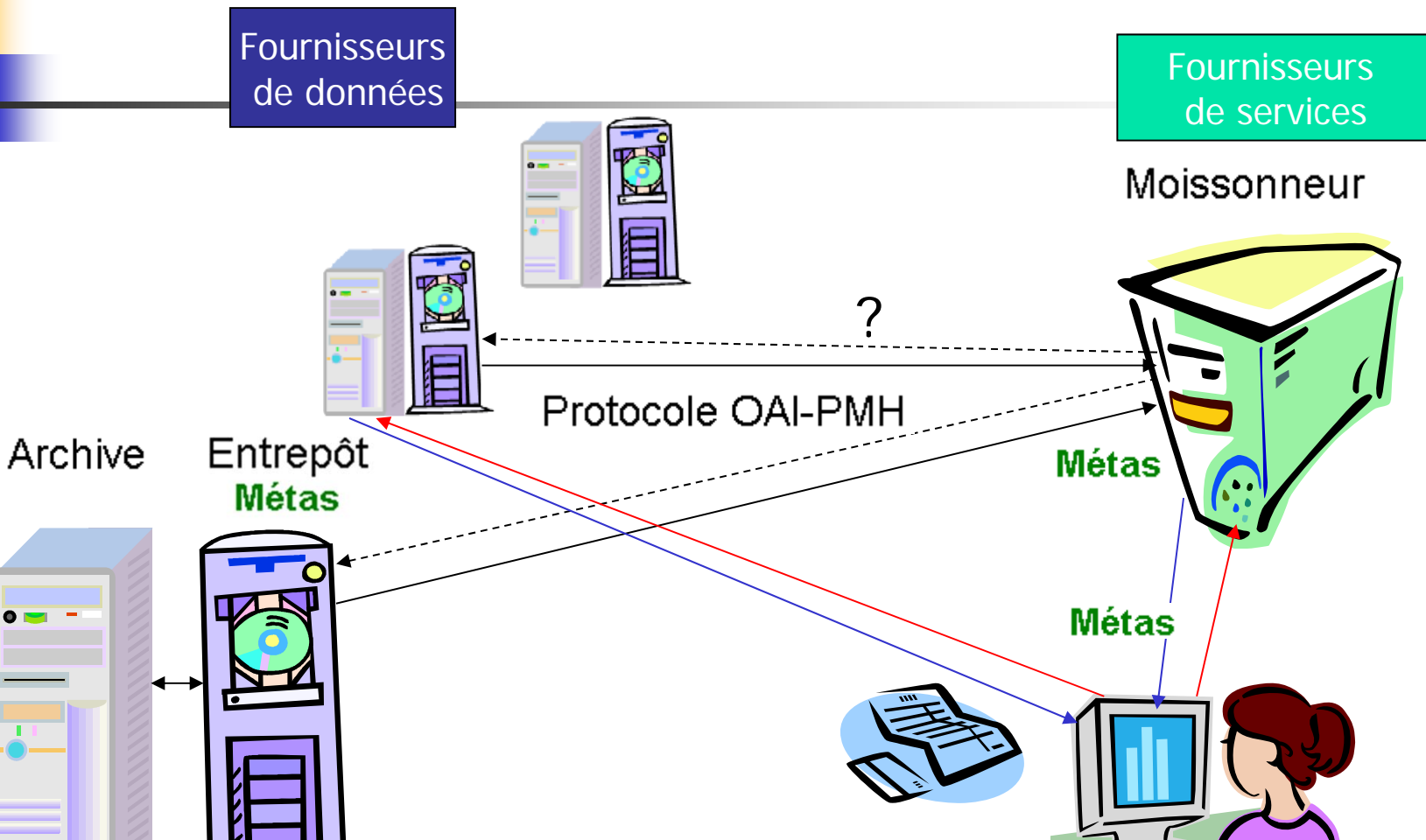
[mettre ici la(les) classification(s) ACM, PACS, MSC, OCIS, etc.]

[Empty text area for classification]

# OAI-PMH, les acteurs

- Fournisseurs de données
  - Créent les archives et leur “entrepôt” de métadonnées
  - Outils (de fonctions variées au-delà OAI-PMH !)
- Fournisseurs de services
  - Moteurs spécialisés, “moissonneurs
  - OAIster, Citebase search, ARC, Perseus, my.OAI
  - Minimum : moissonnage, exposition
    - Interrogation des entrepôts par des “verbes”
  - Valeur ajoutée diverse : espace personnel, impact ...
- Voir le site Open Archives Initiative,  
<http://www.openarchives.org/>, community

# Protocole OAI-PMH



# OAI-PMH, les entrepôts

- Contenu minimal d'un entrepôt de métadonnées
  - Enregistrements compatible OAI-PMH
    - Dublin Core – XML
  - Dans une base de données (\*SQL, ODBC ...)
- Autres formats moissonnables
  - MARC-XML, MODS, ETD-MS
  - METS, EAD, DIDL ...
  - ... Tout format XML de métadonnées

< />

Header : identifiant, date création, set

Metadata, ici dublin Core

oai:documents.irevues.inist.fr:2042/28830</oai:identifier>  
<dc:stamp>2009-11-20T06:28:18Z</dc:stamp>  
<dc:spec>hdl\_2042\_28804</dc:spec>

'creator'

'date'(s)

'identifier'

<dc:dc: xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai\_dc/ http://www.openarchives.org/OAI/2.0/oai\_dc.xsd">  
<dc:creator>BOE, Julien</dc:creator>  
<dc:date>2009-10-13T12:15:49Z</dc:date>  
<dc:date>2009-10-13T12:15:49Z</dc:date>  
<dc:date>2009</dc:date>  
<dc:identifier>http://hdl.handle.net/2042/28830</dc:identifier>

La complexité des mécanismes liés au cycle de l'eau continental joue un rôle important dans la disparité actuelle des résultats des modèles climatiques sur la France et donc dans les incertitudes quant aux impacts du changement climatique d'origine anthropique. Cet article présente les rétroactions qui sont en jeu et souligne l'importance pour les modèles de climat de représenter correctement le cycle hydrologique du climat présent. Cela est particulièrement critique pour prévoir l'effet du changement climatique en été sur la France et plus largement sur une large zone intermédiaire entre l'Europe du Nord et l'Europe du Sud. Pour ces régions, une meilleure compréhension des changements de circulation atmosphérique sur l'Atlantique nord est conjointement nécessaire. Complété par une étude d'impact, l'article souligne que, en France, dans l'état actuel des connaissances, des impacts importants sont à attendre sur la ressource en eau.

'description'(s)

Discrepancies between climate models results over France are largely due to the complexity of mechanisms involved in the continental water cycle and thus on the ensuing uncertainties in the impact of the anthropological climate changes. This article presents the feedback mechanisms involved and emphasizes the importance for climate models to properly represent the hydrological cycle of today's climate. This is particularly critical to forecast the effect of climate change in summer over France, and more broadly, over the wide transition zone between Northern and Southern Europe. For these areas, a better understanding of the changes in the atmospheric circulation over the North Atlantic is also necessary. Ending with an impact study, this paper highlights that with today's understanding, important consequences are to be expected on water supply over France.

'format'

'language'

'publisher'

'rights'

<dc:format>1678447 bytes</dc:format>  
<dc:format>application/pdf</dc:format>  
<dc:language>fr</dc:language>  
<dc:publisher>Société météorologique de France, Paris (FRA)</dc:publisher>

<dc:rights> Document réservé aux abonnés - Licence d'utilisation : http://documents.irevues.inist.fr/utilisation.jsp

'source'

<dc:source> La Météorologie [ISSN 0026-1181], 2009, Série 8, N° 66 ; p. 38-48



# Fin de l'enregistrement OAI-PMH

`<dc:subject>réponse des modèles</dc:subject>`

'subject'(s)

`<dc:subject>bilan hydrique</dc:subject>`

`<dc:subject>bilan thermique</dc:subject>`

`<dc:subject>rétroaction</dc:subject>`

`<dc:title>`

'title'

Prix Prud'homme 2008 : Cycle hydrologique et changements climatiques sur la France : mécanismes, incertitudes, impacts sur la ressource en eau

`<dc:title>`

`<dc:title>`

Hydrological cycle and climate changes over France: mecanisms, uncertainties, impacts on water resources

`<dc:title>`

`<dc:type>Article</dc:type>`

'type'

`<i_dc:dc>`

`<data>`

'about'

'provenance'

# De la base de donnée interne à l'archive à l'entrepôt OAI

- DEMO
- La base de données interne à l'application gère ses propres éléments
  - HAL, <http://hal.archives-ouvertes.fr/> , “services”
    - Eléments propres, formats d'exports divers
  - DSpace, application INIST I-Revues, <http://irevues.inist.fr>
    - Profil d'application Dublin Core qualifié – cf Hermès, 2005, 42
- ... et les expose en OAI-PMH
  - Une requête dans l'entrepôt de HAL
  - Hermès,num 42 dans l'entrepôt de I-revues

# OAI-PMH

## le pouvoir de la simplicité ...

- **Au-delà des “archives ouvertes”, d’autres usages ...**
- **Interopérabilité des revues en libre accès** (Revue.org, Persée ...)
- **Création simple de services collaboratifs**
  - **Portails**
    - Mutualisation des services culturels et patrimoniaux
    - ORI-OAI pour l’Université Numérique ... etc ...
  - **Catalogues**
    - SomRev, revue de sommaires sciences politiques et économiques
    - Catalogue collectif d’archives sonores ... etc ...
- **Editeurs commerciaux** de revues et d’ouvrages : fourniture d’entrepôts moissonnables
  - **métadonnées pour intégration à des applications locales**
    - projets BNE INIST



## 2 - ORI-OAI, Outil de Référencement et d'Indexation - Réseau de portails OAI

Travaux de plusieurs établissements d'enseignement supérieur, 2001-2008

- adhérents à divers groupes de travail et réseaux

Mise en place d'un système ouvert pour

- **gérer et publier** tous les documents numériques produits par les établissements universitaires
  - Pédagogiques, travaux de recherche, administratifs
- les **partager** avec d'autres établissements
- les **valoriser** par une indexation de qualité
- les **rendre accessibles à distance**, selon les droits définis, dans des interfaces ergonomiques

<http://www.ori-oai.org/> + site Médiaguilaine



# ORI-OAI

## principes d'un système

Un référentiel unique de ressources connecté au SI

- “Forme canonique” des ressources

Recherche avancée et accès thématique

Gestion et publication des ressources numériques

- Web, droits d'accès, métadonnées standard
- Work-flow impliquant auteur, et processus validation-suivi paramétrables
- Archivage

Partage avec communauté(s)

- Réseau de portails : **OAI-PMH**
- Interopérabilité avec autres “systèmes” : ENT, UN, HAL, STAR, listes



# ORI-OAI, choix techniques

- Des métadonnées interopérables pour la recherche et la gestion
  - XML, Dublin Core, LOM et LOM-fr, TEF ...
  - Classifications paramétrables articulées sur des standards comme Dewey
- Des applications existantes
  - OAI-PMH, JSR 168, Shibboleth ...
- Des outils libres
  - Lucene ...
- 8 modules interopérables

# Aquitaine patrimoines

Le portail culturel des patrimoines de l'Aquitaine

[Accueil](#)

[Mode d'emploi](#)

[Programme](#)

[Partenaires](#)

[Ressources](#)

[Professionnels](#)

[Technique](#)

Le programme de la Banque numérique du savoir d'Aquitaine (BNSA), le Conseil régional, l'État et leurs partenaires ont réalisé ce portail culturel des patrimoines de l'Aquitaine. Pour en savoir plus sur le BNSA, cliquez ici >

## Sites sélectionnés

Vous est proposé pour une première sélection aquitaine.

[Voir la sélection](#) >

## Recherche dans l'ensemble des ressources

Ressources disponibles au lundi 7 décembre 2009

Recherche simple

[Trouver](#)

Recherche multicritères >

>

Recherche avancée >

## Recherche géographique

35 645 ressources localisées au lundi 7 décembre 2009

[Afficher la carte des établissements fournisseurs de ressources](#) >

Choisir une commune  [Trouver](#)

[Afficher la liste des communes](#) >

Choisir un département



## Expositions virtuelles

Explorez les richesses patrimoniales de l'Aquitaine à travers une sélection d'expositions virtuelles.

[Voir les expositions](#) >





BASES DE DONNÉES

navigation sur le site Webdoc

## Somrev : base de signalement de sommaires de périodiques

### Interrogation de la base Somrev : sommaires de revues

rechercher dans un titre d'article :

on à rechercher dans un titre d'article :

...vues...

...nées...

...nées produite par l'IEP de Lyon en collaboration avec d'autres Centres de documentation et Bibliothèques en Provence, IEP Grenoble, IEP Lille, IEP Lyon, IEP Rennes, IEP Toulouse, Université Catholique de

...sentation de la base de signalement de contenus de périodiques SOMREV et des projets de partenariat

#### Somrev.info

Nouveau !

▶ Nouveau titre dans Somrev : I internationaux

▶ Nouveau titre dans Somrev : I pratiques

▶ Consultez le détail de [l'actualité base](#)

Sommaires

▶ 9688 sommaires (153375 arti



# 3 -TEF, thèses électroniques françaises, objectifs

- **Projet ministériel** de valorisation et pérennisation
  - Circulaire, directives
  - Visant une **diffusion accrue et des échanges**
    - Accès en ligne
    - Signalement mutualisable dépassant le cadre des bibliothèques
  - AFNOR CG 46 / CN 357, métadonnées des thèses
- Vers une **chaîne de production unique et automatisée**  
document électronique – métadonnées ... ..
  - Métadonnées : compatibilité OAI-PMH + richesse de la description SUDOC
  - Document : à suivre ...

# Métadonnées TEF

## ■ Espace de nom TEF

- Un "profil d'application"
  - Version 1, 2005 : **Dublin Core** qualifié + **MODS** et **ETD-MS**
  - Version 2, mai 2006 : encapsulation **METS** + droits **METS**
- Valeurs contrôlées : vocabulaires, autorités MADS...

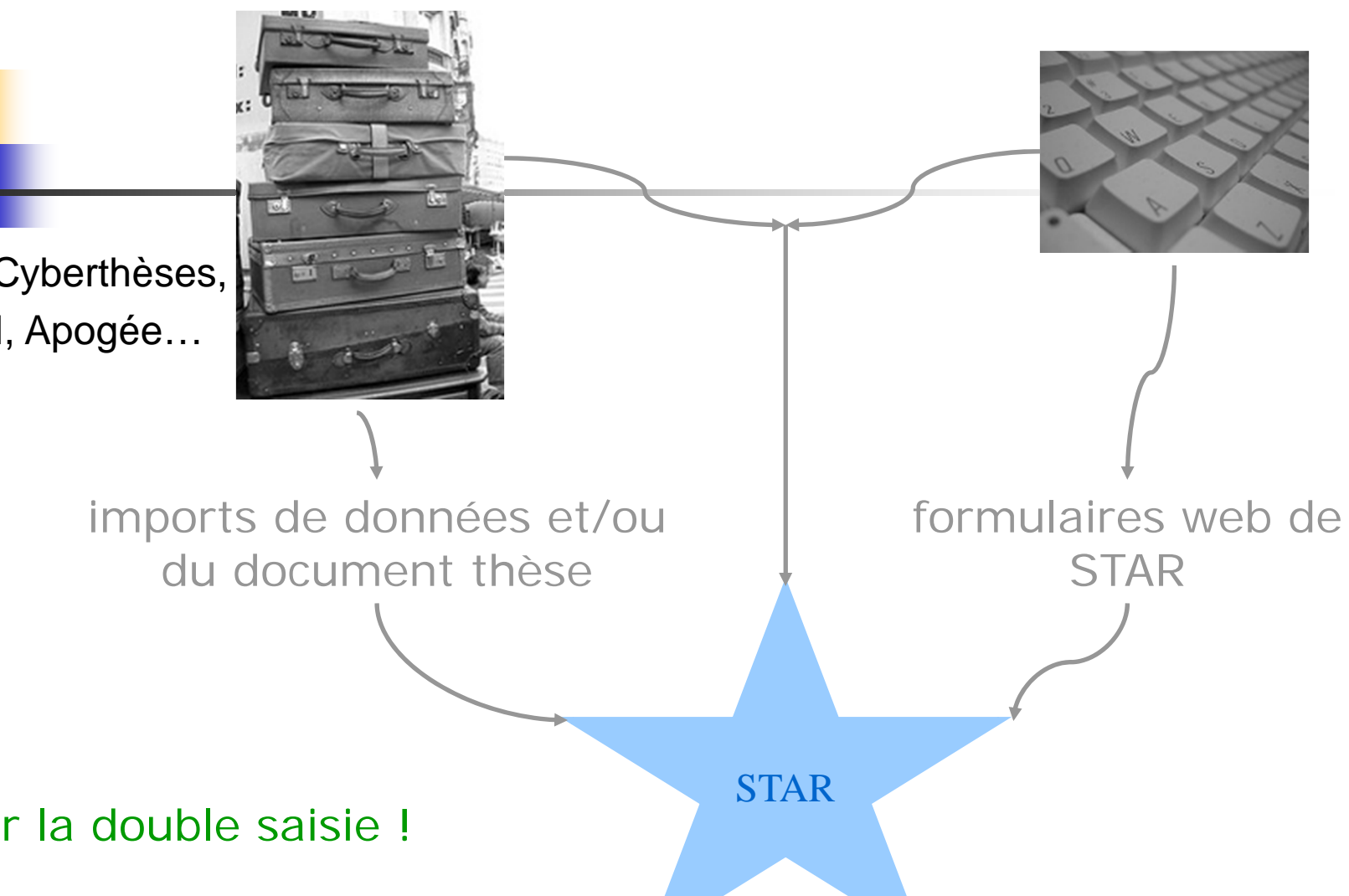
## ■ Registre

- éléments, définitions, hiérarchie, recommandations d'usage, types des valeurs, "mapping"

## ■ Feuilles de style → **OAI-PMH** et UNIMARC (V1)

■ Utilisation des styles **CTAB** (2000)

# Interopérabilité ... ce qui rentre dans STAR



# ... Ce qui sort de STAR

*chaque  
établissement de  
soutenance \**

*Diffusion*



discipline

*coûts national,  
national...*



OPEN ARCHIVES



*diversité pérenne \**

STAR

Nancy-Université



*URL  
pérenne \**

*Signalement \**



e "format"

# METS, Metadata Encoding and Transmission standard

- 2001, maintenu par la Library of Congress,  
<http://www.loc.gov/standards/mets/>
- Une "grosse enveloppe" contenant les métadonnées bien rangées dans des sections structurées
  - Pointant sur, ou contenant, les objets numériques simples
- Pour
  - Créer, organiser et "préparer le fonctionnement" d' "objets numériques complexes"
  - Echanger des masses de données
  - Participer à la pérennisation
    - Objets conformes au paquet de soumission décrit dans la norme **OAIS** de modélisation d'une architecture d'archivage pérenne

# METS, structure

DC,  
MODS  
EAD  
ONLY

Creat. Com.  
METSRights  
PREMIS  
MIX..

metsHdr

Description  
contenu

pointeurs

Administration  
Juridiques  
Techniques  
Préservation \*

Section  
comportements



Liens structure

Section des  
fichiers

Carte(s) de  
structure

# Notices TEF - Un coup d'œil ?

- TEF V1
  - Le registre (historique)  
[http://www.abes.fr/abes/documents/tef/recommandation/tef\\_01.pdf](http://www.abes.fr/abes/documents/tef/recommandation/tef_01.pdf)
- TEF V2, recommandations, schémas exemples
  - [http://www.abes.fr/abes/DesktopDefault.aspx?tabindex=4&tabid=403,](http://www.abes.fr/abes/DesktopDefault.aspx?tabindex=4&tabid=403)

## 4 - eBooks : ePub, OPF et métadonnées sur la publication

```
package version="2.0" xmlns="http://www.idpf.org/2007/opf" unique-
  identifier="BookId">
  <metadata xmlns:dc="http://purl.org/dc/elements/1.1/"
    xmlns:opf="http://www.idpf.org/2007/opf">
    <dc:title>Alice in Wonderland</dc:title>
    <dc:language>en</dc:language>
    <dc:identifiant id="BookId" opf:scheme="ISBN">
3456789X </dc:identifiant>
    <dc:creator opf:role="aut">Lewis Carroll</dc:creator>
    ...
  </metadata>

  <manifest> ... </manifest> ...
package
```



# 5 – Logiciels documentaires et standards XML

- Dans beaucoup de logiciels aujourd'hui, on trouve
  - Exports XML variés des éléments
  - Exposition et moissonnage OAI-PMH via la fonction portail
- ... Et plus ...
  - exemples : Koha, SIM, **couches internes XML, structurant Dublin Core + autres standards ...**
- Zotero "tout simplement"
  - Logiciels de gestion de références bibliographiques Web2
  - <https://addons.mozilla.org/fr/firefox/>
  - Intègre divers standards de métadonnées en import, export et **couche interne XML/ RDF, DC, FOAF ...**

# Zotero, démo

- Sites interopérables avec Zotero
  - Icône spécifique dans la barre d'URL
  - [Revue.org](#) et tous sites Lodel
  - I-Revues et sites DSpace V1.5
  - [Catalogue du Sudoc](#)
    - Sur une page de résultats ou un article
  - [Flick'R](#), Wikipedia, Worldcat, CiteUlike ...
  - [Article de Peccatte sur métadonnées](#) ...
- Intègre directement les métadonnées
  - Standards DC des en-têtes HTML, MODS, MARC-XML, OpenURL COINS
  - ... Dans la bibliothèque choisie – compléter notice si utile
- Exporte en DC, MODS (XML), Zotero (RDF avec DC, FOAF) ...

# TEI, Text Encoding Initiative pour l'encodage de documents

## ■ Un cadre de travail pour l'établissement, l'échange et l'exploitation en XML des textes

- Recherche scientifique sur des textes structurés de façon standard
- Echanges entre labos de corpus standardisés
- Réutilisations diverses → affichages, recherche affinée
- Pérennité des documents

## ■ Manuscrits et textes littéraires, puis plus largement SHS ... autres domaines scientifiques ?

- Source papier ou nativement numérique

## ■ Consortium international TEI

- Une porte d'entrée : <http://www.tei-c.org>
- Réseau français : <https://listes.cru.fr/wiki/tei-res/index>

## ■ Pays anglo-saxons, encore timide en France

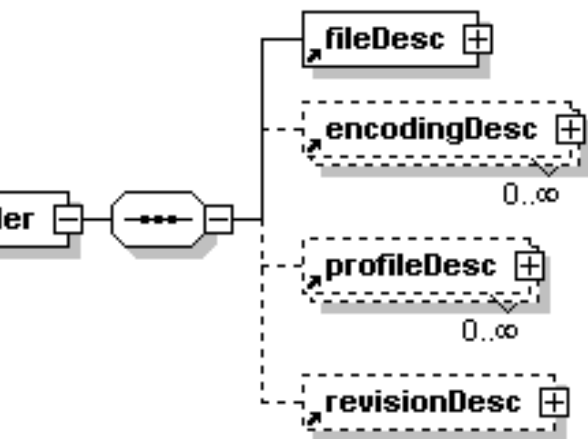
# Document TEI, structure

(voir d'abord [http://www.tei-c.org/Lite/teiu5\\_fr.html](http://www.tei-c.org/Lite/teiu5_fr.html) )

## Élément racine <TEI.2>

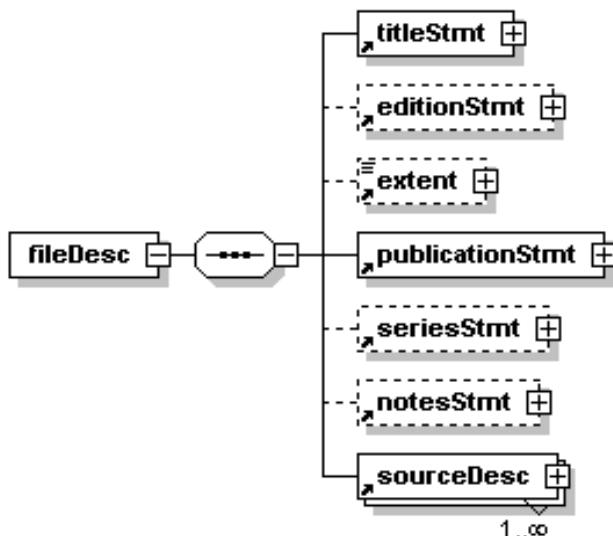
- en-tête TEI <teiHeader>
  - Page de couverture – métadonnées
- Corps du texte <text>
  - **Marquage de la structure** <front><body><back> - récursif, éléments fils selon type de document
    - Ouvrages, articles, manuscrits : sections avec niveaux, titres de section, notes, tableaux ...
    - Ouvrages de référence : structure des notices
    - Oral (interviews ...) : intervenants, timing ...
  - **Marquage sémantique** selon usages

# teiHeader, structure



- description bibliographique de la ressource
- rapport entre un texte électronique et la source ou les sources initiales : description de la manière dont il a été codé
- description "non-bibliographique" : langue(s), mots-clés, contexte de production, participants, environnement
- historique de création et révision du document TEI.

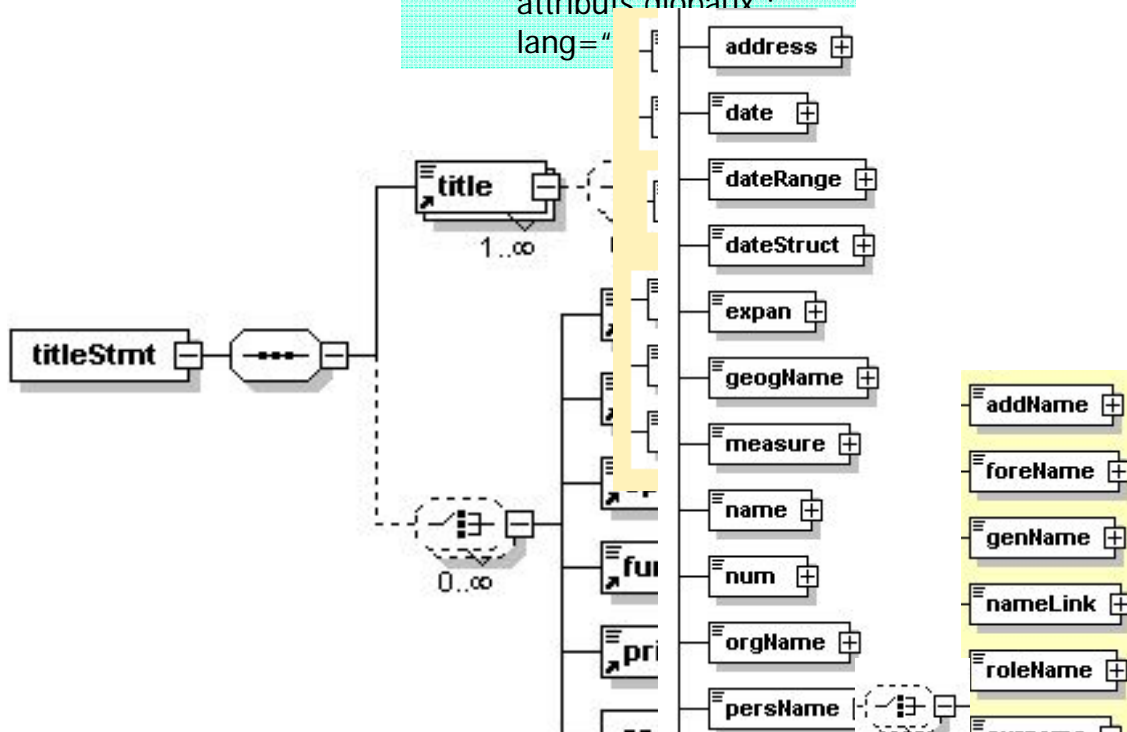
# TEI header, description bibliographique <fileDesc>



- <titleStmt> titres et responsabilité
- <editionStmt> information relative à l'édition de l'instance présente du texte
- <extent> taille approximative du texte électronique
- <publicationStmt> informations concernant la publication ou la diffusion
- <seriesStmt> informations concernant la collection ou série
- <notesStmt> informations complémentaires aux autres parties de la description bibliographique
- <sourceDesc> description bibliographique du ou des

# <titleStmt>

type="main"  
[subordinated,  
parallel,  
abbreviated ...]  
level="a"[s, j, m, u]  
attribs globaux :  
lang="



... éléments  
sémantiques  
valables partout

521 éléments,  
regroupés en 146  
classes

TEI

## quelques exemples français

### ■ Editions critiques de textes littéraires

- [Bibliothèque humaniste virtuelle](#), CERS Tours ... ..
- [les dossiers de Bouvard et Pécuchet de Flaubert](#), ISH Lyon

### ■ Edition critiques d'actes anciens et d'archives

- [TELMA](#), traitement électronique des manuscrits et des archives
  - [Institut de recherche et d'histoire des textes](#)
  - [Ecole des Chartes](#), collections 'elec' et 'du cange' ... ..

### ■ Codage de corpus oraux – [CRDO](#), groupe Catcod

### ■ publication des thèses, notamment à l'université de Lyon 2 <<http://theses.univ-lyon2.fr/>>, [exemple](#)

### ■ Structuration et mise en ligne du [Bulletin des bibliothèques de France](#)

### ■ [portail PERSEE](#) : mise en ligne de revues SHS

- Échanges avec la plateforme de publications ERUDIT



page Historique Marque-pages Outils ?

http://www.cn-telma.fr/

Telma. Centre de ressources nu...



- accueil
- Documentation
- Veille
- Corpus
- A propos
- Contact

## Bienvenue sur le site du centre de ressources numériques Telma

### À la une

#### Nouveaux corpus (publié le 28 mai 2009)

- Mise en ligne des actes royaux de Philippe III et Philippe IV  
[>> Pour en savoir plus](#)
- Ajout de 150 enquêtes au corpus des "Enquêtes menées sous les derniers Capétiens"  
[>> Pour en savoir plus](#)

### Actualité de Telma

#### Ecole thématique : 19 - 24 octobre 2008 (publié le 28 août 2008)

Telma, le CN2SV et le CRDO (centres de ressources numériques) organisent une école thématique sur le thème : Préservation et diffusion numérique des sources de la recherche en sciences humaines et sociales. [>> Lire la suite](#)

#### Journée d'étude - CBMA (publié le 23 janvier 2008)

Telma participera à la journée d'étude : CBMA - Chartae Burgundiae Medii Aevi - Les fonds diplomatiques bourguignons (II), qui se déroulera à l'Université de Bourgogne le 25 janvier 2008. [>> Lire la suite](#)

#### Le Glossarium de Du Cange en ligne (publié le 22 janvier 2008)

L'Ecole nationale des Chartes vient de mettre en ligne à l'adresse suivante :

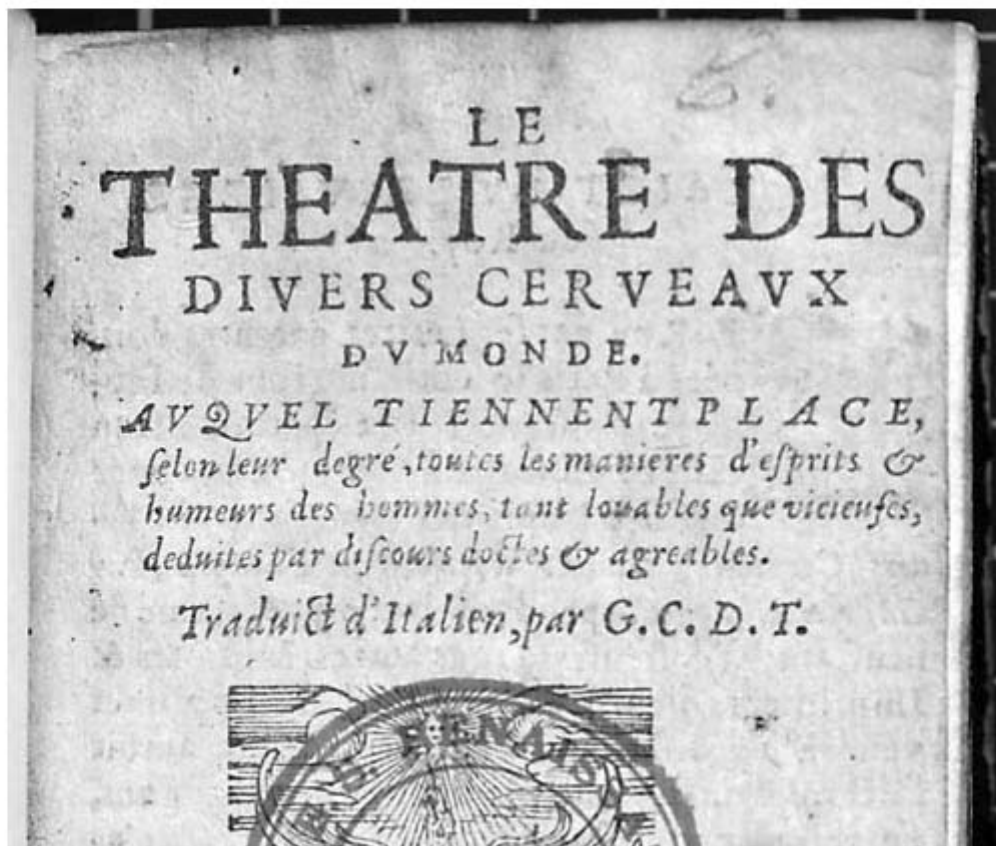
**ES DIVERS  
MONDE.**

ACT EXTRAICT DU PRIV  
ET TRES-VER tueux, Pierre  
EFACE, LE THEATRE D  
BLE DES DIS- COURS.  
BLE DES AUTEURS alleg

THEATRE DES DIVERS C  
S CERVEAUX tranquilles é  
S CERVEAUX BRA- ves, v  
S CERVEAUX alaires & jc  
S CERVEAUX facetieux D  
S CERVEAUX arguts. DI  
S CERVEAUX accorts, fins  
S CERVEAUX vifs, promp  
S CERVEAUX subtils, aigu  
S CERVEAUX sçavants &  
S CERVEAUX vertueux &  
S CERVEAUX vains. DIS  
S CERVEAUX LE- gers, ren  
S CERVEAUX curieux. D  
S CERVEAUX DE- daigner  
S CERVEAUX passionnez &



Garzoni, Tommaso



LE  
THEATRE  
DES  
DIVERS  
CERVEAUX  
DU MONDE.

AUQUEL  
TIENNENT  
PLACE,  
*selon leur degré,  
toutes les manieres  
d'esprits &  
humeurs des  
hommes, tant  
louables que*

[< Minute précédente](#)

Edition diplomatique du texte

Récupérer le texte de la minute

[Minute suivante](#)

3

<Date>1437 (n. st.), 5 janvier.</Date>

<Analyse><NatureJuridique>Procuration</NatureJuridique>. – <Paragraphe><Personne>Pierre de Villebresme</Personne>, écuyer, donne procuratic  
<Personne>Nau Grandet</Personne> pour la vente du lieu de <NomDeLieu>Ramoulu</NomDeLieu> qu'il possède près  
<NomDeLieu>Pithiviers</NomDeLieu>. </Paragraphe></Analyse>

<Transcription>

<Paragraphe><SautDePage>56v</SautDePage> Et le samedi cinquieme jour de janvier.</Paragraphe>

<Paragraphe>Noble homme Pierre de Villebresme, esquier seigneur de <NomDeLieu>Cormeray leez Bloys</NomDeLieu>, fait ses procure  
Nau Grandet seul et pour le tout. Auquel procureur seul et pour le tout portant et monstrant ces lettres, ledit constituant a donné et donne plain pov  
autorité et mandement especial de vendre et adenerer a touz jours a tele personne ou personnes et pour tel pris et somme de deniers comme l  
lui semblera, ung lieu et ses appartenances, appartenant audit constituant, nommé Ramolu estant pres  
<NomDeLieu>Puthivers</NomDeLieu><sup>1</sup>, et generalement tout ce que ledit constituant a en ladicte parroisse de Ramolu et ou pays d'environ, c  
recevoir les deniers et donner quittance du receu. De soy etc. dessaisir etc. et en saisir etc., et aussi de la foy, hommaige, respit ou souffrance  
generalment de faire tout autant et aussi comme se il estoit present en personne. Et a ladicte vente faire promect faire consentir et accor  
damoiselle <Personne>Michelete de la Boissiere</Personne> sa femme etc.</Paragraphe>

<Paragraphe><NoteEditoriale>Dans la marge gauche : </NoteEditoriale><AjoutEnMarge>P. de Villebresme. </AjoutEnMarge></Paragraphe>

</Transcription>

<NoteEditoriale>1. Selon toute vraisemblance <Citation>Pithiviers</Citation>. </NoteEditoriale>

[< Minute précédente](#)

Edition diplomatique du texte

Récupérer le texte de la minute

[Minute suivante](#)

# TEI : des outils spécifiques



- Plateforme Millefeuille, <http://millefeuille.gforge.inria.fr/>
- Projet MutEC - mutualisation d'outils numériques pour les éditions critiques et les corpus, <http://www.mutec-shs.fr/>
- **ROMA** : créer des profils personnalisés

# EAD

## Encoding Archival Description

Objectif : standardiser et mettre en ligne les **inventaires archivistiques**, avec des fonctions de recherche et de navigation → **“instruments de recherche”**

... Dans un format XML pour permettre

- **Pérennisation : conservation et migrations**
- **Accessibilité accrue**
  - Présentations diverses adaptées au contexte d'édition - public cible
  - Fonctions de recherche et navigation
- **Echanges et mutualisation de données standardisées**
  - Vastes réservoirs archivistiques complémentaires
- **Transformations** vers d'autres applications

# EAD

## 'Le' standard archivistique

- Basé sur la norme **ISAD(G)**, fonctions et structure des instruments de recherche

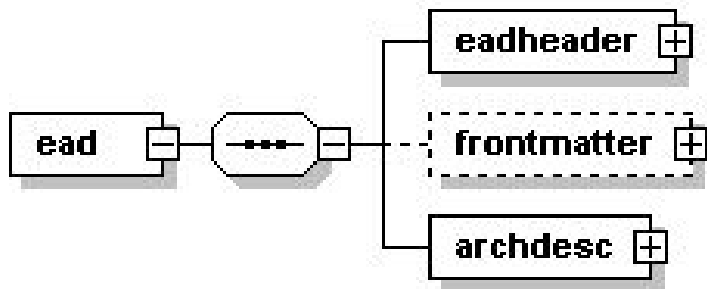
- Respect des fonds, de leur contexte de constitution et de leur structuration
  - Un enregistrement par "unité documentaire"
  - Logique de production → composants physiques hiérarchisés
  - Description du général au particulier - héritage
    - fonds/collection/série/"boîte"

- Maintenu par la Bibliothèque du Congrès

<http://www.loc.gov/standards/>, en association avec

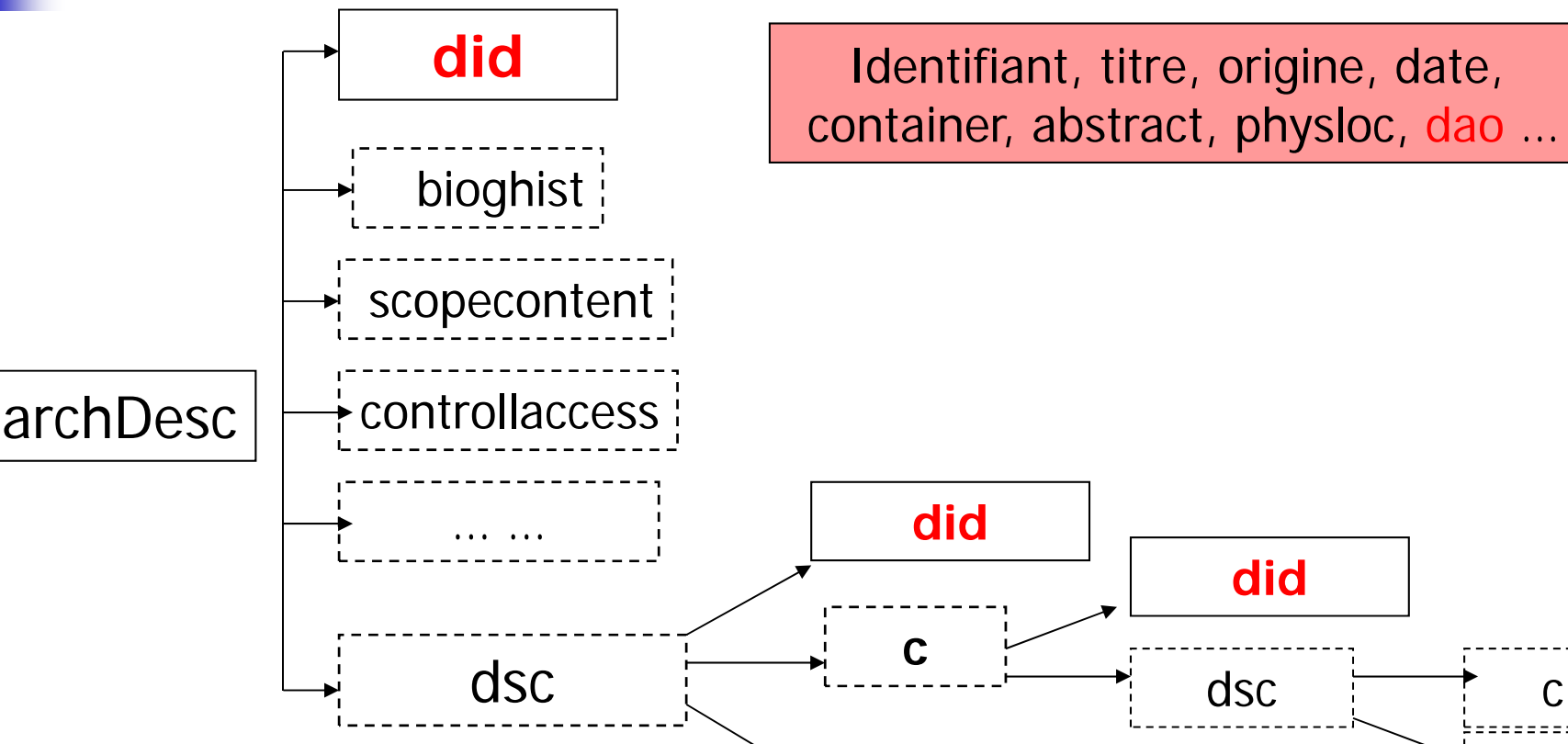
l'American Archivist Society

# EAD, structure



- **<eadheader>** métadonnées TEI décrivant le **document EAD**
- **<archdesc>** description structurée du fonds

# EAD, archDesc une vue simplifiée ...





# EAD

## ... bien présent sur le Web

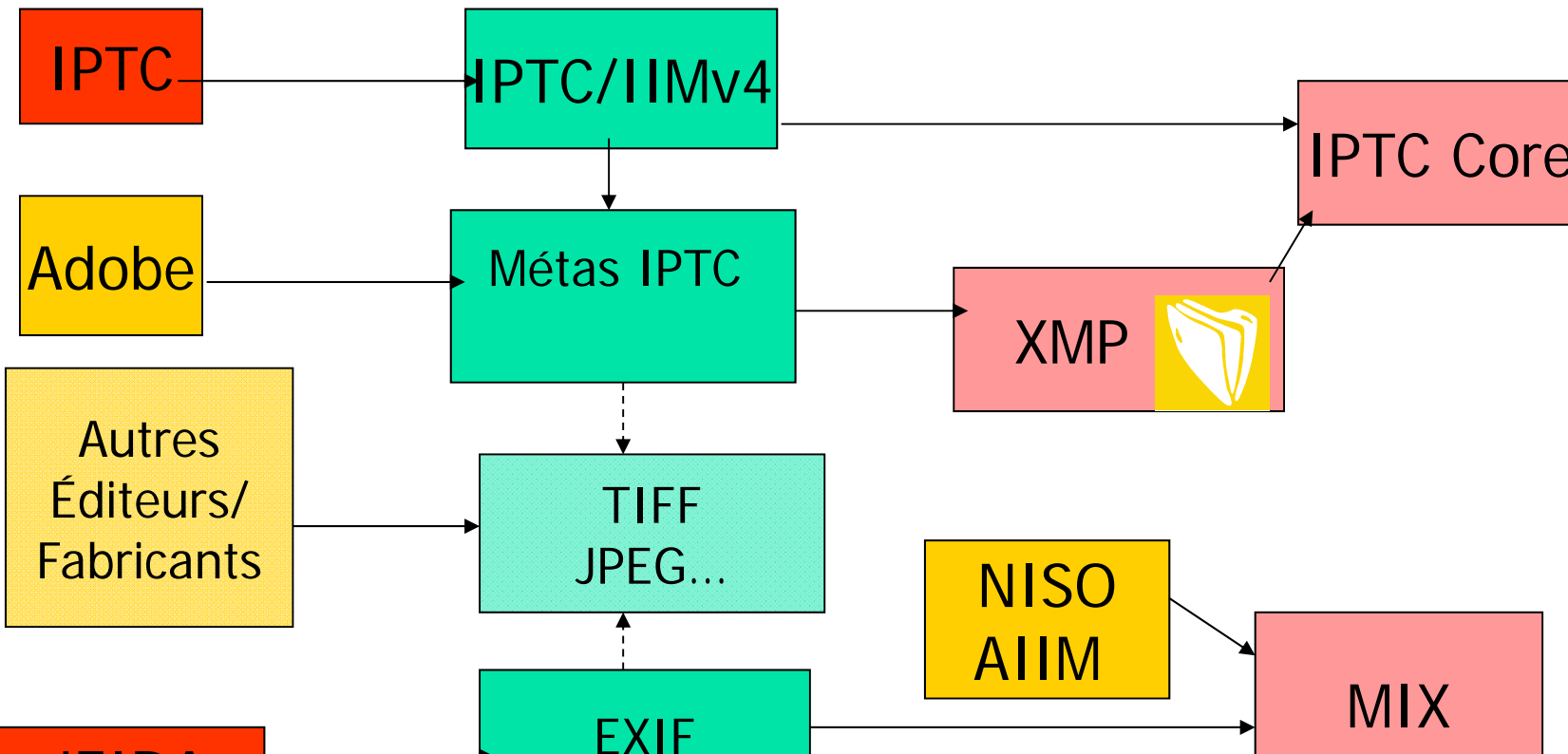
### ■ Souvent conversion d'inventaires archivistiques papier

- Centres d'archives nationales, régionales, municipales ... tous pays
  - Ex : Genève, <http://www.ville-ge.ch/archives/> ... pour la possibilité de voir le fichier EAD brut

### ■ Catalogues de collections spécialisées, et mutualisation

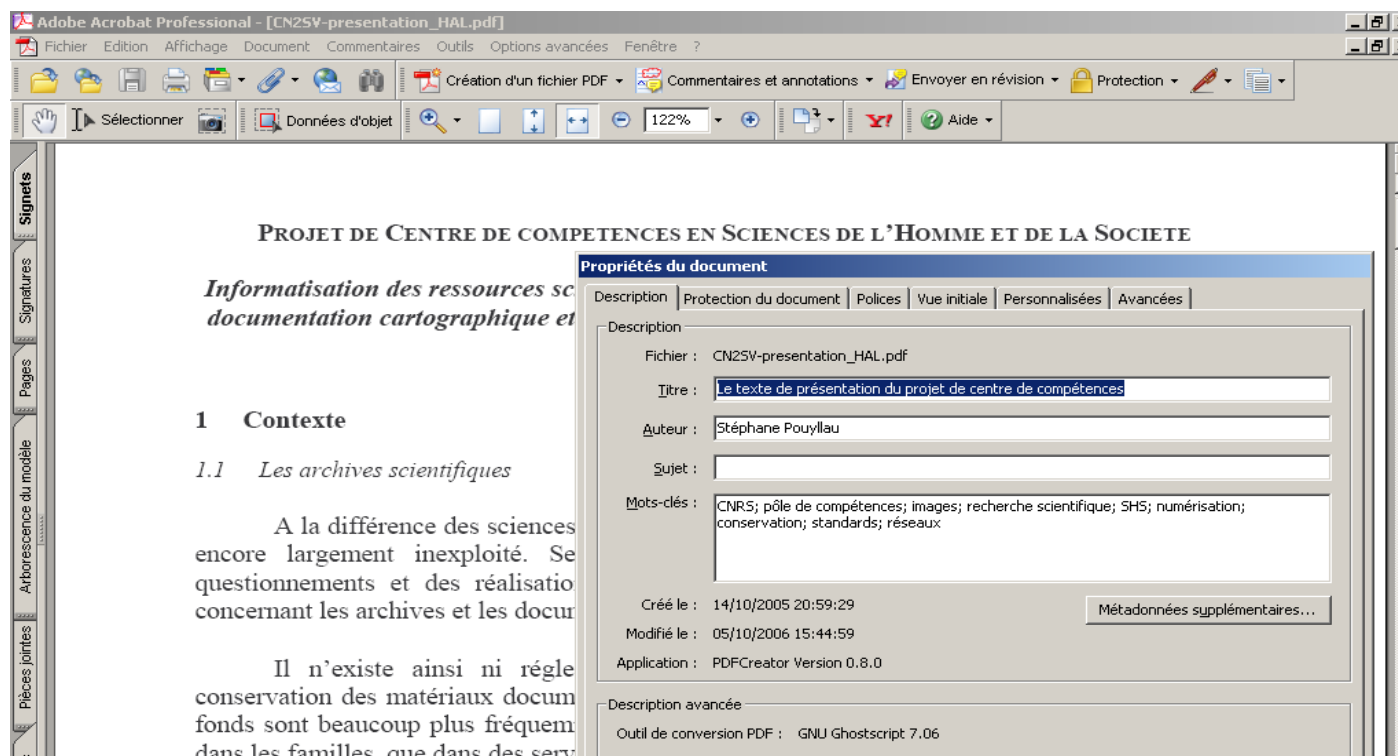
- ArchiWebture, archives du centre national de l'architecture (200-350 fonds) <http://archiwebture.citechailot.fr/awt/>
- CALAMES, catalogue collectif des archives et manuscrits de l'enseignement supérieur <http://www.calames.abes.fr/pub/>

# Les images, un monde de métadonnées



# Métadonnées "XMP" dans un PDF ou une image

chier,  
ropriétés"



# Document pdf et XMP

données  
mentaires,  
tion

Métadonnées du document pour CN25V-presentation\_HAL.pdf

Description

Avancé

Description

Titre du document : Le texte de présentation du projet de centre de compétences

Auteur : Stéphane Pouyllau

Description :

Auteur de la description : Pouyllau Stéphane ; Morel Catherine

Mots-clés : CNRS; pôle de compétences; images; recherche scientifique; SHS; numérisation; conservation; standards; réseaux

Les virgules peuvent être utilisées comme séparateurs de mots-clés.

Etat du copyright : Inconnu

Notice de copyright :

URL d'informations sur le copyright :

Atteindre l'URL

# Document pdf et XMP

## De nombreux espaces de nom

ropriétés,  
cé"



# Propriétés images en XMP partie Dublin Core

Export ouvert avec éditeur XML ... ou pdf ouvert dans le bloc note

...]

```
<rdf:Description rdf:about=" http://...  
  xmlns:dc="http://purl.org/dc/elements/1.1/">  
  <dc:format>application/pdf</dc:format>  
  <dc:title>  
    <rdf:Alt><rdf:li xml:lang="fr">Le texte de présentation du  
projet de centre de compétences</rdf:li></rdf:Alt>  
  </dc:title>  
  <dc:creator>  
    <rdf:Seq><rdf:li>Stéphane Pouyllau</rdf:li></rdf:Seq>  
</dc:creator>  
<dc:subject>
```

# Métadonnées au CN2SV

## “XML à tous les étages”

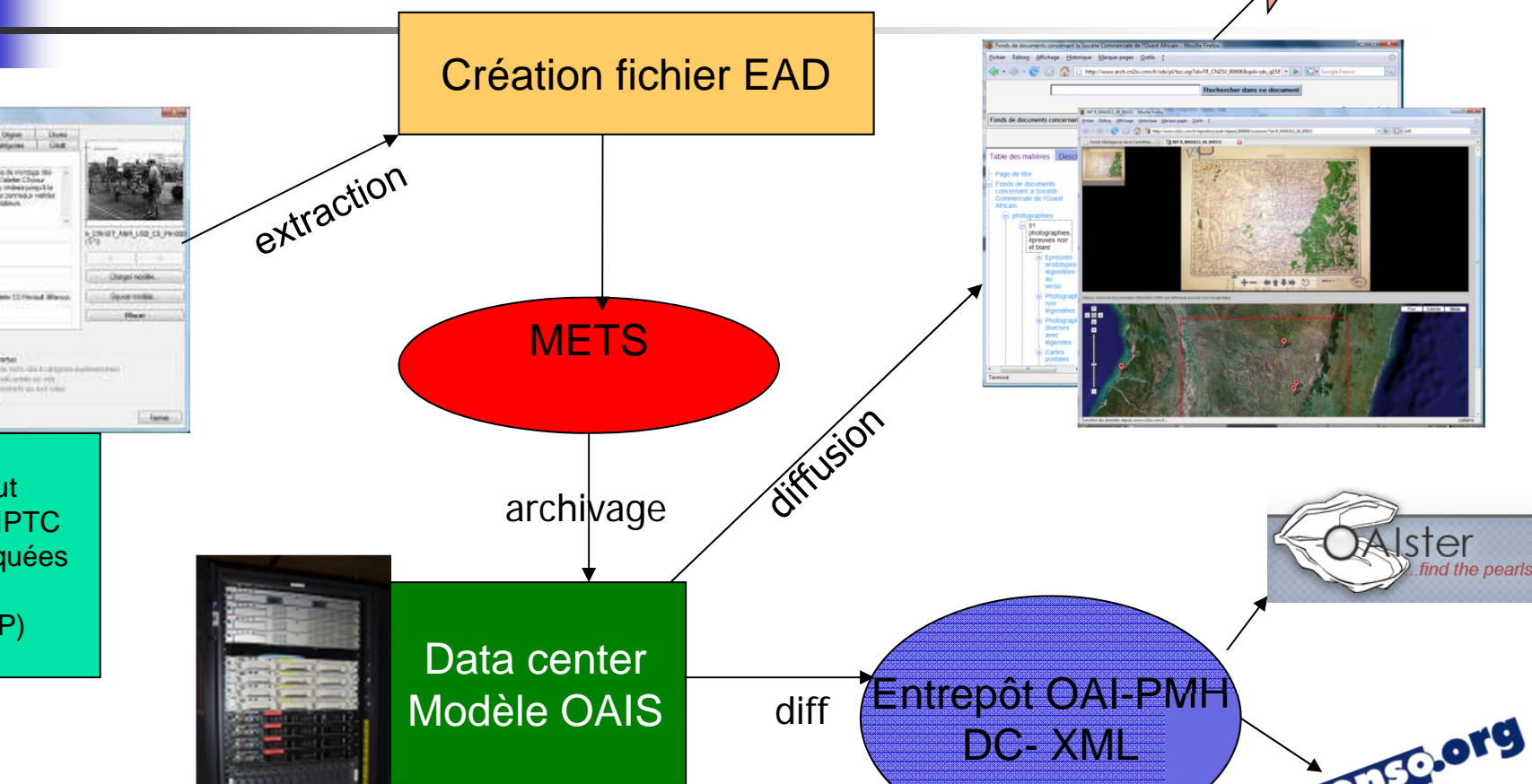
Centre National de Numérisation des Sources Visuelles,  
CNRS,

- Février 2005, adossé au CAK/CRHST, <http://www.cn2sv.fr>
- Un des Centres de Ressources Numériques SHS (TELMA ...)
- Missions :
  - Conservation numérique et mise à disposition Web des archives de labos, chercheurs - documents images
  - Veille, expertise, conseil

Démo : [www.arch.cn2sv.cnrs.fr/ead](http://www.arch.cn2sv.cnrs.fr/ead)

Des standards et des normes

# La plateforme du CN2SV pour les fonds d'archives des sciences





## III - Des outils Complément ...

- Fonctions diverses
  - **Créer et modifier** : éditeur XML, formulaires, logiciels applicatifs divers
  - **Extraire** : Acrobat Pro, Photoshop, programmes maison (php) ...
  - **Afficher** : DC viewer, feuilles de style CSS, XSL ...
  - **Convertir** : DC-DOT, XSL ...
- Outils libres et commerciaux variés
  - Dont logiciels "documentaires", archivistiques, photo
  - > 600 réponses dans SourceForge !
  - Un exemple de base : transformation XSL

# Feuilles de style XSL

- Ensemble de règles qui associent aux éléments des "formes de sortie"
- Dans chaque règle
  - on définit un élément à traiter (chemin)
  - on donne une/des instruction(s) de traitement
- Une feuille XSL est un document XML

```
<xsl:template match="nom_élément">  
  <'action'> [diverses]
```

# XSL

Résultat : éléments Dublin Core

```
<dc:title>Récepteur particulière déterministe en  
communications radiomobiles pour canaux multi-trajets  
évanescents</dc:title>
```

```
<dc:creator>BEN SALEM (Farah)</dc:creator>
```

ce INIST (CPI-XML), zones titre et auteur

```
cord ...> ...
```

```
<fA08 dir=011FRE><s1>Récepteur particulière  
déterministe en communications radiomobiles pour canaux multi-  
trajets évanescents</s1></fA08>
```

```
<fA11 dir=011000><s1><xsl:template match="fA08">
```

```
</fA11> ...
```

```
cord>
```

Feuille de style XSL, extrait

```
<xsl:element name="dc:title">  
  <xsl:value-of select="s1"/>  
</xsl:element>  
</xsl:template>  
<xsl:template match="fA11">  
  <xsl:element  
    name="dc:author">
```

# Transformation XSL

## Deux exemples

### ■ 'En famille', suite

- Fichier source : un arbre généalogique XML
- Un fichier, X sorties ...
  - Avec axe2.xsl et axe2b.xsl : résultats HTML, format de présentation
  - Avec axe3.xsl : résultat XML, répertoire avec noms d'éléments anglais = conversion de métadonnées



### ■ Transformation d'un fichier thèse vers un fichier Dublin Core ...

- Pour moissonnage OAI-PMH ...

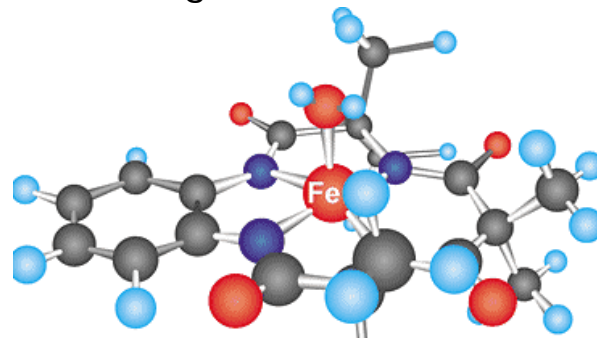
onus !

# Métadonnées et Web sémantique

- D'un Web de "documents" à un Web de "données"
  - "libérées", du document, de la BD, ...et liées
  - Compréhensibles par les machines
    - génèrent d'autres données "automatiquement" (inférence)
- Structuration des données par standards
  - Triplets RDF : sujet-propriété-objet ...
    - Dérivant personnes, créations, lieux, connaissances, droits ...
    - Avec/dans des "ontologies" : OWL, DC, FOAF, geo ...

dia, WordNet, Geonames, Libris...

➤ **10 MILLIARDS triplets !**



mple

## Le fil RSS

# du Web sémantique ... au Web 2

```
<rss version="0.91">
  <channel>
    <title>Libre accès a l'information scientifique et technique. Actualites,
problématiques et perspectives</title>
    <link>http://www.inist.fr/openaccess/</link> [...]
    <item>
      <title> Débat autour des propositions britanniques en faveur de
l'auto-archivage</title>
      <link>http://www.inist.fr/openaccess/breve.php3?id\_breve=311
</link>
      <date>2005-07-13 14:13:15</date>
      <description> La proposition publiée par les conseils de la
recherche britanniques (Resarch Councils ou RCUK) de conditionner le
financement des recherches qu'ils subventionnent à une obligation d'auto-
archivage des publications suscite un vif débat.... </description>
    </item>
  </channel>
</rss>
```

Pipes: editing 'Gina Trapani Master Feed' - Mozilla Firefox

http://pipes.yahoo.com/pipes/pipe.edit?\_id=pncRetu72xGQcIW81BeTaQ

Layout Expand All Collapse All [Pipe Loaded](#) [Run Pipe...](#) [Back to My Pipes](#) [New](#)

**Sources**

- Fetch CSV +
- Feed Auto-Disco +
- Fetch Feed +
- Fetch Data +
- Fetch Page +
- Fetch Site Feed +
- Flickr +
- Google Base +
- Item Builder +
- RSS Item Builder +
- Yahoo! Local +
- YQL +
- Yahoo! Search +

**User inputs**

**Operators**

- Uri
- String
- Date
- Location
- Number

**Fetch Feed**

- URL
- http://del.icio.us/rss/ginatrapani
- http://api.flickr.com/services/feeds
- http://feeds.gawker.com/lifehacke
- http://feeds.feedburner.com/Spun

**Filter**

Block items that match all of the following

- Rules
- item.link Contains lifehacker
- item.description Does not contain Trapani

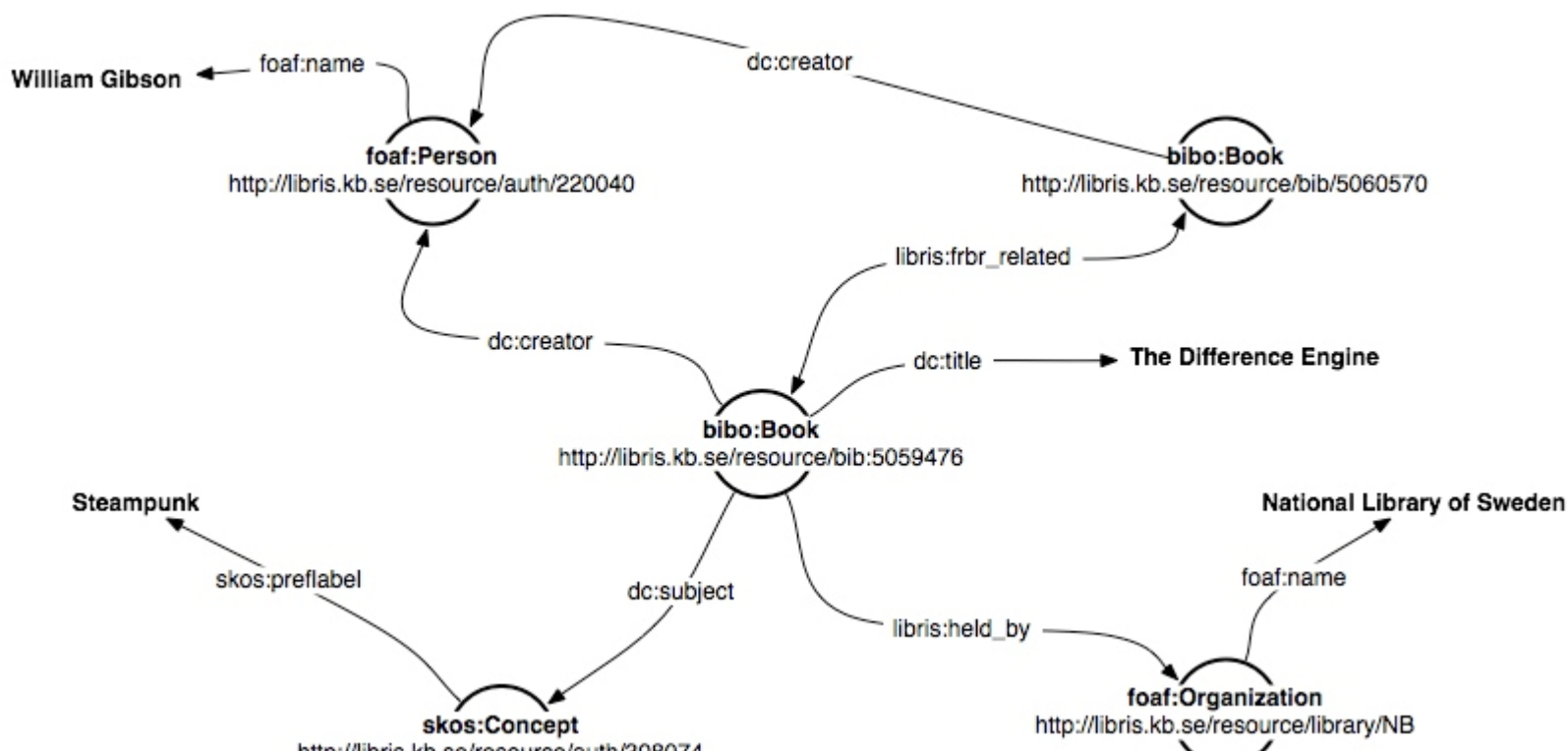
**Sort**

- Sort by
- item.pubDate in descending order

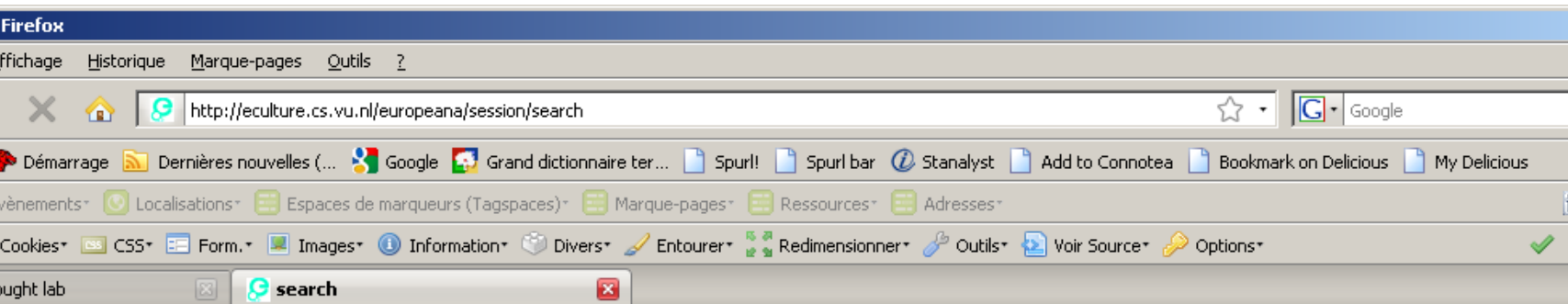
The screenshot shows the Yahoo Pipes editor interface. On the left, there is a sidebar with 'Sources' and 'User inputs' categories. The main workspace contains a workflow diagram with three operators: 'Fetch Feed', 'Filter', and 'Sort', connected by blue lines. The 'Fetch Feed' operator has four URLs listed. The 'Filter' operator is configured to block items that match all of the following rules: 'item.link' contains 'lifehacker' and 'item.description' does not contain 'Trapani'. The 'Sort' operator is configured to sort by 'item.pubDate' in descending order. The top of the browser window shows the URL 'http://pipes.yahoo.com/pipes/pipe.edit?\_id=pncRetu72xGQcIW81BeTaQ' and various browser navigation and bookmark elements.

mp

# LIBRIS, catalogue collectif suédois "mettre les données de l'IST en réseau"







[login](#) | [help](#) | [feedback](#) | [admin](#) | [English](#) ▾

le...



europeana  
think culture


This is a [research prototype](#) of Europeana's semantic search engine.  
Enter a search term, for example: [Egypt](#), [Rembrandt](#), [window](#).

Search input:

concept	
<b>Botticelli</b>	old master
<b>Botticino marble</b>	marble by composition or origin
<b>Sandro Botticelli</b>	

location	
<b>Botticino Sera</b>	inhabited place

person		<a href="#">view all 21 results</a> ▶
<b>Botticini, Francesco</b>	Italian artist 1446 1497	
<b>Botticelli, Sandro</b>		



Louvre  
11,327 artworks



Kandelaar  
Verrocchio, Andrea del

Maria met kind  
Verrocchio, Andrea del

Borstbeeld van de tr...  
Verrocchio, Andrea del

Maria met kind, knie...  
Verrocchio, Andrea del

▼ works related to a student of **Botticelli, Sandro** (3)



Twee naakte jongelin...  
Huber, Wolfgang

Aanbidding van de ko...  
Robetta, Christoforo...

Aanbidding van de ko...  
Robetta, Christoforo...

▼ works related to an artist professionally related to **Botticelli, Sandro** (6)



Wildeman met schild,...



Wildeman met schild,...



Gevecht tussen elf s...  
Claesz., Allaert



Maria met kind  
Anoniem



Engel en putti  
Humbert de Supervill...

Fragment met Hercule...  
Robetta, Christoforo...

▼ works created by an artist professionally related to a student of **Botticelli, Sandro** (1)

ple

# Valorisation sémantique pages HTML RDFa, microformats

---

- Marquage sémantique d'éléments de pages → utilisation machine
  - ... blabla ...par <span property=DC.creator>Gauthier Poupeau</span>, qui ...blabla ...
- Exemple : une publication scientifique
  - <http://imageweb.zoo.ox.ac.uk/pub/2008/plospaper/latest/>
  - Dans le code, éléments sémantiques balisés, nombreux "espaces de nom"
- L'article expliquant l'expérience : [... semantic enhancement of a research article ...](#)

turn all highlighting off

date

disease

habitat

institution

organism

person

place

protein

taxon

[Top](#) | [Abstract](#) | [Author Summary](#) | [Introduction](#) | [Methods](#) | [Results](#) | [Discussion](#) | [Supporting Information](#) | [Acknowledgements](#) | [References](#) | [Data Fusion Supplements](#)

March 2009.

alphabetical order

publication year

frequency of citation (within this paper)

numerical order (original)

typing off

United Nations Human Settlements Programme (2003) The challenge of **slums**: Global report on human settlements 2003. London: Earthscan Publications Ltd. [Link](#) (CITO: *obtains background from, Report, Book, Online Document, not peer reviewed*)

Riley LW, Ko AI, Unger A, Reis MG (2007) Slum health: Diseases of neglected populations. BMC Int Health Hum Rights 7: 2. [DOI PubMed PubMedCentral](#) (CITO: *obtains background from, shares authors with, Opinion, Journal Article, peer reviewed*)

Garau P, Carolini G (2005) The 21st century health challenge of **slums** and **cities**. Lancet 365: 901–903. [DOI PubMed](#) (CITO: *obtains background from, Opinion, Journal Article, peer reviewed*)

United Nations General Assembly of United Nations (2000) United Nations Millennium Declaration. [Link](#). (CITO: *obtains background from, Report, Report Document, Online Document, not peer reviewed*)

Wright A, Lenton R, Lewis K, Frampton J (2005) Focusing on improved water and sanitation for health. Lancet 365: 810–812. [DOI PubMed](#) (CITO: *obtains background from, Opinion, Journal Article, peer reviewed*)

Riley LW, Ko AI, Reis MG, Ribeiro Dourado CM, Johnson WD Jr, Riley LW (1999) Urban epidemic of severe **leptospirosis** in **Brazil**. Salvador Leptospirosis Study Group. Lancet 354: 820–825. [DOI PubMed](#) (CITO: *obtains background from, uses data from, confirms, extends, shares authors with, Research Paper, Journal Article, peer reviewed*)

Ko AI, Athanazio DA, Reis MG, Ko AI (2005) **Leptospirosis**. Curr Opin Infect Dis 18: 376–386. [DOI PubMed](#) (CITO: *obtains background from, shares authors with, Review, Journal Article, peer reviewed*)

Nally JE, Ricaldi JN, Matthias MA, Diaz MM, et al. (2003) **Leptospirosis**: A **zoonotic disease** of global importance. Lancet Infect Dis 3: 767–771. [DOI PubMed](#) (CITO: *obtains background from, Review, Journal Article, peer reviewed*)

Reis MG, Ko AI, Nally JE, Ricaldi JN, et al. (2001) **Leptospirosis**. Clin Microbiol Rev 14: 296–326. [DOI PubMed PubMedCentral](#) (CITO: *obtains background from, uses method in, Review, Journal Article, peer reviewed*)

Collins-Richards D, Brouwer KC, Cunningham CB, et al. (2006) Determining risk for severe **leptospirosis** by molecular analysis of environmental surface waters for pathogenic **Leptospira**. PLoS Med 3: e308. [DOI PubMed](#) (CITO: *obtains background from, confirms, extends, shares authors with, Research Paper, Journal Article, peer reviewed*)

Sabroza PC, Barcellos C (2000) Socio-environmental determinants of the **leptospirosis** outbreak of 1996 in **western Rio de Janeiro**: A geographical

occupational disease  
zoonotic disease

tropical rain forest

city

household

open drainage system

open rainwater drainage system

open sewer drainage system

refuse deposit

slum

alley

workplace

[Brazilian National Commission for Ethics in Research](#)

[Brazilian National Research Council](#)

[Centers for Diseases Control and Prevention](#)

[Centro de Pesquisas Gonçalo Moniz, Fundação Oswaldo Cruz, Ministério da Saúde](#)

[Company for Urban Development of the State of Bahia \(CONDER\)](#)

[Division of International Medicine and Infectious Diseases, Weill Medical College of Cornell University](#)

[Environmental Systems Research Institute](#)

[Escola Nacional da Saúde Pública, Fundação Oswaldo Cruz, Ministério da Saúde](#)

[Image Bioinformatics Research Group, Department of Zoology, University of Oxford](#)

[MRC Biostatistics Unit](#)

[Oswaldo Cruz Foundation](#)

[Oswaldo Cruz Foundation](#)



# Conclusion

Merci de votre attention !

- Une question incontournable pour les ressources numériques
  - “discovery” autant que gestion, quotidienne et future
  - Interopérabilité et mutualisation
- Une réalité
  - Des standards et des normes ... Des réflexions (modélisation)
  - Des réalisations en bibliothèques
  - Des outils ... de plus en plus !
  - Des documents et des blogs-métiers : Figoflog, lespetitescases ...
- Une interopérabilité et une intégration croissantes
  - Via XML et RDF

Via les “travaux en cours” de W3C, DCS, RDF et “travaux en cours”