College of Information Studies

University of Maryland Hornbake Library Building College Park, MD 20742-4345

Metadata

Week 4

LBSC 671

Creating Information Infrastructures

Tonight

- Finishing Preservation
- Metadata

Preserving Behavior

- Word processors
  - Formatting, track changes, undo deleted text
- Spreadsheets
  - Formulas, visualizations
- Databases
  - Queries, forms, derived values
- Computer-Assisted Design (CAD)
  - Display, modification, manufacturing
- Software
  - Simulation, games, embedded systems, …

Behavior Preservation Strategies

- Format migration
  - For example, convert Word Perfect to PDF
- Emulation
  - Allows running old software on newer systems

Apollo Guidance Computer Emulation

AGC Simulation Type

Guidance Computer (AGC) software

- ○ Apollo 1 Command Module
- ○ Apollo 7 Command Module
- ◉ Apollo 8 Command Module
- ○ Apollo 9 Command Module
- ○ Apollo 9 Lunar Module
- ○ Apollo 10 Command Module
- ○ Apollo 10 Lunar Module
- ○ Apollo 11 Command Module
- ○ Apollo 11 Lunar Module
- ○ Apollo 12 Command Module
- ○ Apollo 12 Lunar Module
- ○ Apollo 13 Command Module
- ◉ Apollo 13 Lunar Module
- ○ Apollo 14 Command Module
- ○ Apollo 14 Lunar Module
- ○ Apollo 15-17 Command Module
- ○ Apollo 15-17 Lunar Module
- ○ Apollo Skylab/Soyuz Command Module
- ○ Validation Suite
- ○ Custom: [ ] [...]

Interfaces

- ☑ Guidance Computer
- ☑ DSKY (AGC display and keypad)
- ☑ Attitude Controller Assembly [Handler]
- ☑ Telemetry Downlink Monitor
- ☑ LM Abort Computer (AEA)
- ☑ DEDA (AEA display and keypad)
- ☐ AGC CPU Bus/Input/Output Monitor
  - ☐ Inertial Monitor Unit / FDAI (8-ball)
  - ☐ Discrete Outputs
  - ☐ Discrete Inputs (crew)
  - ☐ Discrete Inputs (LM system)
  - ☐ Propulsion/Thrust/Fuel Monitor

[Novice] [Expert]

Browse Source Code

[AGC] [AEA]

Options

AGC Startup

- ◉ Restart program, wiping memory
- ○ Restart program, preserving memory
- ○ Resume from ending point of prior run
- ○ Custom: [ ] [...] [Save]

Interface styles

DSKY:	◉ Full	○ Half	○ "Lite"
Downlink:	◉ Normal	○ "Retro"	
DEDA:	◉ Full	○ Half	

Use AGC/AEA debugger?

AGC code:	◉ Normal	○ Debugger
AEA code:	◉ Normal	○ Debugger

LM Abort Computer (AEA) software

- ○ Apollo 9 (Flight Programs 3, 4)
- ○ Apollo 10 (Flight Program 5)
- ◉ Apollo 11 (Flight Program 6)
- ○ Apollo 12-14? (Flight Program 7)
- ○ Apollo 15-17 (Flight Program 8)
- ○ Custom: [ ] [...]

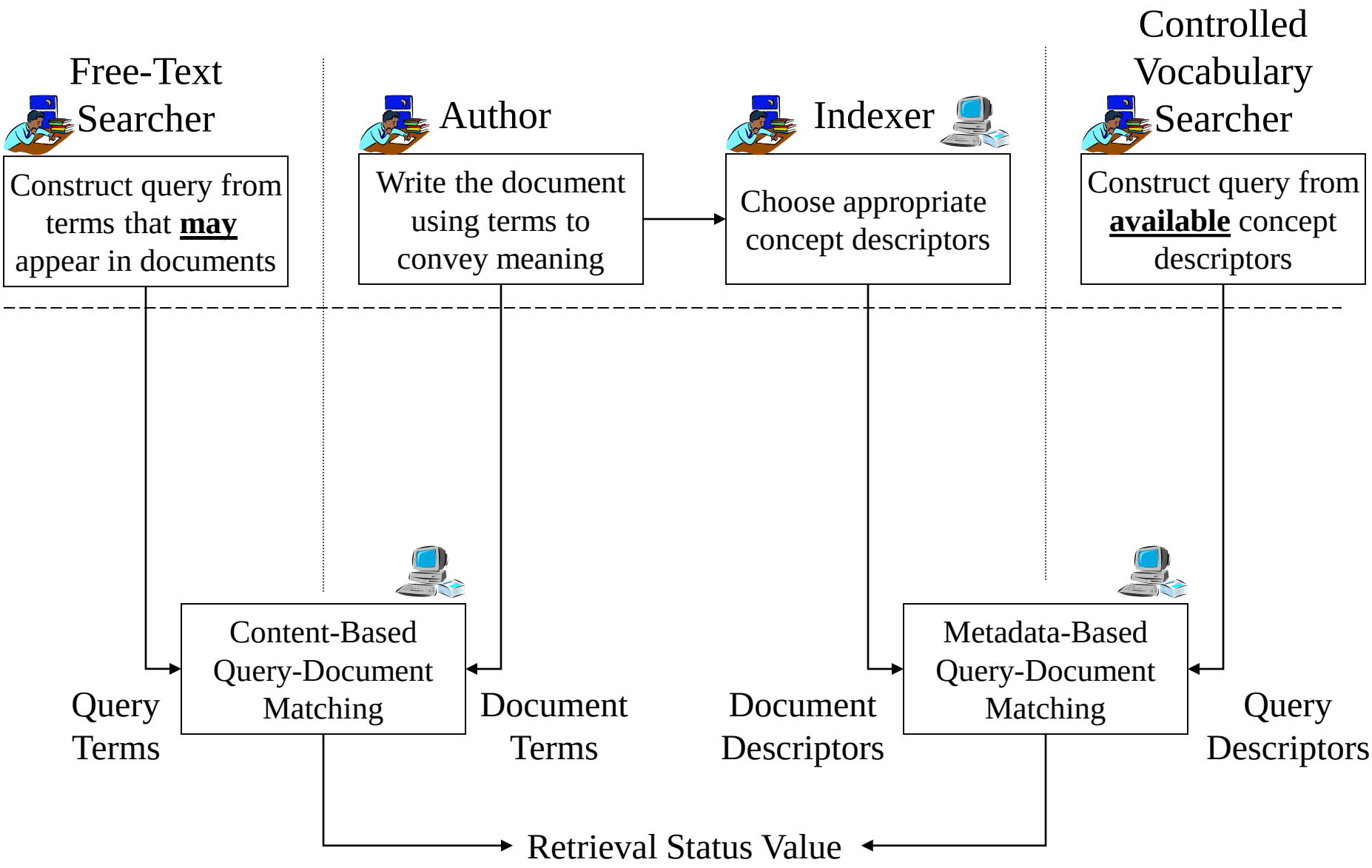
[Run!] [Defaults] [Exit]

http://www.ibiblio.org/apollo/

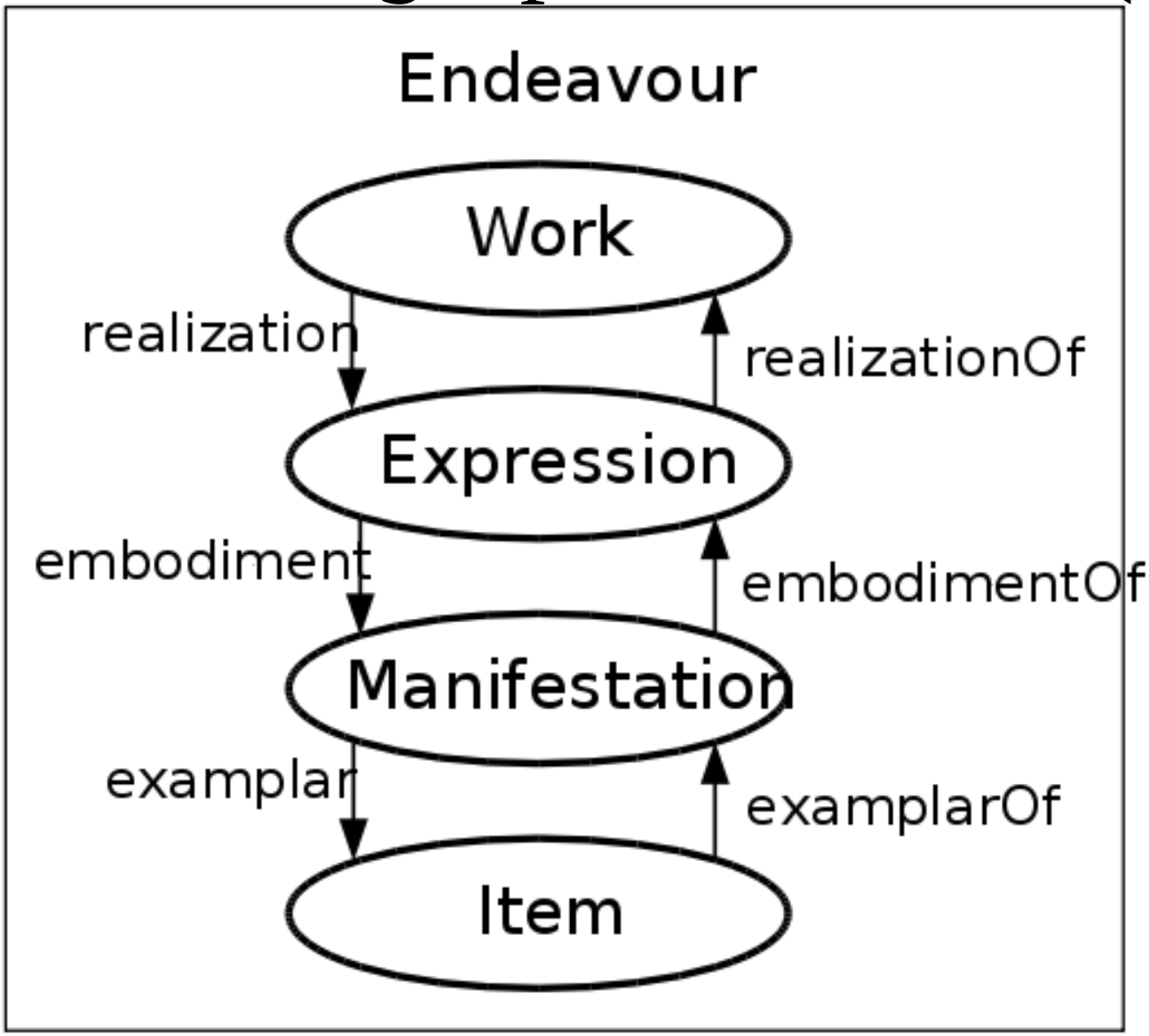
An Integrated Strategy

- Delay decay of organic materials
  - But balance costs and benefits
- Balance quality and scale
  - Preservation: rescue at-risk collections
  - Access: Quantity has a quality all its own
- Design in diversity
  - Technologies, risk exposure, institutions
- Adequately resource the process

Two Ways of Searching



Functional Requirements for *Bibliographic* Records (FRBR)



Midsummer Night's Dream

2005 Free for All

August 23 Performance

N/A

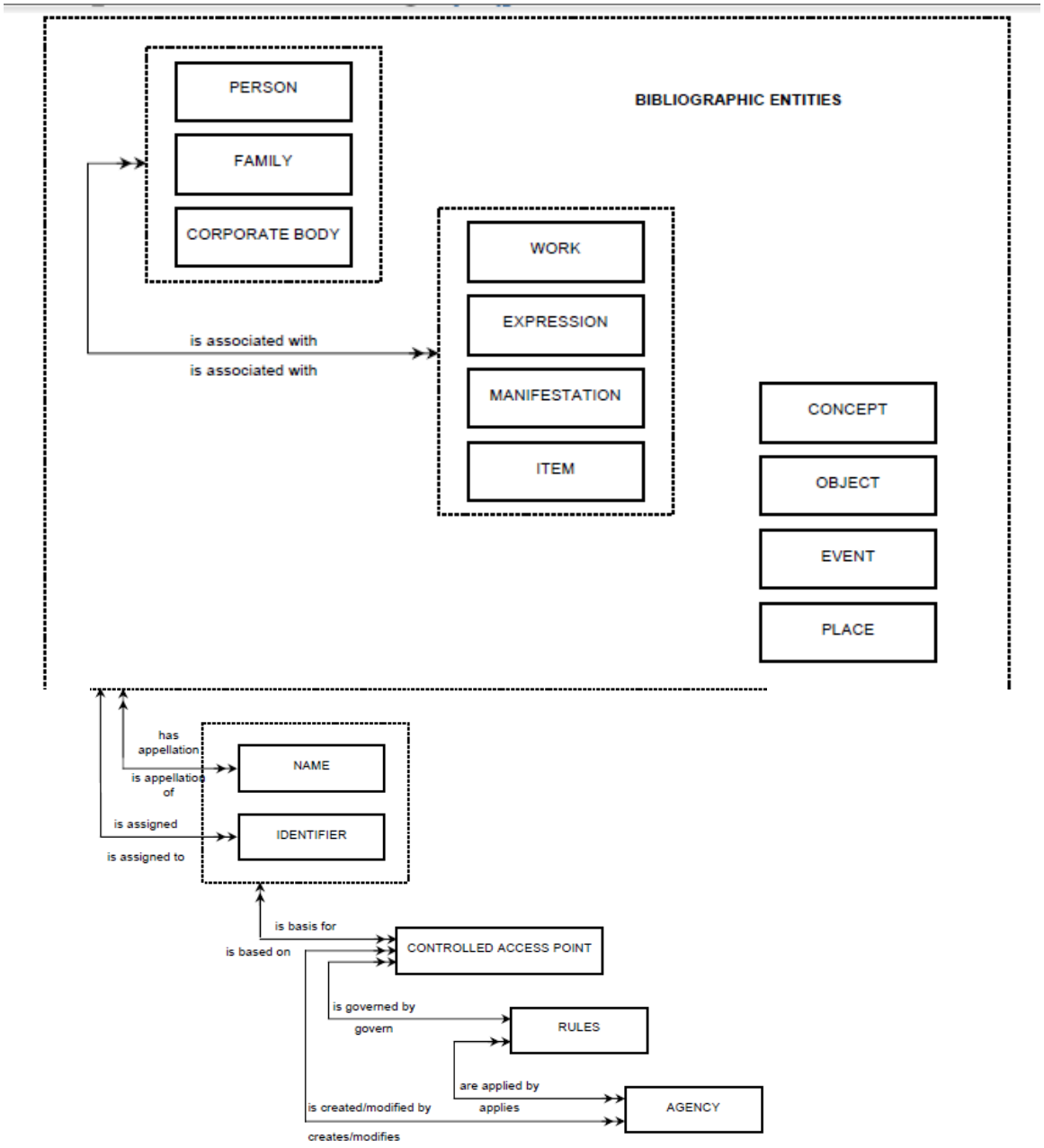
FRBR Bibliographic User Tasks

- Find it
  - Search (“to find”)
  - Recognize (“to identify”)
  - Choose (“to select”)
- Serve it
  - Location (“to obtain”)

FRBR Entity Types

- Subject-Only Entities
  - (abstract) Concepts
  - (tangible) Objects
  - (any kind of) Places
  - Events
- Subject or Responsibility Entities
  - Persons
  - (any kind of) “Corporate” Bodies
  - Families (technically, only in FRAD)
- Product Entities
  - Works, Expressions, Manifestations, Items

Functional Requirements for Authority Data



IFLA, 2013

FRBR in OCLC's FictionFinder

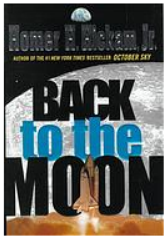
FictionFinder beta
OCLC Research

Project Page | Feedback | Known Problems | Exit

Browse | Search GO [Advanced]

You searched: Basic Index for **nasa**

Back to Results << Previous Work **5** of **43** Next > Find Any edition



Back to the moon : a novel
Hickam, Homer H., 1943-

6 editions, in 2 languages, held by 1324 libraries

Summary: A cosmic romp in which Jack Medaris, a former astronaut who is a widower, saves the female crew of a NASA spaceship from an attack by chauvinist astronauts. The ladies take him to the moon, he finds romance and brings back a rare helium element which will help his business.

Genres: Adventure fiction | Adventure stories | Science fiction

Settings: Moon

Subjects: Lunar exploration | Helium —Isotopes [+]

Audience:

Kids General Special

Editions | Genres | Characters | Settings

Narrow by Languages: All (6)

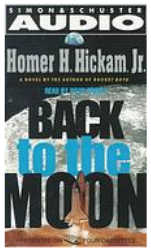
	Title / Author	OCLC #	Date	Langu
1.	Back to the moon : Homer H. Hickam, Jr	40979898	1999	Englis
2.	Back to the moon Homer H. Hickam	41713450	1999	Englis
3.	Back to the moon : Homer H. Hickam, Jr	43890225	2000	Englis
4.	Back to the moon Homer H. Hickam	42765643	1999	Englis
5.	Ben yue zhui qi ling / Xi kan mu zhu ; Wu Hong yi = Back to the moon / by Homer H. Hickam, Jr	47716097	2000	Chine
6.	Back to the moon Homer H. Hickam, Jr	49832301	1999	Englis

FictionFinder beta
OCLC Research

Project Page | Feedback | Known Problems | Exit

Browse | Search GO [Advanced]

Back to Work < Previous Edition 4 of 6 Next > Find This Edition



Back To the Moon A Novel
Hickam, Homer H., 1943- / Homer H. Hickam. [sound recording] :

Edition: Library ed.

Date: 1999.

Language: English

Publisher: Prince Frederick, MD : Recorded Books, 1999.

ISBN: 0671046403

OCLC: 42765643

Citations | Details | Excerpt | Reviews | Table of Contents

Details

Summary: Jack Medaris, a man of science driven by the memory of the woman who once inspired him, risks everything to sidetrack the space shuttle Columbia and take it on an unscheduled detour to the moon. When the meticulously plotted launch goes fatally wrong, and payload specialist Penny High Eagle further complicates Jack's plan, he must confront unforeseen challenges both in space and on the ground.

Settings: Moon

Performer: Read by Boyd Gaines.

Dublin Core

- Goals:
  - Easily understood, implemented and used
  - Broadly applicable to many applications
- Approach:
  - Intersect several standards (e.g., MARC)
  - Suggest only “best practices” for element content
- Implementation:
  - Initially 15 optional and repeatable “elements”
    - Refined using a growing set of “qualifiers”
  - Now extended to 22 elements

Dublin Core Elements (version 1.1)

Content

- Title
- Subject [LCSH, MeSH, …]
- Description
- Type
- Coverage [spatial, temporal, …]
- Related resource
- Rights

Instantiation

- Date [Created, Modified, Copyright, …]
- Format
- Language
- Identifier [URI, Citation, …]

Responsibility

- Creator
- Contributor
- Source
- Publisher

Resource Description Framework

- XML schema for describing resources
- Can integrate multiple metadata standards
  - Dublin Core, P3P, PICS, vCARD, …
- Dublin Core provides a XML “namespace”
  - DC Elements are XML “properties
    - DC Refinements are RDF “subproperties”
  - Values are XML “content”

Dublin Core in RDF XML

```
<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://media.example.com/audio/guide.ra">
    <dc:creator>Rose Bush</dc:creator>
    <dc:title>A Guide to Growing Roses</dc:title>
    <dc:description>Describes process for planting and nurturing
                    different kinds of rose bushes.</dc:description>
    <dc:date>2001-01-20</dc:date>
  </rdf:Description>
</rdf:RDF>
```

Five “Levels” of Metadata

- Framework
  - Functional Requirements for Bibliographic Records (FRBR)
- Schema (“Data Fields and Structure”) 
  - Dublin Core
- Vocabulary (“Data Content and Values”) 
  - Resource Description and Access (RDA)
  - Library of Congress Subject Headings (LCSH)
- Representation (abstract “Data Format”)
  - Resource Description Framework (RDF)
- Serialization (“Data Format”) 
  - RDF in eXtensible Markup Language (RDF/XML)

Adapted from Elings and Waibel, *First Monday*, (12)3, 2007

Types of “Metadata”

- Descriptive
  - Content, creation process, relationships
- Technical
  - Format, system requirements
- Usage
  - Display, derivative works
- Administrative
  - Acquisition, authentication, access rights
- Preservation
  - Media migration

Adapted from Introduction to Metadata, Getty Information Institute (2000)

Metadata Encoding and Transmission Standard (METS)

- Descriptive metadata (e.g., subject, author)
- Administrative metadata (e.g., rights, provenance)
- Technical metadata (e.g., resolution, color space)
- Behavior (which program can render this?)
- Structural map (e.g., page order)
  - Structural links (e.g., Web site navigation links)
- Files (the raw data)
- Root (meta-metadata)

Aspects of Metadata

- What kinds of objects can we describe?
  - MARC, Dublin Core, FRBR, …
- How can we convey it?
  - MODS, RDF, OAI-PMH, METS
- What can we say?
  - LCSH, MeSH, PREMIS, …
- What can we do with it?
  - Discovery, description, reasoning

FRBR Bibliographic User Tasks

- Find it
  - Search (“to find”)
  - Recognize (“to identify”)
  - Choose (“to select”)
- Serve it
  - Location (“to obtain”)

Broader View of Metadata Uses

- Have it
  - Preservation (e.g., PREMIS)
  - Validation
  - Disposition
- Find it
  - Search/Recognize/Choose
  - Browse (“Navigation”)
- Serve it
  - Persistent location
  - Structure
  - Surrogates
- Use it
  - Context
  - Rights management
  - User behavior capture
  - Reasoning (“Semantic Web”)

Metadata Sources

- Automated
  - Capture
  - Extraction
  - Classification
- Manual
  - Professional
  - Community
  - Personal

Metadata Capture: Exchangeable Image Format (EXIF)

- Time
- Location
- Camera manufacturer and model
- Camera orientation
- Exposure information (shutter speed, f stop)
- Thumbnail versions
  - Altering the image may not change the thumbnail!

Inconsistent Metadata

http://www.umiacs.umd.edu/~oard/rtw/

Metadata Capture: Email

- Message metadata
  - Times
    - Sent
    - Resent
    - Received
  - Route
  - In-reply-to
  - Attachment file type
- System metadata
  - Folder

Metadata Capture: Windows File System (NTFS)

- Time file created (or copied)
  - Most recent one; optionally “journaled”
- Time file content changed (or made changeable)
  - Most recent one; optionally “journaled”
- Time file renamed (or moved)
  - Most recent one
- Time file metadata created or changed
  - Most recent one
- Time file accessed (content or metadata)
  - Most recent one; optionally disabled

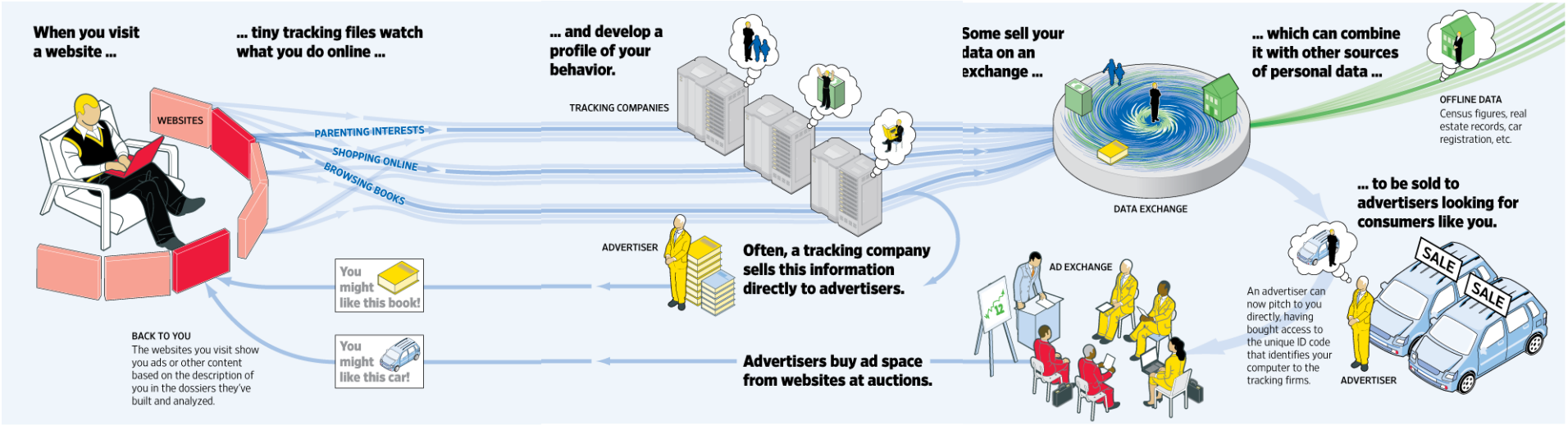
Metadata Capture: Microsoft Word

- Author
- Title
- Dates (may not agree with file system)
  - Created
  - Modified
  - Accessed
  - Printed
  - Each tracked change

Metadata Capture: User Behavior

Behavior Category	Minimum Scope: Segment	Minimum Scope: Object	Minimum Scope: Class
Examine	View Listen	Select	
Retain	Print	Bookmark Save Purchase Delete	Subscribe
Reference	Copy / paste Quote	Forward Reply Link Cite	
Annotate	Mark up	Tag Publish	Organize
Create	Type Edit		

Exploiting Behavioral Metadata

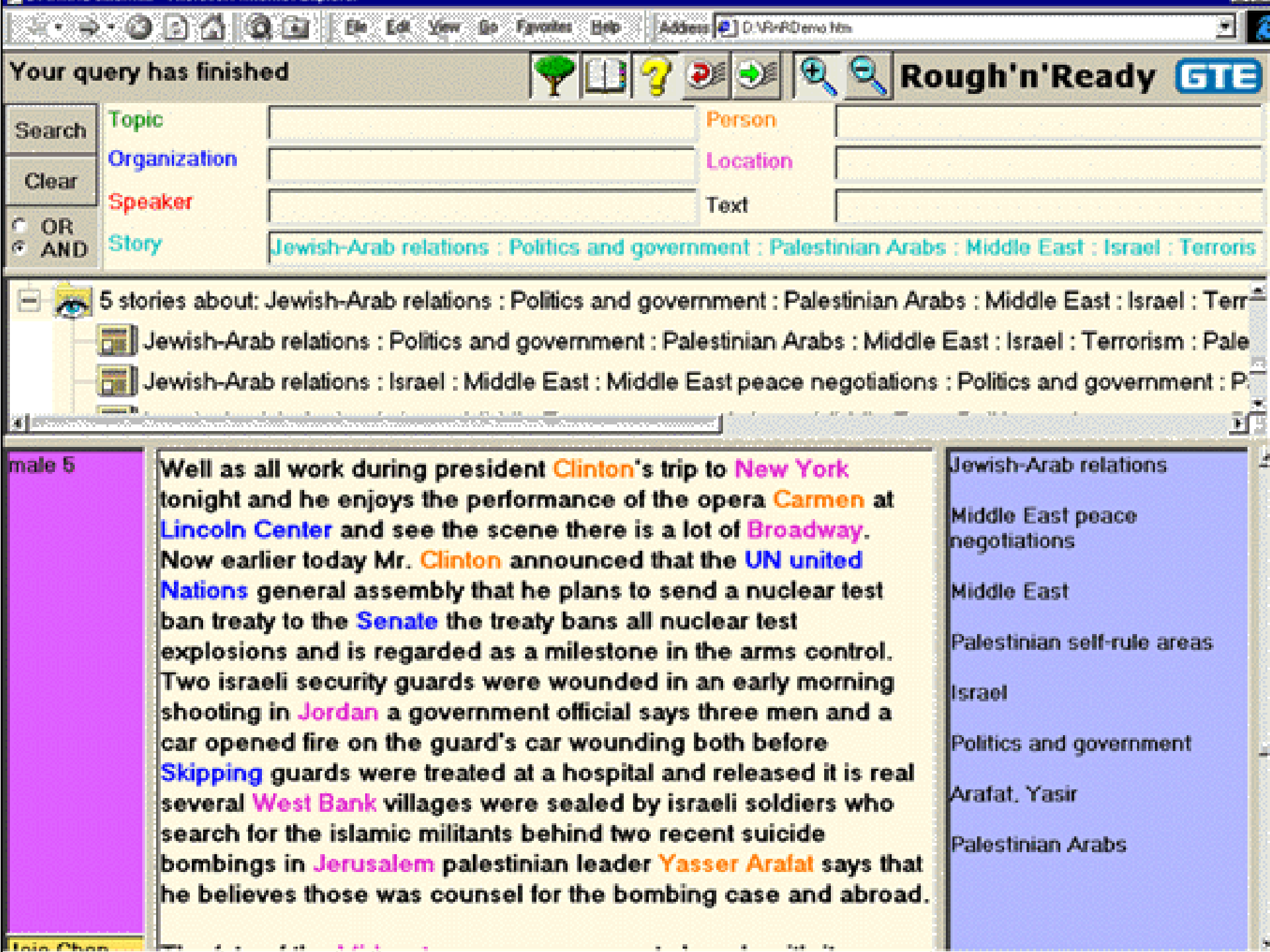


http://wsj.com/wtk

Metadata Extraction: Named Entity "Tagging"

- Machine learning techniques can find:
  - Location
  - Extent
  - Type

- Two types of features are useful
  - Orthography
    - e.g., Paired or non-initial capitalization
  - Trigger words
    - e.g., Mr., Professor, said, …


Your query has finished
Rough'n'Ready
GTE
Search
Clear
OR
AND
Topic
Person
Organization
Location
Speaker
Text
Story
Jewish-Arab relations : Politics and government : Palestinian Arabs : Middle East : Israel : Terroris
5 stories about: Jewish-Arab relations : Politics and government : Palestinian Arabs : Middle East : Israel : Terr
Jewish-Arab relations : Politics and government : Palestinian Arabs : Middle East : Israel : Terrorism : Pale
Jewish-Arab relations : Israel : Middle East : Middle East peace negotiations : Politics and government : P
male 5
Well as all work during president Clinton's trip to New York tonight and he enjoys the performance of the opera Carmen at Lincoln Center and see the scene there is a lot of Broadway. Now earlier today Mr. Clinton announced that the UN united Nations general assembly that he plans to send a nuclear test ban treaty to the Senate the treaty bans all nuclear test explosions and is regarded as a milestone in the arms control. Two israeli security guards were wounded in an early morning shooting in Jordan a government official says three men and a car opened fire on the guard's car wounding both before Skipping guards were treated at a hospital and released it is real several West Bank villages were sealed by israeli soldiers who search for the islamic militants behind two recent suicide bombings in Jerusalem palestinian leader Yasser Arafat says that he believes those was counsel for the bombing case and abroad.
Jewish-Arab relations
Middle East peace negotiations
Middle East
Palestinian self-rule areas
Israel
Politics and government
Arafat, Yasir
Palestinian Arabs

Rule Based Classification: Machine-Assisted Indexing

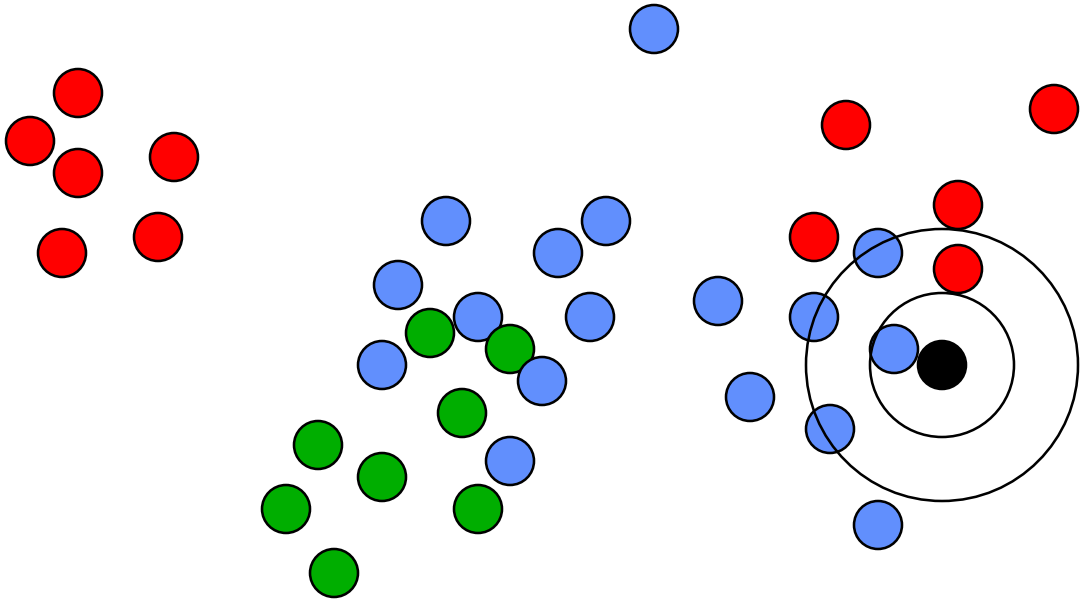
Access Innovations system:

```
//TEXT: science
IF (all caps)
  USE research policy
  USE community program
ENDIF
IF (near “Technology” AND with “Development”)
  USE community development
  USE development aid
ENDIF
```

near: within 250 words

with: in the same sentence

Machine Learning for Classification: The k-Nearest-Neighbor Classifier



Metadata Sources

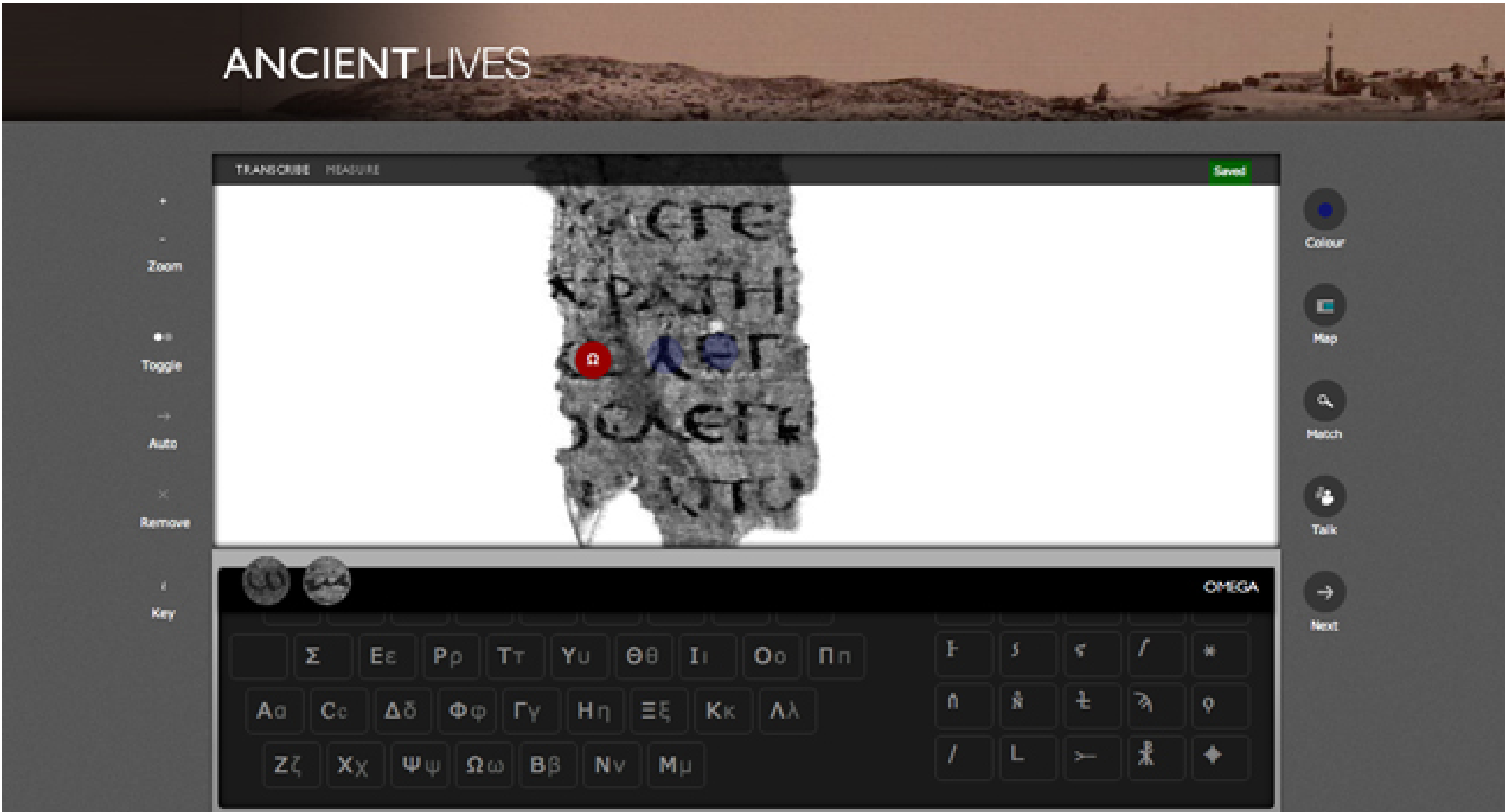
- Automated
  - Capture
  - Extraction
  - Classification
- Manual
  - Professional
  - Community
  - Personal

Community Metadata: “Folksonomies”

Community Metadata: Games With a Purpose

van Ahn and Dabbish, CHI 2004

Community Metadata: Crowdsourcing



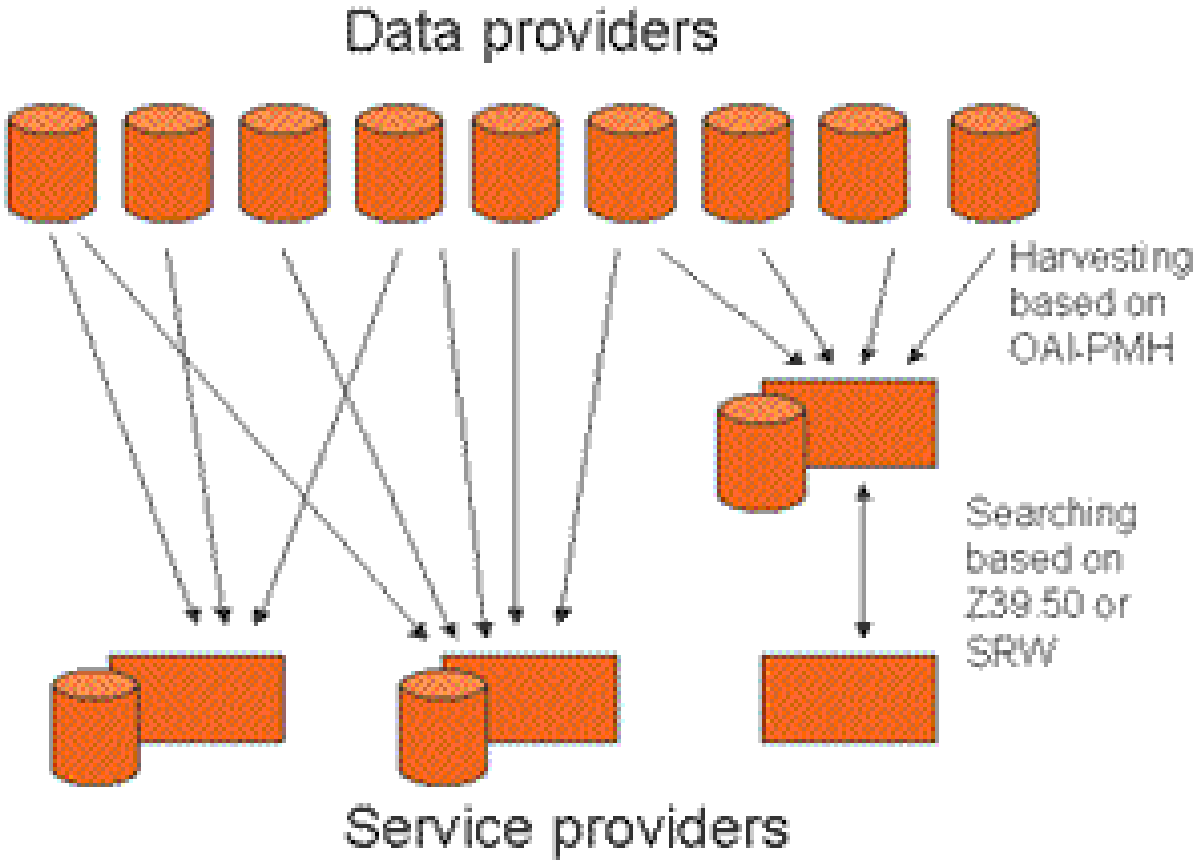
Sources of File Type Metadata

- Capture:
  - MyDocument.xls
  - Attachment MIME type
- Extraction
  - “Magic bytes”
- Classification
  - Machine learning on byte sequences
- Manual
  - Mechanical Turk

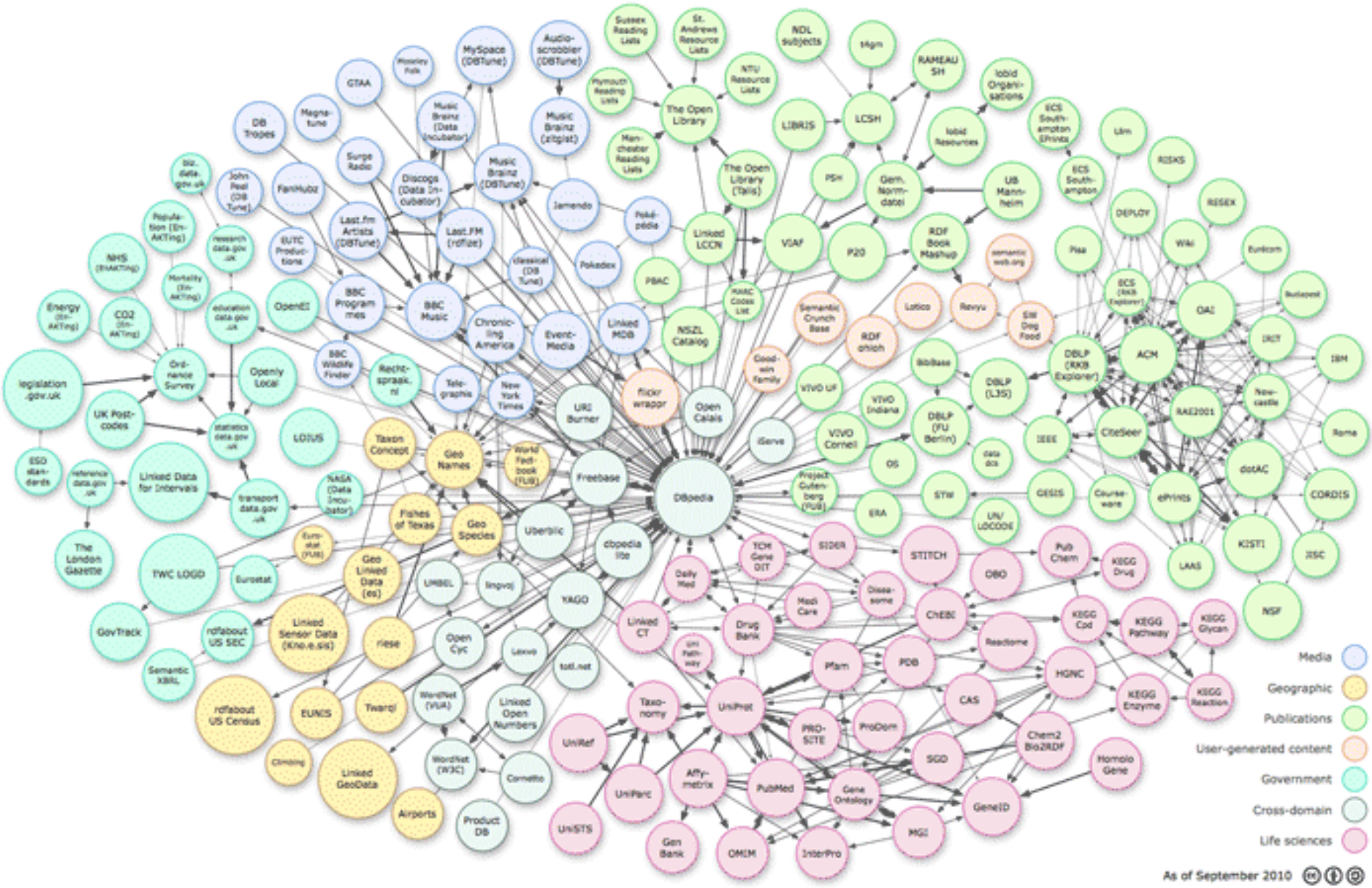
Metadata Challenges

- Balancing cost and benefit
- Accommodating dynamic factors
  - Content
  - Location
- Reuse for unanticipated purposes
- Remaining interpretable in the far future

Open Archives Initiative-Protocol for Metadata Harvesting (OAI-PMH)



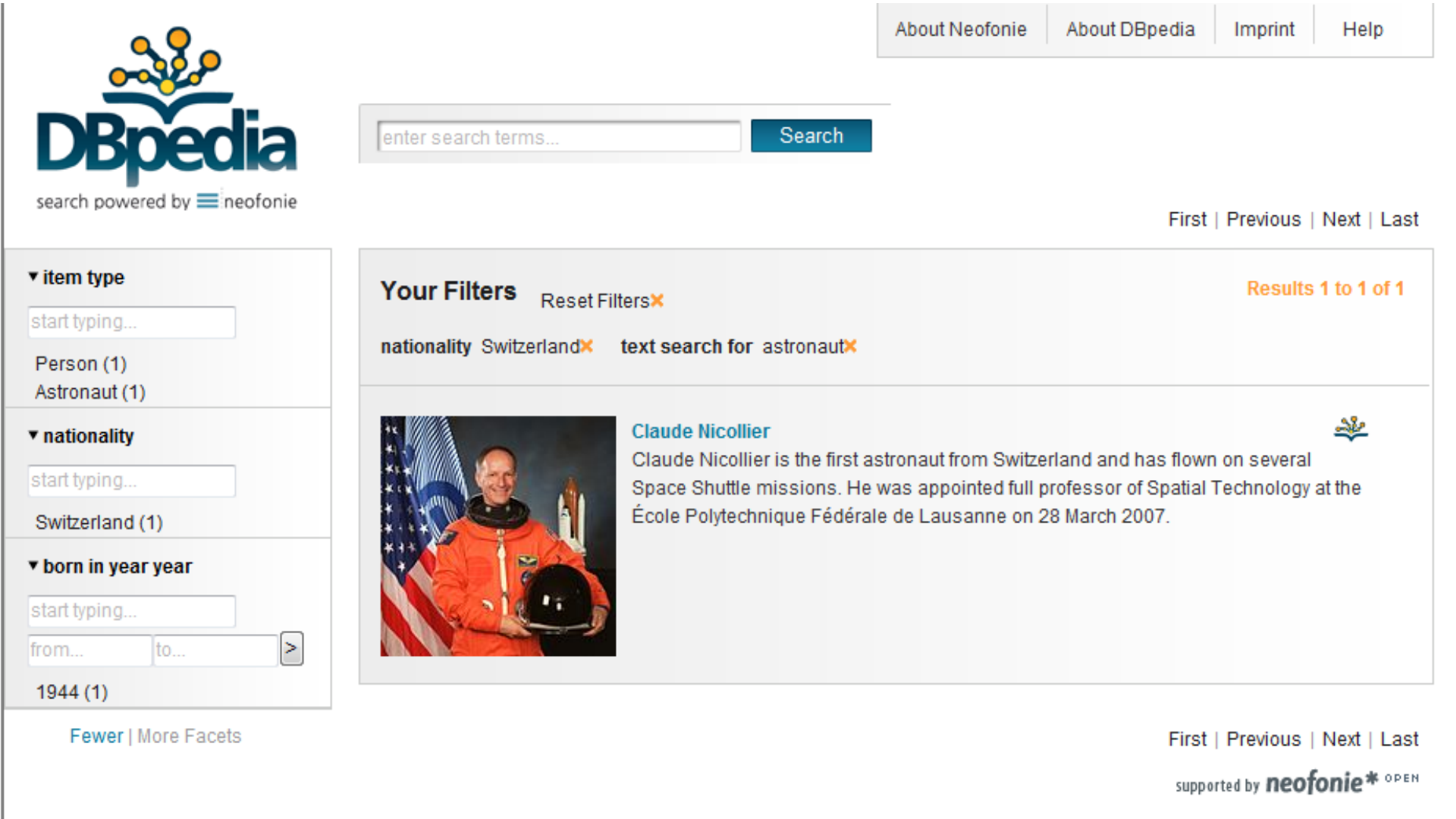
Linked Open Data



Web Ontology Language (OWL)

```
<owl:Class rdf:about="http://dbpedia.org/ontology/Astronaut">
        <rdfs:label xml:lang="en">astronaut</rdfs:label>
        <rdfs:label xml:lang="de">Astronaut</rdfs:label>
        <rdfs:label xml:lang="fr">astronaute</rdfs:label>
        <rdfs:subClassOf
                rdf:resource="http://dbpedia.org/ontology/Person">
        </rdfs:subClassOf>
</owl:Class>
```

“Semantic Web” Search



Putting It All Together

	Material Culture *Libraries* *Archives* *Museums*	*Bibliographic* *Libraries* *Archives* *Museums*	*Archival* *Libraries* *Archives* *Museums*
 **Data Structure**	CDWA	MARC	EAD
 **Data Content**	CCO	AACR2 (RDA)	DACS
 **Data Format**	XML	XML/ISO2709	XML
 **Data Exchange**	OAI	OAI Z39.50 SRU/SRW	OAI

Adapted from Elings and Waibel, *First Monday*, (12)3, 2007

Before You Go!

- On a sheet of paper (no names), answer the following question:

  What was the muddiest point in today's class?