



National Research
Council Canada

Conseil national
de recherches Canada

CISTI

Canada Institute for Scientific
and Technical Information

Metadata Demystified

Canadian Metadata Forum 2005

Presented by:

Karen Morgenroth, Data Management Coordinator

2005-09-27

Canada 

Overview

- Metadata--what is it, anyway?
- Real world examples
- Why use metadata?
- The web and metadata
- Interoperability
- How do you “do” metadata?
- How does metadata work?

Metadata defined

- Derived from Greek (*meta*) and Latin (*data*)
- Data about data
- Data that describes the content, format or attributes of information resources

Metadata in the real world (1)

The Yellow Pages



...include services, products, companies,
addresses, phone numbers ...

Yellow Pages standardized elements

Name: form of name defined by company usage

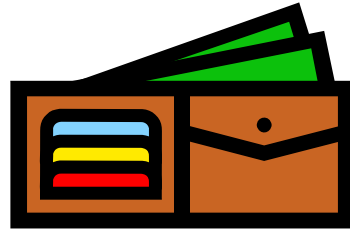
Address: format specified by Yellow Pages publisher

Phone number: format specified by international standards

Service/product: listing includes cross-referencing from synonyms

Metadata in the real world (2)

Your wallet



...contains driver's license, health card,
ATM card, Visa, AMEX, MC, SIN, phone
card...

Wallet standardized elements

Layout: credit cards and bank cards have account number information in the same location

Name: driver's license and health card require full form of name

ID numbers: presented in a standard format

Metadata in the real world (3)

Archival Finding Aids and Library Catalogues



...include creators, titles, publishers, provenance, dates, subjects, locations, availability...

Library catalogues/finding aids standardized elements

Names: name authorities

Descriptions: AACR, RAD (content standards)

Subjects: controlled vocabularies such as LCSH, LCCN, DDC, AAT (standardized value sets)

Communication formats: MARC (structural standard)

What metadata can tell us about a resource

- **Who** created it?
- **Who** maintains it?
- **When** was it created?
- **When** was it published?
- **What** is the content about?
- **Where** is it stored?

Benefits of metadata

The availability of good, descriptive metadata improves our ability to find and access the information we need.

If they're properly tagged, they're far easier to find.



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 'TECHNOLOGY' header, followed by the article title "'Tagging' helps unclutter data" and a sub-headline "Online search categorizes how humans label things". The article text discusses the concept of tagging digital photos and videos. An inset photo shows a man, Noah Brier, using a laptop. The browser's status bar at the bottom indicates the page is from CNN.com.

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

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

classmates.com
I GRADUATED 1994
IN: 1984 1974
Continue

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

Canada Institute for Scientific and Technical Information

The web and metadata

Web content

- is stored in many formats and presented in many ways
- consists of text, images, music, photographs, video
- appears in multiple standardized formats: HTML, XML, PDF, WAV, JPEG, GIF, MPEG

Web metadata standards

Content standards:

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

Value sets:

- LCSH, AAT, GoC Core Subject Thesaurus, etc.

Syntaxes

- HTML, XML, RDF, EAD, etc.

Dublin Core

- Pre-eminent standard for description of web resources
- 15 core elements

Dublin Core Metadata Element Set

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

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

Government Information

Classification, Thesauri, Controlled Vocabularies (IMRC)

Thesauri and Controlled Vocabularies

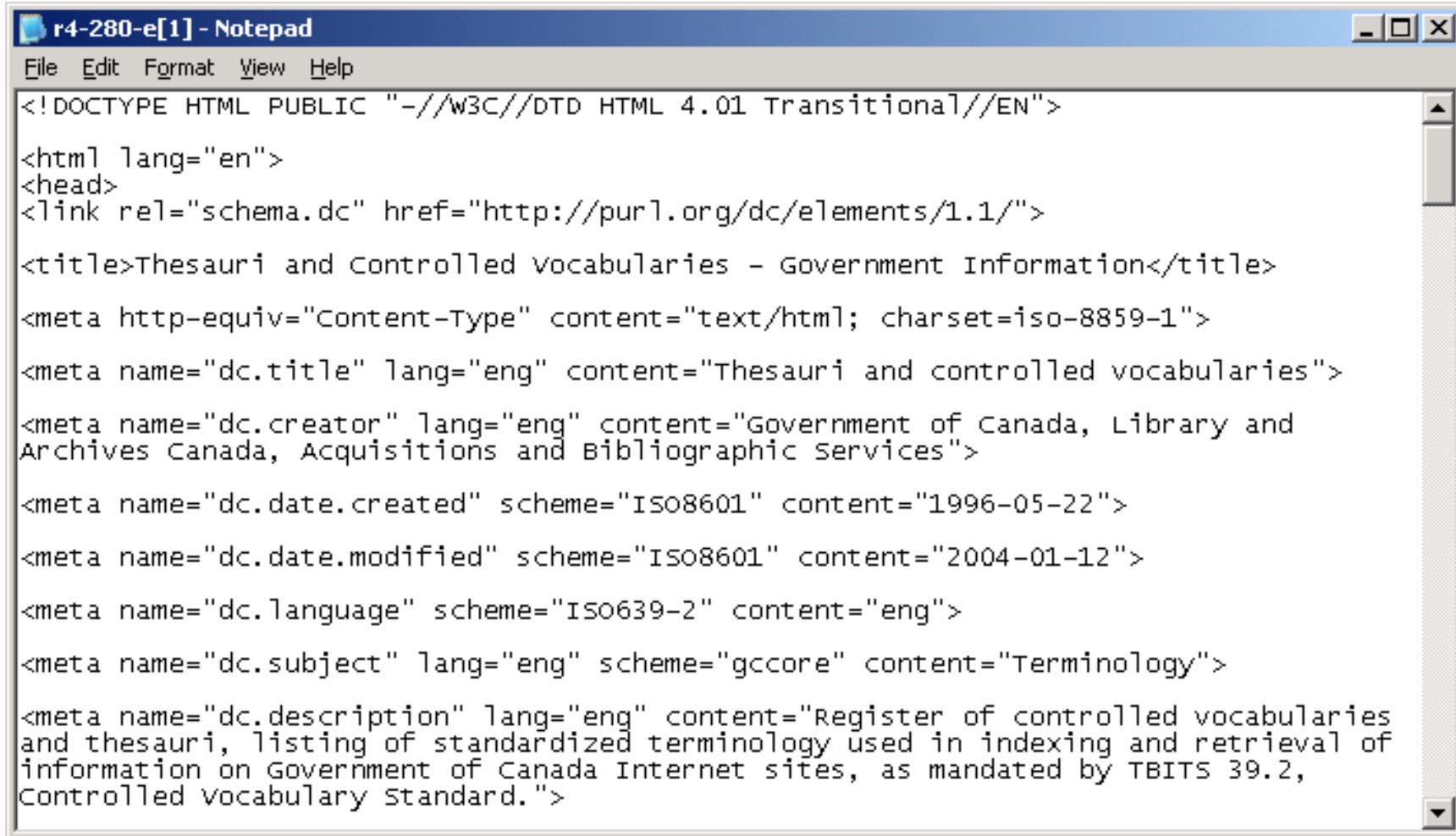
General

- [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

DC metadata in 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.">
```

Interoperability

For information exchange and
resource sharing across domains

- Union catalogues
- Institutional repositories
- Records management

Crosswalks

- Are a method of reconciling metadata created for different purposes in different environments
- Identify equivalent elements from different schema

DC to MARC crosswalk

Date

A date associated with an event in the life cycle of the resource.

MARC 21:

Unqualified:

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

Qualified:

- Available: 307 ##\$a (Hours, Etc.)

[etc.]

How do you “do” metadata?

- Manually
- With metatagging tools

```
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.">
```

Commercial metataggers

- Normally part of a web content management system producing metadata-driven sites
- Examples: Interwoven, OpenText, and many more vendors

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

How does metadata work?

Engines index full text and metadata.

Engines search that index in response to queries and return lists of items that relate in some way to the query

➔ More precise search results

Windows Explorer window showing the NRC-CIARC website search page.

NRC-CIARC

Search

Gateways for
Choose one... Go

Printable version

Home >

Search

Quick Search all NRC Web sites

Specify a keyword or set of keywords and click "Search". Place quotation marks around exact phrases.

Search

Advanced Search

To search the following fields, specify a keyword or set of keywords and click "Search". Place quotation marks around exact phrases.

Anywhere:

Title:

Subject:

Language:

Condensed results: (no summaries)

Select the NRC Web site(s) you would like to search:

Windows taskbar: Start, Met..., Inbo..., FW..., dem..., NR..., Mic..., Se..., Unti..., Mic..., 2:54 AM



Français	Contact Us	Help	Search	Canada Site
NRC Research Press Home	What's New	CISTI Home	NRC Home	

NRC Research Press

- [Journals](#) ▶
- [Affiliated journals](#) ▶
- [Monographs](#) ▶
- [Conference Proceedings](#) ▶
- [Customer service](#) ▶
- [Publishing services](#) ▶

Search Journals

To search the following fields, please specify a keyword or set of keywords and click 'Search'. [Search help](#).

Anywhere:

Author:

To search for a full name, use quotation marks and begin with the last name ("Smith John A"). Use a wildcard if unsure of full name or initials ("Smith John*" or "Smith J*").

Title:

Abstract:

Published:

After (exclusive) Before (exclusive)

Volume: **Pages:** (enter first or last page number)

Journal:

Use CTRL-click for multiple selections

- Biochemistry and Cell Biology
- Canadian Geotechnical Journal
- Canadian Journal of Botany

Condensed results (no summaries)

Search

Reset

Questions?

Thank you

Many thanks to Deane Zeeman!

karen.morgenroth@nrc-cnrc.gc.ca