

- ▶ **Cours /élément de module** : M 2.1.1 – Metadata et Description de documents
- ▶ **Niveau**: 1^{ère} année – Cycle Ingénieur des données et des connaissances
- ▶ **Professeur** : M. Amine SENNOUNI
- ▶ **Année universitaire**: 2020–2021
- ▶ **Partie 1/2 du cours**: la Recherche d'Information et les métadonnées : une vue d'ensemble

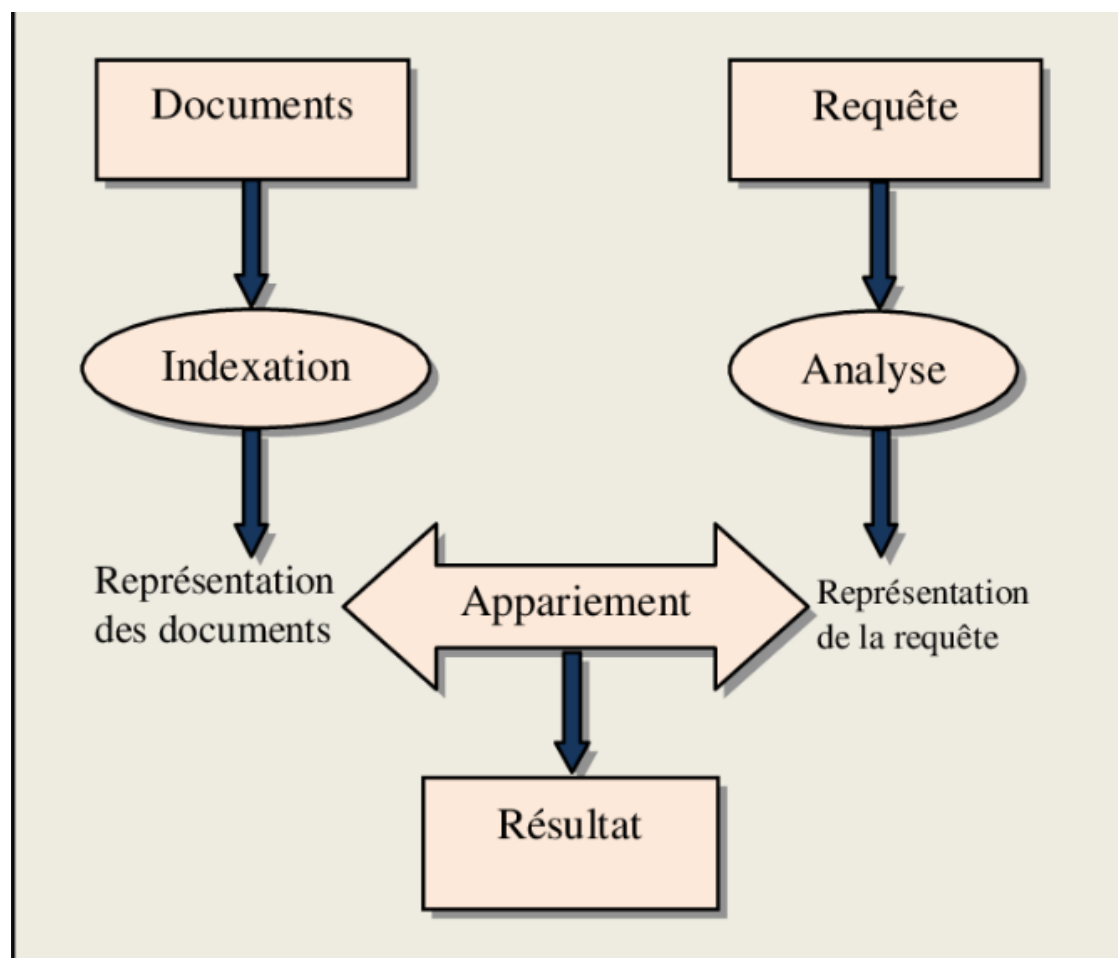
Rappel du pacte du bon déroulement du cours

- ▶ 1 – Etre à temps (5 minutes comme seuil de retard tolérable);
- ▶ 2 – Etre attentif durant le cours magistral, TP et TD (le support n'est qu'un ordre du jour);
- ▶ 3 – Corréler l'utile du cours au parcours de la formation et le dissocier un peu de la validation;
- ▶ 4 – Toute conduite nuisible au bon déroulement du cours est passible à des mesures disciplinaires prises en soft (soustraction de points, faire sortir de la classe, un rapport soumis à l'administration etc.);
- ▶ 5 – Faire partie d'un environnement interactif où la communication est bidirectionnelle et croire qu'en dehors de la classe le cours continue;
- ▶ 6 – Le rappel pour tenir le bon fil conducteur, rendant le contenu un ensemble cohérent, lié et continu.

En guise de préambule...

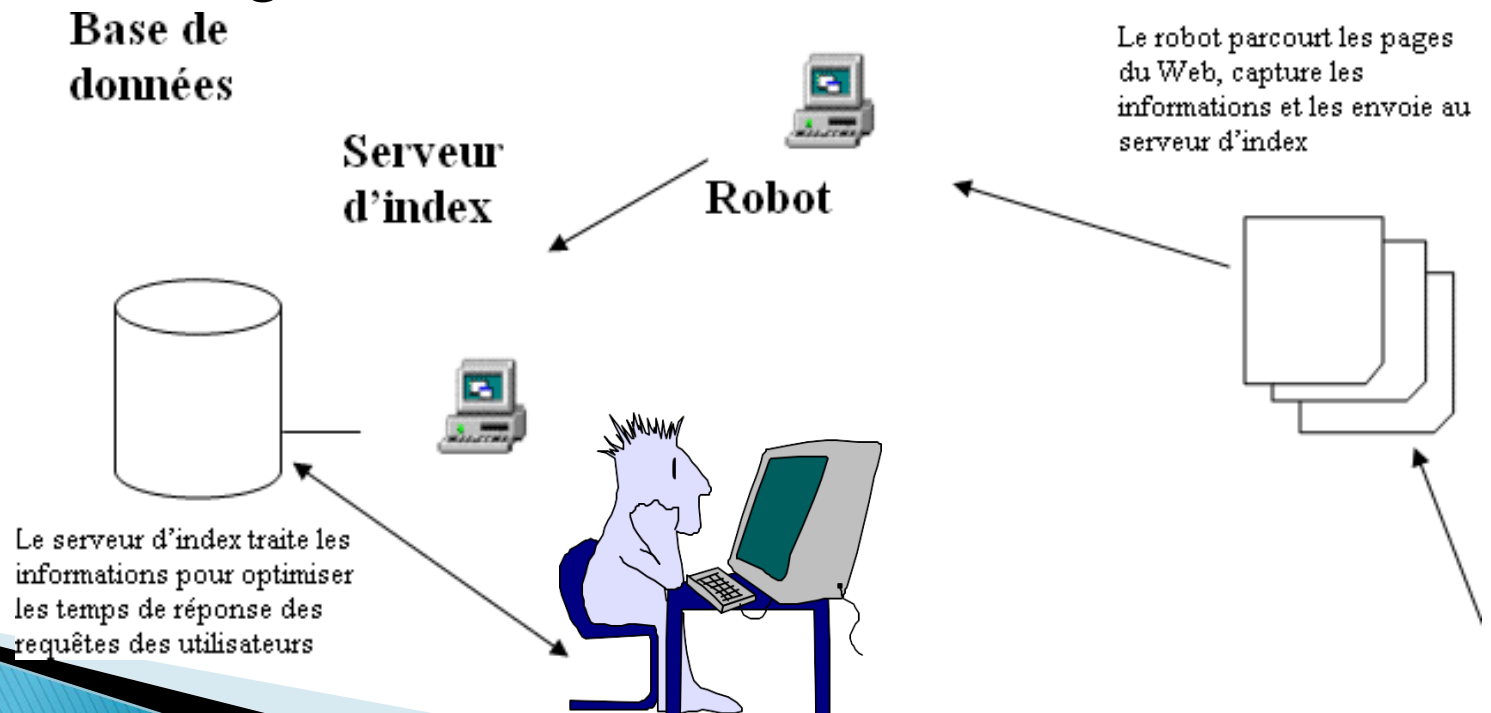
Recherche d'information et métadonnées

La Recherche d'Information Documentaire: Modèle U



Composants d'un moteur de recherche

- Robot d'indexation
- Index
- Interface d'interrogation



Le « logiciel » de recherche

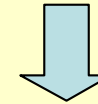
Interface d'interrogation

⇒ *Simple, mode avancé (expert) ou en mode spécialisé (types de fichiers)*

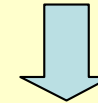
Visualisation des résultats

⇒ *Classement et affichage des résultats*

Préparation de la requête



Interrogation de la



base Affichage des

réponses

Recherche « filtrée »

Géographie (mondial, francophone)

Langue (français, anglais, espagnol...) Type (texte, images, son, vidéo)

Format (HTML, DOC, DOC, PPT, RTF...)

...

Les techniques de base

2 modes d'indexation de documents



Balises *META* pour un document HTML
description, keywords, author ...



Indexation en mode plein texte

Le contenu de la page est indexé intégralement

Images
? Sons
?
Vidéos ?
...

Mise à jour de l'index

- ① automatisée (robot logiciel : spider ou crawler) par repérage et visite
- ① manuelle (par l'auteur du site ou la personne chargée du référencement)
- ① par des professionnels (bases de données professionnelles)

Avant les métadonnées...



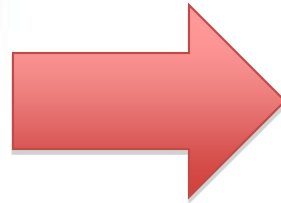
Source: Germano-fil. <http://germano-fil.hypotheses.org/147>



Source: Photographie annotée servant à la reconstitution d'une scène du crime (Tourves, 1878), Archives départementales du Var, <http://www.archives.var.fr/article.php?laref=2784&titre=photographie-annotee-servant-a-la-reconstitution-d-une-scene-du-crime-tourves-1878->

Qu'est-ce que les métadonnées?

Une image vaut mille mots



Nutrition Facts

Serving Size: 1/2 cup condensed

Amount Per Serving

Calories 90 Calories from Fat 54

% Daily Value*

Total Fat 6 g 9%

Saturated Fat 0.5 g 2%

Trans Fat 0 g

Cholesterol 5 mg 2%

Sodium 860 mg 36%

Potassium

Total Carbohydrate 9 g 3%

Dietary Fiber 3 g 12%

Sugars 1 g

Sugar Alcohols

Protein 1 g

Vitamin A 300 IU 6%

Vitamin C 0 mg 0%

Calcium 20 mg 2%

Iron 0 mg 0%

Définition

- Du grec "meta"
 - Ce qui dépasse, englobe un objet, une science.
(Le Robert)
 - Le terme "meta" vient du grec et dénote quelque chose de nature plus élevée ou plus fondamentale.
(Bibliothèque de l'Université Laval, Canada)
- Donnée qui renseigne sur la nature de certaines autres données et qui permet ainsi leur utilisation pertinente
- Anglais : metadata

Définition

- Dans sa plus simple expression, on réfère aux métadonnées comme étant:
 - « des données servant à décrire d'autres données » ou encore, « de l'information servant à décrire l'information ».
- Concrètement,
 - Les métadonnées consiste en de l'information qui décrit, explique, permet la localisation de ressources afin d'en faciliter le repérage, l'utilisation et la gestion.

À quoi servent-elles?

- Identifier et organiser les ressources;
- Repérer l'information:
 - Repérage de ressources en utilisant des critères de recherche;
 - Information sur la localisation des ressources.
- Assurer la gestion et préservation à long terme des documents.

À quoi servent-elles?

- Permettre la gestion des droits d'accès, d'utilisation, de copie;
- Certifier l'autorité intellectuelle du contenu;
- Contribuer à l'intégrité des documents numériques;
- Conserver le contexte de création de la ressource;
- Informer sur le mode de lecture;
- Etc.

Description

Les métadonnées servent à décrire:

- **Contenu:** ce que la ressource contient; ce qui est intrinsèque à la ressource.
- **Contexte:** informations extrinsèques à une ressource.
- **Structure:** organisation de la ressource.

Description des ressources

Pour décrire une ressource, on doit d'abord reconnaître ses propriétés (attributs ou éléments) et sélectionner celles qui nous semblent pertinentes en fonction des besoins.

- Un document numérique
 - Un titre
 - Un ou des auteur(s)
 - Un format
- Un objet dans un musée
 - Les dimensions
 - Les matériaux
 - L'âge

Description des ressources

La description des documents se fait par le biais des métadonnées portant sur **la forme (auteur, titre etc.) et le fond (mots clés, résumé, sujet etc.)**= toutes les opérations du traitement intellectuel et d'analyse de contenu.

- Pour les documents physiques (papiers ou électronique)

=> catalogage



- Pour les documents numériques et pages web =>

Balisage/marquage/taguage



Description des ressources

La description des documents se fait de trois façons distinctes et en amont de l'opération du RI ou IR:



- **Automatique** (spiders/crawlers=robots d'indexation ou logiciels dédiés);



- **Humaine manuelle établie par des professionnels** sur la base des normes, format et standards dédiés (soient des langages contrôlés) après analyse de forme et de fond des ressources cibles de la description et attribution des métadonnées et leurs valeurs selon la structure du norme/format/ standard;



- **Humaine manuelle établie par tout le monde** dans le cadre du web social/collaboratif (soit un langage libre), où le choix des métadonnées et leurs valeurs ne repose sur aucune norme, standard ou format.

Description



Nom Prénom
Adresse
Code Postal
Ville
Numéro de client
Numéro de téléphone

Nom du FAI
Adresse du service client

A (ville), le (date)

Objet : Demande de résiliation de mon contrat d'abonnement pour cause de fin d'engagement

Madame, Monsieur,

Je vous manifeste par la présente mon souhait de résilier mon abonnement actuel, souscrit auprès de votre société.

Je bénéficie en effet, depuis le (date), du forfait (dénomination du forfait).

La raison de cette demande est que j'arrive au terme de mon engagement qui s'achève le (date) et que je ne souhaite plus disposer de vos services. En application de l'article 1135 du Code Civil et des causes légales de révocation afférentes, je suis dans mon droit de vous réclamer une résiliation unilatérale. Par conséquent, je vous demande de mettre un terme au contrat d'abonnement qui nous lie.

Je tiens également à vous faire savoir que je viens de révoquer l'autorisation de prélèvement automatique, comme prévu par l'article 2004 du Code Civil. Cette procédure ne me soustrait pas à mes engagements et aux éventuelles sommes dues. Je décide juste par ce moyen de vous régler par le mode de paiement qui me convient.

Cette résiliation prendra effet au terme du délai de préavis défini dans les conditions générales applicables au contrat.

Je vous demande de m'envoyer une facture de clôture de compte ainsi qu'une confirmation écrite m'indiquant la date effective de résiliation.

Je vous demande en outre de bien vouloir supprimer de vos fichiers toutes les informations relatives à mes coordonnées personnelles, postales et bancaires. Conformément à la Loi Informatique et Libertés du 6 janvier 1978, je m'oppose à toute utilisation ou mise à disposition de mes coordonnées personnelles à des fins commerciales.

Dans l'attente d'une réponse de votre part, je vous prie d'agréer, Madame, Monsieur, l'expression de mes salutations distinguées.

Signature.



Métadonnées

- ▶ Typologie des métadonnées :
 - Métadonnées administratives
(comprenant notamment : métadonnées de gestion des droits et des accès, et métadonnées de préservation)
 - Métadonnées structurelles
(explicitation des relations entre les composants de la ressource, liens entre ces composants). Facilitent la navigation et la présentation des ressources électroniques
 - Métadonnées descriptives
(but : découverte et identification des ressources ; seront les seules traitées dans la présente intervention)
- ▶ Ces catégories n'ont pas de frontières bien définies et se recouvrent souvent partiellement



Types de métadonnées

- Métadonnées administratives en sous-catégories:
 - (1) Les **métadonnées sur la gestion des droits** qui permet la gestion de la propriété intellectuelle et des droits d'accès aux ressources;
 - (2) Les **métadonnées sur la préservation** qui contiennent certaines information pour assurer l'archivage et la préservation à long terme des ressources.

Une définition parmi tant d'autres

Métadonnées

- ▶ « Les métadonnées sont des informations structurées qui décrivent, expliquent, localisent ou encore facilitent la découverte, l'utilisation ou la gestion d'une ressource informationnelle. »
- ▶ NISO (National Information Standards Organization), *Understanding metadata*, 2004, ISBN 1-880124-62-9
<http://www.niso.org/standards/resources/UnderstandingMetadata.pdf>



Métadonnées : essai de définition



Un **document numérique / électronique**, c'est une suite de fichiers sans lien entre eux

+ identifiant unique (localisation / identification)



Métadonnées descriptives

- rattacher le document à l'original / différentes versions d'un document
- donner accès à la copie numérique / électronique



Métadonnées de structure

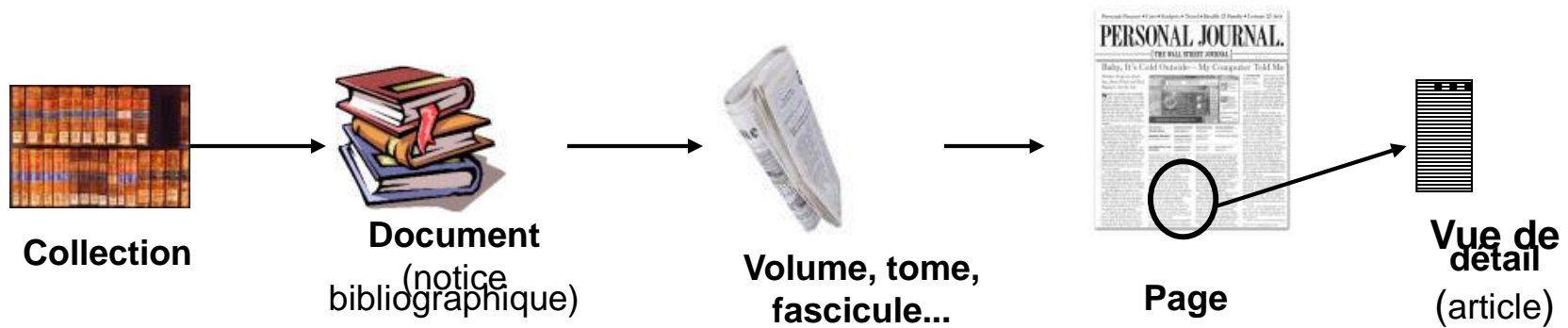
- rattacher les fichiers entre eux
- reconstituer la structure du document



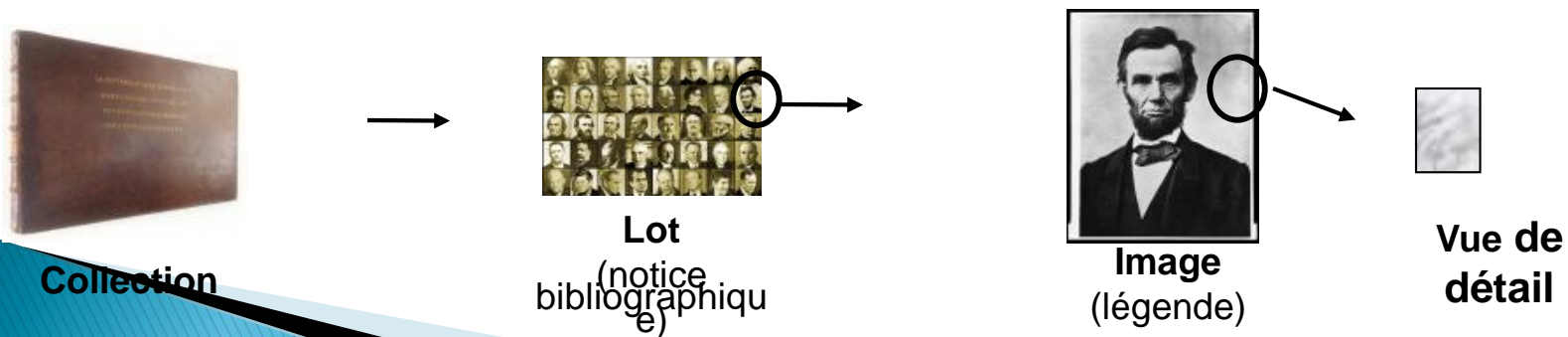
Métadonnées administratives

- gérer la collection, c'est-à-dire
- gérer les droits d'accès
- préserver les informations techniques nécessaires à la lecture des fichiers
- garantir l'intégrité des fichiers et le suivi de leurs éventuelles modifications

Le rôle des métadonnées de structure



- ▶ Un document complexe peut avoir plusieurs niveaux d'accès ou **niveaux de granularité**
- ▶ Document simple : pas de problème de structure en soi, mais peut tout de même avoir plusieurs **niveaux d'accès**



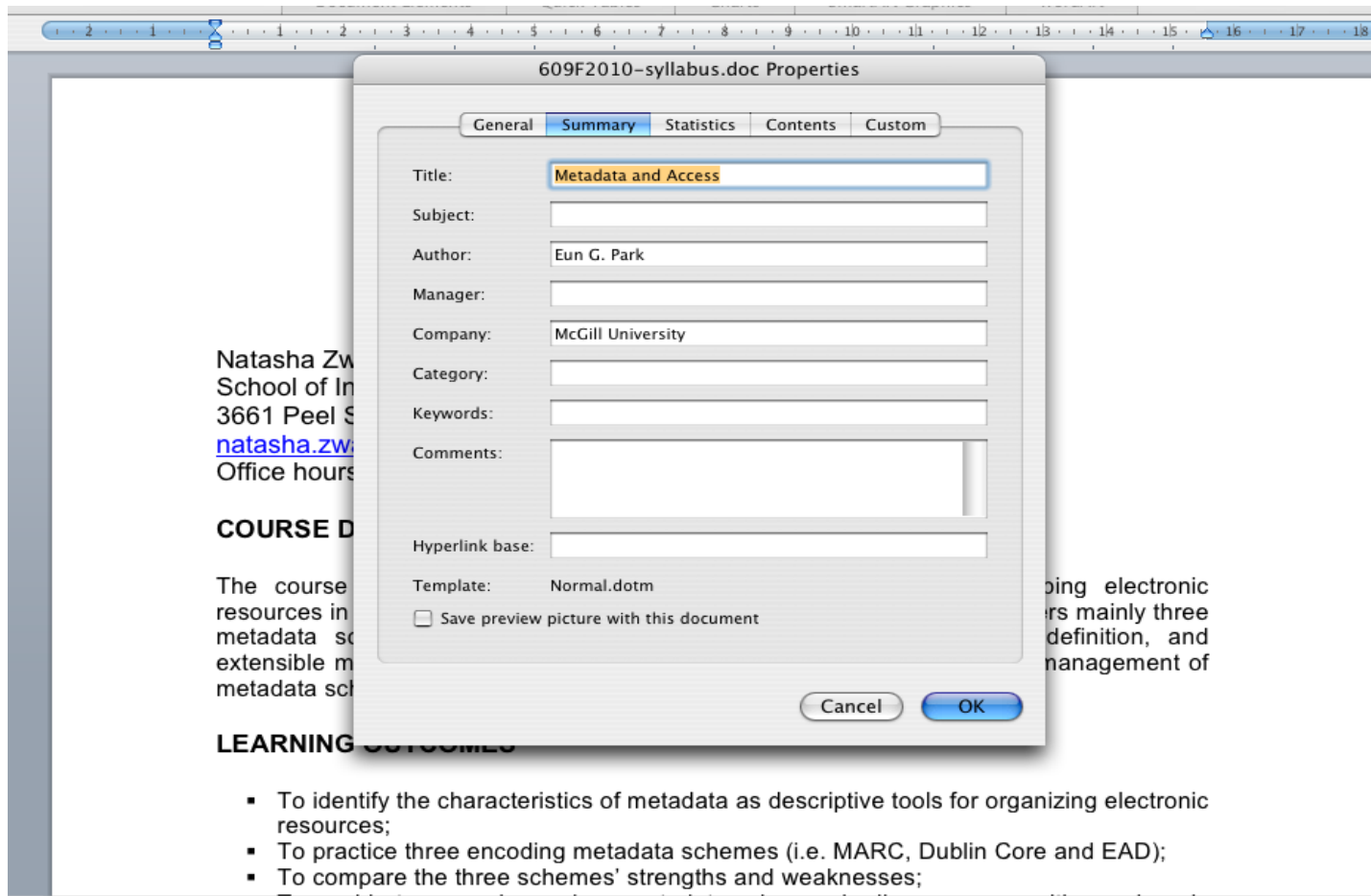
Où sont les métadonnées?

Métadonnées internes

- Les métadonnées peuvent être intégrées dans la ressource numérique elle-même.
- Le document, s'il est déplacé, transporte ses métadonnées avec lui.

► Métadonnées d'un fichier Word

Fichier → Propriétés



► Métadonnées d'une page Web

Search this site [Members' Login](#)

Association of Canadian Archivists

Advocacy Outreach Communications Professional Development Membership About Us

Providing professional leadership and representing the needs of the archival profession.

SUBMIT A PROPOSAL
ACA 2011 TORONTO

ARCHIVARIA ACA Foundation

Latest News

August 10, 2010
ACA Institute on Personal Archives, Oct. 29 & 30, 2010
City of Toronto Archives, Toronto, ON. Canada
Registration Form

ACA Institutes offer an intense learning experience featuring topics or curriculum at an advanced level. Institutes typically last two to three days and involve preparatory readings and case studies. [more...](#)

Welcome

This website is maintained by the Association of Canadian Archivists (ACA), for the benefit of any individuals interested in archives and the profession.

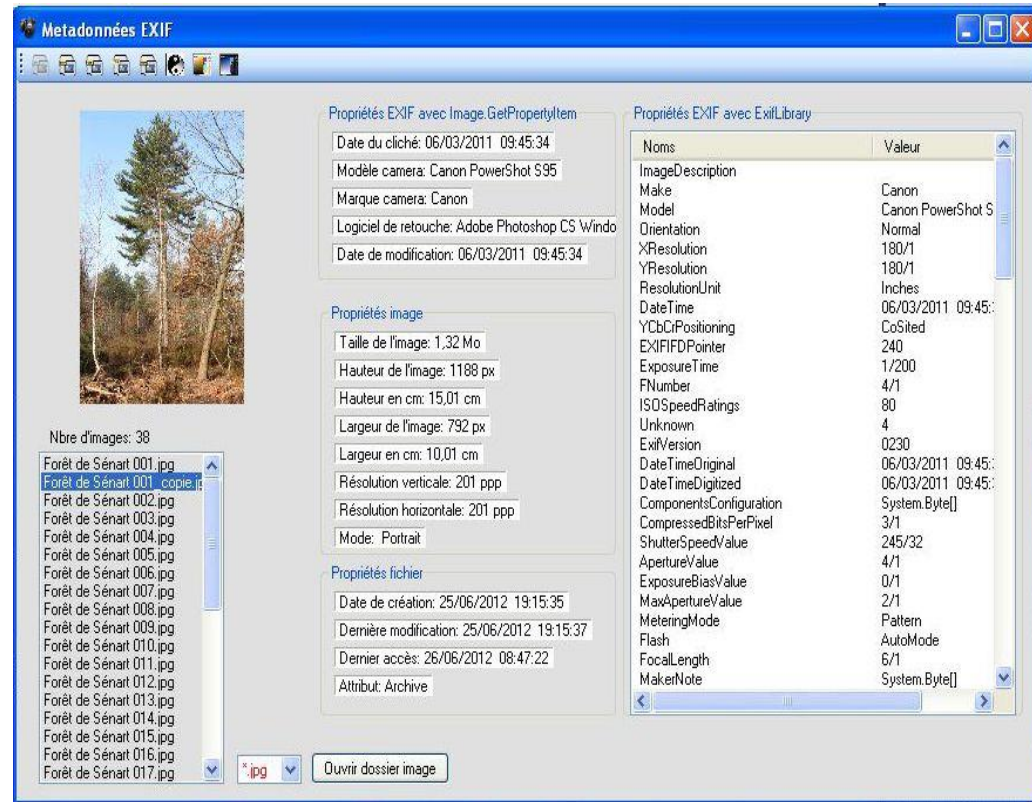
[Click to Contact Us](#)

Source of: <http://archivists.ca/>

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">
<html xmlns="http://www.w3.org/1999/xhtml" xml:lang="en" lang="en" dir="ltr">
<head>
<meta http-equiv="Content-Type" content="text/html; charset=utf-8" />
<title>Welcome | The Association of Canadian Archivists</title>
<meta http-equiv="Content-Type" content="text/html; charset=utf-8" />
<link rel="schema.dc" href="http://purl.org/dc/elements/1.1/" />
<link rel="canonical" href="http://archivists.ca/" />
<meta name="copyright" content="© The Association of Canadian Archivists" />
<meta name="dc.creator" content="The Association of Canadian Archivists" />
<meta name="dc.publisher" content="The Association of Canadian Archivists" />
<link rel="shortcut icon" href="/sites/default/themes/archivists/favicon.ico" type="image/x-icon" />
<link type="text/css" rel="stylesheet" media="all" href="/sites/default/modules/peaceworks/badge/stylesheet/peaceworl
<link type="text/css" rel="stylesheet" media="all" href="/sites/default/files/css/css_92e5b4f5e1006f54f72267b2bd1ac30a.d
<link type="text/css" rel="stylesheet" media="print" href="/sites/default/files/css/css_a34a298f230d6c4c84274b9c8ac6339
</if lt IE 8>
<link type="text/css" rel="stylesheet" media="all" href="/sites/default/themes/archivists/Styles/iefixes.css?W" />
</endif-->
<!--if lte IE 6>
<link type="text/css" rel="stylesheet" media="all" href="/sites/default/themes/archivists/Styles/ie6.css?W" />
</endif-->
<script type="text/javascript" src="/sites/default/files/js/js_f49e8dfb314a8589d91d0f2a5b7ebf21.js"></script>
<script type="text/javascript">
<!--//--><![CDATA[//><!--
jQuery.extend(Drupal.settings, { "basePath": "/" });
//--><![>
</script>
<!--if lt IE 8><script src="http://ie7-js.googlecode.com/svn/version/2.0(beta3)/IE8.js" type="text/javascript"></sc
</head>
<body class="front not-logged-in page-node node-type-page no-sidebars">
<div class="skiplinks">
<a class="skip" href="#navigation" title="Skip to the navigation">Skip to the navigation</a>
<a class="skip" href="#content" title="Skip to the content">Skip to the content</a>
<a class="skip" href="#search" title="Skip to the search">Skip to search</a>
</div>
<div id="pageContainer" class="clear-block">
<div id="pageHeader" class="clear-block">
<h1 id="mastheadLogo"><a href="/" title="Association of Canadian Archivists"><a href="https://www.members-archivists.ca/">
<div><div id="SearchBar">
<a class="skip" name="search" id="search"></a>
<div class="form-item" id="edit-search-theme-form-1-wrapper">
```

Définit les paramètres de prise de vue et les réglages de l'appareil au moment de la capture numérique. Ce sont des métadonnées de type interne.

- ✓ fabricant et mode de l'appareil
- ✓ hauteur et largeur de l'image
- ✓ date et heure de la prise de vue
- ✓ orientation
- ✓ résolution, temps d'exposition, ouverture
- ✓ présence d'un flash
- ✓ coordonnées GPS



Où sont les métadonnées?

Métadonnées externes

- À l'extérieur des ressources numériques
- Les métadonnées peuvent être contenues dans une notice séparée du document.

► Métadonnées dans une base de données

Record displayed at KentLINK OPAC

KentLINK CATALOG

START OVER REQUEST ITEM MARK FOR EXPORT MARC DISPLAY RETURN TO BROWSING ONIOLINK CENTRAL ANOTHER SEARCH

(Search History)

AUTHOR Caplan, Priscilla System Sorted Search

Record 3 of 4
Record: [Prev](#) [Next](#)

Author [Caplan, Priscilla](#)
Title **Metadata fundamentals for all librarians / Priscilla Caplan**
Publish info Chicago : American Library Association, 2003

LOCATION	CALL #	STATUS
MAIN see bldg guide	Z666.5 .C37 2003	MISSING

Description ix, 192 p. ; 28 cm
Note Includes bibliographical references and index
Contents Metadata basics -- Syntax, creation, and storage -- Vocabularies, classification, and identifiers -- Approaches to interoperability -- Metadata and the Web -- Library cataloging -- The TEI header -- The Dublin Core -- Archival description and the EAD -- Metadata for art and architecture -- GILS and government information -- Metadata for education -- ONIX International -- Metadata for geospatial and environmental resources -- The Data Documentation Initiative -- Administrative metadata -- Structural metadata -- Rights metadata
Subjects [Metadata](#)
[Information organization](#)
OCLC # 50906019
Isn/std # GBA3-Y7095
ISBN 0838908470
LCCN 2002151683

Record 3 of 4

MARC record display

START OVER REGULAR DISPLAY RETURN TO BROWSING ANOTHER SEARCH (Search History)

```
LEADER 00000cam 2200000 a 4500a
001 50906019
003 OCoLC
005 20030715092633.0
008 021023s2003 ilu b 001 0 eng
010 2002151683
015 GBA3-Y7095
020 0838908470
040 DLC|cDLC|dUKM|dC#P|dXFF|dKSU|dCoLC
049 KSUU
050 00 Z666.5|b.C37 2003
082 00 025.3|221
100 1 Caplan, Priscilla
245 10 Metadata fundamentals for all librarians /|cPriscilla
Caplan
260 Chicago :|bAmerican Library Association,|c2003
300 ix, 192 p. ;|c28 cm
504 Includes bibliographical references and index
505 00 |tMetadata basics --|tSyntax, creation, and storage --
|tVocabularies, classification, and identifiers --
|tApproaches to interoperability --|tMetadata and the Web
--|tLibrary cataloging --|gThe|tTEI header --|gThe|tDublin
Core --|tArchival description and the EAD --|tMetadata for
art and architecture --|tGILS and government information -
-|tMetadata for education --|tONIX International --
|tMetadata for geospatial and environmental resources --
|gThe|tData Documentation Initiative --|tAdministrative
metadata --|tStructural metadata --|tRights metadata
650 0 Metadata
650 0 Information organization
```



[enlarge](#)

Title: Wright 1902 glider being launched.

Description: Octave Chanute's photograph of the Wright 1902 glider being launched at Kill Devil Hills. One of the Wright Brothers is running holding a wing strut. The pilot is not identified but is one of the Wright Brothers. Note on reverse by Chanute "7. Good speed attained."

Place: [Outer Banks \(N.C.\)](#)

Date: October, 1902

Creator: [Octave Chanute](#)

Subject: [Wright, Wilbur, 1867-1912](#); [Wright, Orville, 1871-1948](#); [Wright 1902 glider](#); [Gliders](#); [Airplane, take off](#); [Flight](#); [Gliding and soaring](#); [Sand dunes](#); [Sand hills](#); [Kill Devil Hills \(N.C.\)](#)

Work Type: Photographic print, black and white

Copyright: Wright State University

Terms of Use: www.libraries.wright.edu/special

Contributor: Wright State University Libraries

Collection: Wright Brothers Collection

Repository Name: Special Collections and Archives, Wright State University Libraries

Repository Place: Dayton, Ohio

Repository Number: 16-1-13; N2991

To bookmark this page use: <http://worlddmcc.ohiolink.edu/History/Details?oid=4924>

Enjeux des métadonnées

Pour être pleinement efficaces, les métadonnées doivent absolument faire l'objet d'une normalisation.

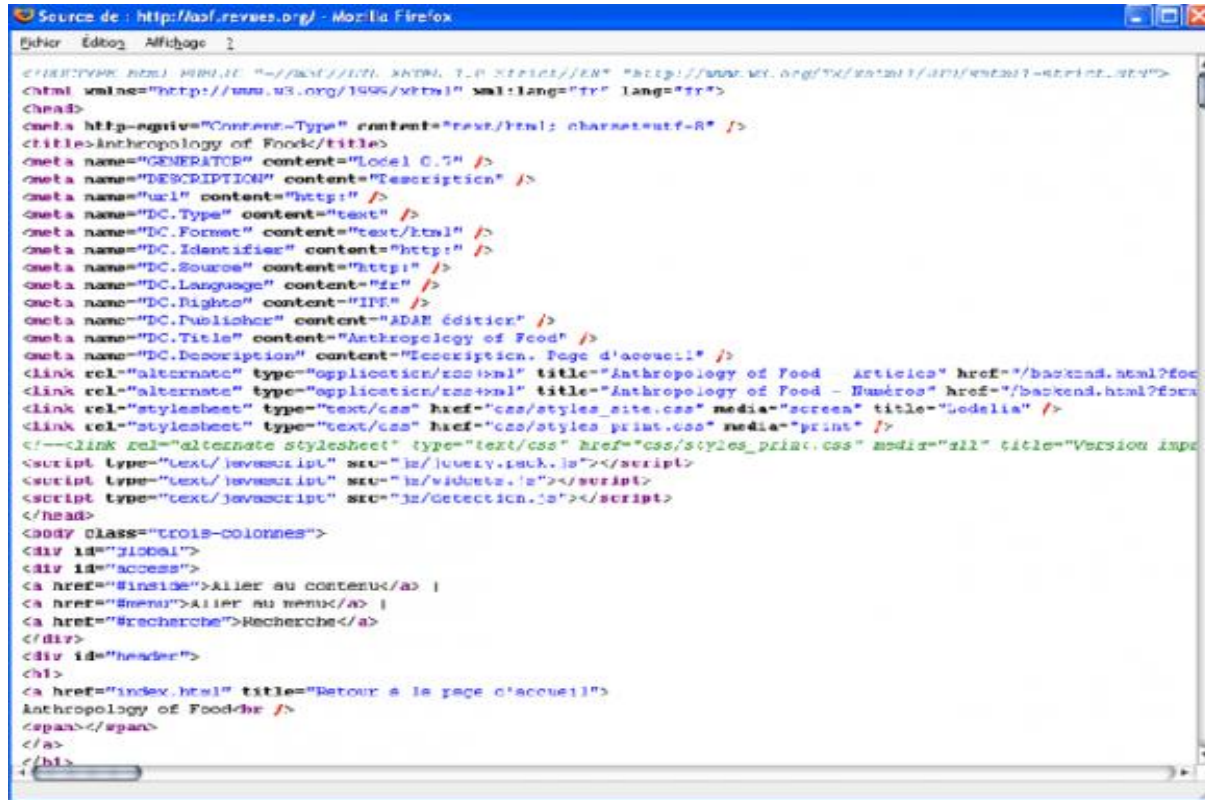
Considérations générales

Norme & Standard (Définition)

- ▶ **Normes** : caractère *officiel, légal* (« de jure »). Ensemble de règles de conformité, édictées et élaborés par des organismes officiels de normalisation (ISO, AFNOR...)
- ▶ **Standards** : caractère *consensuel* (« de facto »). Ensemble de recommandations émanant d'un groupe représentatif d'utilisateurs : *sociétés* ou *consortia* (XML : W3C,...) ; *groupements de professionnels* (Dublin Core : Dublin Core Metadata Initiative) ; *institutions* (MarcXML : Library of Congress)
- ▶ Parfois un standard s'officialise en norme (Dublin Core)



Les métadonnées et le Web



```
Source de : http://aaf.revues.org/ - Mozilla Firefox
Fichier  Éditeur  Affichage  ?

<?xml:namespace prefix="http://www.w3.org/1999/xhtml" xml:lang="fr" lang="fr">
<html>
<head>
<meta http-equiv="Content-Type" content="text/html; charset=utf-8" />
<title>Anthropology of Food</title>
<meta name="GENERATOR" content="Model 0.7" />
<meta name="DESCRIPTION" content="Description" />
<meta name="url" content="http:" />
<meta name="DC.Type" content="text" />
<meta name="DC.Format" content="text/html" />
<meta name="DC.Identifier" content="http:" />
<meta name="DC.Source" content="http:" />
<meta name="DC.Language" content="fr" />
<meta name="DC.Rights" content="IPE" />
<meta name="DC.Publisher" content="ADAM éditeur" />
<meta name="DC.Title" content="Anthropology of Food" />
<meta name="DC.Description" content="Description. Page d'accueil" />
<link rel="alternate" type="application/rss+xml" title="Anthropology of Food Articles" href="/backend.html?feed" />
<link rel="alternate" type="application/rss+xml" title="Anthropology of Food - Numéros" href="/backend.html?form" />
<link rel="stylesheet" type="text/css" href="css/styles_site.css" media="screen" title="Modelia" />
<link rel="stylesheet" type="text/css" href="css/styles_print.css" media="print" />
<!--<link rel="alternate stylesheet" type="text/css" href="css/styles_print.css" media="all" title="Version impr" />
<script type="text/javascript" src="js/jquery.js"></script>
<script type="text/javascript" src="js/vidueta.js"></script>
<script type="text/javascript" src="js/detection.js"></script>
</head>
<body class="trois-colonnes">
<div id="model">
<div id="access">
<a href="#inside">Allier au contenu</a> |
<a href="#menu">Allier au menu</a> |
<a href="#recherche">Recherche</a>
</div>
<div id="header">
<h1>
<a href="index.html" title="Retour à la page d'accueil">
Anthropology of Food</a>
<span></span>
</h1>
</div>
```

Le contenu des pages web est structuré à l'aide de balises en langage HTML. Les différentes balises employées permettent de structurer les informations selon différents thèmes (mots-clés, description, auteur, titre, sujet...).

Les métadonnées et le web collaboratif

Folksonomie

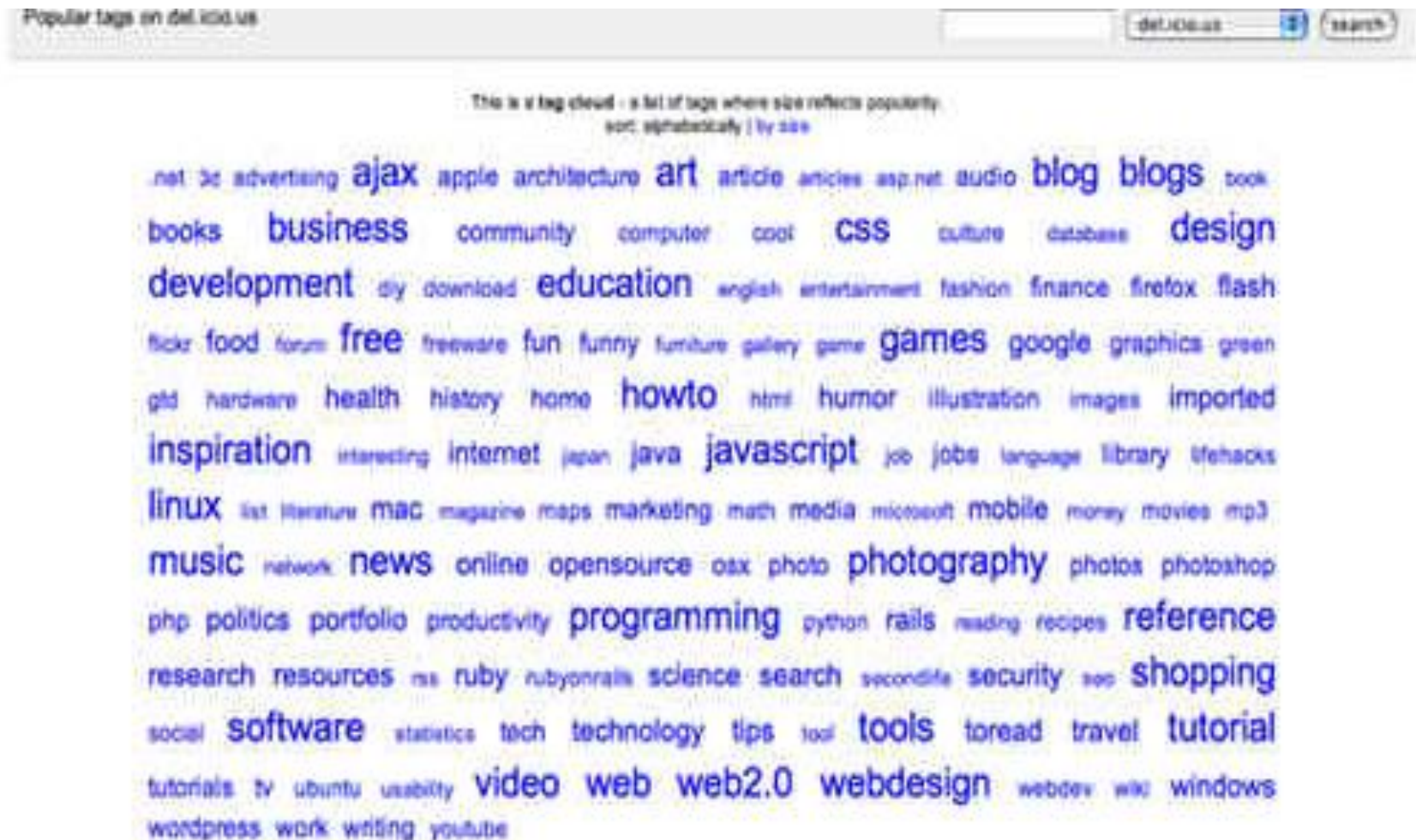
Système de classification collaborative et spontanée de contenus Internet, basé sur l'attribution de mots-clés librement choisis par des utilisateurs non spécialistes, qui favorise le partage de ressources et permet d'améliorer la recherche d'information.

Les métadonnées et le web collaboratif

Folksonomie

- Attribution d'étiquettes textuelles (tags) qui facilitent le partage et le repérage.
- Système de classification (taxonomie) composés de mots-clés libres attribués de manière spontanée et collaborative (folks).
- Indexation en vocabulaire libre du contenu.

Les métadonnées et le web collaboratif

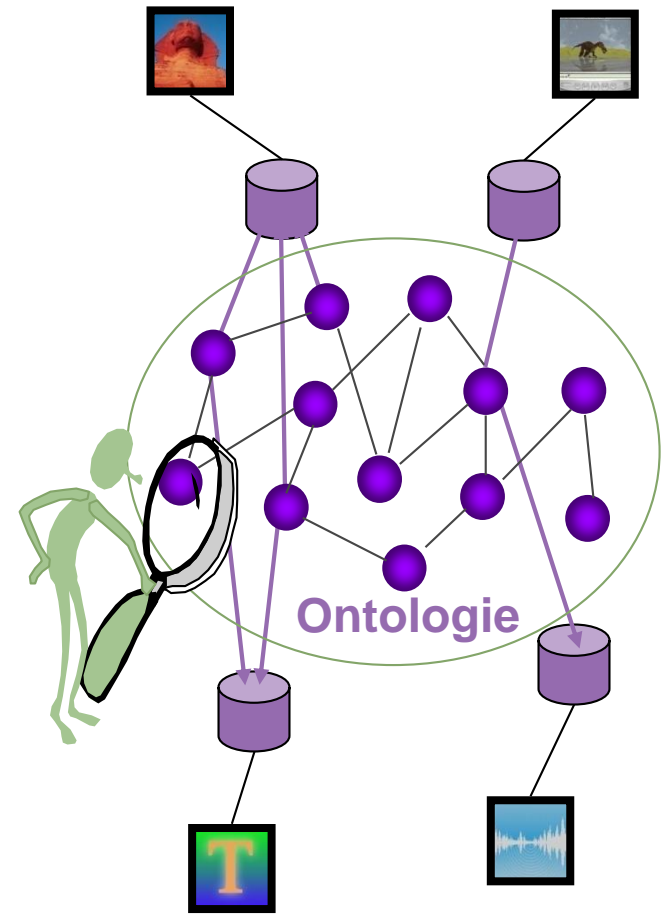


Les métadonnées et le web sémantique

- ▶ Les métadonnées trouvent leur plein potentiel dans le web sémantique.
- ▶ Avec l'organisation XML des informations, le web est en train de passer d'un ensemble de page à un ensemble de données d'une gigantesque base de données (data web) dans laquelle chaque mot important peut être mis en valeur par l'auteur de la page.
- ▶ Le fichier RDF (Resource Description Framework) décrit la structure des données de chaque page web XML et le langage SPARQL (Protocol and RDF Query Language) collecte l'information pour les moteurs de recherche ou les agents personnels intelligents qui peuvent explorer le web pour nous.

Les métadonnées et le web sémantique

- ▶ Faire face à la croissance exponentielle des informations. en associant aux ressources ...
- ▶ des **métadonnées** décrivant leurs propriétés sémantiques ou autres...
- ▶ liées aux **ontologies** décrivant la structure des connaissances d'un domaine,....
- ▶ permettant aux **agents informatiques** de traiter les ressources par leurs propriétés (**déclarées ou déduites**)



En conclusion

- ▶ Les métadonnées sont une composante essentielle du web.
- ▶ Les possibilités de plus en plus grandes qu'offrent le web, notamment grâce au collaboratif, permettent une plus grande diffusion des ressources numériques.

Merci pour votre attention