



National Research
Council Canada

Conseil national
de recherches Canada

CISTI

Canada Institute for Scientific
and Technical Information

Les métadonnées démystifiées

*Forum sur les métadonnées au
Canada 2005*

Exposé de :

Karen Morgenroth, coordonnatrice de la gestion des données

Le 27 septembre 2005

Canada 

Un aperçu

- Les métadonnées : que sont-elles exactement?
- Des exemples concrets
- Pourquoi utiliser les métadonnées?
- Le Web et les métadonnées
- L'interopérabilité
- Comment fait-on pour « créer » des métadonnées?
- Comment les métadonnées fonctionnent-elles?

Les métadonnées : définition

- Du grec (*meta*) et du latin (*data*)
- Données à propos des données
- Renseignements descriptifs à propos du format ou des attributs des ressources d'information

Les métadonnées dans le monde réel (1)

Les pages jaunes



comprennent les services, les produits, les entreprises, les adresses, les numéros de téléphone...

Les éléments normalisés des pages jaunes

Nom : formule d'usage du nom définie par l'entreprise

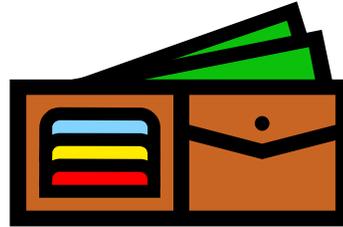
Adresse : format utilisé par l'éditeur des pages jaunes

Numéro de téléphone : format selon les normes internationales

Service ou produit : inclusion de renvois à partir de synonymes

Les métadonnées dans le monde réel (2)

Votre portefeuille



renferme permis de conduire, carte d'assurance maladie, carte de débit, cartes Visa, AMEX et MC, carte d'assurance sociale, télécarte...

Les éléments normalisés du portefeuille

Présentation : L'information relative au numéro de compte sur les cartes de crédit et les cartes de débit est au même endroit.

Nom : Le nom sur le permis de conduire et la carte d'assurance maladie doit être écrit au long.

Numéros d'identification : La présentation est normalisée.

Les métadonnées dans le monde réel (3)

Instruments de recherche d'archives
et catalogues
de bibliothèques



y compris les créateurs, les titres, les éditeurs, la
provenance, les dates, les sujets,
l'emplacement, les disponibilités...

Les éléments normalisés des catalogues et des outils de recherche

Noms : autorités

Descriptions : RCAA, RDDA (contenu normalisé)

Sujets : vocabulaires contrôlés tels que LCSH, LCCN, DDC, AAT (ensemble de valeurs normalisé)

Formats de communication : MARC (structure normalisée)

Ce que les métadonnées nous révèlent au sujet d'une ressource

- **Qui** a créé la ressource
- **Qui** gère la ressource
- **Quand** la ressource a été créée
- **Quand** la ressource a été publiée
- **Quel** est son contenu
- **Où** la ressource est archivée

Les avantages des métadonnées

La disponibilité de métadonnées descriptives et de qualité améliore notre capacité à trouver et à consulter l'information dont nous avons besoin.

Un balisage approprié améliore grandement leur repérage



The screenshot shows a Microsoft Internet Explorer browser window displaying the CNN.com website. The address bar shows the URL: CNN.com - 'Tagging' helps unclutter data - May 3, 2005 - Microsoft Internet Explorer. The browser's menu bar includes File, Edit, View, Favorites, Tools, and Help. The CNN.com logo is prominently displayed at the top left, with 'MEMBER SERVICES' and 'International' links to the right. A search bar is located below the logo, with radio buttons for 'The Web' and 'CNN.com'. A vertical navigation menu on the left lists various categories, with 'Technology' highlighted in red. The main content area features a large red heading 'TECHNOLOGY' followed by the article title 'Technology: 'Tagging' helps unclutter data'. The sub-headline reads 'Online search categorizes how humans label things'. The article is dated Tuesday, May 3, 2005, and posted at 11:23 AM EDT (1523 GMT). The text of the article begins with 'NEW YORK (AP) -- Here's how we tend to organize our digital photos: We stick them into a folder on our computer and label it "Hawaii trip," or whatever.' An image of a man, Noah Brier, is shown sitting at a desk and using a laptop. The caption below the image reads 'Noah Brier surfs the "flickr" site'. The browser's status bar at the bottom shows the page title and navigation arrows.

CNN.com - 'Tagging' helps unclutter data - May 3, 2005 - Microsoft Internet Explorer

File Edit View Favorites Tools Help

CNN.com MEMBER SERVICES International

SEARCH The Web CNN.com Search

Home Page
World
U.S.
Weather
Business at CNNMONEY
Sports at SI.com
Politics
Law
Technology
Science & Space
Health
Entertainment
Travel
Education
Special Reports

classmates-com
I GRADUATED 1994
IN: 1984 1974
Continue

SERVICES
Video
E-mail Newsletters
Your E-mail Alerts
RSS

TECHNOLOGY

'Tagging' helps unclutter data

Online search categorizes how humans label things

Tuesday, May 3, 2005 Posted: 11:23 AM EDT (1523 GMT)

NEW YORK (AP) -- Here's how we tend to organize our digital photos: We stick them into a folder on our computer and label it "Hawaii trip," or whatever.

Here's a new way: Forget folders or albums. Just "tag" the photos based on what's actually in each frame.

Now, extrapolate this concept to the ideas, images, videos -- and people -- you meet or wish to find online. If they're properly tagged, they're far easier to find.



AP PHOTO

Noah Brier surfs the "flickr" site

Canada Institute for Scientific and Technical Information

Le Web et les métadonnées

Le contenu Web :

- est sauvegardé et présenté sous plusieurs formes;
- est composé de texte, d'images, de musique, de photographies, de séquences vidéo;
- est affiché dans un grand nombre de formats normalisés (HTML, XML, PDF, WAV, JPEG, GIF, MPEG).

Normes sur les métadonnées pour ressources Web

Normes de contenu

- Dublin Core, IEE-LOM, CSDGM, etc.

Ensembles de valeurs

- LCSH, AAT, Thésaurus des sujets de base du Gouvernement du Canada, etc.

Syntaxes

- HTML, XML, RDF, EAD, etc.

Dublin Core

- Norme par excellence pour la description de ressources Web
- 15 éléments fondamentaux

Ensemble des éléments de métadonnées Dublin Core

Le titre

Le créateur

La date

La langue

Le sujet

La description

L'éditeur

Le collaborateur

Le type

Le format

L'identifiant

La source

La relation

La couverture

Les droits

<http://www.collectionscanada.ca/8/4/r4-280-f.html>

Government Information

Classification, Thesauri, Controlled Vocabularies (IMRC)

Thesauri and Controlled Vocabularies

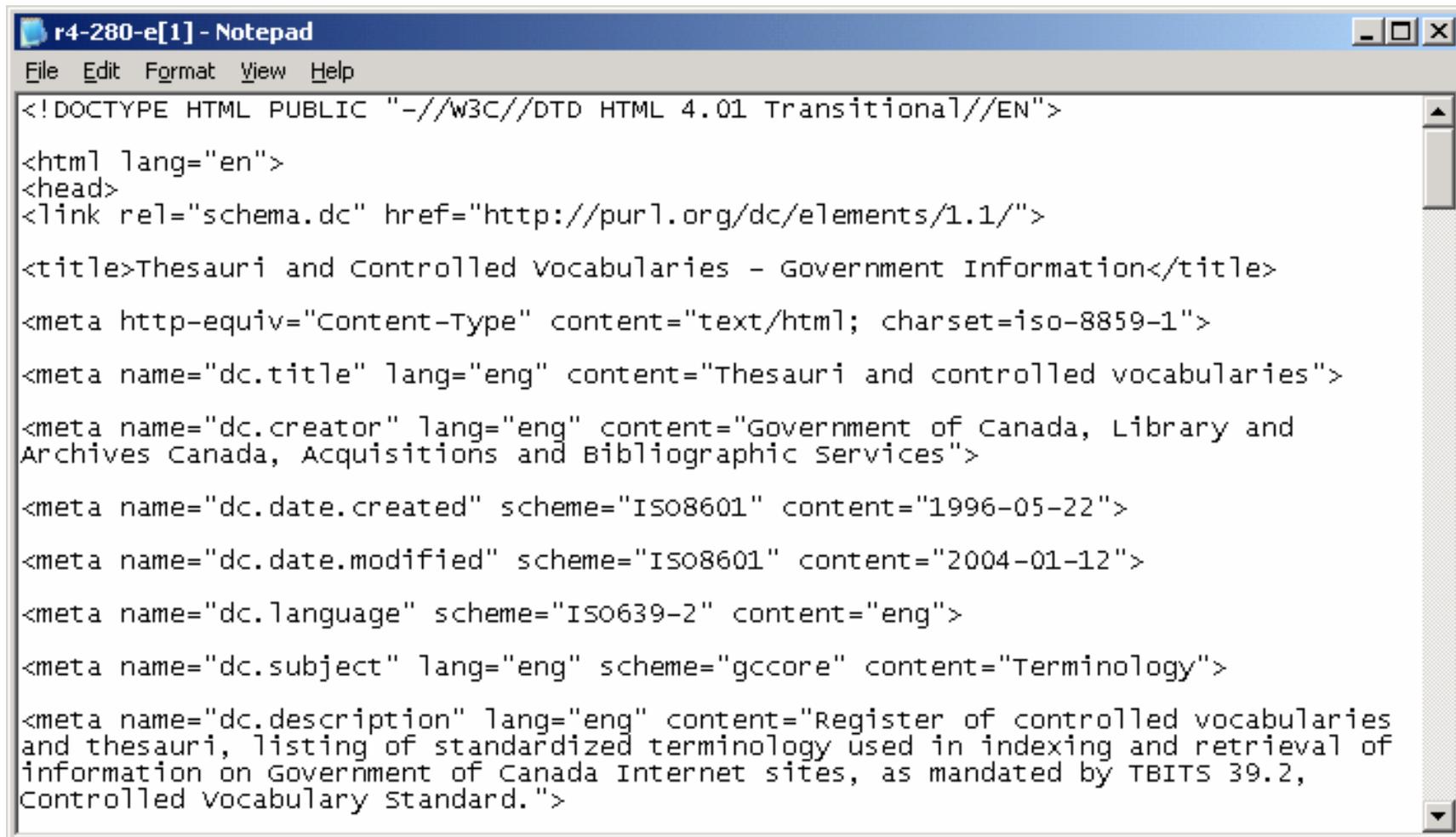
General

- [Bibliography](#).
- [Definitions](#).
- [Implementing Controlled Vocabulary on Government of Canada Web Sites - FAQ](#).
- [Registering a Standardized Vocabulary](#)
 - [Vocabulary Registration Form](#) - **NEW**

Terminology Used in the Federal Government

- Canadian Government-maintained Controlled Vocabularies and Thesauri.
 - [dc.audience Vocabularies](#)
 - [dc.coverage Vocabularies](#)
 - [dc.format Vocabularies](#)
 - [dc.subject Vocabularies](#)
 - [dc.type Vocabularies](#)
- [Externally-maintained Controlled Vocabularies and Thesauri](#).
General classifications

Les métadonnées DC en format HTML



```
r4-280-e[1] - Notepad
File Edit Format View Help
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN">
<html lang="en">
<head>
<link rel="schema.dc" href="http://purl.org/dc/elements/1.1/">
<title>Thesauri and Controlled vocabularies - Government Information</title>
<meta http-equiv="Content-Type" content="text/html; charset=iso-8859-1">
<meta name="dc.title" lang="eng" content="Thesauri and controlled vocabularies">
<meta name="dc.creator" lang="eng" content="Government of Canada, Library and
Archives Canada, Acquisitions and Bibliographic Services">
<meta name="dc.date.created" scheme="ISO8601" content="1996-05-22">
<meta name="dc.date.modified" scheme="ISO8601" content="2004-01-12">
<meta name="dc.language" scheme="ISO639-2" content="eng">
<meta name="dc.subject" lang="eng" scheme="gccore" content="Terminology">
<meta name="dc.description" lang="eng" content="Register of controlled vocabularies
and thesauri, listing of standardized terminology used in indexing and retrieval of
information on Government of Canada Internet sites, as mandated by TBITS 39.2,
Controlled Vocabulary Standard.">
```

Interopérabilité

Pour l'échange d'information et de ressources dans différents domaines

- Catalogues collectifs
- Dépôts institutionnels
- Gestions de dossiers

Tables de concordance

- Méthodes de rapprochement de métadonnées créées pour différentes raisons dans des environnements variés
- Identification d'éléments équivalents provenant de schémas différents

Table de concordance : DC vers MARC

Date

Une date associée à une étape du cycle de vie d'une ressource

MARC 21

Sans qualificatifs

- 260 ##\$c (Date de publication, de distribution, etc.)

Avec qualificatifs

- Disponibilité : 307 ##\$a (Heures, etc.)

[etc.]

Comment fait-on pour créer des métadonnées?

- Manuellement
- À l'aide d'outils de balisage

```
index-e[1] - Notepad
File Edit Search Help
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN">
<html lang="en">
<head>
<link rel="schema.dc" href="http://purl.org/dc/elements/1.1/">
<meta http-equiv="Content-Type" content="text/html; charset=iso-8859-1">

<title>Library and Archives Canada</title>

<meta name="dc.title" lang="eng" content="Library and Archives Canada">

<meta name="dc.creator" lang="eng" content="Government of Canada, Library
and Archives Canada, ">

<meta name="dc.date.created" scheme="IS08601" content="1995-12-20">

<meta name="dc.date.modified" scheme="IS08601" content="2005-05-06">

<meta name="dc.language" scheme="IS0639-2" content="eng">

<meta name="dc.subject" lang="eng" scheme="gccore" content="National
archives, National libraries">

<meta name="dc.description" lang="eng" content="Access online resources,
services and the collection of Library and Archives Canada: Canadian
documentary heritage (publications and records) providing knowledge about
Canadian society and serving as the collective memory of the Government of
Canada.">
```

Les métabaliseurs offerts sur le marché

- Font habituellement partie d'un système de gestion de contenu Web pour la création de sites centrés sur les métadonnées
- Exemples : Interwoven, OpenText et de nombreux autres fournisseurs

Edit generic

Help

aero_main_e

Save | Save As | Form Settings | Preview | Generate

Metadata

Document title * Aerodynamics Laboratory

Creator * Government of Canada, National Research Council Canada

Owner - 1 item

Owner - 1

Name * Gagnon, Michelle

Subject - 0 items

- Currently there are no entries -

Date created * 2005-03-15 Today's Date

Date modified * 2005-03-24 Today's Date

Description

Keywords

Type

abstract
administrative page
agreement

Choose types

Finish Cancel

Comment les métadonnées fonctionnent-elles?

Les moteurs de recherche procèdent à l'indexation des textes en version intégrale et des métadonnées.

Les moteurs de recherche dépouillent cet index en fonction des requêtes et fournissent des listes d'éléments se rapportant à la requête.

➔ Meilleurs résultats de recherche

Recherche

Les portails pour
Choisir...
Aller

Version imprimable

[Accueil](#) >

Recherche

Recherche rapide de tous les sites Web du CNRC

Précisez le ou les mots clés et cliquez sur « Chercher ». Pour chercher une expression exacte, utilisez des guillemets.

Chercher

Recherche avancée

Pour chercher les champs suivants, précisez le ou les mots clés et cliquez sur « Chercher ». Pour chercher une expression exacte, utilisez des guillemets.

Partout :

Titre :

Sujet :

Langue : Pages en français et en anglais

Résultats abrégés : (sans résumé)

Choisissez le(s) site(s) Web du CNRC que vous désirez chercher :

Tous les sites Web du CNRC
Site Web principal du CNRC



English	Contactez-nous	Aide	Recherche	Site du Canada
Accueil Presses scientifiques	Quoi de neuf	Accueil ICIST	Accueil CNRC	

Presses scientifiques du **CNRC**

Recherche des revues

Pour chercher les champs suivants, veuillez préciser le ou les mots clés et cliquer sur « Chercher ». [Aide pour la recherche](#).

- Revues ▶
- Revues affiliées ▶
- Monographies ▶
- Comptes rendus de conférences ▶
- Service à la clientèle ▶
- Services d'édition ▶

Partout :

Auteur :

Pour un nom intégral, utilisez des guillemets et commencez par le nom de famille ("Huot Jean P"). Utilisez la troncature si vous n'êtes pas certain du nom ou des initiales ("Huot Jean*" ou "Huot J*").

Titre :

Résumé :

Publiée :

après (exclusif) avant (exclusif)

Volume : **Pages :** (première ou dernière page)

Revue :

CTRL-clique pour sélections multiples

- Toutes les revues
- Biochimie et biologie cellulaire
- Dossier environnement
- Génome

Résultats abrégés (sans résumé)

Questions?

Merci!

Mille fois merci à Deane Zeeman!

karen.morgenroth@nrc-cnrc.gc.ca