University of Massachusetts Boston

From the SelectedWorks of Joanne M. Riley

Spring March 6, 2014

Working with Data in Archival Settings

Joanne M. Riley, *University of Massachusetts - Boston*

Available at: https://works.bepress.com/joanne_riley/18/

Working with Data in Archival Settings

Joanne Riley
University Archivist
Joseph P. Healey Library
University of Massachusetts Boston
617-287-5927
Joanne.riley@umb.edu

Mass. Memories Road Show - Mozilla Firefox

File Edit View History Delicious Bookmarks Tools Help

openarchives.umb.edu/cdm/search/collection/p15774coll6/searchterm/snow/order/title

life detective reese

Share page:

Search the Web

Mass. Memories Road Show | CONTENTdm Administration | life detective reese - Google Search

Open Archives: Digital Collections
at the University of Massachusetts Boston

Open Archives Home | Browse This Collection | About the Road Show | Open Archives Blog | Log in | Favorites | ? Help

Search: snow — new search — Search — Advanced Search

Add or remove other collections to your search:

- [x] Mass. Memories Road Show
- [ ] AFSC Vietnam Curriculum Project - Children's Drawings
- [ ] Boston State College Photograph Collection
- [ ] Descriptions of UMass Boston's Collections (Finding Aids)
- [ ] Doug Clifford Photograph Collection, 1988

Show 13 more

Narrow your search by:

- Location Names
- Decade
- Road Show Name

Browse suggested topics

- Boston Harbor Islands Mass. Memories Road Show (186)
- Halifax Mass. Memories Road Show (163)
- Lowell Mass. Memories Road Show (245)
- Milton Mass. Memories Road Show (333)
- Sharon Mass. Memories Road Show (175)

You've searched: Mass. Memories Road Show

All fields: snow

(15 results)

Sort by: Title | Display Options | Save to favorites

Display: 50 | Check/uncheck all

[ ] ""Farm in Snow"" - Circa 1835 farm manager's cottage at the Governor Gore estate

[ ] A ""Snow Day"" on Spectacle Island

[ ] Easter Sunday

[ ] Edward Rowe Snow with Friends of the Boston Harbor Islands Volunteers

[ ] Enjoying snow in our garden

[ ] Gramma and Grampa go wild with snowfolk

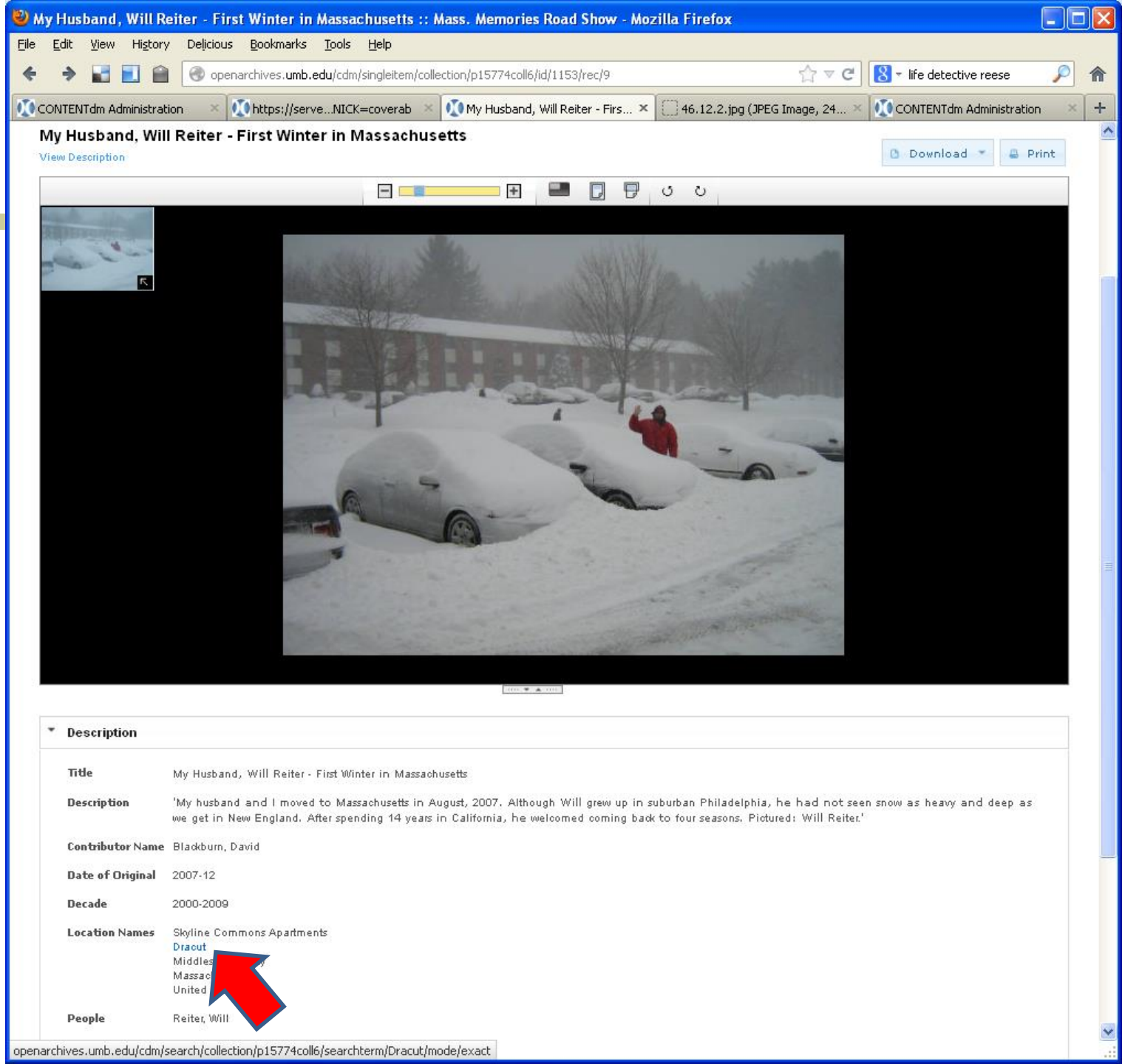
[ ] Hockey

[ ] My First Winter in Massachusetts

[ ] My Husband, Will Reiter - First Winter in Massachusetts

[ ] My Snow Day in Halifax

zotero


My Husband, Will Reiter - First Winter in Massachusetts :: Mass. Memories Road Show - Mozilla Firefox
File Edit View History Delicious Bookmarks Tools Help
openarchives.umb.edu/cdm/singleitem/collection/p15774coll6/id/1153/rec/9
life detective reese
CONTENTdm Administration
https://serve...NICK=coverab
My Husband, Will Reiter - Firs...
46.12.2.jpg (JPEG Image, 24...
CONTENTdm Administration
My Husband, Will Reiter - First Winter in Massachusetts
View Description
Download
Print
Description
Title
My Husband, Will Reiter - First Winter in Massachusetts
Description
'My husband and I moved to Massachusetts in August, 2007. Although Will grew up in suburban Philadelphia, he had not seen snow as heavy and deep as we get in New England. After spending 14 years in California, he welcomed coming back to four seasons. Pictured: Will Reiter.'
Contributor Name
Blackburn, David
Date of Original
2007-12
Decade
2000-2009
Location Names
Skyline Commons Apartments
Dracut
People
Reiter, Will
openarchives.umb.edu/cdm/search/collection/p15774coll6/searchterm/Dracut/mode/exact

Mass. Memories Road Show - Mozilla Firefox

File Edit View History Delicious Bookmarks Tools Help

openarchives.umb.edu/cdm/search/collection/p15774coll6/searchterm/Dracut/mode/exac

life detective reese

CONTENTdm Administra... | https://ser...ICK=coverab | Mass. Memories Road S... | 46.12.2.jpg (JPEG Imag... | CONTENTdm Administra...

Open Archives: Digital Collections

at the University of Massachusetts Boston

Open Archives Home | Browse This Collection | About the Road Show | Open Archives Blog | Log in | Favorites

Search within results Search Advanced Search

Add or remove other collections to your search:

- [x] Mass. Memories Road Show
- [ ] AFSC Vietnam Curriculum Project - Children's Drawings
- [ ] Boston State College Photograph Collection
- [ ] Descriptions of UMass Boston's Collections (Finding Aids)
- [ ] Doug Clifford Photograph Collection, 1988

Show 13 more

Narrow your search by:

- Location Names
- Road Show Name
- Decade

Browse suggested topics

- Boston Harbor Islands Mass. Memories Road Show (186)
- Halifax Mass. Memories Road Show (163)
- Lowell Mass. Memories Road Show (245)
- Milton Mass. Memories Road Show (333)
- Sharon Mass. Memories Road Show (175)
- Waltham Mass. Memories Road Show (569)

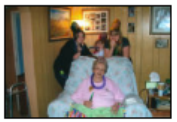
You've searched: Mass. Memories Road Show

All fields: Dracut ✖

Sort by: Title | Display Options | Save to favorites

Display: 20 | Check/uncheck all

Centerville Square, Bridge Street



Grammy's Last New Year's

My Dad and Me

My Father's Palonia Baking Company Truck

My First Winter in Massachusetts

My Husband, Will Reiter - First Winter in Massachusetts

My Mom and Me

Relatives at the Beach

What is Metadata

Standard definitions:

- "Data About Data"
- "Information About Information"

But what, really, does that MEAN?!

CDs store audio information

Information about information

Tchaikovsky composed this piece

This song is 1'30" long

I own every Big Mama Thornton's recording

That album has 17 tracks

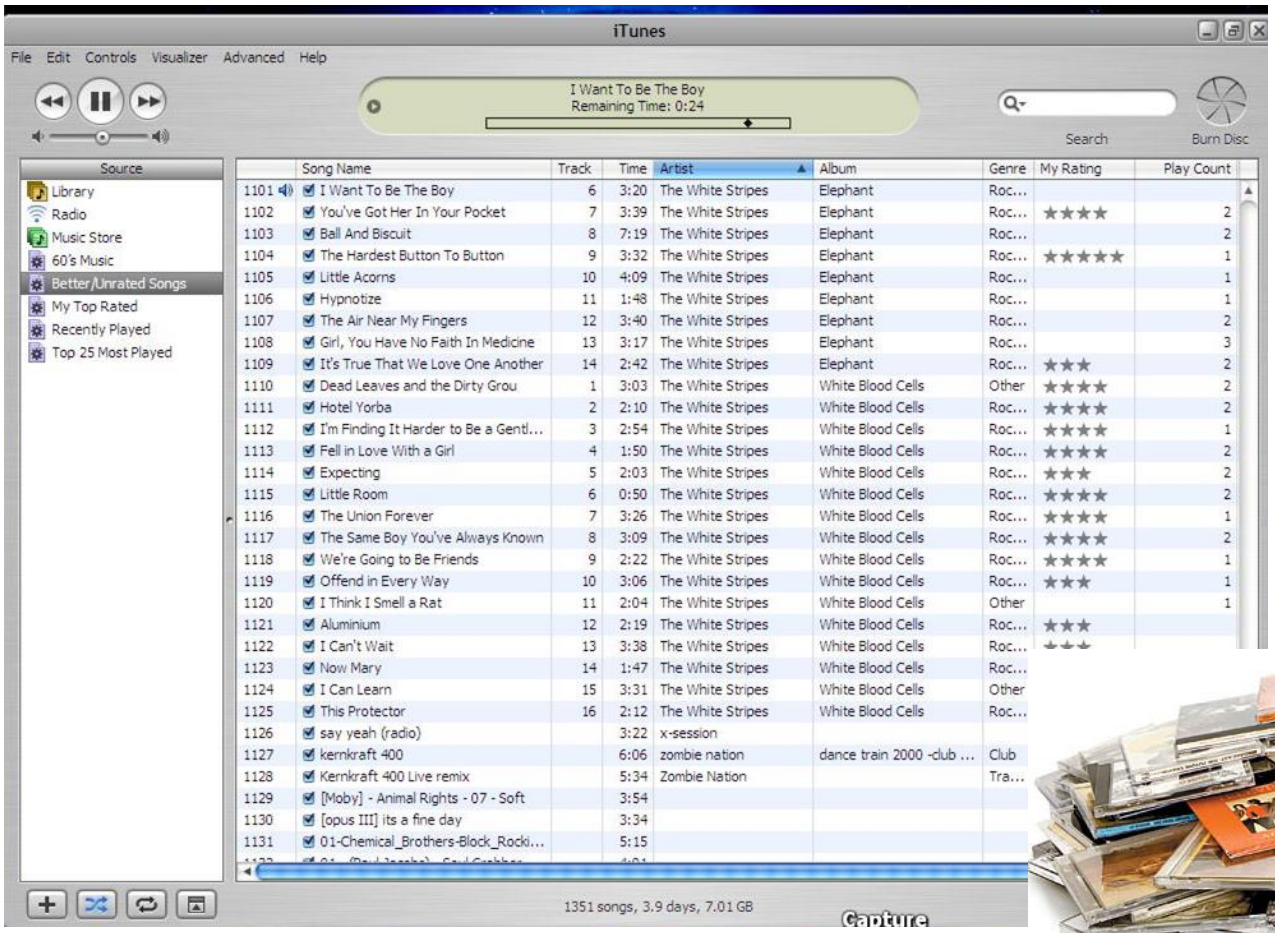
This CD is Rock 'n Roll

That CD is jazz

This is my favorite song

Organize this information in a structured way and it becomes "metadata"

iTunes stores Metadata…



…about audio resources

What is Metadata?

"Metadata is structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource."

***Understanding Metadata*, NISO Press 2004**

Benefits of Metadata

- Allows resources to be found, identified, described, annotated, collocated, shared and disambiguated

- "Describing a resource with metadata allows it to be understood by both humans and machines in ways that promote interoperability."

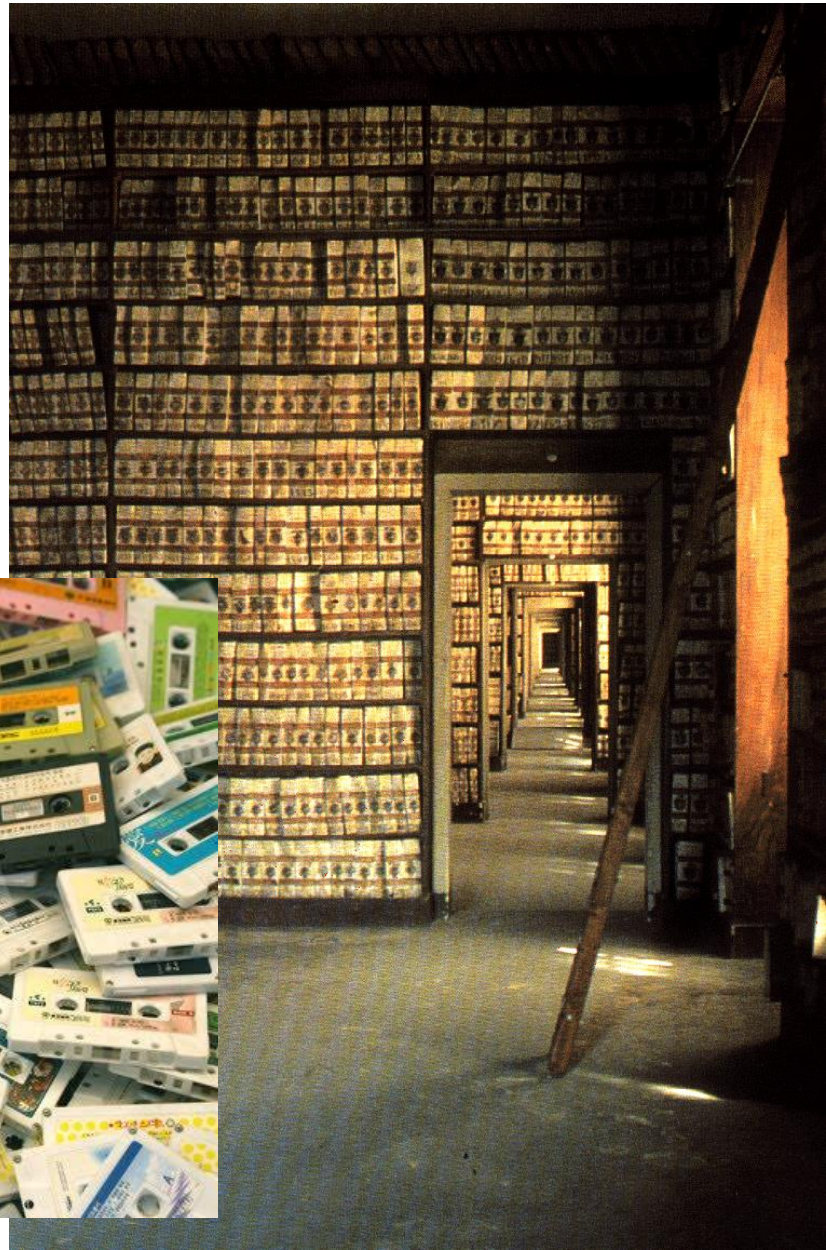
- Offers "ability to search resources seamlessly across a network"

Source: Understanding Metadata, NISO Press. 2004.
http://www.niso.org/publications/press/

Archives are chock full of resources in need of metadata

Visual Information

Textual Information



Audio Information

Semantic information on the web can be expressed in three different ways:

1. Inferred
2. Free-form tagging
3. Structured data

These go in order from the **consumer** of data doing most of the work, to the **producer** of data doing most of the work.

Source: “A brief introduction to the Semantic Web”
Yaron Koren
National Library of Israel
August 11, 2011
http://wikiworks.com/nli-seminar/intro-to-semantic-web.pdf

Semantic information on the web can be expressed in three different ways:

1. Inferred
2. Free-form tagging
3. Structured data

These go in order from the **consumer** of data doing most of the work, to the **producer** of data doing most of the work.

Source: “A brief introduction to the Semantic Web”
Yaron Koren
National Library of Israel
August 11, 2011
http://wikiworks.com/nli-seminar/intro-to-semantic-web.pdf

Three Case Studies

1. *Medici Archive Project: Documentary Sources for the Arts and Humanities 1537 – 1743*

   (Relational database)

2. *The History of the Accademia di San Luca, c. 1590 – 1635: Documents from the Archivio di Stato di Roma*

   (Text markup – TEI)

3. Healey Library's OpenArchives (login server15774.contentdm.oclc.org)

   (Dublin Core Schema in a proprietary data system)

The History of the Accademia di San Luca, c. 1590–1635: Documents from the Archivio di Stato di Roma

www.nga.gov/casva/accademia/index.shtm

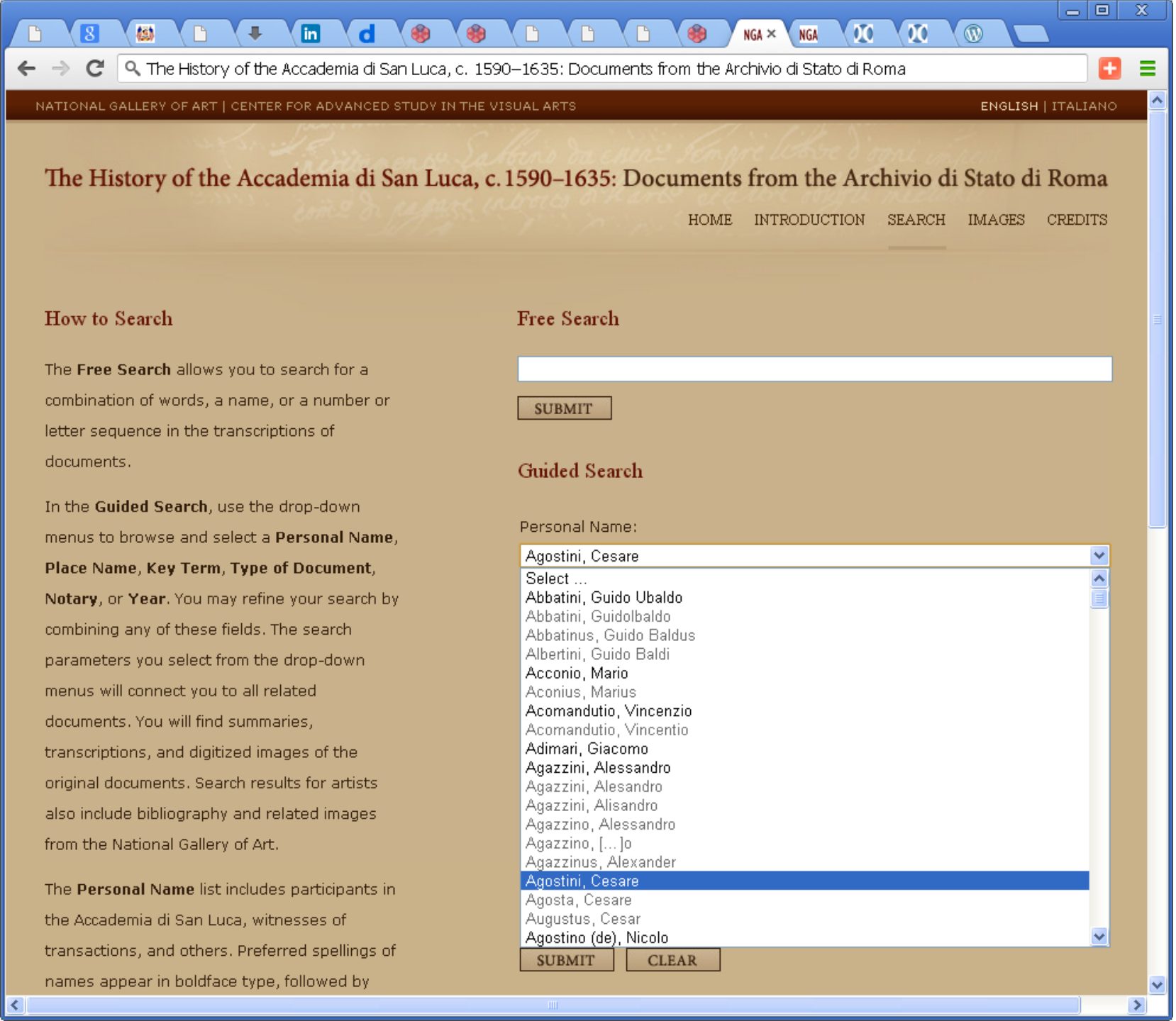
NATIONAL GALLERY OF ART | CENTER FOR ADVANCED STUDY IN THE VISUAL ARTS

ENGLISH | ITALIANO

The History of the Accademia di San Luca, c. 1590–1635: Documents from the Archivio di Stato di Roma

A project of the National Gallery of Art, Center for Advanced Study in the Visual Arts, in association with the Archivio di Stato di Roma and the Accademia di San Luca

INTRODUCTION SEARCH IMAGES CREDITS


The History of the Accademia di San Luca, c. 1590–1635: Documents from the Archivio di Stato di Roma
NATIONAL GALLERY OF ART | CENTER FOR ADVANCED STUDY IN THE VISUAL ARTS
ENGLISH | ITALIANO
The History of the Accademia di San Luca, c.1590–1635: Documents from the Archivio di Stato di Roma
HOME INTRODUCTION SEARCH IMAGES CREDITS
How to Search
The Free Search allows you to search for a combination of words, a name, or a number or letter sequence in the transcriptions of documents.
In the Guided Search, use the drop-down menus to browse and select a Personal Name, Place Name, Key Term, Type of Document, Notary, or Year. You may refine your search by combining any of these fields. The search parameters you select from the drop-down menus will connect you to all related documents. You will find summaries, transcriptions, and digitized images of the original documents. Search results for artists also include bibliography and related images from the National Gallery of Art.
The Personal Name list includes participants in the Accademia di San Luca, witnesses of transactions, and others. Preferred spellings of names appear in boldface type, followed by
Free Search
SUBMIT
Guided Search
Personal Name:
Agostini, Cesare
Select ...
Abbatini, Guido Ubaldo
Abbatini, Guidolbaldo
Abbatinus, Guido Baldus
Albertini, Guido Baldi
Acconio, Mario
Aconius, Marius
Acomandutio, Vincenzio
Acomandutio, Vincentio
Adimari, Giacomo
Agazzini, Alessandro
Agazzini, Alesandro
Agazzini, Alisandro
Agazzino, Alessandro
Agazzino, [...]o
Agazzinus, Alexander
Agostini, Cesare
Agosta, Cesare
Augustus, Cesar
Agostino (de), Nicolo
SUBMIT
CLEAR

www.nga.gov/casva/accademia/html/eng/ASRTNCUff1516110925.shtm#

Document Transcription

[fol. 210r]

Congregatio

Die 25 Mensis septembris 1611

IN Mei etc. Constituti personaliter, et Congregati pro secunda ut moris est in loco infrascritto omnes infrascritti Videlicet

Ill.res DD. Cherubinus Albertus Princeps

Gaspar Cellius Consiliarius

Durantes Albertus

Horatius Borgiannus

Io'es Bap'ta de Novaria

Carolus Oldradus

Antivedutus Gramatica

Hieronimus Masseus

Marcellus Provenzanus, et

Hironimus Nanni

omnes de Congregatione Academica Pictorum, et sculptorum

^in Ecclesia sancti Lucae^

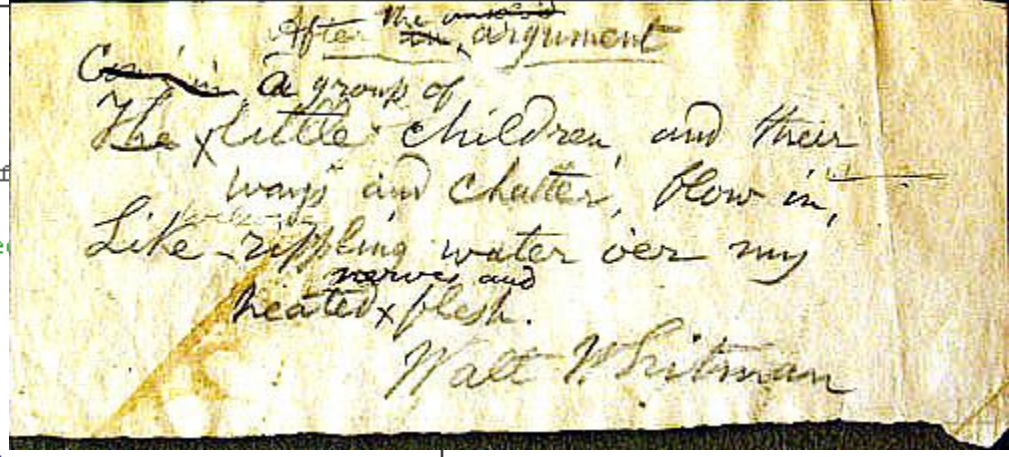
asserentes se esse numerum sufficientem ad infrascritta et alia dictae Academiae negotia pertractanda, et nihilominus pro absentibus quatenus opus sit de rato etc. promittentes ita quod etc. alias etc. sponte etc. omni meliori modo etc.

In primis decretarunt quod fieret memorialis Ill.mo D. Io'i

ASR, TNC, uff. 15, 1611, pt. III, vol. 52, fol. 210r

iipimage.sourceforge.net

```
<body>
   <pb xml:id="leaf01r" type="recto"/>
   <lg type="poem">
      <head rend="underline" type="main-authorial">Af
         <del type="overstrike" seq="1">an</del>
         <add place="supralinear" type="insertion" se
         type="overstrike">unsolv'd</del> </add>
      </subst> argument</head>
      <l>
         <seg>
            <del type="overstrike">The</del>
            <add place="supralinear" type="insertion">
               <del type="overstrike">Coming in,</del>
               <subst>
                  <del type="overwrite" seq="1">a</del>
                  <add place="over" type="overwrite" seq="2">A </add>
               </subst>
            group of </add>
         little children, and their</seg>
         <seg>ways and chatter, flow
            <add place="inline" type="unmarked">in, </add>
            <del type="overstrike">
               <add place="supralinear" type="unmarked">upon me</add>
            </del>
         </seg>
      </l>
      <l>
         <seg>Like <add place="supralinear" type="insertion">welcome </add>
         rippling water o'er my </seg>
         <seg>heated <add place="supralinear" type="insertion">nerves and
         </add> flesh.</seg>
      </l>
      <closer>
         <signed>Walt Whitman</signed>
      </closer>
   </lg>
</body>
```



Accademia di San Luca

The transcriptions are encoded in Extensible Markup Language (XML) following the guidelines of the Text-Encoding Initiative (TEI) for digital encoding of literary and linguistic texts (http://www.tei-c.org).

The transcriptions may be searched by personal name (under all known variant spellings), place name, key term, document type, notary name, and year. The documents included are not only newly available to students and scholars of early modern Italy, but are also accessible in a way that both promotes their use and allows for the identification of additional archival material.

Implementation description: We used TEI tagsets specific to manuscript material, personal names, and place names. Our programmer implemented Perl script to batch process the documents from MSWord to XHTML to XML to HTML. We tagged the XML files in TEI according to the standards of the Text Encoding Initiative (TEI) P5 Guidelines (November 2007).

Source: http://www.tei-c.org/Activities/Projects/hi01.xml

Organizing Metadata

STRUCTURE

- Which fields
- What they're called
- Whether they're required or not
- Best to base the structure on a standardized schema

FIELD CONTENTS

- Guidelines
  ("in our project we enter names as "Last, First MI")
- Controlled vocabularies
  ("in our project we use LOCSH for subject terms and TGN for Places")

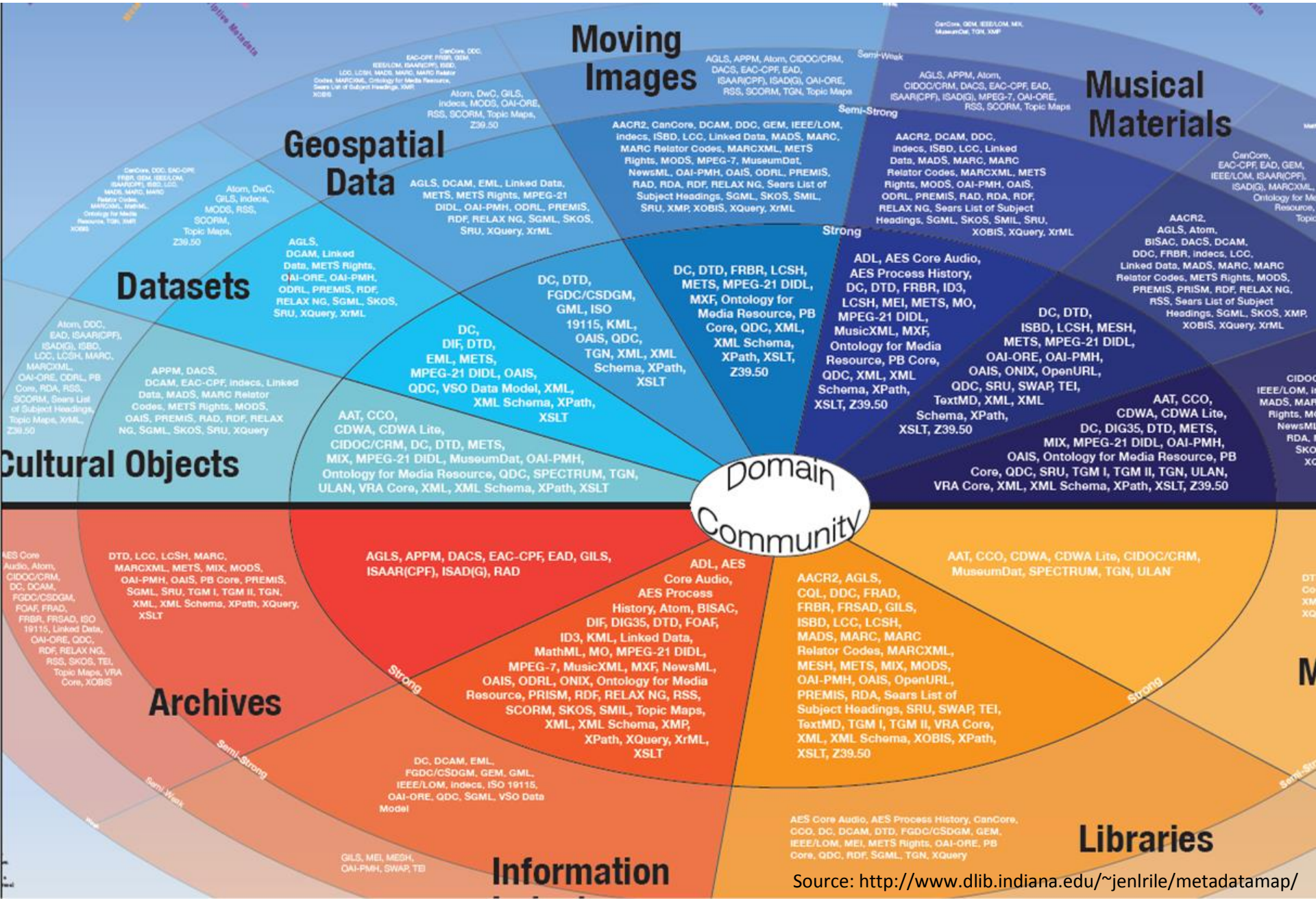
Widely used schemas

- MARC 91% *Used by librarians for library collections*
- EAD 84% *Used by archivists for finding aids*
- Dublin Core 78% *Used by cultural heritage practitioners for items*

Choose a schema (set of fields)

- A Metadata Schema is a *"set of metadata elements designed for a specific purpose, e.g. describing a particular type of info resource for a specific community of practice."*
- Many possible schemas!
  - Always best to adopt a schema already in use by many others
  - But you can also make up your own… if you must…
- **Choose** which to employ based on:
  1. **your resource types**
  2. **your audience**
  3. **your community of practice**

Seeing Standards: A Visualization of the Metadata Universe



Source: http://www.dlib.indiana.edu/~jenlrile/metadatamap/

Simple Dublin Core (metadata schema)

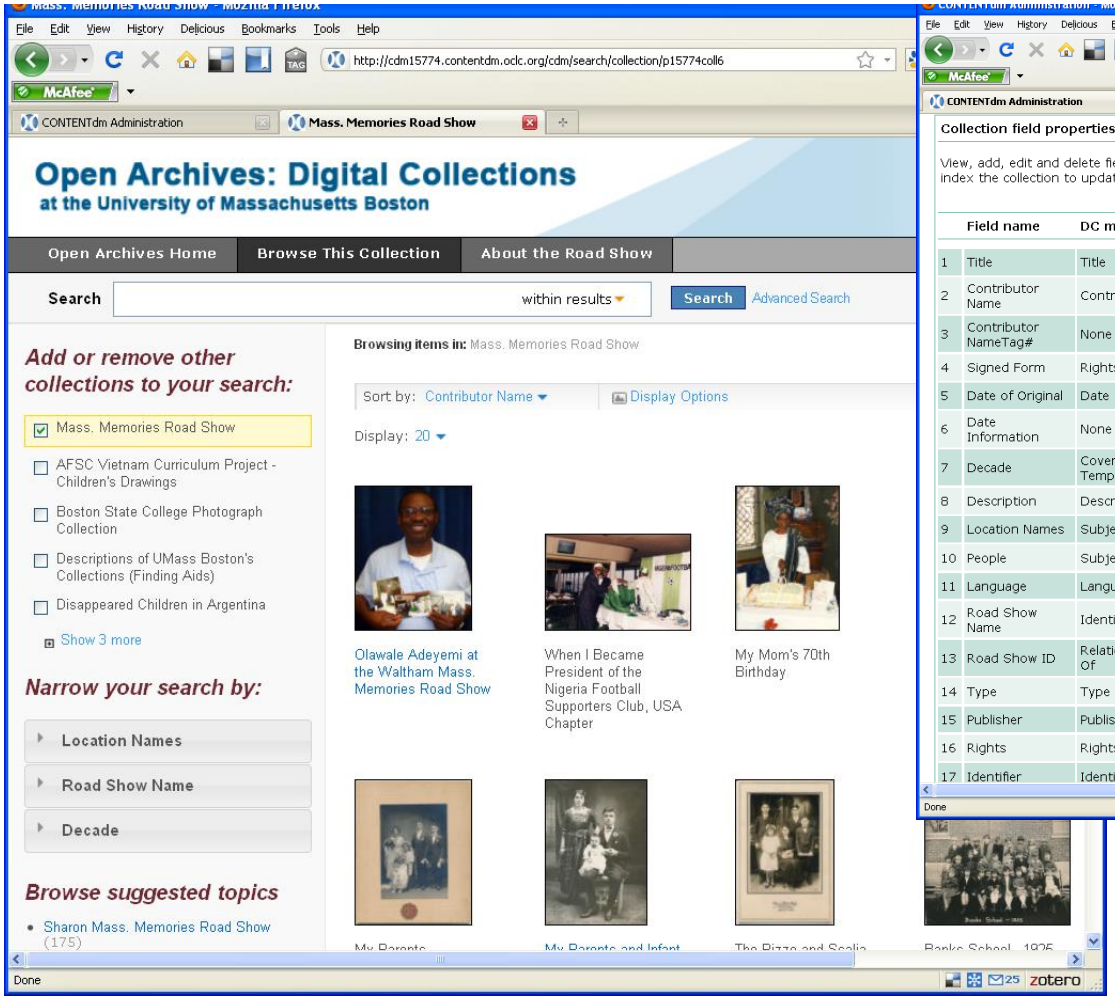
15 metadata elements or fields:

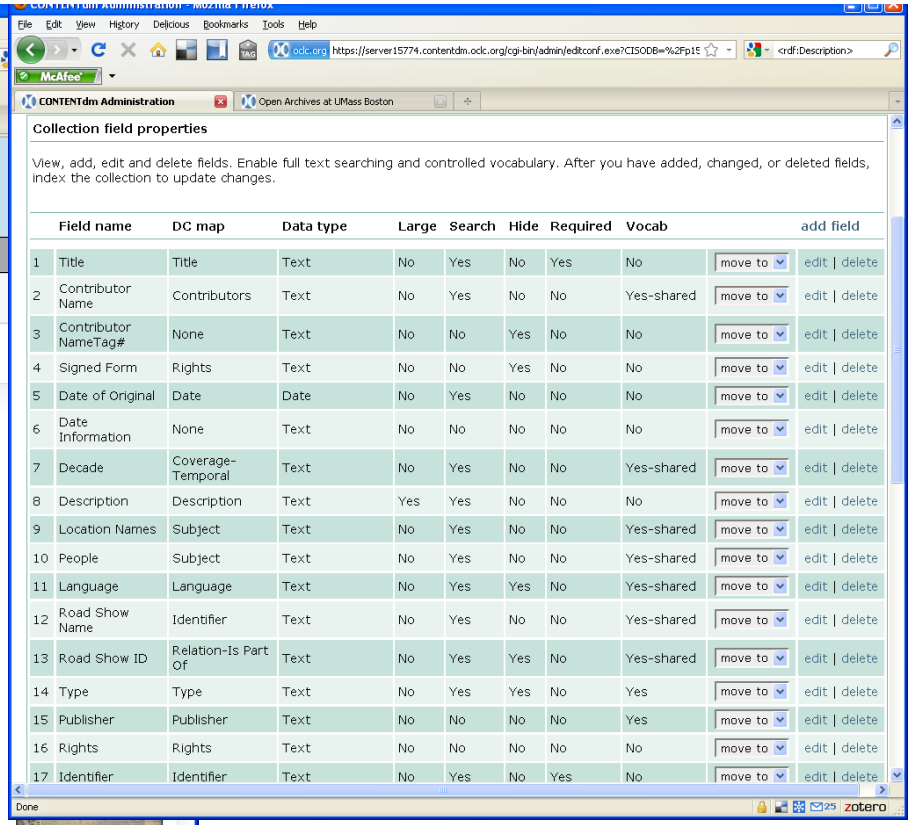
1. Title
2. Creator
3. Subject
4. Description
5. Publisher
6. Contributor
7. Date
8. Type
9. Format
10. Identifier
11. Source
12. Language
13. Relation
14. Coverage
15. Rights

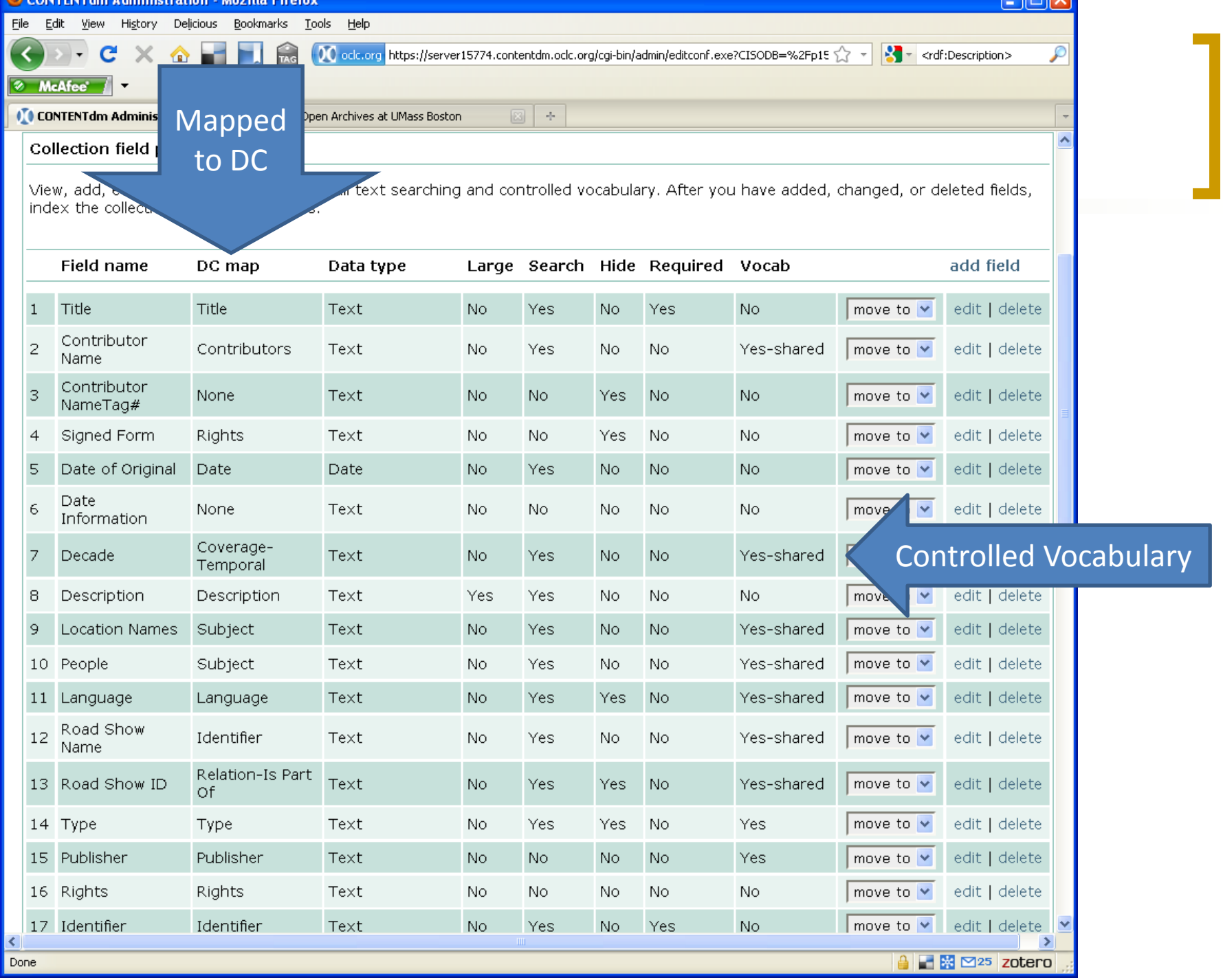
←Structural Elements

But what controls the content?

Metadata in Action at UMB






Mapped to DC
Controlled Vocabulary
File Edit View History Delicious Bookmarks Tools Help
https://server15774.contentdm.oclc.org/cgi-bin/admin/editconf.exe?CISODB=%2Fp15
Open Archives at UMass Boston
Collection field
text searching and controlled vocabulary. After you have added, changed, or deleted fields,
Field name DC map Data type Large Search Hide Required Vocab add field
1 Title Title Text No Yes No Yes No move to edit | delete
2 Contributor Name Contributors Text No Yes No No Yes-shared move to edit | delete
3 Contributor NameTag# None Text No No Yes No No move to edit | delete
4 Signed Form Rights Text No No Yes No No move to edit | delete
5 Date of Original Date Date No Yes No No No move to edit | delete
6 Date Information None Text No No No No No edit | delete
7 Decade Coverage-Temporal Text No Yes No No Yes-shared
8 Description Description Text Yes Yes No No No edit | delete
9 Location Names Subject Text No Yes No No Yes-shared move to edit | delete
10 People Subject Text No Yes No No Yes-shared move to edit | delete
11 Language Language Text No Yes Yes No Yes-shared move to edit | delete
12 Road Show Name Identifier Text No Yes No No Yes-shared move to edit | delete
13 Road Show ID Relation-Is Part Of Text No Yes Yes No Yes-shared move to edit | delete
14 Type Type Text No Yes Yes No Yes move to edit | delete
15 Publisher Publisher Text No No No No Yes move to edit | delete
16 Rights Rights Text No No No No No move to edit | delete
17 Identifier Identifier Text No Yes No Yes No move to edit | delete
Done
zotero

The Case for Content Control

LOCATION:

- Massachusetts?
- MA?
- Mass?

PEOPLE:

- Mary Smith?
- Mary C. Smith?
- Smith, Mary?

DECADE:

- 1950s?
- Early 1950s?
- 1950-1959?

FORMAT:

- Sound?
- Audio?
- MP3?

The Case for Content Control

Chekhov, Anton Pavlovich, 1860-1904

Čechov, Anton P. 1860-1904

Čehov, Anton Pavlovič (1860-1904).

Čechov, Anton Pavlovič [forme avant 2007]

Chejov, Anton Pavlovich 1860-1904

Čechov, Anton Pavlovič, 1860-1904

Tjechov, Anton, 1860-1904

Čechov, Anton Pavlovič

Чехов, Антон Павлович, 1860-1904

צ'כוב, אנטון פבלוביץ', 1860-1904

Tchekhoff, Anton, 1860-1904

تشيكوف، انطون، 1860-1904 م

Tchekhov, Anton Pavlovitch, 1860-1904

Source: [LOC Names Authority](/)

Widely Used Content Vocabularies:

- AAT - *Art & Architecture Thesaurus – Getty Institute*
- LCSH - *Library of Congress Subject Headings*
- LC Name Authority File – *Library of Congress Name Auth.*
- MeSH – *Medical Subject Headings*
- TGM - *Thesaurus of Graphic Materials (LOC)*
- TGN - *Thesaurus of Geographic Names – Getty Institute)*
- ULAN -*Union List of Artists' Names – Getty)*

Project Profile – which fields should contain controlled content?

STRUCTURAL ELEMENTS:

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

POSSIBLE CONTROLLED CONTENT:

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

“Decade” element: Controlled Vocabulary – pick one!

1700-1709
1710-1719
1720-1729
1730-1739
1740-1749
1750-1759
1760-1769
1770-1779
1780-1789
1790-1799
1800-1809
1810-1819
1820-1829
1830-1839
1840-1849
1850-1859
1860-1869
1870-1879
1880-1889
1890-1899
1900-1909
1910-1919
1920-1929
1930-1939
1940-1949
1950-1959
1960-1969
1970-1979
1980-1989
1990-1999
2000-2009
2010-2019
2020-2029
2030-2039
2040-2049
2050-2059

Mass. Memories Road Show - Mozilla Firefox
File Edit View History Delicious Bookmarks Tools Help
http://cdm15774.contentdm.oclc.org/cdm/search/collection/p15774coll6
<rdf:Description>
McAfee
CONTENTdm Administration
Mass. Memories Road Show

Open Archives: Digital Collections
at the University of Massachusetts Boston

Add or remove other collections to your search:

- [x] Mass. Memories Road Show
- [ ] AFSC Vietnam Curriculum Project - Children's Drawings
- [ ] Boston State College Photograph Collection
- [ ] Descriptions of UMass Boston's Collections (Finding Aids)
- [ ] Disappeared Children in Argentina

Show 3 more

Narrow your search by:

- Location Names
- Road Show Name
- Decade

Browse suggested topics

- Sharon Mass. Memories Road Show (175)

Browsing items in: Mass. Memories Road Show

Sort by: Contributor Name | Display Options | 1

Display: 20

Olawale Adeyemi at the Waltham Mass. Memories Road Show

When I Became President of the Nigeria Football Supporters Club, USA Chapter

My Mom's 70th Birthday

Friend's Party

My Parents

My Parents and Infant

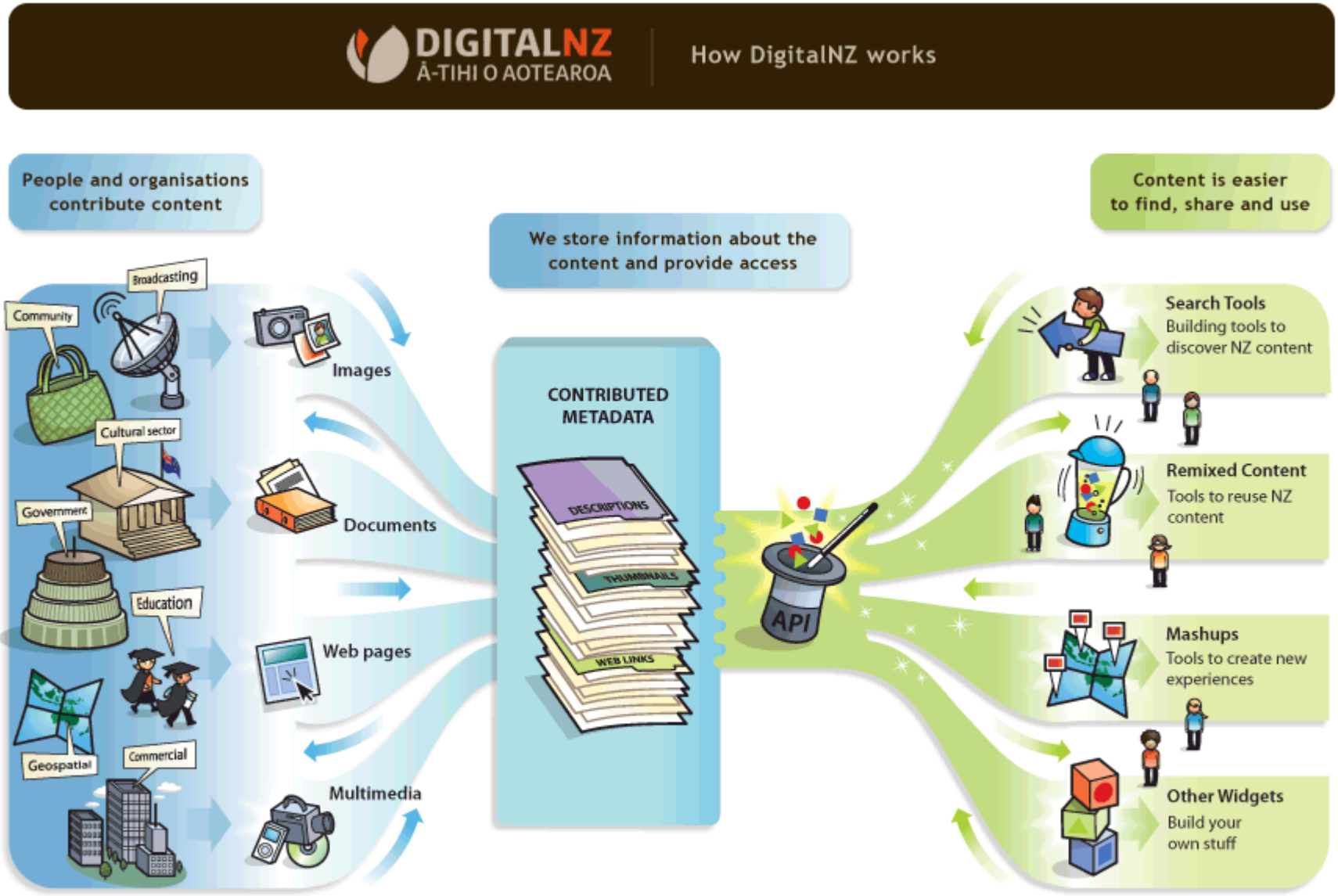
The Pizza and Scalia

Banks School - 1925

Done
25 zotero

Digital Repositories

- There are decisions to be made before creating any digital resources project
  - Which schema to use (fields)?
  - Which content standards to apply?
- Those decisions are never clear-cut
  - Think them through, commit to standards and move on
- **Standards** (schemas and content) allow for sharing, finding, interoperability



Source: https://natlib.govt.nz/about-us/open-data/digitalnz-api

Some Useful Resources

- *Understanding Metadata*, NISO Press. 2004.
- *Metadata*. SPEC Kit #298: Executive Summary. 2007.
- *A Practical Application of Dublin Core: Worthington Memory*. 2002. Allen and Davis.
- This presentation: Metadata Lecture_Riley
https://docs.google.com/leaf?id=0B68-DLZjnv5INmI2MzU2NDYtZjA3NC00YzE2LWJlZjYtOWZjMDAwOTk5ZGE3&hl=en&authkey=CJDIgfgK