

Basics of Dublin Core Metadata

Tutorial

Dublin Core – Building blocks for interoperability

17 December 2009

Thomas Baker



Basics of Dublin Core metadata

- **Dublin Core in historical context**
 - Core metadata on a new World-Wide Web
 - Towards a Web of Data
 - Structured Data and search engines
- **Interoperability options in a Web of Data**
 - Levels of Interoperability
 - Data Integration and Structured Search
 - Metadata Records and DCMI Abstract Model
 - Web-enabled controlled vocabularies
 - Linking legacy data to the Web of Data



1994: The Need

- **Web had arrived:** WWW conference, Chicago
- Clear: librarians would not “scale” to what was coming...
- “We need a **simple template** for describing Web pages”
- Librarians and computer people met in Dublin, Ohio...



1995: the “Dublin Core”

Elements

1. Identifier
2. Title
3. Creator
4. Contributor
5. Publisher
6. Subject
7. Description
8. Coverage
9. Format
10. Type
11. Date
12. Relation
13. Source
14. Rights
15. Language

Simple enough for non-experts to understand

A “library catalog card” for Web objects



NSDL Portal Prototype - Microsoft Internet Explorer

File Edit View Favorites Extras ?

Zurück Vorwärts Abbrechen Aktualisieren Startseite Suchen Favoriten Verlauf E-Mail Drucken

Adresse <http://nsdlib.nsd.comell.edu/nsdl/portal/index2.html?page=11> Wechseln zu Links >>

Site for science

[Register/Login](#) SEARCH:

[HOME](#) [SEARCH](#) [HELP](#) [ADD](#)

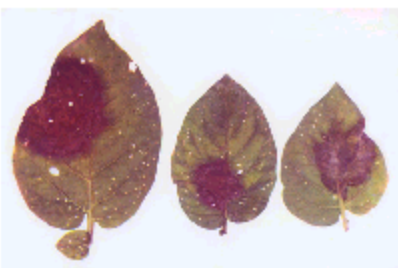
[SITE for Students](#) [SITE for Teachers](#) [My SITE](#) [News](#) [Tours](#) [Tools](#) [Science Pictures](#)

INDEX OF
Science Pictures

IMAGE COLLECTIONS

- ▶ PLANT PATHOLOGY HISTORICAL IMAGES
- ▶ HUBBLE TELESCOPE VIEWS
- ▶ BIRDS OF AUSTRALIA
- ▶ BIRD SLIDE OF THE WEEK
- ▶ COOL SCIENCE IMAGES
- ▶ EARTH SCIENCE IMAGE GALLERY
- ▶ SCIENCE IMAGE ONLINE
- ▶ SMITHSONIAN SCIENCE SERVICE
- ▶ COMET HALE-BOPP IMAGES
- ▶ SCIENCE AND IMAGE ARCHIVE

EXHIBIT
CU Plant Pathology Image Collection


 751 of 887

22-Jul-40
"Phytophthora infestans, late blight on Solanum tuberosum."
a. Kodachrome of a single blight spot b. Kodachrome of blighted plant c. Separation negatives of blight spot on a leaf

[More information](#)

ABOUT THE COLLECTION

"SITE for Science" is a core integration system prototype for a digital library. The National Science Foundation provides funding for this project which is part of the National Science, Mathematics, Engineering, and Technology Education Digital Library Program.

[ABOUT](#) [CONTACT](#) 

<http://siteforscience.nsd.comell.edu/nsdl/content/plant-pathology/29157.jpg> Internet



NSDL Portal Prototype - Microsoft Internet Explorer

Datei Bearbeiten Ansicht Favoriten Extras ?

Zurück Vorwärts Abbrechen Aktualisieren Startseite Suchen

Adresse http://nsdlib.nsdlib.cornell.edu/nsdl/portal/index2.html?page=11

Site for science

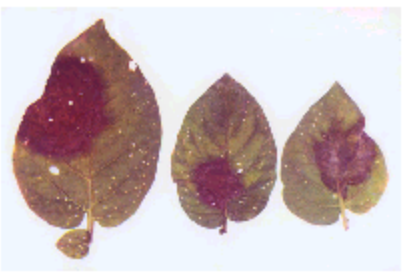
SITE for Students SITE for Teacher

INDEX OF
Science Pictures

IMAGE COLLECTIONS

- ▶ PLANT PATHOLOGY HISTORICAL IMAGES
- ▶ HUBBLE TELESCOPE VIEWS
- ▶ BIRDS OF AUSTRALIA
- ▶ BIRD SLIDE OF THE WEEK
- ▶ COOL SCIENCE IMAGES
- ▶ EARTH SCIENCE IMAGE GALLERY
- ▶ SCIENCE IMAGE ONLINE
- ▶ SMITHSONIAN SCIENCE SERVICE
- ▶ COMET HALE-BOPP IMAGES
- ▶ SCIENCE AND IMAGE ARCHIVE

EXHIBIT
CU Plant Pathology Image Collection



22-Jul-40
"Phytophthora infestans, late blight on Sol
 a. Kodachrome of a single blight spot b. Kc blight spot on a leaf

[More information](#)

ABOUT THE COLLECTION

Catalog View - Microsoft Internet ...

Site for science

Label	Information
Title	"Phytophthora infestans, late blight on Solanum tuberosum."
Subject	
Description	a. Kodachrome of a single blight spot b. Kodachrome of blighted plant c. Separation negatives of blight spot on a leaf
Creator	Photographer W. R. Fisher
Contributor	
Publisher	
Identifier	29157
Date	22-Jul-40
Type	Image
Language	en
Relation	
Source	Cornell University Plant Pathology Historical Photographs
Coverage	
Rights	Dept of Plant Pathology, Cornell University

Links >>

ictures

of

"SITE for Science" is a core integration system prototype for a digital library. The National Science Foundation provides funding for this project which is part of the National Science, Mathematics, Engineering, and Technology Education Digital Library Program.

[ABOUT](#) [CONTACT](#) 

http://siteforscience.nsdlib.cornell.edu/nsdl/content/plant-pathology/29157.jpg Internet



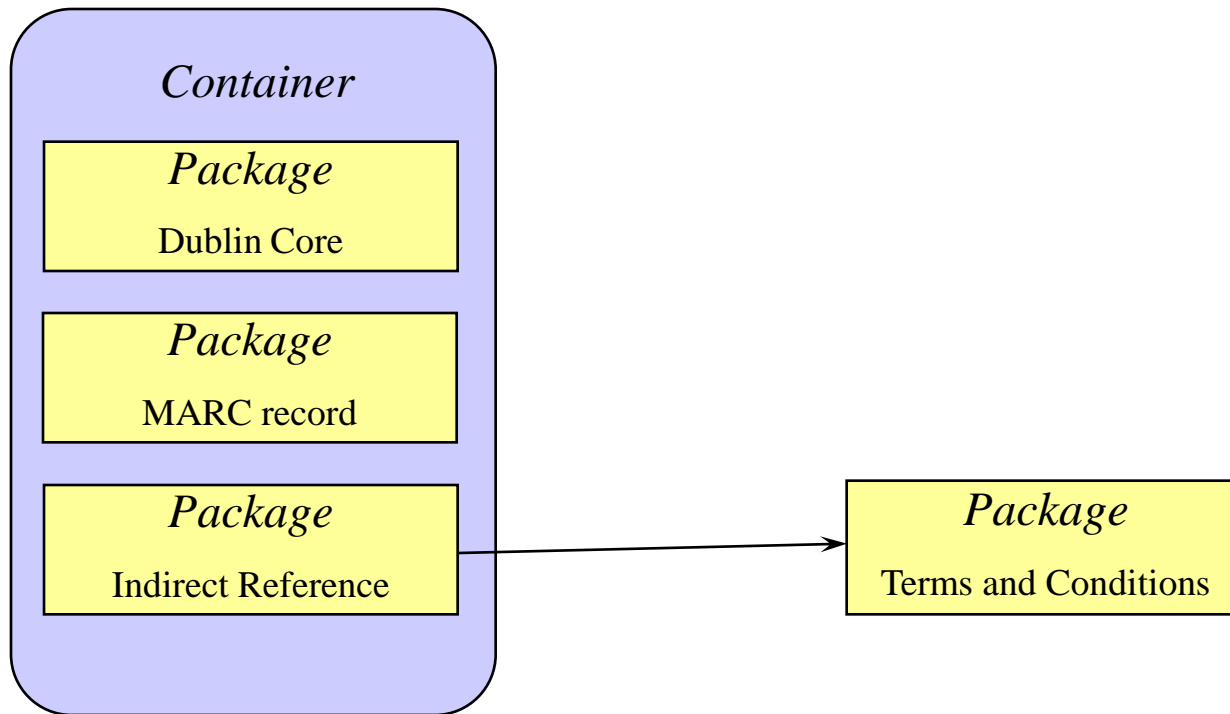
1996: Modular metadata

- Not: “One size fits all”.
- Different ways may be needed to describe one object:
 - MARC records for library catalogs
 - Dublin Core for simpler descriptions
 - Specialized metadata for terms and conditions of use
- Recognized need for a general framework for different types of metadata



1996: “Warwick Framework”

- Metadata “Packages” associated with a resource





1997: Qualification to add precision

- Not just any Date, but a *Date Created*
- Not just any Subject, but a *Library of Congress Subject Heading*
- Dumb-down: ignore extra details to see just a “core” description



1997: W3C starts work on Resource Description Framework (RDF)





1998: DC elements get URIs

- <http://purl.org/dc/elements/1.1/title:> identifiers rooted in Web infrastructure
- New assumptions
 - How can machines and software process this automatically?
- Work on the underlying data model...



2000: Idea of “Application Profiles”

- **Customized implementations**
 - Use “the Dublin Core” with **other vocabularies**
 - **Local rules** and guidelines
- Application profile provides documentation so that others can follow
- Not “take it or leave it”, but “**take what you want, create what you need**”



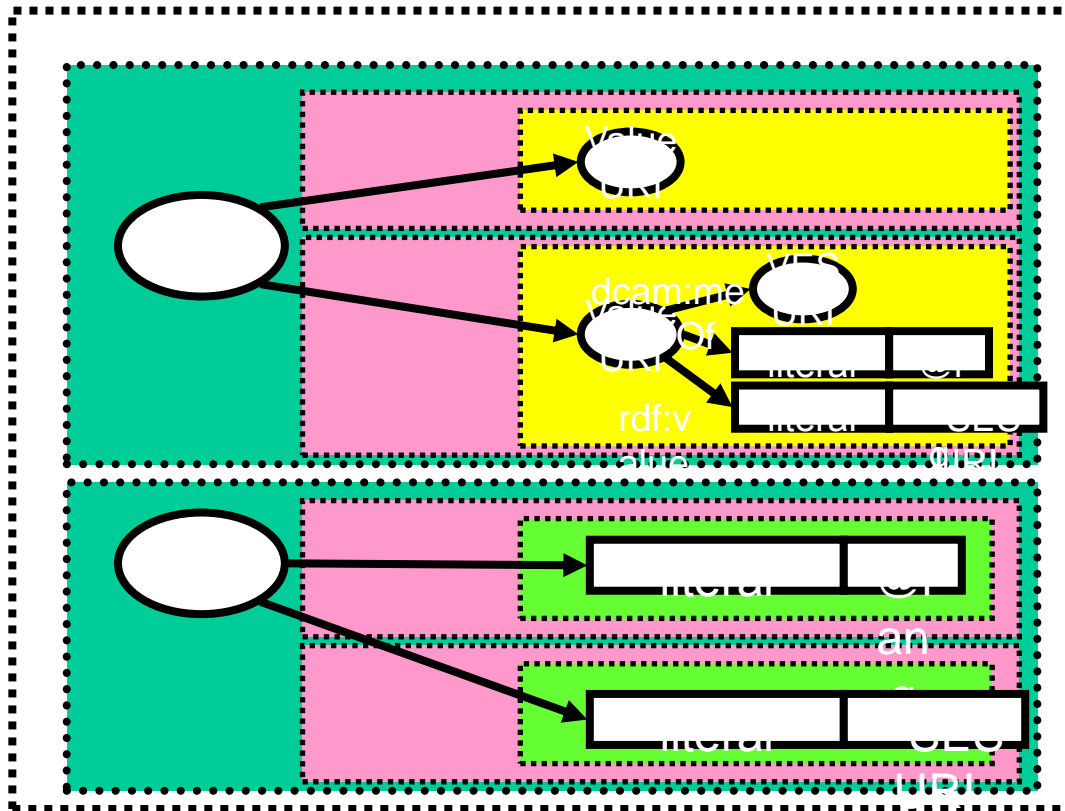
2000: Growing the vocabulary

Elements	Refinements		Encodings	Types
1. Identifier	Abstract	Is referenced by	Box	Collection
2. Title	Access rights	Is replaced by	DCMIType	Dataset
3. Creator	Alternative	Is required by	DDC	Event
4. Contributor	Audience	Issued	IMT	Image
5. Publisher	Available	Is version of	ISO3166	Interactive
6. Subject	Bibliographic citation	License	ISO639-2	Resource
7. Description	Conforms to	Mediator	LCC	Moving Image
8. Coverage	Created	Medium	LCSH	Physical Object
9. Format	Date accepted	Modified	MESH	Service
10. Type	Date copyrighted	Provenance	Period	Software
11. Date	Date submitted	References	Point	Sound
12. Relation	Education level	Replaces	RFC1766	Still Image
13. Source	Extent	Requires	RFC3066	Text
14. Rights	Has format	Rights holder	TGN	
15. Language	Has part	Spatial	UDC	
	Has version	Table of contents	URI	
	Is format of	Temporal	W3CTDF	
	Is part of	Valid		



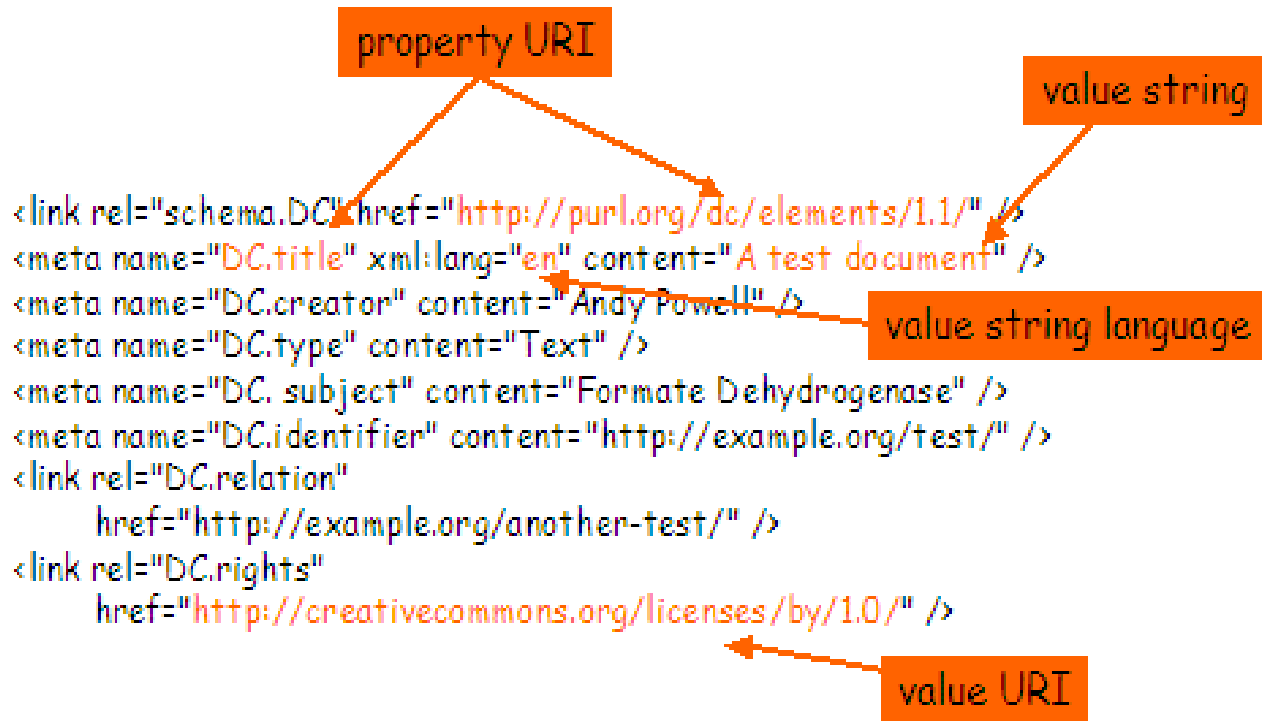
2003-2007: DCMI Abstract Model

- Moving the informal model of Dublin Core early years onto a solid RDF foundation





Motivation: One model – many syntax options





DCMI Model today

- **Core semantics**
 - The Dublin Core and related vocabularies
- **Data model**
 - RDF (and related DCMI Abstract Model)
- **Application Profiles**
 - Use mixed vocabularies
 - Based on underlying data model
 - Customized for specific purposes

Today: Dublin Core and Structured Search





Levels of Interoperability



Interoperability Levels for Dublin Core metadata

Shared validatable
constraints

➤ 4: Description Set

- Shared formal-semantic model

Shared model for “records”

➤ 3: Description

- Shared formal-semantic model

Shared formal-semantic model

➤ 2: Shared

- Shared

Shared (natural-language) definitions



Open- and closed-world

Shared constraints

Open-world data optimized for specific environments.

Shared model for “records”

Open-world data captured in manageable records.

Shared formal-semantic model

“Open-world” data.

Shared (natural-language) definitions

Data in silos. “Intra-operability” within silos.



Supporting technologies

Shared constraints

**DCMI Description Set Profile.
SPARQL Query Patterns.**

Shared model for “records”

**DCMI Abstract Model. DC-DS-XML.
SPARQL Named Graphs.**

Shared formal-semantic model

**Linked data. RDF data. Extracted triples.
DC-RDF. DC-HTML. RDFa!**

Shared (natural-language) definitions

**Closed systems. Proprietary systems. Web of APIs.
DC-XML/2003 and other early DCMI specs.**



Deployed base

Shared constraints

Shared “records”

Shared formal-semantic model

Shared (natural-language) definitions



Rate of growth

Shared constraints

Shared "records"

Shared formal-semantic model

Shared (natural-language) definitions



Which level do you require?

Shared constraints

Pro: Validation. Quality.

Contra: It is “constraining”...

Shared model for “records”

Pro: Provenance. Trust.

Contra: Lack of mature, deployed models.

Shared formal-semantic model

Pro: Easier to integrate and migrate data.

Contra: Harder to design, less tools.

Shared (natural-language) definitions

Pro: Easier to deploy. Validatable records.

Contra: Closed-world. Interoperability by (thousands of) ad-hoc agreements.

Level-1 applications s interoperate with shared or mapped schemas

