



# Introduction aux métadonnées



Jean-Marc LECARPENTIER

Université de Caen Basse-Normandie  
GREYC, CRNS UMR 6072  
Équipe HULTECH

19 décembre 2012



# Plan

Metadonnées ?  
Resource Description Framework  
Vocabulaires  
Dublin Core  
Métadonnées embarquées  
Conclusion

- 1 Metadonnées ?
- 2 Resource Description Framework
- 3 Vocabulaires
- 4 Dublin Core
- 5 Métadonnées embarquées
- 6 Conclusion



# Metadata

## Metadonnées ?

Resource Description Framework

Vocabulaires

Dublin Core

Métadonnées embarquées

Conclusion

- "data about data"
- "Machine-understandable information about Web resources or other things"

Tim Berners-Lee, W3C (1997)



# Metadata

Metadonnées ?  
Resource Description Framework  
Vocabulaires  
Dublin Core  
Métadonnées embarquées  
Conclusion

- "Structured data about resources that can be used to help support a wide range of operations  
Michael Day, 2001
- "Structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use or manage"  
information objects  
NISO, 2004



# Comment définir les métadonnées

Metadonnées ?  
Resource Description Framework  
Vocabulaires  
Définition  
Métadonnées embarquées  
Conclusion

## Diverses fonctions

- "Data associated with objects which relieves their potential users of having to have full advance knowledge of their existence or characteristics"

Dempsey & Heery, 1998

- Catégories de métadonnées :
  - Descriptive
  - Structure
  - Administrative
  - Technique



# Quelles fonctionnalités ?

Metadonnées ?  
Resource Description Framework  
Vocabulaires  
Dublin Core  
Métadonnées embarquées  
Conclusion

- Publication et découverte de ressources
- Utilisation de ressources
- Gestion de la propriété intellectuelle
- Commerce
- Localisation
- Évaluation de contenu
- ...



# À quoi appliquer les métadonnées ?

Metadonnées ?  
Resource Description Framework  
Vocabulaires  
Droits  
Métadonnées embarquées  
Conclusion

- "Web resources or other things," c'est-à-dire :
- sites Web, pages Web, images, bases de données, objets de musées, localisations géographiques, événements, concepts, archives, etc.
- Aux métadonnées elles-mêmes



# Où trouver les métadonnées

Metadonnées ?  
Resource Description Framework  
Vocabulaires  
Dublin Core  
Métadonnées embarquées  
Conclusion

- Dans une ressource :
  - titre et table des matières (livre), balises Meta (pages web), tags ID3 (MP3), etc.
- Liée à la ressource :
  - balises link rel="meta" dans les pages web
- Gestion indépendante





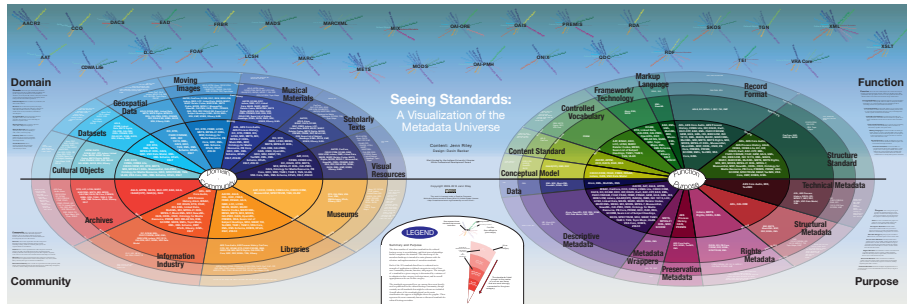
# Quelle importance ?

- ... "is recognised as a critically important, and yet increasingly problematic and complex concept with relevance for information objects as they move through time and space"  
Gilliland-Swetland (2004)
- MAIS...
- il existe énormément de formats, schémas de métadonnées



# L'univers des métadonnées

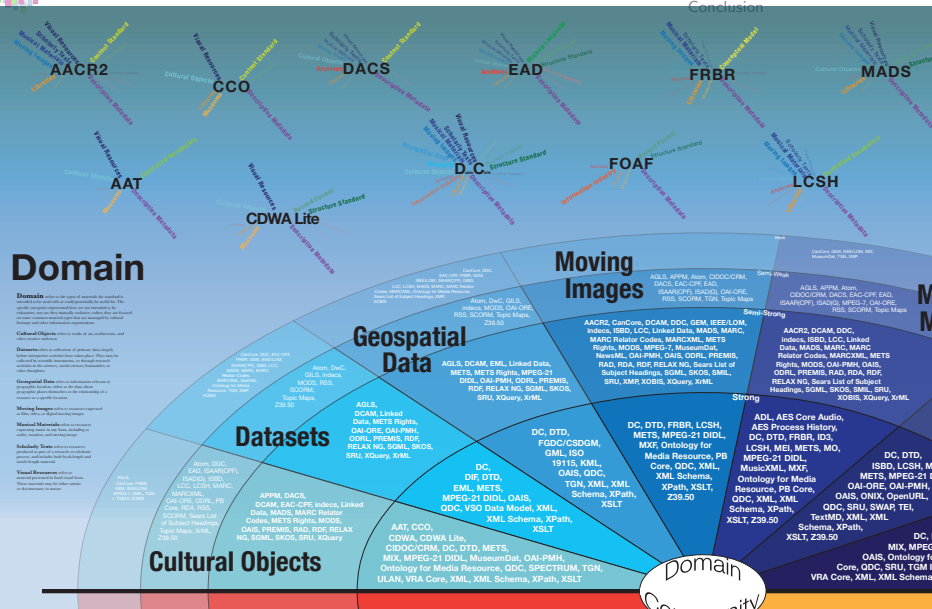
Metadonnées ?  
Resource Description Framework  
Vocabulaires  
Dublin Core  
Métadonnées embarquées  
Conclusion



<http://www.dlib.indiana.edu/~jenlrile/metadatamap/>

# L'univers des métadonnées

Metadonnées ?  
Resource Description Framework  
Vocabulaires  
Dublin Core  
Métadonnées embarquées  
Conclusion





# Standards pour les métadonnées

Metadonnées ?  
Resource Description Framework  
Vocabulaires  
Dublin Core  
Métadonnées embarquées  
Conclusion

- "resource management communities"
- inclut en général
  - Modèle conceptuel
  - Ensemble de composants nommés ("terms", "elements", etc.)
  - Spécification de la représentation ("binding")



# En résumé

Metadonnées ?  
Resource Description Framework  
Vocabulaires  
Dublin Core  
Métadonnées embarquées  
Conclusion

- Métadonnées omniprésentes
- Pas uniquement pour trouver des objets ("discovery")
- Échanges d'informations (ex. des images)
- Problématique de la diversité des standards



# Resource Description Framework

Metadonnées ?  
Resource Description Framework  
Vocabulaires  
Dublin Core  
Métadonnées embarquées  
Conclusion

- Spécification W3C, 1999
- Représenter des informations à propos de ressources sur le Web
- Adapté à la description des métadonnées
- Adapté au traitement de l'information par des machines





# Principe

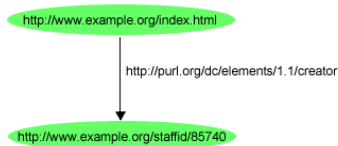
- Utilisation des URI pour désigner des choses
- Ressources décrites par leurs propriétés et leurs valeurs
- Exemple : description d'une personne





# Concepts

- `http://www.example.org/index.html` a un créateur (creator) dont la valeur est John Smith
- Modéliser le fait que John Smith a créé une page web
- Sujet : `http://www.example.org/index.html`
- Prédicat : creator
- Objet : John Smith
- *machine-processable* :
  - identificateurs pour sujet, prédicat et objet : URI
  - langage pour représenter ces relations : RDF/XML







# Notations

- Sous forme de triplets (*triples*) :

`<http://www.ex.org/index.html> <http://purl.org/dc/elements/1.1/creator> <http://www.ex.org/staffid/85740>`

- Définition des *namespace* XML :

```
prefix dc:, namespace URI: http://purl.org/dc/elements/1.1/  
prefix ex:, namespace URI: http://www.example.org/  
prefix exstaff:, namespace URI: http://www.example.org/staffid/
```

Notation :

`ex:index.html`                      `dc:creator`                      `exstaff:85740` .



# Notation RDF/XML

Metadonnées ?  
Resource Description Framework  
Vocabulaires  
Dublin Core  
Métadonnées embarquées  
Conclusion

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
          xmlns:exterm="http://www.example.org/terms/"

  <rdf:Description rdf:about="http://www.example.org/index.html">
    <exterm:creation-date>August 16, 1999</exterm:creation-date>
  </rdf:Description>

</rdf:RDF>
```



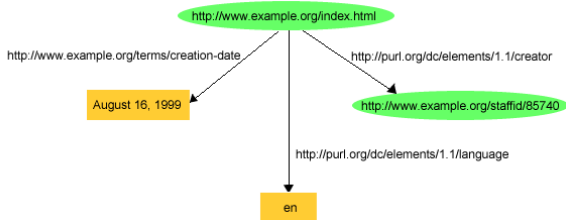
# RDF/XML avec plusieurs propriétés

Metadonnées ?  
Resource Description Framework  
Vocabulaire  
DUPP Code  
Métadonnées embarquées  
Conclusion

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
          xmlns:dc="http://purl.org/dc/elements/1.1/"
          xmlns:exterms="http://www.example.org/terms">

  <rdf:Description rdf:about="http://www.example.org/index.html">
    <exterms:creation-date>August 16, 1999</exterms:creation-date>
    <dc:language>en</dc:language>
    <dc:creator rdf:resource="http://www.example.org/staffid/85740"/>
  </rdf:Description>

</rdf:RDF>
```

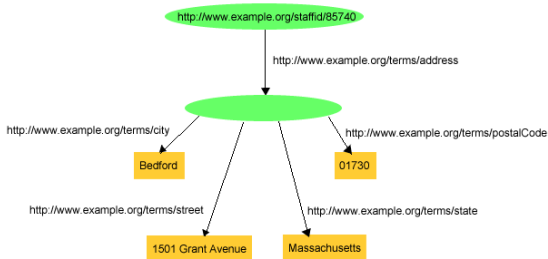




# Noeuds anonymes

Metadonnées ?  
Resource Description Framework  
Vocabulaires  
Dublin Core  
Métadonnées embarquées  
Conclusion

- Exemple : objet de type adresse sans URI
- Créer un URI  
→ lourdeurs
- Utilisation de noeud anonyme





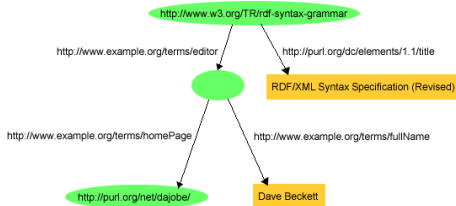
# Noeuds anonymes

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:exterms="http://example.org/stuff/1.0/">

  <rdf:Description rdf:about="http://www.w3.org/TR/rdf-syntax-grammar">
    <dc:title>RDF/XML Syntax Specification (Revised)</dc:title>
    <exterms:editor rdf:nodeID="abc"/>
  </rdf:Description>

  <rdf:Description rdf:nodeID="abc">
    <exterms:fullName>Dave Beckett</exterms:fullName>
    <exterms:homePage rdf:resource="http://purl.org/net/dajobe/">
  </rdf:Description>

</rdf:RDF>
```





# Littéraux typés

- Valeurs scalaires
- Utilisation de XML-Shema

```
<rdf:Description rdf:about="http://www.example.org/index.html">  
  <extms:creation-date rdf:datatype="http://www.w3.org/2001/XMLSchema#date">1999-08-16  
  </extms:creation-date>  
</rdf:Description>
```

http://www.example.org/staffid/85740

http://www.example.org/terms/age

27

http://www.example.org/staffid/85740

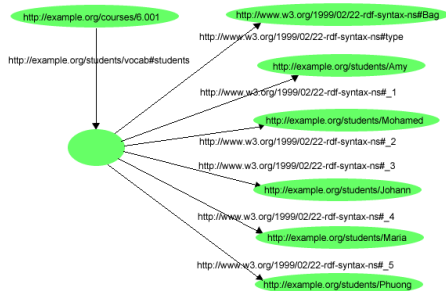
http://www.example.org/terms/age

"27"^^http://www.w3.org/2001/XMLSchema#integer



# Exemple avec rdf :Bag

- Description de *groupes* de choses
- 3 types de conteneurs :
  - "sac" : `rdf:Bag`
  - liste : `rdf:Seq`
  - alternative : `rdf:Alt`
- Utilisation de `rdf:type` pour spécifier le type de conteneur



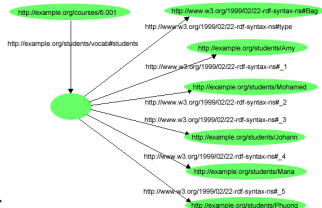


# Exemple avec rdf :Bag

Metadonnées ?  
Resource Description Framework  
Vocabulaires  
Dublin Core  
Métadonnées embarquées  
Conclusion

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:s="http://example.org/students/vocab#">

  <rdf:Description rdf:about="http://example.org/courses/6.001">
    <s:students>
      <rdf:Bag>
        <rdf:li rdf:resource="http://example.org/students/Amy"/>
        <rdf:li rdf:resource="http://example.org/students/Mohamed"/>
        <rdf:li rdf:resource="http://example.org/students/Johann"/>
        <rdf:li rdf:resource="http://example.org/students/Maria"/>
        <rdf:li rdf:resource="http://example.org/students/Phuong"/>
      </rdf:Bag>
    </s:students>
  </rdf:Description>
</rdf:RDF>
```







# RDF en bref

Metadonnées ?  
Resource Description Framework  
Vocabulaires  
Dublin Core  
Métadonnées embarquées  
Conclusion

- Possibilités immenses
- Principes très simples
- Mais utilisation pas si simple !
- Peu lisible par humains
- Traitement par les machines
- Notation basée sur XML (URI, namespaces, etc)



# Problématique

Metadonnées ?  
Resource Description Framework  
Vocabulaires  
Dublin Core  
Métadonnées embarquées  
Conclusion

- S'assurer que l'on parle de la même chose

- ```
<rdf:Description rdf:ID="item10245">  
  <rdf:type rdf:resource="http://www.example.com/terms/Tent"/>  
  <exterm:model rdf:datatype="&xsd:string">Overnighter</exterm:model>  
  <exterm:sleeps rdf:datatype="&xsd:integer">2</exterm:sleeps>  
  <exterm:weight rdf:datatype="&xsd:decimal">2.4</exterm:weight>  
</rdf:Description>
```
- ```
<rdf:Description rdf:ID="item10245">  
  <rdf:type rdf:resource="http://www.decathlon.fr/terms/Tentes"/>  
  <decat:modele rdf:datatype="&xsd:string">Overnighter</decat:modele>  
  <decat:nbPers rdf:datatype="&xsd:integer">2</decat:nbPers>  
  <decat:poids rdf:datatype="&xsd:decimal">2.4</decat:poids>  
</rdf:Description>
```



# Schéma ou Vocabulaire ?

Metadonnées ?  
Resource Description Framework  
**Vocabulaires**  
Dublin Core  
Métadonnées embarquées  
Conclusion

- Schéma : ensemble de règles pour encoder l'information, définies par une communauté d'utilisateurs
- Vocabulaire : ensemble des valeurs possibles pour une information
- Exemple : Type de reliure d'un livre → broché, relié, etc



# Définition d'un vocabulaire

Metadonnées ?  
Resource Description Framework  
**Vocabulaires**  
Dublin Core  
Métadonnées embarquées  
Conclusion

Dépend du niveau de contrôle souhaité

- RDF-Schema (RDFS)
- Classification (SKOS)
- Ontologies (OWL)



# Choix d'un schéma

- Dépend de :
  - domaine d'application
  - rôle
  - usage
  - public
- difficulté du choix
- difficulté de l'interopérabilité entre schémas
- Création d'un schéma basique satisfaisant un cadre très généraliste : Dublin Core



# Application Profile

Metadonnées ?  
Resource Description Framework  
**Vocabulaires**  
Dublin Core  
Métadonnées embarquées  
Conclusion

- *Mix and match* d'éléments de divers schémas
- Combinés pour un type d'utilisation spécifique
- Permet la réutilisation de standards existants
- Sans avoir à créer de nouveaux schémas



# Dublin Core

Metadonnées ?  
Resource Description Framework  
Vocabulaires  
**Dublin Core**  
Métadonnées embarquées  
Conclusion

- Dublin Core Metadata Initiative (DCMI)
- Création en 1995 suite à un *workshop* à Dublin, Ohio
- Schéma généraliste
- Renseigner de façon universelle des informations de base
- Mission : fournir des standards simples pour faciliter la découverte, le partage et la gestion de l'information
- Consensus interdisciplinaire sur 15 éléments



# Dublin Core

Metadonnées ?  
Resource Description Framework  
Vocabulaires  
**Dublin Core**  
Métadonnées embarquées  
Conclusion

- Consensus sur 15 éléments pour la découverte d'informations
- Éléments optionnels, répétables
- Axé sur la simplicité → ne convient pas pour des descriptions complexes
- Pas de règles de catalogage
- Fournit une interopérabilité sémantique de base au travers des domaines et des communautés





# Éléments de Dublin Core

Metadonnées ?  
Resource Description Framework  
Vocabulaires  
**Dublin Core**  
Métadonnées embarquées  
Conclusion

- Title
- Subject
- Description
- Creator
- Publisher
- Contributor
- Date
- Type
- Format
- Identifier
- Source
- Language
- Relation
- Coverage
- Rights



# Dublin Core Application Profiles

Metadonnées ?  
Resource Description Framework  
Vocabulaires  
Dublin Core  
Métadonnées embarquées  
Conclusion

- Utilisation de DC enrichi d'autres schémas
- Prévus pour des applications de types spécifiques
- Exemple : UK eGovernment Metadata Standard (eGMS)



# Dublin Core en RDF

Metadonnées ?  
Resource Description Framework  
Vocabulaires  
**Dublin Core**  
Métadonnées embarquées  
Conclusion

```
<?xml version="1.0"?>

<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/">

  <rdf:Description rdf:about="http://www.w3schools.com">
    <dc:description>W3Schools - Free tutorials</dc:description>
    <dc:publisher>Refsnes Data as</dc:publisher>
    <dc:date>2008-09-01</dc:date>
    <dc:type>Web Development</dc:type>
    <dc:format>text/html</dc:format>
    <dc:language>en</dc:language>
  </rdf:Description>

</rdf:RDF>
```



# DC en bref

Metadonnées ?  
Resource Description Framework  
Vocabulaires  
**Dublin Core**  
Métadonnées embarquées  
Conclusion

- Simplicité
- Usage très répandu
- **Le** schéma à utiliser



# Métadonnées embarquées

Metadonnées ?  
Resource Description Framework  
Vocabulaires  
Dublin Core  
Métadonnées embarquées  
Conclusion

- Métadonnées inscrites dans le document  
→ informations transférées avec le fichier
- Autres types :
  - pages Web (meta tags, etc)
  - tags ID3 pour les MP3
  - documents PDF
  - audio, vidéo avec MPEG-7
  - livres au format ePub



# Pages web

- Métadonnées générales de la page
  - Balises meta, link
  - Utiliser RDF avec `<rdf:RDF ...>` : quid de HTML5 ?
- Métadonnées insérées dans le contenu de la page
  - Micro-formats : utilisation des attributs class de façon détournée
  - RDFa1.1 (<http://www.w3.org/TR/xhtml1-rdfa-primer/>)
  - Microdata (<http://dev.w3.org/html5/md/>)
  - Quel choix ? Quelle utilisation par les moteurs ?



# Métadonnées et images

Metadonnées ?  
Resource Description Framework  
Vocabulaires  
Dublin Core  
**Métadonnées embarquées**  
Conclusion

- Enjeux importants
- Droits d'utilisation, de reproduction
- Utilisation par les logiciels
- Interopérabilité



# Métadonnées et images

Metadonnées ?  
Resource Description Framework  
Vocabulaires  
Dublin Core  
Métadonnées embarquées  
Conclusion

- Deux schémas principaux
  - Exchangeable Image Format (EXIF) : données techniques fournies par l'appareil
  - International Press Telecommunications Council (IPTC) : données sur le contenu de l'image, ses ayants-droits, etc
- XMP (Extensible Metadata Platform) : mécanisme générique d'inclusion de métadonnées dans les fichiers

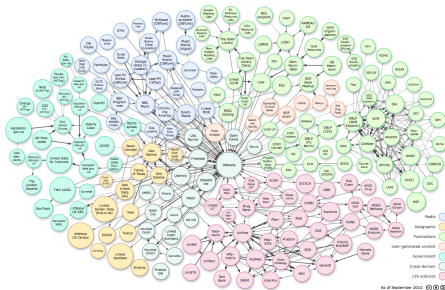




# Linked data

Metadonnées ?  
Resource Description Framework  
Vocabulaires  
Dublin Core  
Métadonnées embarquées  
Conclusion

- Web de données
- Inclure dans les pages web des relations vers des ressources (internes ou externes)
- Ressources identifiées par des URIs HTTP qui existent
- URIs correspondent à des données lisibles par les humains mais aussi par les machines avec RDF/XML
- Données pouvant elles-aussi renvoyer vers d'autres URIs





# Conclusion

- Sujet vaste et complexe
- Solutions dépendent du contexte et des communautés de pratique
- Problématique pour le développeur : quid des métadonnées ?
  - lesquelles ? quel schéma ? quel vocabulaire ?
  - métadonnées le plus souvent non gérées par les applications
- Comment apporter des solutions "génériques" qui peuvent être mises en application dans divers cas ?