

Metadata for the Web Beyond Dublin Core?

CS 431 - March 9, 2005

Carl Lagoze - Cornell University

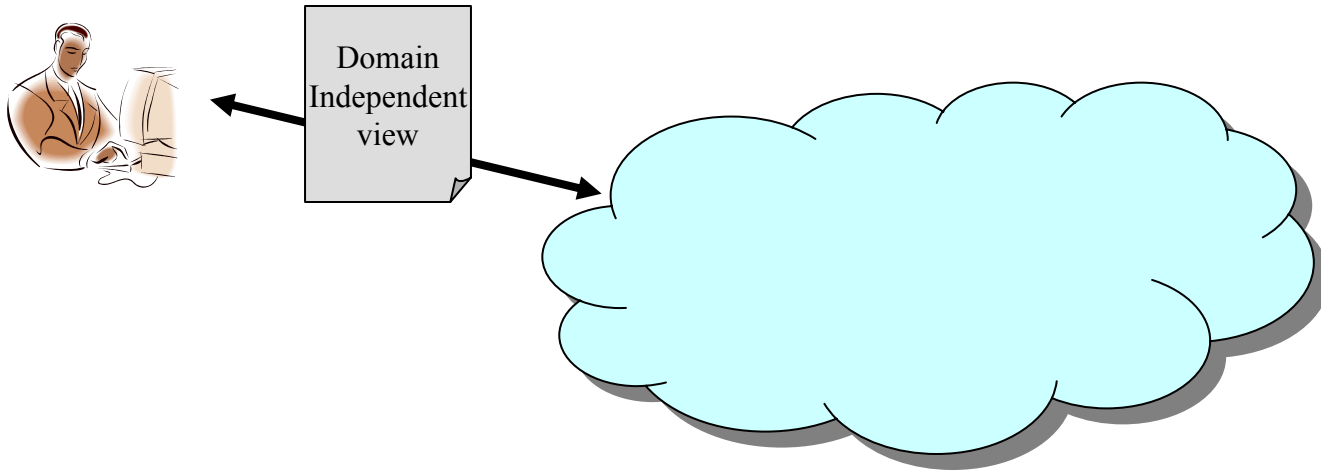
Acknowledgements to Liz Liddy and Geri Gay

Components of the Dublin Core Standard

- Core 15 elements
 - <http://www.dublincore.org/documents/dces/>
- Element refinements/qualifiers
 - "is-a" relationships
 - <http://www.dublincore.org/documents/dcmi-terms/>
- Type vocabulary
 - Genre for type element
 - <http://dublincore.org/documents/dcmi-type-vocabulary/>
- URIs for terms above
 - E.g., <http://purl.org/dc/elements/1.1/contributor>
- Encoding guidelines
 - XHTML
 - XML/RDF

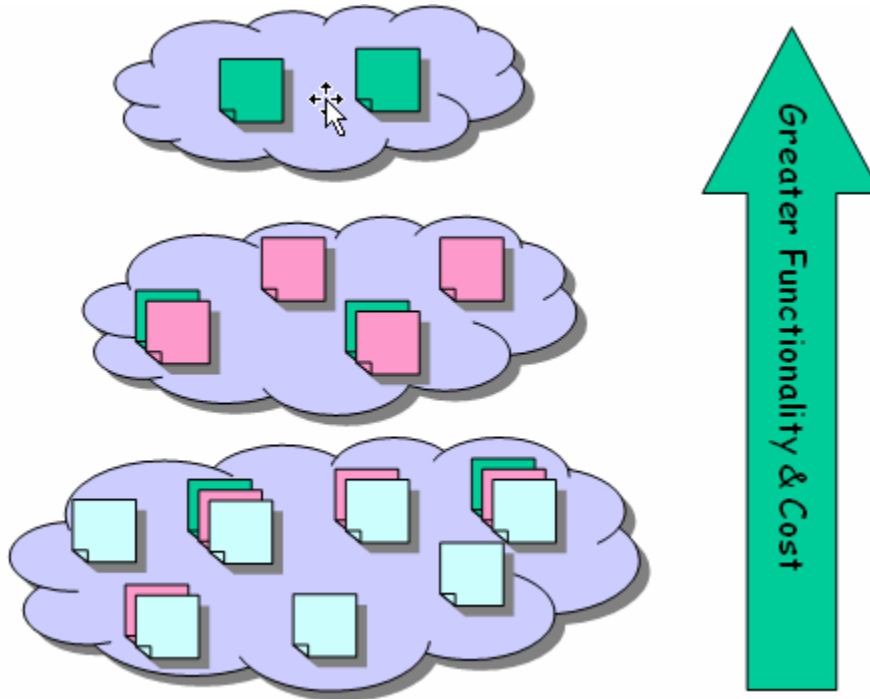
What is the Dublin Core (1)

- A simple set of properties to support resource discovery on the web (fuzzy search buckets)?
- Questions
 - Necessary
 - Possible (spam, expertise, uncontrolled vocabulary)



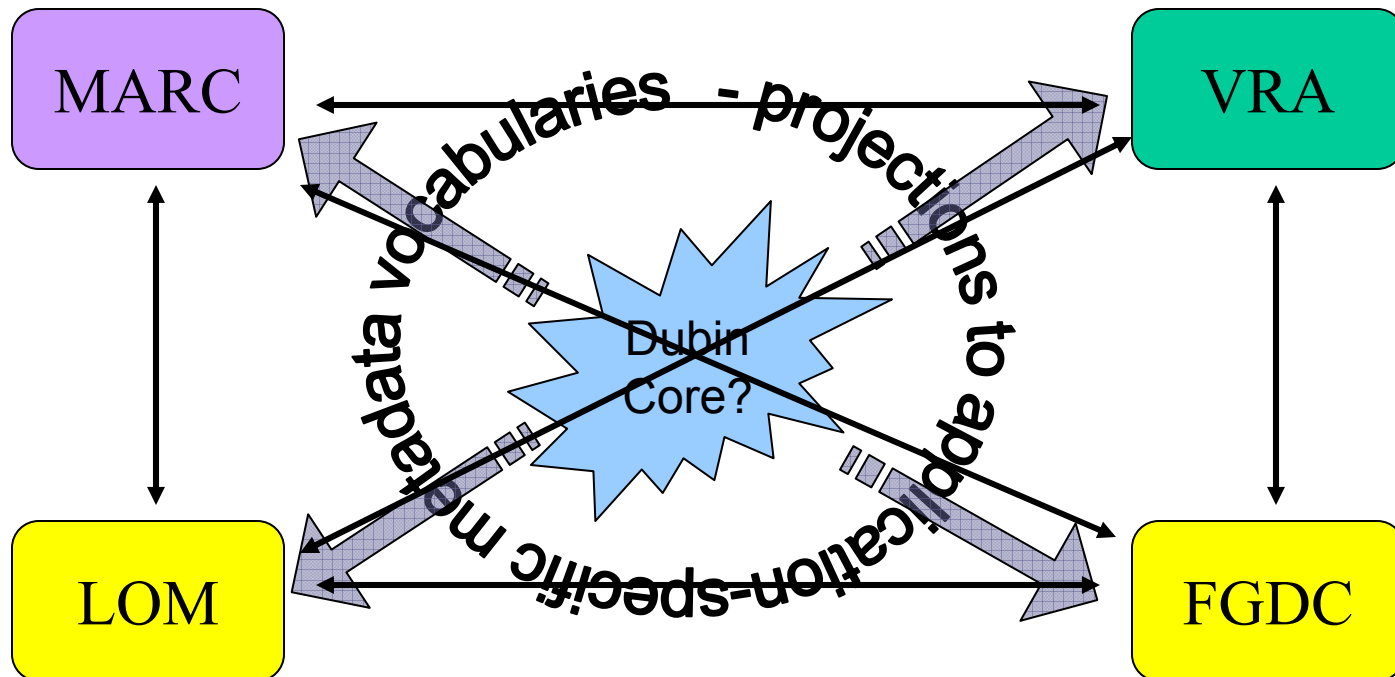
What is Dublin Core (2)?

- An extensible ontology for resource description?
- Questions:
 - Are these the right primitive classes?
 - Is the attribute/value data model rich enough?



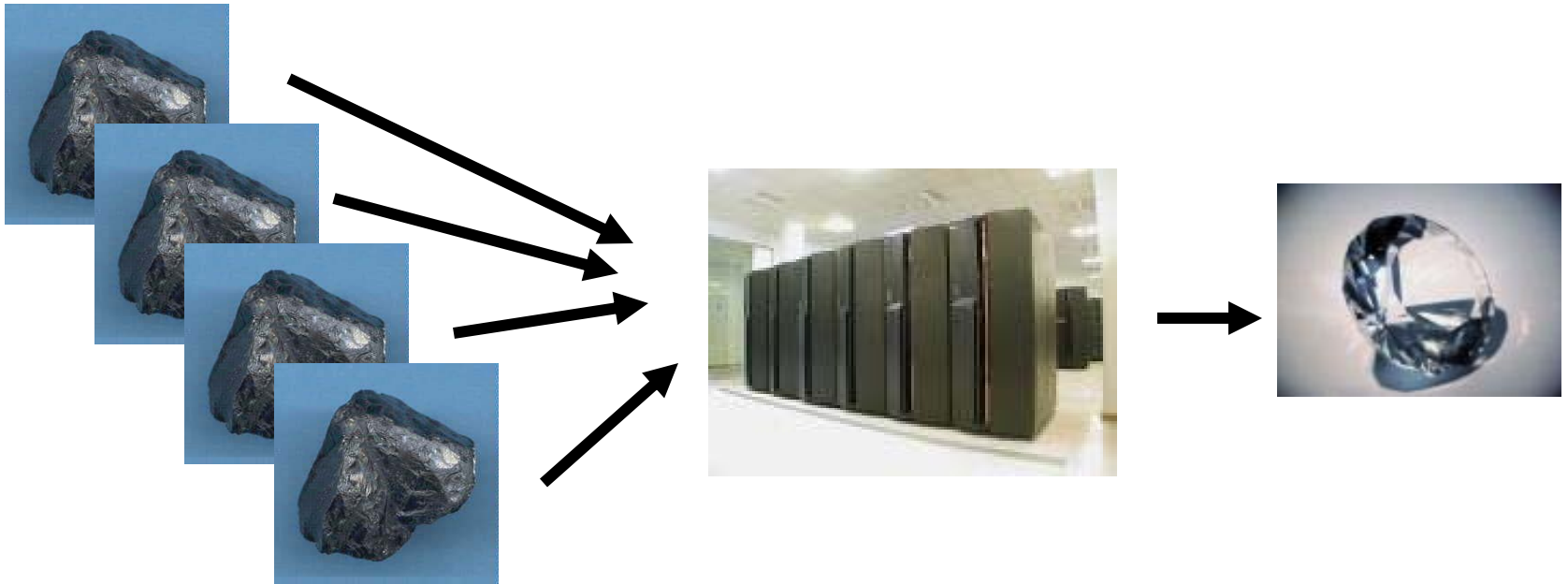
What is the Dublin Core (3)?

- A cross-domain switchboard for combining heterogeneous formats?
- Same modeling and class problems

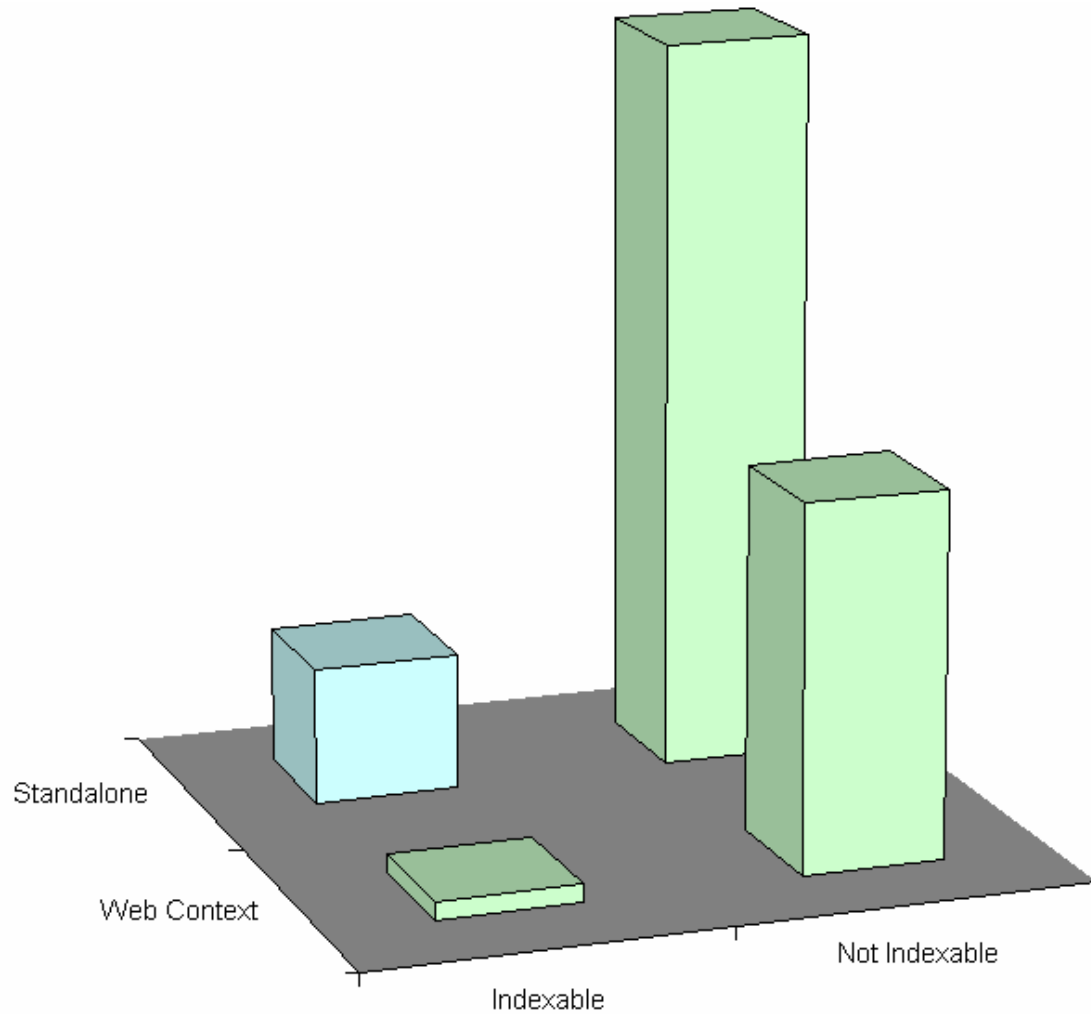


What is the Dublin Core (4)?

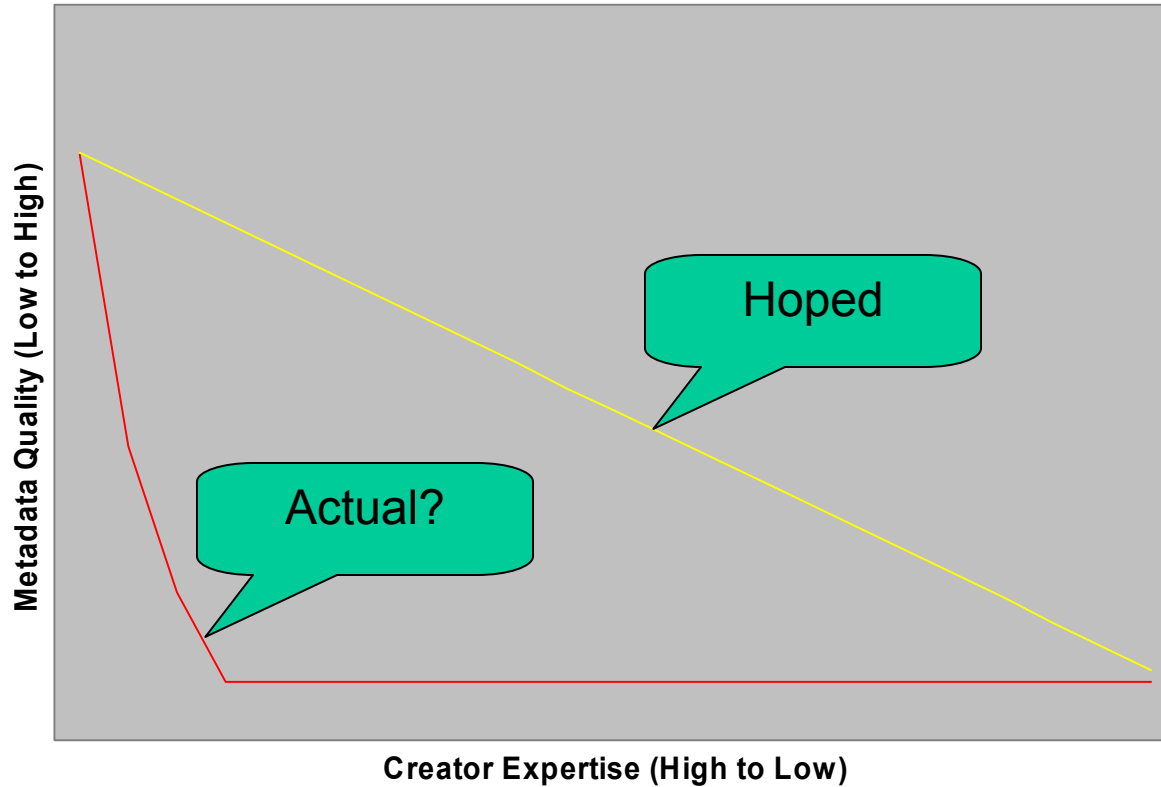
- Raw materials for generating refined descriptions



Metadata question 1: What types of resources?



Metadata question 2: What level of expertise?



Metadata question 2: How important is quality?



Metadata question 3: Machine Generation?



Metadata question 4: User needs

- This is not the only discovery model:



- What about:
 - Collocation
 - Topic browsing
 - Known item searching
 - Other needs for metadata

User Studies: Methods & Questions

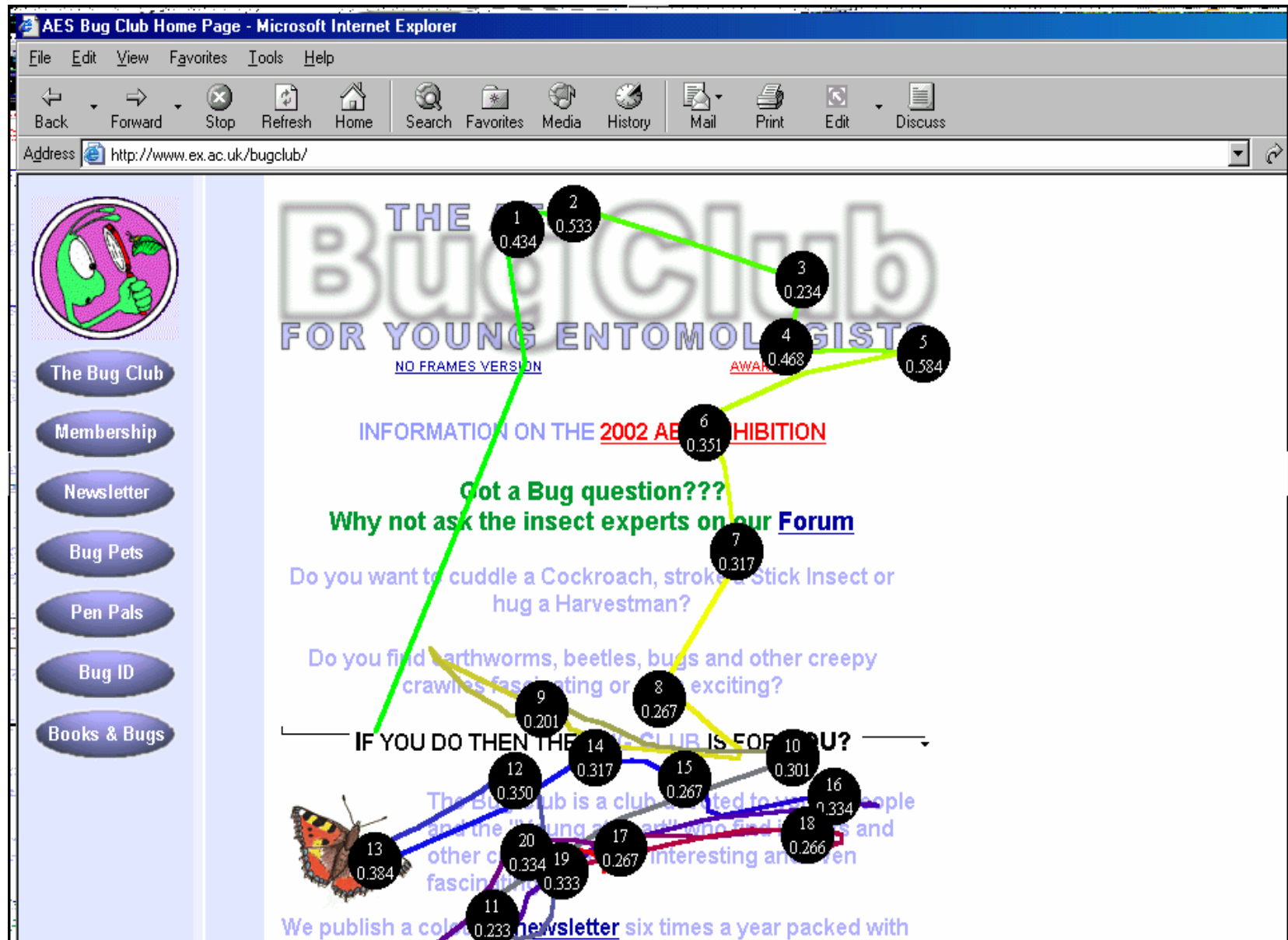
1. Observations of Users Seeking DL Resources

- How do users search & browse the digital library?
- Do search attempts reflect the available metadata?
- Which metadata elements are the most important to users?
- What metadata elements are used most consistently with the best results?

User Studies: Methods & Questions (cont'd)

- **2. Eye-tracking with Think-aloud Protocols**
 - Which metadata elements do users spend most time viewing?
 - What are users thinking about when seeking digital library resources?
 - Show correlation between what users are looking at and thinking.
 - Use eye-tracking to measure the number & duration of fixations, scan paths, dilation, etc.
- **3. Individual Subject Data**
 - How does expertise / role influence seeking resources from digital libraries?

Eye Scan Path For Bug Club Document




Eye Scan Path For Sigmund Freud Document

A Science Odyssey: People and Discoveries: Sigmund Freud - Microsoft Internet Explorer

Address: <http://www.pbs.org/wgbh/aso/databank/entries/bhfreu.html>

PEOPLE AND DISCOVERIES

Sigmund Freud 1939



Sigmund Freud was born in the Austro-Hungarian Empire in 1856. His father was a small time merchant, and Freud's mother was his second wife. Freud had two half-brothers some 20 years older than himself. His family moved to Vienna when he was four years old. Although he claimed he hated the city, he lived there until it was occupied by Germany in 1938. Freud's family background was Jewish, though his father was a freethinker and Freud himself an avowed atheist.

Freud was a good student and very ambitious. Medicine and law were the professions then open to Jewish men, and in 1873 he entered the University of Vienna medical school. He was interested in science above all; the idea of practicing medicine was slightly repugnant to him. He hoped to go into neurophysiological research, but pure research was hard to manage in those days unless you were independently wealthy. Freud was engaged and needed to be able to support a family before he could marry, and so he determined to go into private practice with a specialty in neurology.

During his training he befriended Josef Breuer, another physician and physiologist. The two men discussed medical cases together and one of Breuer's would have a lasting effect on Freud. Known as Anna O., this patient was a young woman suffering from what was then called hysteria. She had temporary paralysis, could not speak her native German but could speak French and English, couldn't drink water even when thirsty, and so on. Breuer discovered that if he hypnotized her, she would talk of things she did not remember in her conscious state, and afterwards her symptoms were relieved -- thus it was called "the talking cure." Freud went to Paris for further study under [Jean-Martin Charcot](#), a neurologist known all over Europe for his studies of hysterics and use of hypnosis.

In 1886, Freud returned to Vienna, opened a private practice specializing in nervous and brain disorders, and married. He tried hypnosis with his hysteric and neurotic patients, but gradually discarded the practice. He found he could get patients to talk just by putting them in a relaxing position (the couch) and encouraging them to say whatever came into their heads (free association). He then analyzed what they had remembered or expressed and determined what traumatic events in their past had caused their current

Point Number	Approximate Value
1	0.251
2	0.450
3	0.483
4	0.483
5	0.472
6	0.234
7	0.472
8	0.233
9	0.251
10	0.433
11	0.217
12	0.200
13	0.217
14	0.234
15	0.233

Evaluating MetaData

Blind Test of Automatic vs. Manual Metadata

Expectation Condition – Subjects reviewed:

1st - metadata record

2nd – lesson plan

and then judged whether metadata provided an accurate preview of the lesson plan on 1 to 5 scale

Satisfaction Condition– Subjects reviewed:

1st – lesson plan

2nd – metadata record

and then judged the accuracy and coverage of metadata on 1 to 5 scale, with 5 being high

Qualitative Study Results

Expec Satis Comb

# Manual Metadata Records			153
571	724		
# Automatic Metadata Records			139
532	671		

Qualitative Study Results

Comb Expec Satis

~~# Manual Metadata Records~~ ~~153~~

571 724

Automatic Metadata Records 139

532 671

Manual Metadata Average Score 4.03

Qualitative Study Results

Comb Expec Satis

Manual Metadata Records 153

571 724

Automatic Metadata Records 139

~~532 671~~

Manual Metadata Average Score 4.03

3.81 3.85

Automatic Metadata Average Score 3.76

Models for Deploying Metadata

- Embedded in the resource
 - low deployment threshold
 - Limited flexibility, limited model
- Linked to from resource
 - Using xlink
 - Is there only one source of metadata?
- Independent resource referencing resource
 - Model of accessing the object through its surrogate
 - Resource doesn't 'have' metadata, metadata is just one resource annotating another

Syntax Alternatives:

HTML

- **Advantages:**
 - Simple Mechanism - META tags embedded in content
 - Widely deployed tools and knowledge
- **Disadvantages**
 - Limited structural richness (won't support hierarchical, tree-structured data or entity distinctions).

Dublin Core in XHTML

- <http://www.dublincore.org/documents/dcq-html/>
- <link> to establish pseudo-namespace
 - <link rel="schema.DC" href="http://purl.org/dc/elements/1.1/" />
 - <link rel="schema.DCTERMS" href="http://purl.org/dc/terms/" />
- <meta> for metadata statements
 - Use of attributes
 - name attribute for DC element
 - content attribute for element value
 - scheme attribute for encoding scheme or controlled vocabulary
 - lang attribute for language of element value
 - Examples
 - <meta name="DC.date" scheme="DCTERMS.W3CDTF" content="2001-07-18" />
 - <meta name="DC.type" scheme="DCTERMS.DCMIType" content="Text" />
 - <meta name="DC.subject" xml:lang="fr" content="fruits de mer" />

Dublin Core in xHTML example

```
<head profile="http://dublincore.org/documents/dcq-html/">
<title>Expressing Dublin Core in HTML/XHTML meta and link elements</title>
<link rel="schema.DC" href="http://purl.org/dc/elements/1.1/" />
<link rel="schema.DCTERMS" href="http://purl.org/dc/terms/" />

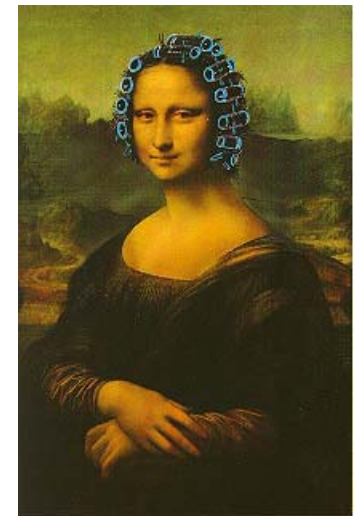
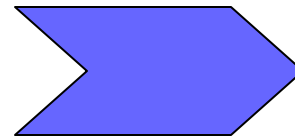
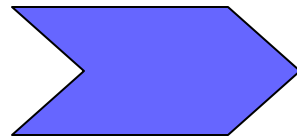
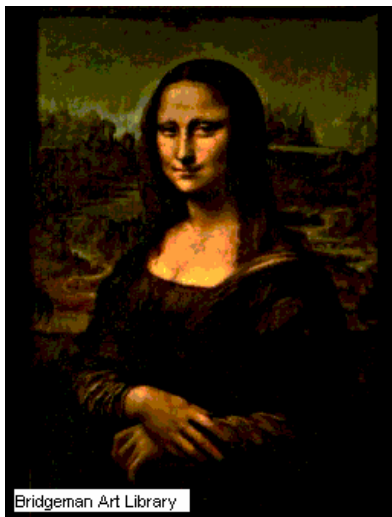
<meta name="DC.title" lang="en" content="Expressing Dublin Core
in HTML/XHTML meta and link elements" />
<meta name="DC.creator" content="Andy Powell, UKOLN, University of Bath" />
<meta name="DCTERMS.issued" scheme="DCTERMS.W3CDTF" content="2003-11-01" />
<meta name="DC.identifier" scheme="DCTERMS.URI"
content="http://dublincore.org/documents/dcq-html/" />
<link rel="DCTERMS.replaces" hreflang="en"
href="http://dublincore.org/documents/2000/08/15/dcq-html/" />
<meta name="DCTERMS.abstract" content="This document describes how
qualified Dublin Core metadata can be encoded
in HTML/XHTML &lt;meta&gt; elements" />
<meta name="DC.format" scheme="DCTERMS.IMT" content="text/html" />
<meta name="DC.type" scheme="DCTERMS.DCMIType" content="Text" />
</head>
```

Unqualified Dublin Core in RDF/XML

```
<?xml version="1.0"?>
<!DOCTYPE rdf:RDF PUBLIC "-//DUBLIN CORE//DCMES DTD 2002/07/31//EN"
  "http://dublincore.org/documents/2002/07/31/dcmes-xml/dcmes-xml-dtd.dtd">
<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://dublincore.org/">
  <dc:title>Dublin Core Metadata Initiative - Home Page</dc:title>
  <dc:description>The Dublin Core Metadata Initiative Web site.</dc:description>
  <dc:date>2001-01-16</dc:date>
  <dc:format>text/html</dc:format>
  <dc:language>en</dc:language>
  <dc:contributor>The Dublin Core Metadata Initiative</dc:contributor>
  <!-- guesses for the translation of the above titles -->
  <dc:title xml:lang="fr">L'Initiative de métadonnées du Dublin Core</dc:title>
  <dc:title xml:lang="de">der Dublin-Core Metadata-Diskussionen</dc:title>
</rdf:Description>
</rdf:RDF>
```

<http://www.dublincore.org/documents/2002/07/31/dcmes-xml/>

Multi-entity nature of object description



Photographer



Camera type



Computer artist

Software

Attribute/Value approaches to metadata...

The playwright of Hamlet was Shakespeare

subject
Hamlet

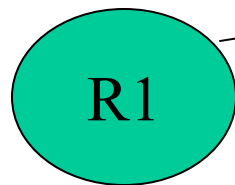
implied verb
has a

metadata noun
creator

literal
Shakespeare

metadata adjective

Playwright



dc:creator.playwright → "Shakespeare"

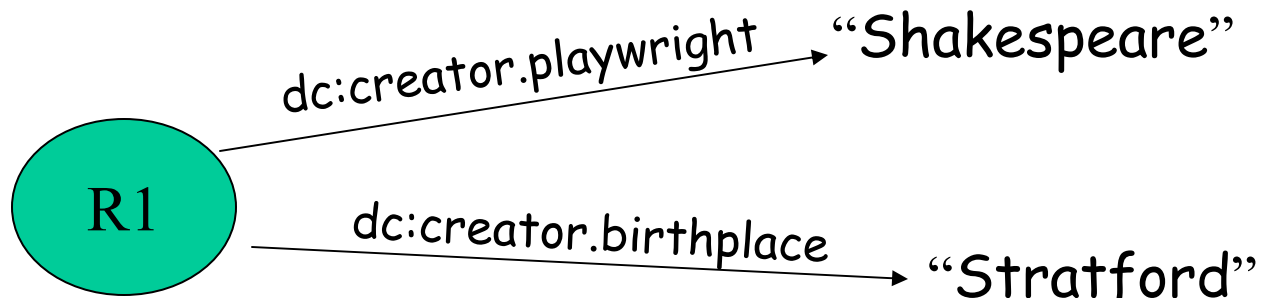
dc:title → "Hamlet"

...run into problems for richer descriptions...

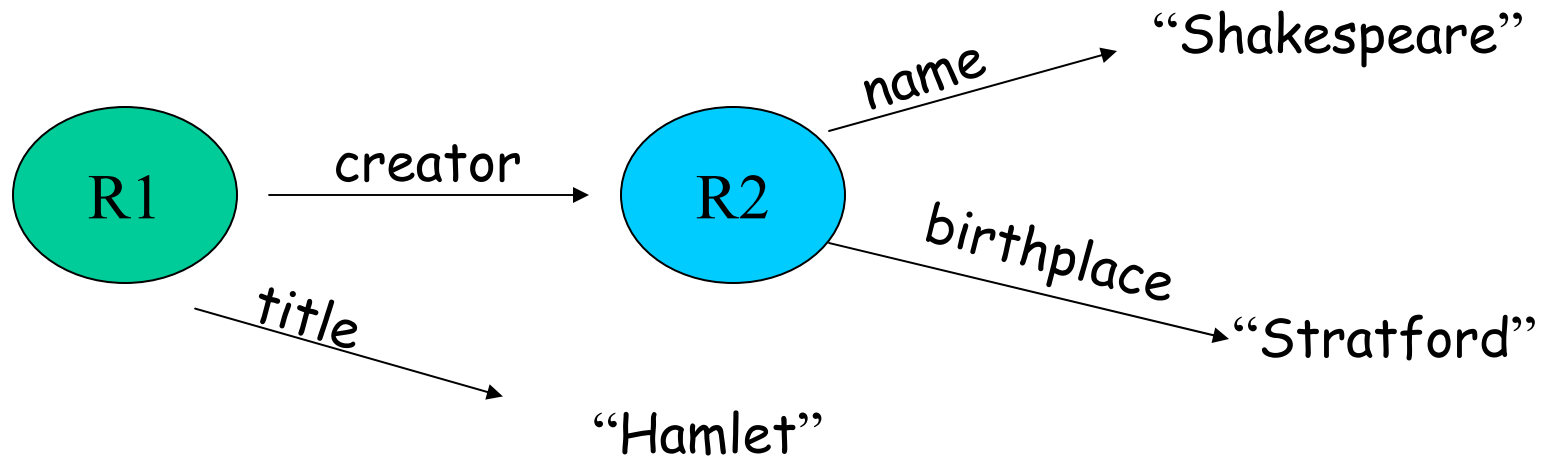
The playwright of Hamlet was Shakespeare,
who was born in Stratford

Hamlet has a creator Stratford

birthplace



...because of their failure to model entity distinctions ...



... and their failure to associate attributes with temporal semantics

- What happened when
- In what sequence did things happen
- Concepts
 - Discreet events
 - Parallelism
 - Dependencies
- Temporal semantics are notoriously difficult and face tractability problems

Applying a Model-Centric Approach

- Formally define common entities and relationships underlying multiple metadata vocabularies
- Describe them (and their inter-relationships) in a simple logical model
- Provide the framework for extending these common semantics to domain and application-specific metadata vocabularies.

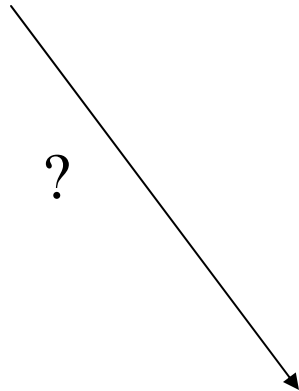
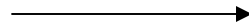
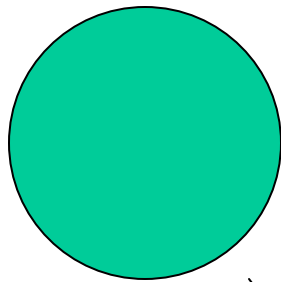
Events are key to understanding resource complexity?

- Events are implicit in most metadata formats (e.g., 'date published', 'translator')
- Modeling implied events as first-class objects provides attachment points for common entities - e.g., agents, contexts (times & places), roles.
- Clarifying attachment points facilitates understanding and querying "*who was responsible for what when*".

ABC/Harmony Event-aware metadata ontology

- <http://jodi.ecs.soton.ac.uk/Articles/v02/i02/Lagoze/>
- Recognizing inherent lifecycle aspects of description (esp. of digital content)
- Modeling incorporates time (events and situations) as first-class objects
 - Supplies clear attachment points for agents, roles, existential properties
- Resource description as a “**story-telling**” activity

Resource-centric Metadata



Title	Anna Karenina
Author	Leo Tolstoy
Illustrator	Orest Vereisky
Translator	Margaret Wettlin
Date Created	1877
Date Translated	1978
Description	Adultery & Depression

Birthplace	Moscow
Birthdate	1828

