Virtual Collections

Challenges in Harvesting and Transforming Metadata from Harvard Catalogs for Topical Collections

Michael Vandermillen, Randy Stern Harvard University

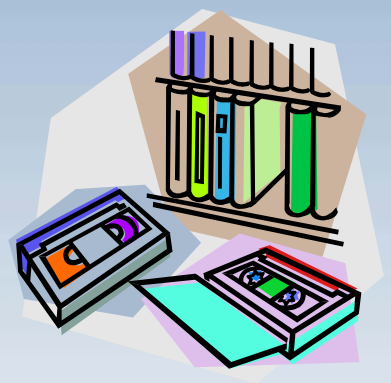
Nov 9, 2006

Overview

- What is the Virtual Collections Service?
- Creating A Virtual Collection
- Architecture
- Implementation Challenges

What is the Virtual Collections Service?

Virtual Collections Is …



A library service that allows Harvard curators, librarians and archivists to create a web-based collection view of thematically related resources which exist in a number of systems or within a single system.

Features of VC

- Scheduled harvesting of metadata from HOLLIS (OPAC), VIA (Visual Information Access) and Harvard Geospatial Library

- Web-based search, browse, and display interface; short and full record views
- Search results include thumbnails and links to digital objects

- Flexible configuration
  - “Out of the box” stand alone public interface
  - “Embedded” option for integrating into web sites, such as Harvard Open Collections sites

Virtual Collections is not…

- A new catalog – records must exist in another catalog
- A repository for digital objects
- A platform for short term, non-”library-like” projects

A stand alone collection

Latin American Pamphlets

Virtual Collections - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Search Favorites Media

Address http://fiona.hul.harvard.edu:11080/vc/deliver/home?_collection=LAP Go Links

Latin American Pamphlet Digital Collection

Latin American Pamphlets Virtual Collection

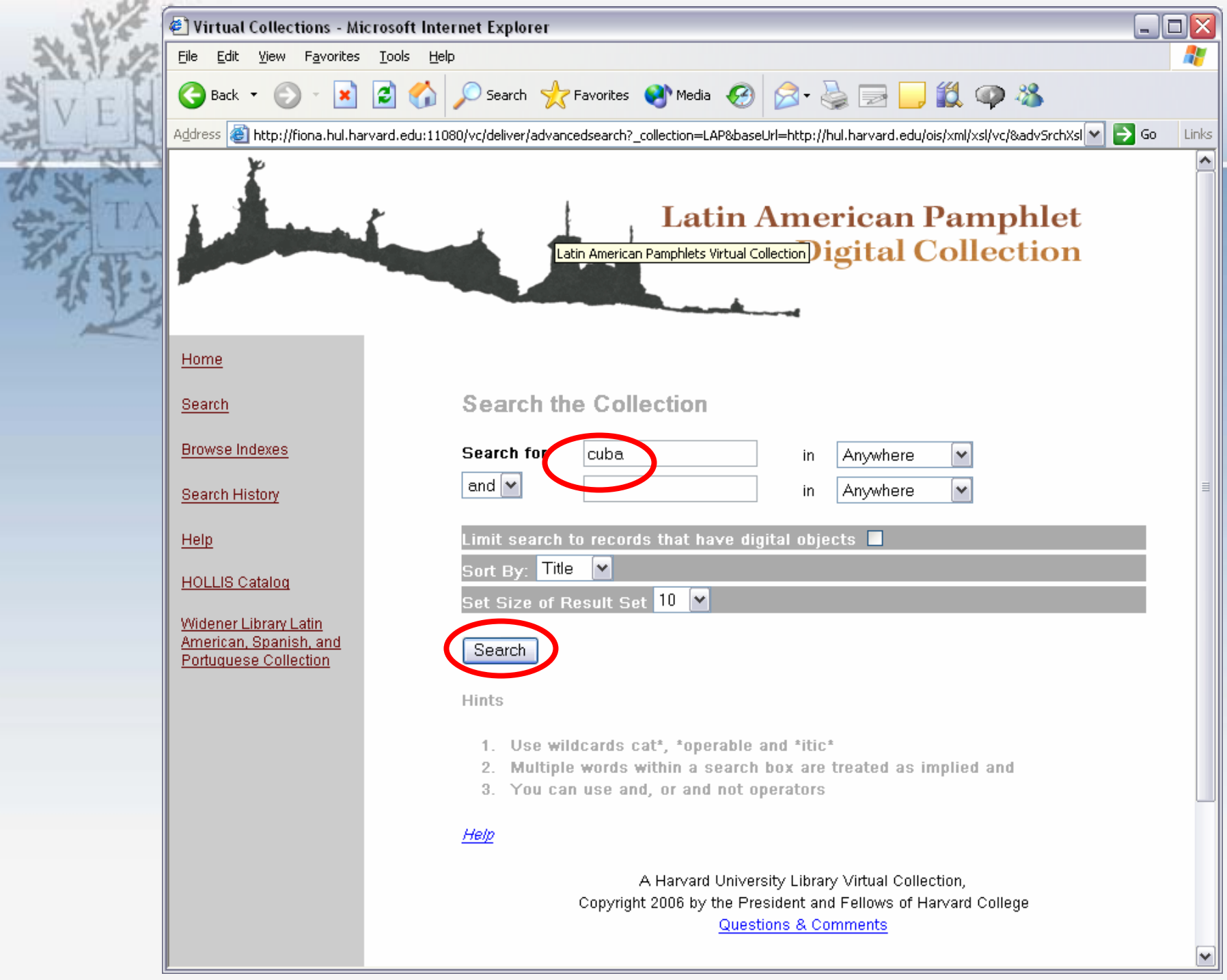
- Home
- Search
- Browse Indexes
- Search History
- Help
- HOLLIS Catalog
- Widener Library Latin American, Spanish, and Portuguese Collection

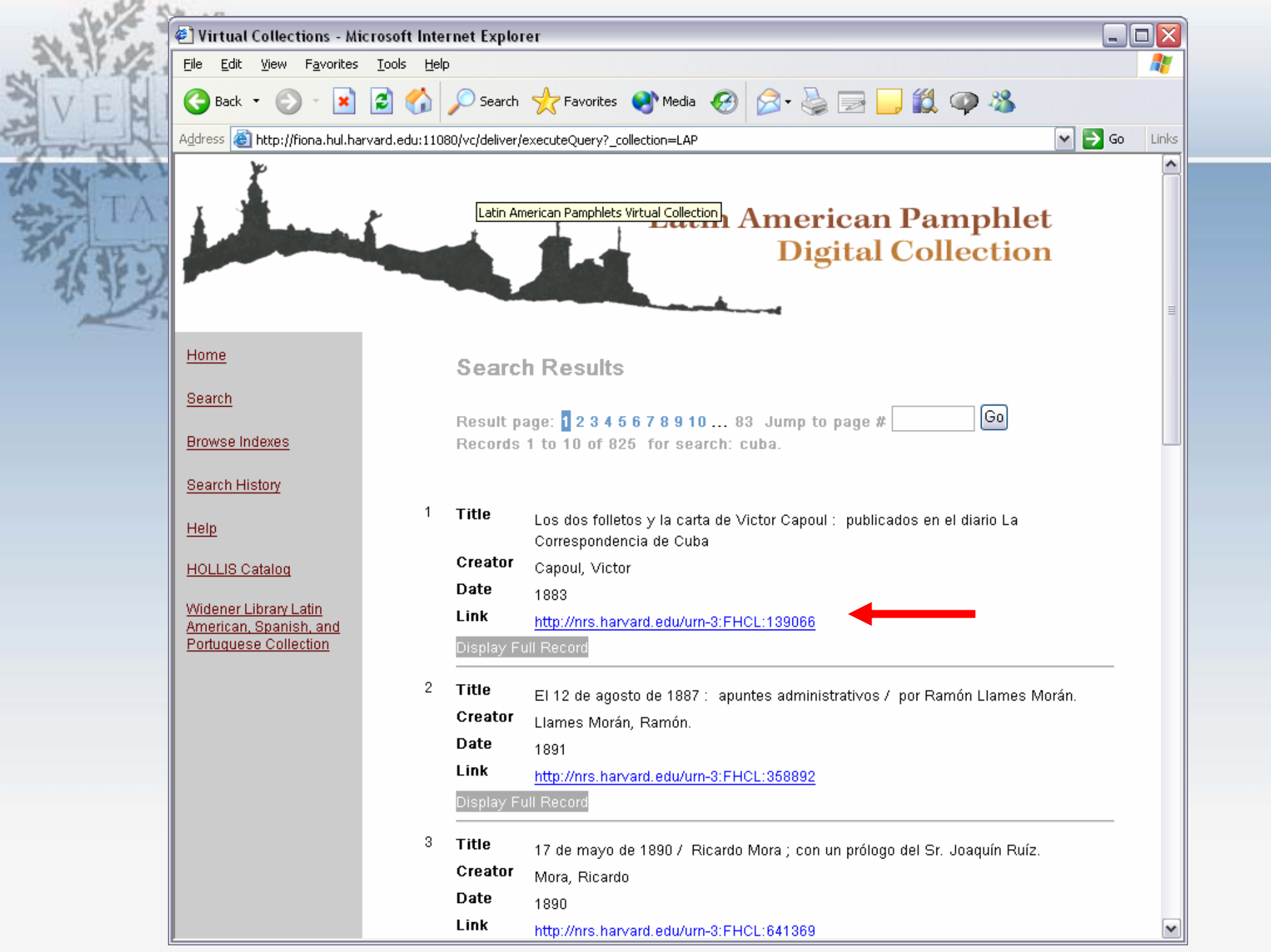
Harvard's Widener Library is the repository of many scarce and unique Latin American pamphlets published during the 19th and the early 20th centuries. One of the few institutions to have consistently collected Latin American pamphlets, Harvard has benefited from collections formed by Luis Montt (Chile), Nicolás Acosta (Bolivia), Manuel Segundo Sánchez (Venezuela), José Augusto Escoto (Cuba), Blas Garay (Paraguay), Charles Sumner, John B. Stetson and others. Chile, Cuba, Bolivia and Mexico are the countries most heavily represented in this collection.

These pamphlets are valuable primary resources for students and researchers working on Latin American history. They document the emergence of the Latin American colonies as independent states, and illuminate many aspects of their populations' social and cultural life. Many pamphlets are devoted to boundary disputes, territorial expansion, the description of unexplored territories and the relationship between Church and State.

This collection of more than 5,000 titles was largely uncataloged and virtually inaccessible to researchers until a cataloging and digitization project was initiated in 2002. The Latin American Pamphlet Digital Collection contains catalog records with links to page images of the digitized pamphlets. As additional pamphlets are cataloged and digitized they will be added to this Collection. For more complete bibliographic data, please see the HOLLIS catalog.

Questions & Comments


Virtual Collections - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Address http://fiona.hul.harvard.edu:11080/vc/deliver/advancedsearch?_collection=LAP&baseUrl=http://hul.harvard.edu/ois/xml/xsl/vc/&advSrchXsl
Go Links
Latin American Pamphlets Virtual Collection
Latin American Pamphlet
Digital Collection
Home
Search
Browse Indexes
Search History
Help
HOLLIS Catalog
Widener Library Latin American, Spanish, and Portuguese Collection
Search the Collection
Search for
cuba
in
Anywhere
and
in
Anywhere
Limit search to records that have digital objects
Sort By:
Title
Set Size of Result Set
10
Search
Hints
1. Use wildcards cat*, *operable and *itic*
2. Multiple words within a search box are treated as implied and
3. You can use and, or and not operators
Help
A Harvard University Library Virtual Collection,
Copyright 2006 by the President and Fellows of Harvard College
Questions & Comments


Virtual Collections - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Address http://fiona.hul.harvard.edu:11080/vc/deliver/executeQuery?_collection=LAP
Go Links
Latin American Pamphlets Virtual Collection
Latin American Pamphlet
Digital Collection
Home
Search
Browse Indexes
Search History
Help
HOLLIS Catalog
Widener Library Latin American, Spanish, and Portuguese Collection
Search Results
Result page: 1 2 3 4 5 6 7 8 9 10 ... 83 Jump to page # Go
Records 1 to 10 of 825 for search: cuba.
1 Title Los dos folletos y la carta de Victor Capoul : publicados en el diario La Correspondencia de Cuba
Creator Capoul, Victor
Date 1883
Link http://nrs.harvard.edu/urn-3:FHCL:139066
Display Full Record
2 Title El 12 de agosto de 1887 : apuntes administrativos / por Ramón Llames Morán.
Creator Llames Morán, Ramón.
Date 1891
Link http://nrs.harvard.edu/urn-3:FHCL:358892
Display Full Record
3 Title 17 de mayo de 1890 / Ricardo Mora ; con un prólogo del Sr. Joaquín Ruíz.
Creator Mora, Ricardo
Date 1890
Link http://nrs.harvard.edu/urn-3:FHCL:641369

PDS Los dos folletos y la carta de Victor Capoul. Habana: Imp. de Obra La Correspondencia de Cu - Microsoft Interne...

File Edit View Favorites Tools Help

Back Search Favorites Media

Address http://pds.harvard.edu:8080/pdx/servlet/pds?id=2588795&n=1&s=6&medium.x=5&medium.y=5

Go Links

Harvard University Library
Page Delivery Service

Los dos folletos y la carta de Victor Capoul. Habana: Imp. de Obra La Correspondencia de Cuba, 1883.

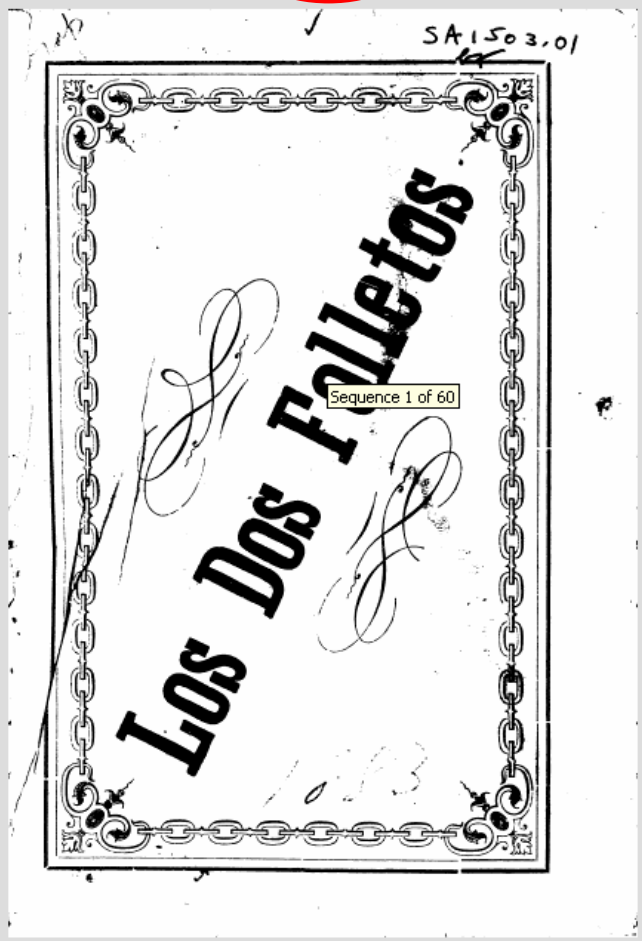
Search | View Text | Printable Version | Related Links | Help | Copyright

Page: [1] Go to Sequence: 1 Go to

Image Size:

Expand All | Collapse All

Los dos folletos y la carta de Vic



LOS DOS FOLLETOS

Y LA CARTA DE

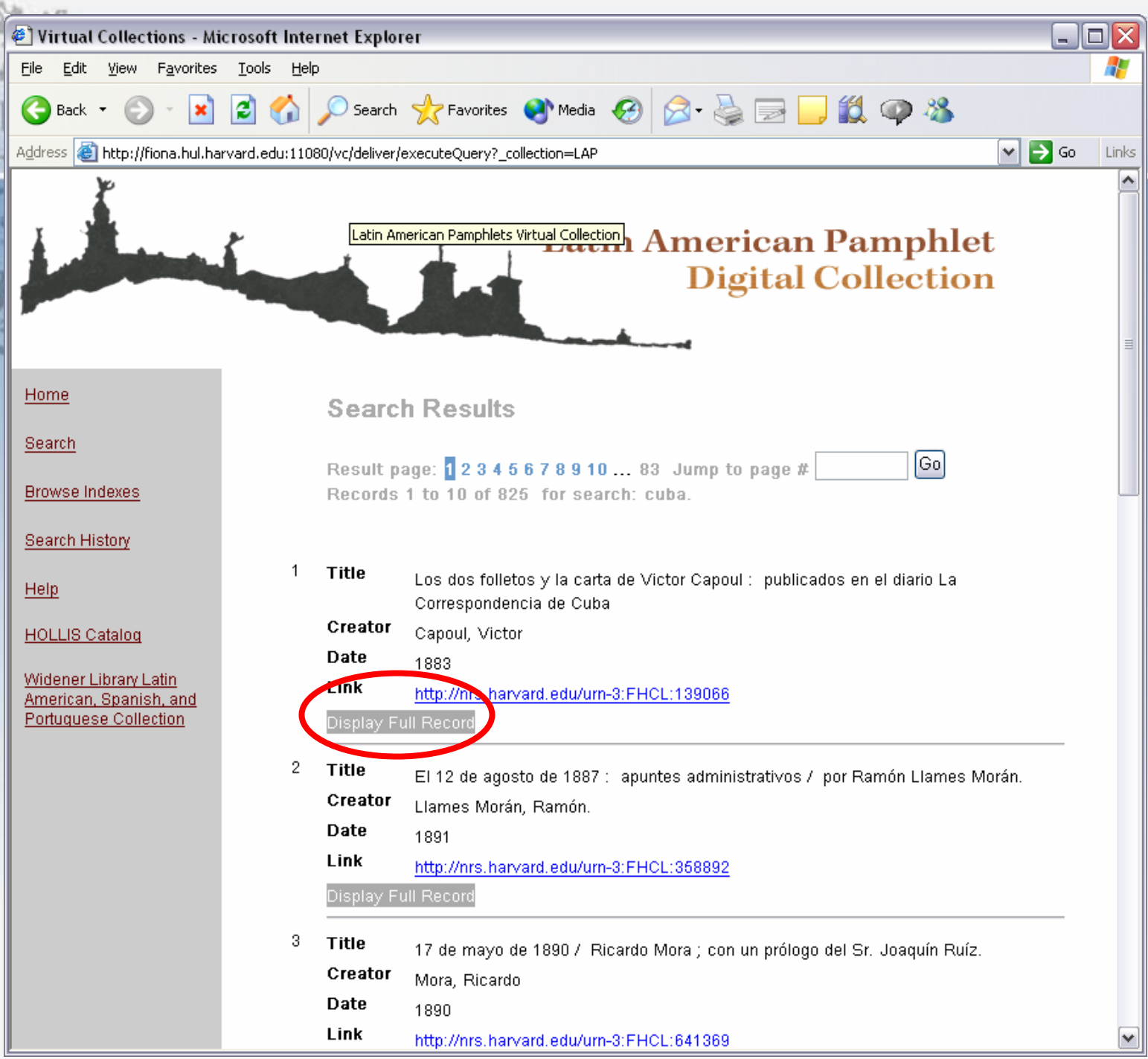
VICTOR CAPOUL,

PUBLICADOS EN EL DIARIO

LA CORRESPONDENCIA DE CUBA.

HABANA:
IMP. DE OBRA LA CORRESPONDENCIA DE CUBA,
CALLE DE LA OBRAPÍA NUMERO 24.

1883.


Virtual Collections - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Address http://fiona.hul.harvard.edu:11080/vc/deliver/executeQuery?_collection=LAP
Go Links
Latin American Pamphlets Virtual Collection
Latin American Pamphlet
Digital Collection
Home
Search
Browse Indexes
Search History
Help
HOLLIS Catalog
Widener Library Latin American, Spanish, and Portuguese Collection
Search Results
Result page: 1 2 3 4 5 6 7 8 9 10 ... 83 Jump to page # Go
Records 1 to 10 of 825 for search: cuba.
1 Title Los dos folletos y la carta de Victor Capoul : publicados en el diario La Correspondencia de Cuba
Creator Capoul, Victor
Date 1883
Link http://nrs.harvard.edu/urn-3:FHCL:139066
Display Full Record
2 Title El 12 de agosto de 1887 : apuntes administrativos / por Ramón Llames Morán.
Creator Llames Morán, Ramón.
Date 1891
Link http://nrs.harvard.edu/urn-3:FHCL:358892
Display Full Record
3 Title 17 de mayo de 1890 / Ricardo Mora ; con un prólogo del Sr. Joaquín Ruíz.
Creator Mora, Ricardo
Date 1890
Link http://nrs.harvard.edu/urn-3:FHCL:641369

Virtual Collections - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Search Favorites Media

Address http://fiona.hul.harvard.edu:11080/vc/deliver/vcFullRecord?collection=LAP&inoId=2776&recordNumber=18&method=view Go Links

Latin American Pamphlet Digital Collection

Home

Search

Browse Indexes

Search History

Help

HOLLIS Catalog

Widener Library Latin American, Spanish, and Portuguese Collection

Full Record Display

Record 1 of 825 for search: **cuba** . Jump to record # Go

Next Record Back To Results Display

Title	Los dos folletos y la carta de Victor Capoul : publicados en el diario La Correspondencia de Cuba
Name/Creator	Capoul, Victor, 1839-1924
HOLLIS ID	009250537
Digital Object	http://nrs.harvard.edu/urn-3:FHCL:139066
Location	Widener HD
Place of Origin	Habana
Publisher	Imp. de Obra La Correspondencia de Cuba
Creation Date	1883
Language	Spanish
Description	60 p. ; 17 cm.
Form/Genre	text
Subject	Serrano y Domínguez, Francisco , duque de la Torre , 1810-1885 ; Havana (Cuba) -- Description and travel
Note	"Los duques de la Torre y el casamiento de su hijo : extracto del folleto del señor Carreras": p. [5]-45. "Carta de Capoul á 'El Figaro'": p. [47]-60.
Other Title	Duques de la Torre y el casamiento de su hijo
Other Title	Latin American pamphlet digital project at Harvard University

FULL CATALOG - Full View of Record - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Search Favorites Media

Address http://lms01.harvard.edu/F/N3UV92NCS33RL53XGHIYT4DK6DGUAHQFYXQT228S1QTD5P78DK-17952?func=find-b&find_code=SYS&req Go Links

Harvard University HOLLIS CATALOG

- Search
- Expanded Search
- Command Search
- Results List
- Previous Searches
- Display Options

My Account/Renew
Comments/Requests
Start Over
Help

Full Catalog | Find Articles | Journal Titles | Digital Resources | Reserves

Full View of Record : FULL CATALOG

Print/Save/Send | Add to List | View List | Last Browse | Back to Results List

Choose format: Full view | Short view | MARC

Record 1 out of 1 ◀ Previous Record Next Record ▶

Title : Los dos folletos y la carta de Victor Capoul : publicados en el diario La Correspondencia de Cuba.

Published : Habana : Imp. de Obra La Correspondencia de Cuba, 1883.

INTERNET LINK : http://nrs.harvard.edu/urn-3:FHCL:139066 [Provides access to page images of entire work]

Locations/Orders : Availability

Location : Networked Resource Holdings Availability

Location : Master Microforms Harvard Depository Film Mas 35306 Holdings Availability

Location : Widener Harvard Depository SA 1503.01 box 7 Holdings Availability

Description : 60 p. ; 17 cm.

Restrictions : At least one location has information and/or restrictions on access. Click on the holdings link(s) for specific information.

Notes : "Los duques de la Torre y el casamiento de su hijo : extracto del folleto del señor Carreras": p. [5]-45.
"Carta de Capoul á 'El Figaro'": p. [47]-60.

Subject : Serrano y Domínguez, Francisco, duque de la Torre, 1810-1885.

Subject : Havana (Cuba) -- Description and travel.

Authors : Carreras, Luis, 1840-1888. Duques de la Torre y el casamiento de su hijo.
Capoul, Victor, 1839-1924.

Other Titles : Latin American pamphlet digital project at Harvard University.
Latin American pamphlet digital project at Harvard University. Preservation microfilm collection ; 1811.

HOLLIS Number : 009250537

Virtual Collections - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Search Favorites Media

Address http://fiona.hul.harvard.edu:11080/vc/deliver/vcFullRecord?collection=LAP&inoId=2776&recordNumber=18&method=view Go Links

Latin American Pamphlet Digital Collection

Home

Search

Browse Indexes

Search History

Help

HOLLIS Catalog

Widener Library Latin American, Spanish, and Portuguese Collection

Full Record Display

Record 1 of 825 for search: **cuba** . Jump to record # Go

Next Record Back To Results Display

Title	Los dos folletos y la carta de Victor Capoul : publicados en el diario La Correspondencia de Cuba
Name/Creator	Capoul, Victor, 1839-1924
HOLLIS ID	009250537
Digital Object	http://nrs.harvard.edu/urn-3:FHCL:139066
Location	Widener HD
Place of Origin	Habana
Publisher	Imp. de Obra La Correspondencia de Cuba
Creation Date	1883
Language	Spanish
Description	60 p. ; 17 cm.
Form/Genre	text
Subject	Serrano y Domínguez, Francisco , duque de la Torre , 1810-1885 ; Havana (Cuba) – Description and travel
Note	"Los duques de la Torre y el casamiento de su hijo : extracto del folleto del señor Carreras": p. [5]-45. "Carta de Capoul á 'El Figaro'": p. [47]-60.
Other Title	Duques de la Torre y el casamiento de su hijo
Other Title	Latin American pamphlet digital project at Harvard University

Virtual Collections - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Search Favorites Media

Address http://fiona.hul.harvard.edu:11080/vc/deliver/browseCombine?_collection=LAP&searchterm=339&navtype=next&browseinde Go Links

Latin American Pamphlet Digital Collection

Home

Search

Browse Indexes

Search History

Help

HOLLIS Catalog

Widener Library Latin American, Spanish, and Portuguese Collection

Browse the Collection

Title | Name/Creator | Subject | Form/Genre

a b c d e f g h i j k l m n o p q r s t u v w x y z

Prev Next or Jump To: Go

Caba, Jeraldino

Caballero, Andrés S

Caballero, José Agustín

Caballero de Rodas, Antonio

Cabanyes, Manuel de

Cabero de Dorado, María

Cabildo Eclesiástico (Santiago, Chile)

Cabildo Eclesiástico de Santa Cruz (Santa Cruz, Bolivia)

Cabrera, Ladislao

Cacao

Prev Next

Hints

- Click on an index to see all terms for this heading
- After conducting a search select "prev" or "next" to browse

Virtual Collections - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Search Favorites Media

Address http://fiona.hul.harvard.edu:11080/vc/deliver/executeQuery?_collection=LAP&searchterm=341&searchxpath=tf%3Acontain: Go Links

Latin American Pamphlet Digital Collection

Home

Search

Browse Indexes

Search History

Help

HOLLIS Catalog

Widener Library Latin American, Spanish, and Portuguese Collection

Search Results

Result page: 1 Records 1 to 1 of 1 for search: Caballero, José Agustín in (Subject).

1

Title El presbítero D. José Agustín Caballero, su vida y sus obras / por Alfredo Zayas.

Creator Zayas y Alfonso, Alfredo

Date 1891

Link http://nrs.harvard.edu/urn-3:FHCL:534878

Display Full Record

Back to Browse Results

Result page: 1 Records 1 to 1 of 1 for search: Caballero, José Agustín in (Subject).

Questions & Comments

An Integrated Virtual Collection

Harvard Open Collections Program “Women Working”

Open Collections Program: Women Working - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Search Favorites Media

Address http://ocp.hul.harvard.edu/ww/ Go Links

OPEN COLLECTIONS PROGRAM

Women Working, 1800 - 1930

HARVARD UNIVERSITY LIBRARY

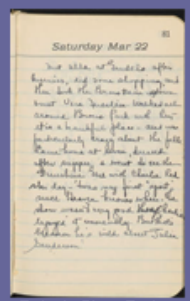
Browse the Collection
Search the Collection
Teacher Resources

Keyword Search
trade cards
Search

FAQ
Help
Site Credits/Contact
Related Links

Open Collections Home
Questions/Comments
Link To Us
Subscribe

Print A Women Working 2006/1893 Calendar!



Today's Diary Entries

THE COLLECTION

Women Working, 1800 - 1930 focuses on women's role in the United States economy and provides access to digitized historical, manuscript, and image resources selected from Harvard University's library and museum collections. The collection features approximately 500,000 digitized pages and images including:

- 7,500 pages of manuscripts
- 3,500 books and pamphlets
- 1,200 photographs

Image at left: Cover illustration of Pinkham Pamphlet, Volume 154. Lydia E. Pinkham Medicine Company Advertising Records, 1873-1968. From the collection of Schlesinger Library, Radcliffe Institute.

FEATURED

Teaching Resources

Educators and students, explore "Women Working" materials by theme! Includes a bibliography and teaching frameworks for educators of all levels.

"The Materials of the New Race": Immigration and Whiteness

Photographs, maps, graphs, texts, and illustrations that trace the shifting boundaries between new immigrant and other "alien" groups, and between these and the "native white" population

Soap and Settlements: "Making a Cleaner Society"

Photographs, advertisements, and other materials depicting Progressive Era assimilation efforts, consistency and

Open Collections Program: Women Working: Photographs - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Search Favorites Media

Address http://ocp.hul.harvard.edu/ww/photos.html Go Links

OPEN COLLECTIONS PROGRAM

Women Working, 1800 - 1930

HARVARD UNIVERSITY LIBRARY

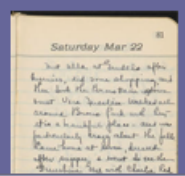
Home
Browse the Collection
Search the Collection
Teacher Resources

Keyword Search
trade cards
Search

FAQ
Help
Site Credits/Contact
Related Links

Open Collections Home
Questions/Comments
Link To Us
Subscribe

Print A Women Working 2006/1893 Calendar!



Search Results

Result page: 1 Records 1 to 49 of 49 for search: trade cards.

1 **Title** Rapid Transit Soap
Creator Colgate & Co.
Date 1875

Display Full Record

2 **Title** Soapine
Creator Kendall Mfg. Co.
Date 1875

Display Full Record

3 **Title** David's Prize Soap
Creator David's Prize Soap Co.
Date 1875

Display Full Record

4 **Title** Scourene
Creator Scourene Mfg. Co.

http://nrs.harvard.edu/urn-3:HBS.BAKER.TC:401?buttons=Y - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Search Favorites Media

Address http://ids.harvard.edu:8080/ids/servlet/imgdelv?id=2065&buttons=Y Go Links

OPEN COLLECTIONS PROGRAM
Women Working, 1800 - 1930
HARVARD UNIVERSITY LIBRARY

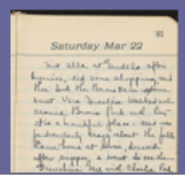
Home
Browse the Collection
Search the Collection
Teacher Resources

Keyword Search
trade cards
Search

FAQ
Help
Site Credits/Contact
Related Links

Open Collections Home
Questions/Comments
Link To Us
Subscribe

Print A Women Working 2006/1893 Calendar!



Search Results

Result page: 1 Records 1 to 49 of 49 for search: trade cards.

1 **Title** Rapid Transit Soap
Creator Colgate & Co.
Date 1875

Display Full Record

2 **Title** Soapine
Creator Kendall Mfg. Co.
Date 1875

Display Full Record

3 **Title** David's Prize Soap
Creator David's Prize Soap Co.
Date 1875

Display Full Record

4 **Title** Scourene
Creator Scourene Mfg. Co.

Open Collections Program: Women Working: Photographs - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Search Favorites Media

Address http://ocp.hul.harvard.edu/ww/photos.html Go Links

OPEN COLLECTIONS PROGRAM

Women Working, 1800 -1930

HARVARD UNIVERSITY LIBRARY

Print A Women Working 2006/1893 Calendar!

Full Record Display

Record 1 of 49 for search: **trade cards** . Jump to record # Go

Next Record | Back To Results Display

Title	Rapid Transit Soap
Name/Creator	Colgate & Co., n.d., New York, New York, United States, creatoradvertiser (manufacturer)
VIA ID	olvwork84422
Digital Object	
Location	Baker Library, Harvard Business School, Historical Collections; TC6240.0031
Publisher	Chas. Shields' Sons
Creation Date	1875
Description	9 x 14.5
Form/Genre	still image advertisements chromolithographs trade cards
Access/Rights	Baker Library, Harvard Business School, Historical Collections:Contact repository for permissions and fees.
Access/Rights	Copyright 2000, President and Fellows of Harvard College; all rights reserved.
Materials Type	chromolithography
Subject	soap ; cleaning compounds ; cleaning compounds industry ; city & town life
Note	Verso: blank. Provenance: Gift of James S. Robinson.

VIA Record - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Search Favorites Media

Address http://via.harvard.edu:9080/via/deliver/deepLink?_collection=via&recordId=olvwork84422 Go Links

Harvard University Library
Visual Information Access

Quick search GO

About VIA | Search | Browse | Search History | Portfolios | Help

Full Record | Grid View | Search Results

Record 1 of 1

From the search

Save entire record...

Work

Recto (verso blank)

Save this...

Title:	Rapid Transit Soap
Work Type:	advertisements; chromolithographs; trade cards
Creator:	Colgate & Co. (n.d.), New York, New York, United States
Production:	New York, New York, United States, Chas. Shields' Sons
Date:	between 1875 and 1899
Description:	Verso: blank.
Dimensions:	9 x 14.5
Topics:	soap; cleaning compounds; cleaning compounds industry; city & town life
Materials/Techniques:	chromolithography
Note:	*Provenance:* Gift of James S. Robinson.
Copyright:	Copyright 2000, President and Fellows of Harvard College; all rights reserved.
Use Restrictions:	Baker Library, Harvard Business School, Historical Collections:Contact repository for permissions and fees.
Repository:	Baker Library, Harvard Business School, Historical Collections TC6240.0031 :624

Record Identifier: olvwork84422

Record 1 of 1

Creating a Virtual Collection

Creating a Virtual Collection

Step 1. Define the collection

Step 2. Define the look and feel requirements

Step 3. Harvest metadata records

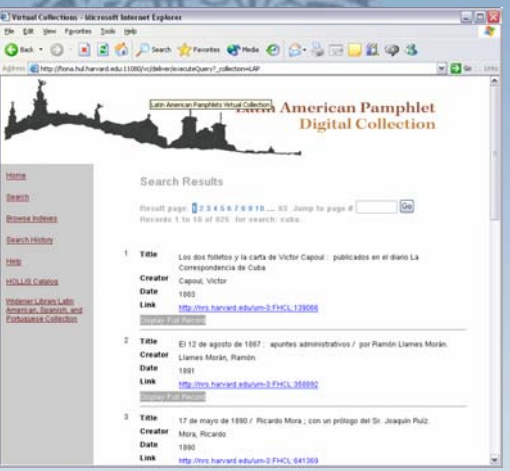
Step 4. Add categories (optional)

Step 1: Define the Collection

- What materials from HOLLIS, VIA, and/or HGL are to make up the collection?
- Is additional cataloging required in the primary catalog?
- Do digital versions of any of the objects need to be created?
- Will any VC-specific subject terms (“categories”) be added to the collection?

Step 2: Define the Look and Feel Requirements

- Curator selects an implementation option:



  - **Stand-alone – minimal work**
    - Data and public “skin” (custom banner, background colors, footer, menu links, home page text) hosted by VC

  - **Integrated – look and feel**
    - Data hosted by VC; public “skin” styled and hosted locally
    - Setup and maintenance under curator’s control

  - Both options offer the same search, browse, and record display features

Step 3: Harvest Metadata Records

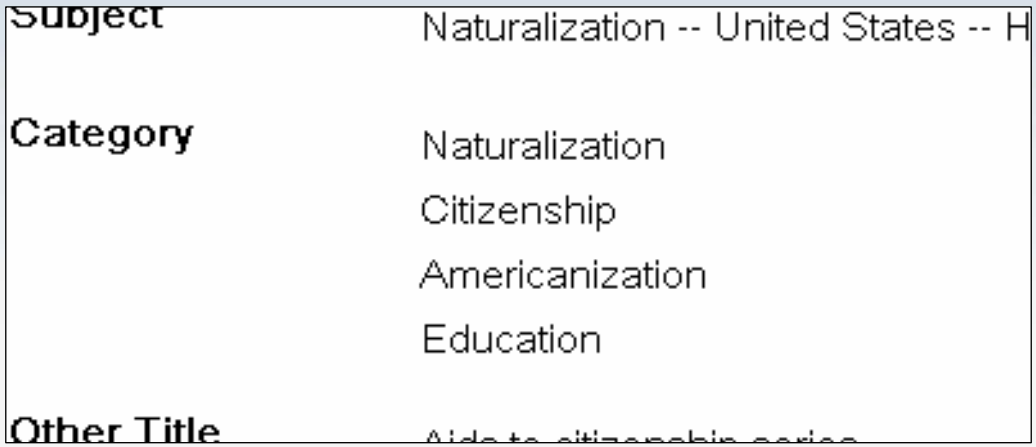
- Sources: HOLLIS OPAC, VIA, Harvard Geospatial Library
- Curator creates a “harvest file”
- Curator FTPs the harvest file to a dropbox; records harvested overnight
  - Automatic weekly harvesting picks up changes in source cataloging

Harvest file

HOLLIS 002223743
VIA olvwork125041
HGL MGISRAILTRA1

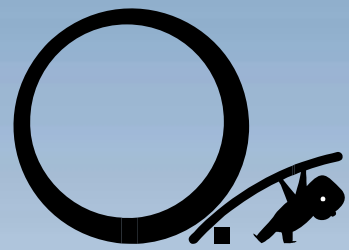
Step 4: (Optional) Add Categories to Records

- Categories are optional subject-like terms added to records in VC
- They supplement but do not replace subject terms from the source system
- They are created using the VC Web Maintenance Interface to a collection
- They can be searched and browsed



Benefits of Virtual Collections

- Minimal setup time for a fully functional web based collection



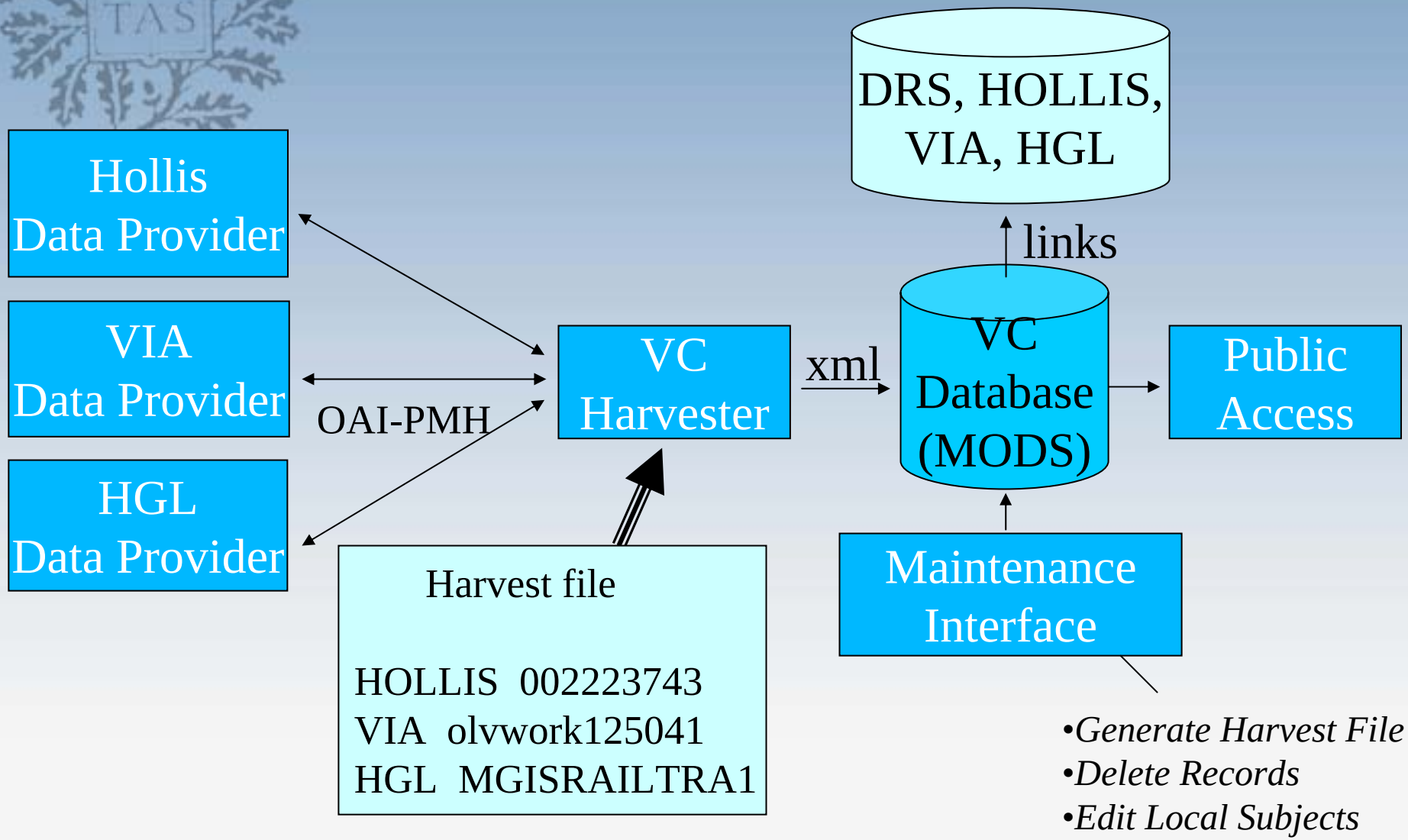
- Centrally supported infrastructure – 24x7 support
- Have our cake and eat it too
  - Capitalize on central cataloging
  - Automated sync with catalog updates
  - Multiple contexts for discovery (union catalog and virtual collection)

Current Collections

- ***Latin American Pamphlets***
  (Harvard College Library Latin American Collection) (3500 records)
- ***Immigration to the United States, 1789-1930***
  (Harvard University Library Open Collections Program) (8400 records)
- ***Women Working***
  (Open Collections Program) (4200 records)
- ***Studies in Scarlet***
  (Harvard Law Library; upcoming)

How Virtual Collections Work
and Some Implementation Challenges

VC Architecture



VC Technologies

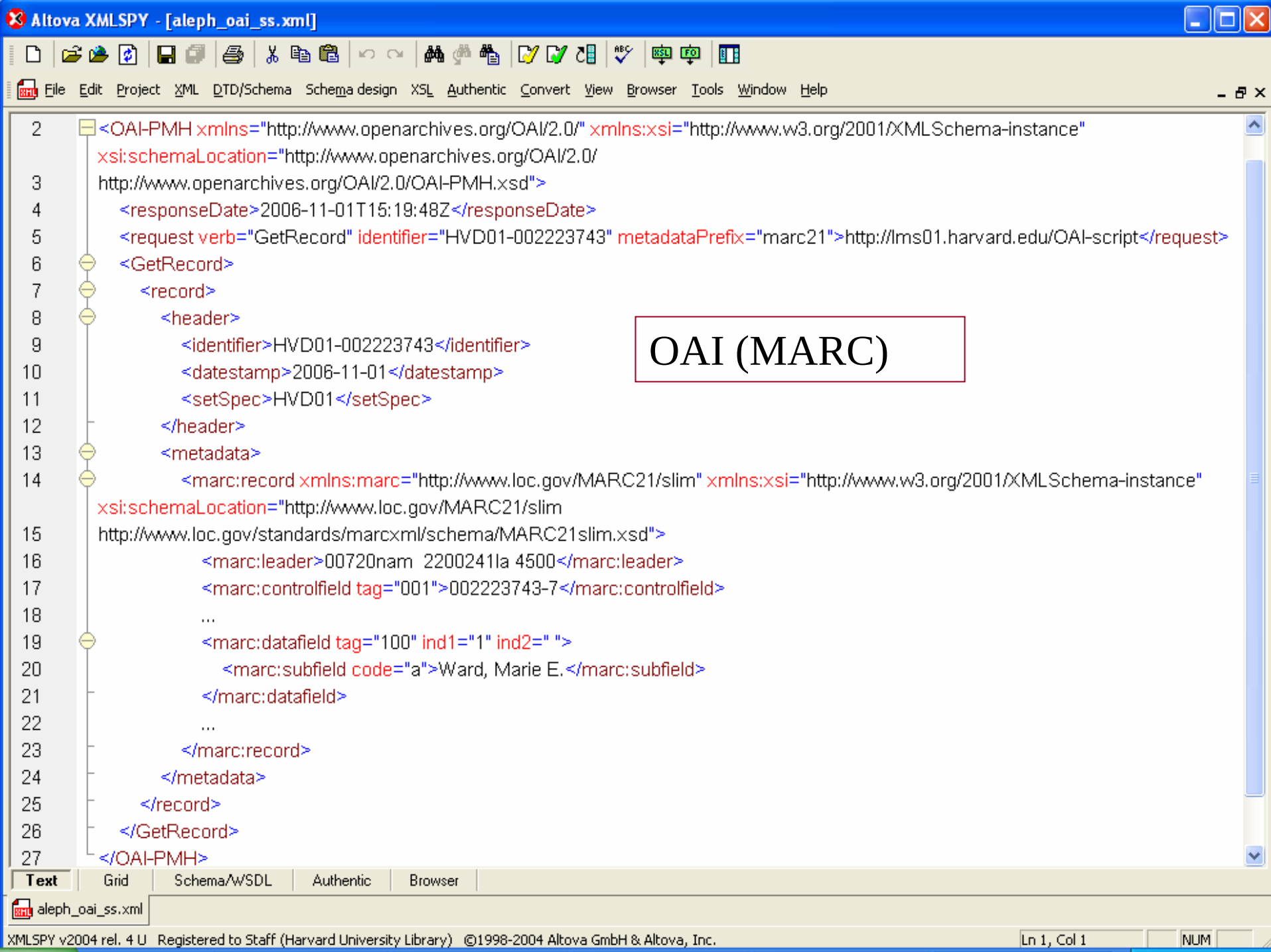
- Java 1.4 (Servlets, JSPs)
- Struts framework
- Tomcat 4.1 (Servlet engine)
- Tamino Native XML database
- XML Schema
- XSLT
- OAI-PMH (OAICAT – OCLC)
- MODS (Metadata Object Description Schema)

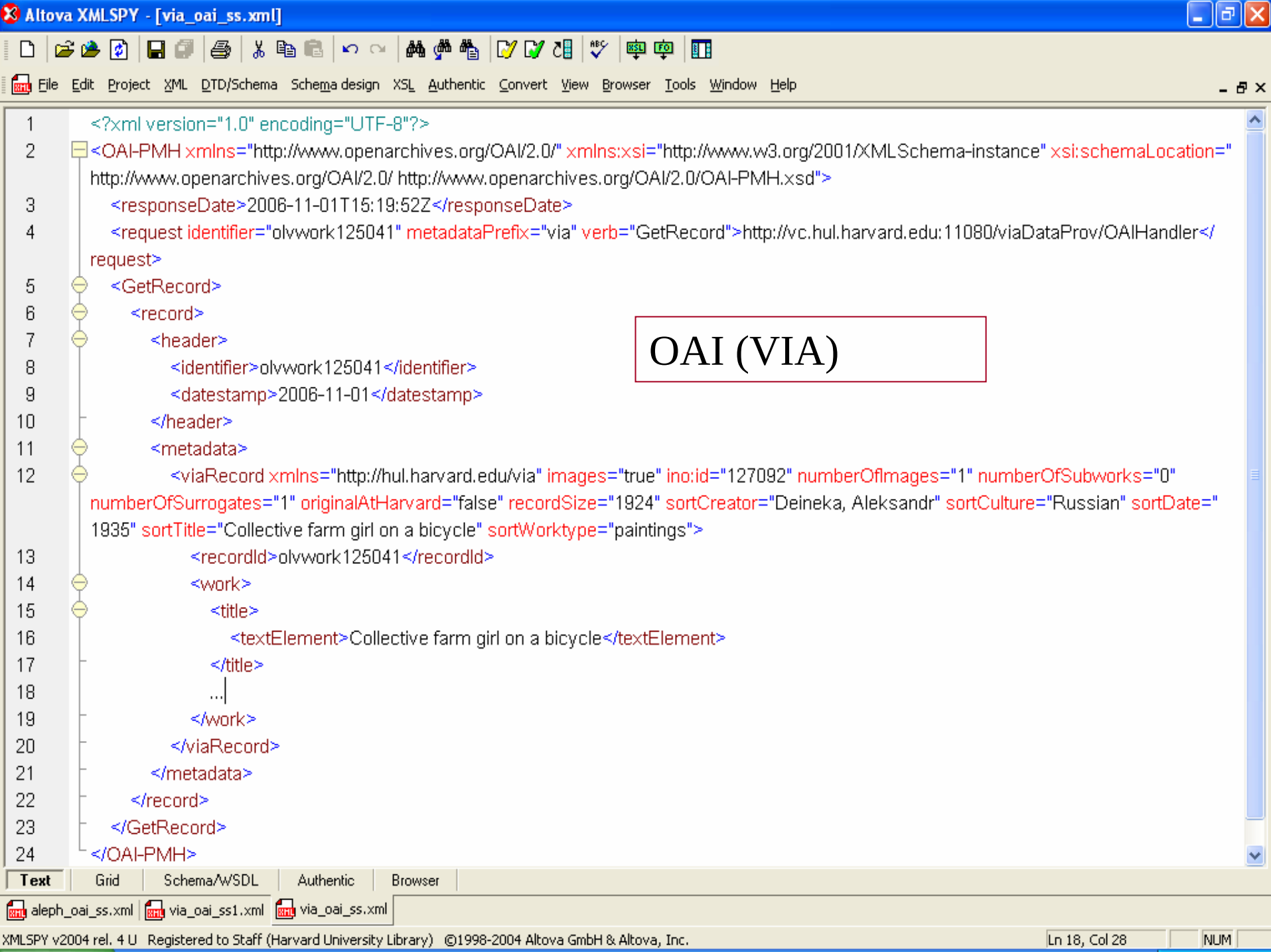
Harvesting: OAI-PMH

- Harvester uses GetRecord verb

  verb=**GetRecord**&identifier=HVD01-002223743& metadataPrefix=marc21

- HOLLIS: Data Provider from ExLibris
- VIA: Custom Data Provider
- HGL: Custom Data Provider


Altova XMLSPY - [aleph_oai_ss.xml]
File Edit Project XML DTD/Schema Schema design XSL Authentic Convert View Browser Tools Window Help
<OAI-PMH xmlns="http://www.openarchives.org/OAI/2.0/" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/
http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
<responseDate>2006-11-01T15:19:48Z</responseDate>
<request verb="GetRecord" identifier="HVD01-002223743" metadataPrefix="marc21">http://lms01.harvard.edu/OAI-script</request>
<GetRecord>
<record>
<header>
<identifier>HVD01-002223743</identifier>
<datestamp>2006-11-01</datestamp>
<setSpec>HVD01</setSpec>
</header>
<metadata>
<marc:record xmlns:marc="http://www.loc.gov/MARC21/slim" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.loc.gov/MARC21/slim
http://www.loc.gov/standards/marcxml/schema/MARC21slim.xsd">
<marc:leader>00720nam 2200241la 4500</marc:leader>
<marc:controlfield tag="001">002223743-7</marc:controlfield>
...
<marc:datafield tag="100" ind1="1" ind2=" ">
<marc:subfield code="a">Ward, Marie E.</marc:subfield>
</marc:datafield>
...
</marc:record>
</metadata>
</record>
</GetRecord>
</OAI-PMH>
OAI (MARC)
Text Grid Schema/WSDL Authentic Browser
aleph_oai_ss.xml
XMLSPY v2004 rel. 4 U Registered to Staff (Harvard University Library) ©1998-2004 Altova GmbH & Altova, Inc.
Ln 1, Col 1
NUM


Altova XMLSPY - [via_oai_ss.xml]
File Edit Project XML DTD/Schema Schema design XSL Authentic Convert View Browser Tools Window Help
1 <?xml version="1.0" encoding="UTF-8"?>
2 <OAI-PMH xmlns="http://www.openarchives.org/OAI/2.0/" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="
http://www.openarchives.org/OAI/2.0/ http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
3 <responseDate>2006-11-01T15:19:52Z</responseDate>
4 <request identifier="olvwork125041" metadataPrefix="via" verb="GetRecord">http://vc.hul.harvard.edu:11080/viaDataProv/OAIHandler</
request>
5 <GetRecord>
6 <record>
7 <header>
8 <identifier>olvwork125041</identifier>
9 <datestamp>2006-11-01</datestamp>
10 </header>
11 <metadata>
12 <viaRecord xmlns="http://hul.harvard.edu/via" images="true" ino:id="127092" numberOfImages="1" numberOfSubworks="0"
numberOfSurrogates="1" originalAtHarvard="false" recordSize="1924" sortCreator="Deineka, Aleksandr" sortCulture="Russian" sortDate="
1935" sortTitle="Collective farm girl on a bicycle" sortWorktype="paintings">
13 <recordId>olvwork125041</recordId>
14 <work>
15 <title>
16 <textElement>Collective farm girl on a bicycle</textElement>
17 </title>
18 ...
19 </work>
20 </viaRecord>
21 </metadata>
22 </record>
23 </GetRecord>
24 </OAI-PMH>
OAI (VIA)
Text Grid Schema/WSDL Authentic Browser
aleph_oai_ss.xml via_oai_ss1.xml via_oai_ss.xml
XMLSPY v2004 rel. 4 U Registered to Staff (Harvard University Library) ©1998-2004 Altova GmbH & Altova, Inc.
Ln 18, Col 28
NUM

Harvesting Challenges

- Aleph OAI provider lacks holdings info
  - Add holdings fields using XServer

```
<oai_marc>
 ...
 <varfield id="856" i1="4" i2="1">
      <subfield label="u">http://nrs.harvard.edu/urn-3:RAD.SCHL:159598</subfield>
      <subfield label="z">Provides access to page images of entire work</subfield>
    </varfield>
    <varfield id="852" i1="7" i2="">
      <subfield label="2">ZHCL</subfield>
      <subfield label="b">WID</subfield>
      <subfield label="c">HD</subfield>
      <subfield label="h">SG 1690.3</subfield>
    </varfield>
 ...
 </oai_marc>
```

Harvesting Challenges

- Local subjects in MARC records
- 600/650 -> MODS:subject

```
<datafield tag="650" ind1="" ind2="0">
    <subfield code="a">Cycling.</subfield>
</datafield>
```

- 690 -> **VC:subject**

```
<datafield tag="690" ind1="" ind2="9">
  <subfield code="a">
    Conduct of life; Education; Living
    conditions; Dress; Health; Recreation.
  </subfield>
</datafield>
```

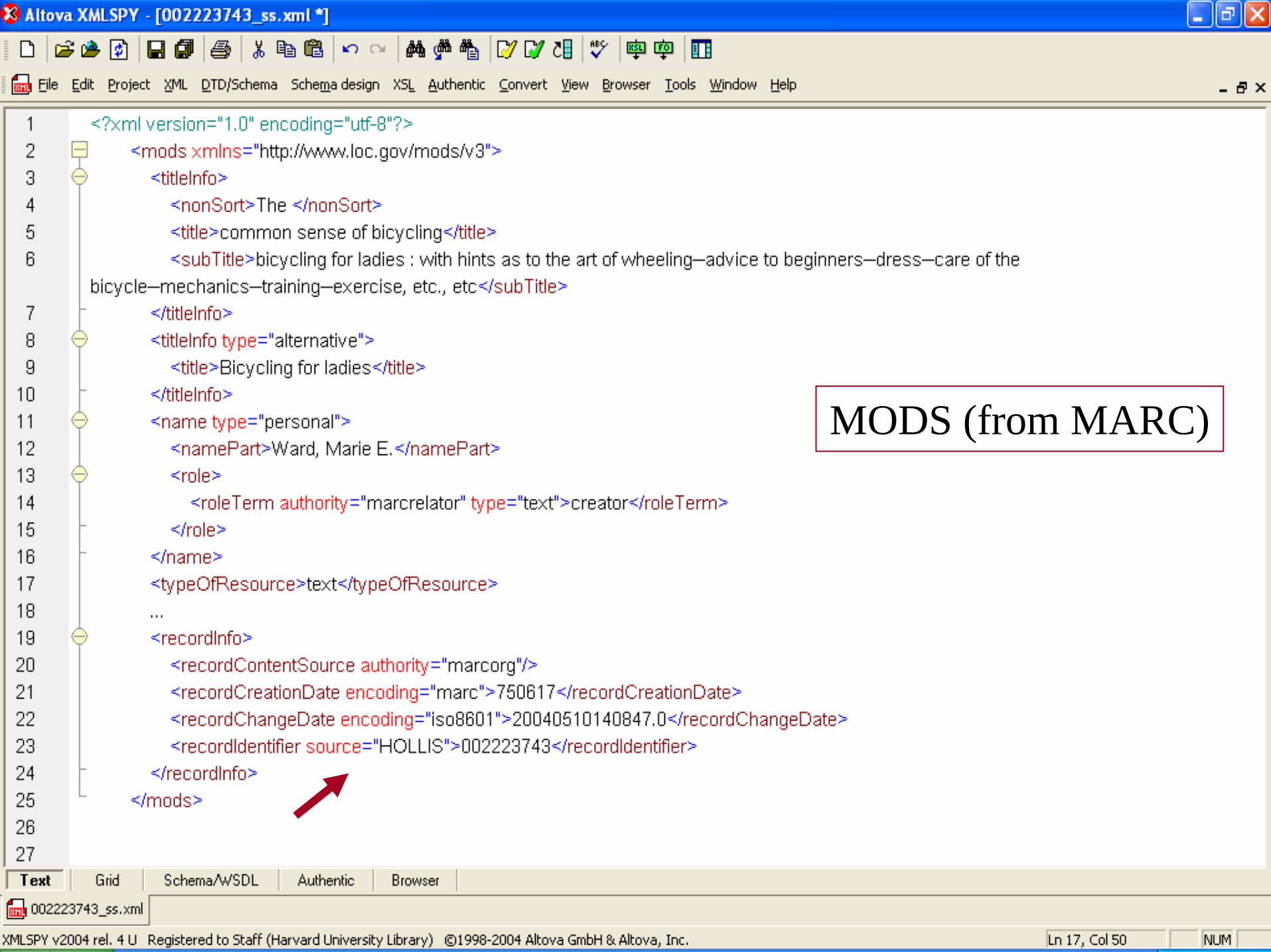
690 Subjects parsed; Loaded into VC if Match found

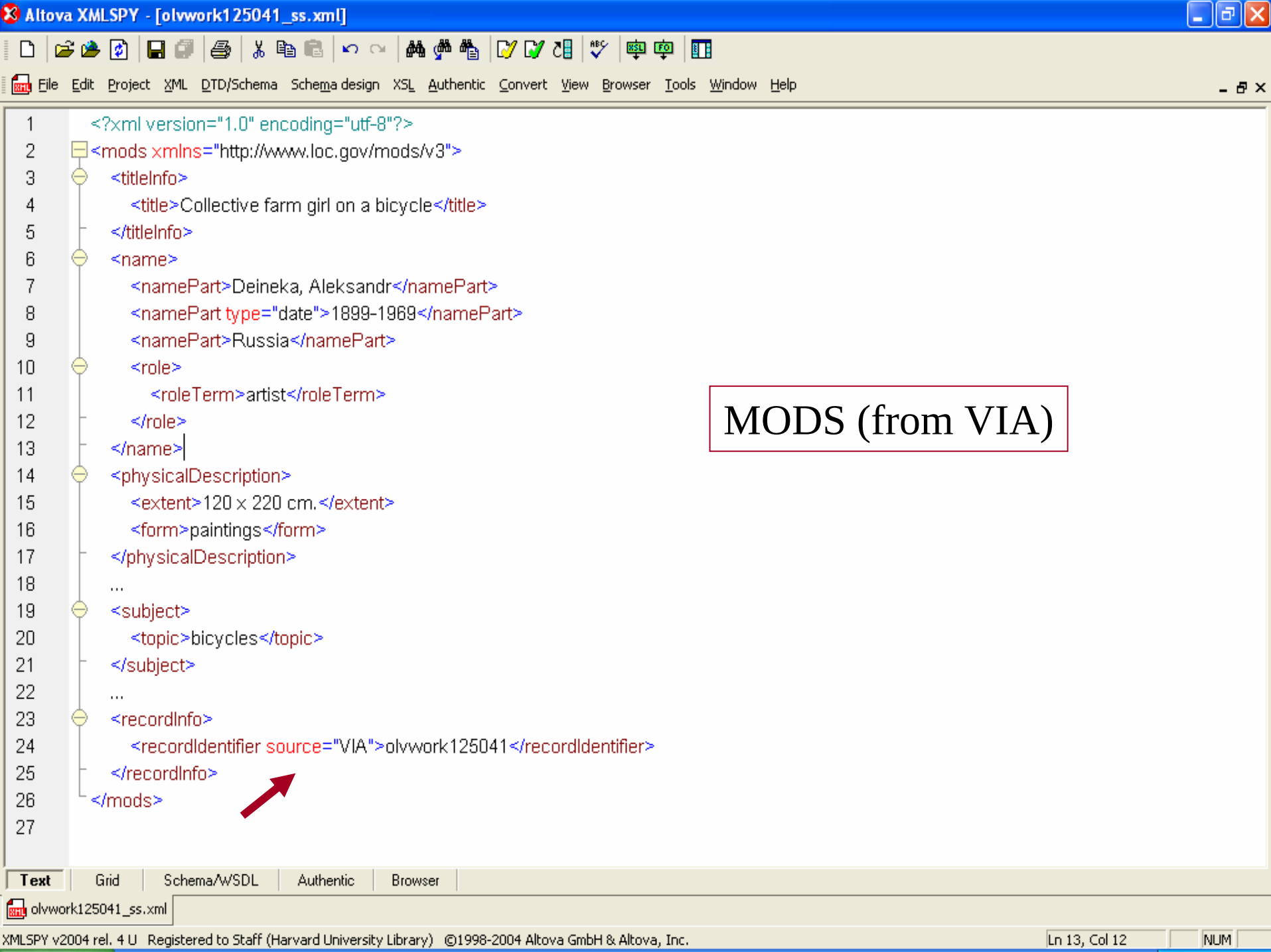
Harvesting Challenges

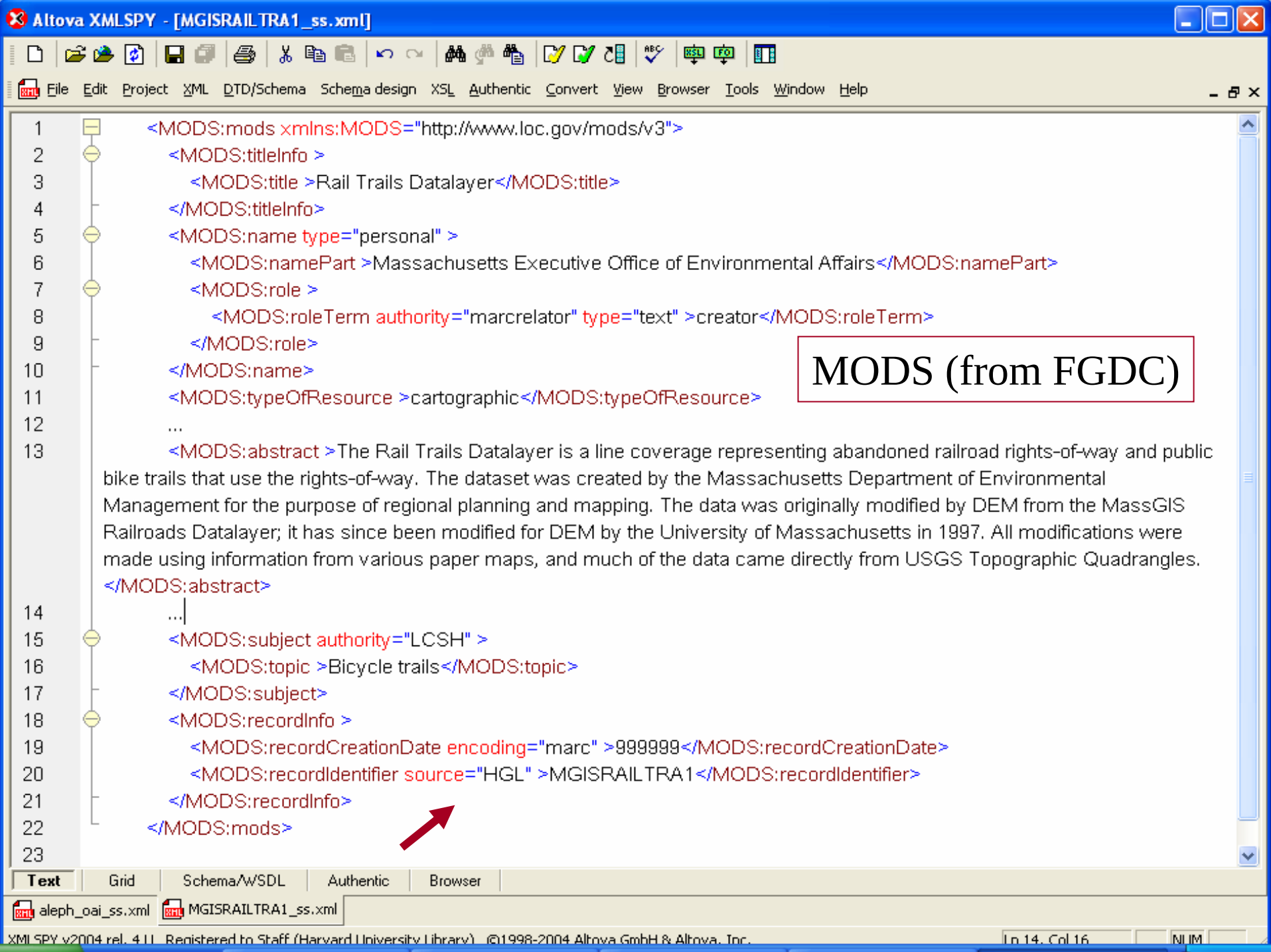
- Seeding and refreshing a collection
  - Aleph report for HOLLIS record id
  - Custom query tool for VIA record id generation (future)

Metadata Mapping: Crosswalks/XSL

Catalog	XML Input	XSL (crosswalk)	XML Ouptut
HOLLIS	marc21slim	**marc21slim2mods.xsl (Lib of Congress)**	MODS
VIA	via native	via2mods.xsl (Custom)	MODS
HGL (Geospatial)	FGDC (Federal Geog. Data Comm.)	1. fgdc 2 marc **2.** marc21slim2mods.xsl	MODS


Altova XMLSPY - [002223743_ss.xml *]
File Edit Project XML DTD/Schema Schema design XSL Authentic Convert View Browser Tools Window Help
<?xml version="1.0" encoding="utf-8"?>
<mods xmlns="http://www.loc.gov/mods/v3">
<titleInfo>
<nonSort>The </nonSort>
<title>common sense of bicycling</title>
<subTitle>bicycling for ladies : with hints as to the art of wheeling—advice to beginners—dress—care of the bicycle—mechanics—training—exercise, etc., etc</subTitle>
</titleInfo>
<titleInfo type="alternative">
<title>Bicycling for ladies</title>
</titleInfo>
<name type="personal">
<namePart>Ward, Marie E.</namePart>
<role>
<roleTerm authority="marcrelator" type="text">creator</roleTerm>
</role>
</name>
<typeOfResource>text</typeOfResource>
...
<recordInfo>
<recordContentSource authority="marcorg"/>
<recordCreationDate encoding="marc">750617</recordCreationDate>
<recordChangeDate encoding="iso8601">20040510140847.0</recordChangeDate>
<recordIdentifier source="HOLLIS">002223743</recordIdentifier>
</recordInfo>
</mods>
MODS (from MARC)
Text Grid Schema/WSDL Authentic Browser
002223743_ss.xml
XMLSPY v2004 rel. 4 U Registered to Staff (Harvard University Library) ©1998-2004 Altova GmbH & Altova, Inc.
Ln 17, Col 50
NUM


Altova XMLSPY - [olvwork125041_ss.xml]
File Edit Project XML DTD/Schema Schema design XSL Authentic Convert View Browser Tools Window Help
1 <?xml version="1.0" encoding="utf-8"?>
2 <mods xmlns="http://www.loc.gov/mods/v3">
3 <titleInfo>
4 <title>Collective farm girl on a bicycle</title>
5 </titleInfo>
6 <name>
7 <namePart>Deineka, Aleksandr</namePart>
8 <namePart type="date">1899-1969</namePart>
9 <namePart>Russia</namePart>
10 <role>
11 <roleTerm>artist</roleTerm>
12 </role>
13 </name>
14 <physicalDescription>
15 <extent>120 x 220 cm.</extent>
16 <form>paintings</form>
17 </physicalDescription>
18 ...
19 <subject>
20 <topic>bicycles</topic>
21 </subject>
22 ...
23 <recordInfo>
24 <recordIdentifier source="VIA">olvwork125041</recordIdentifier>
25 </recordInfo>
26 </mods>
27
MODS (from VIA)
Text Grid Schema/WSDL Authentic Browser
olvwork125041_ss.xml
XMLSPY v2004 rel. 4 U Registered to Staff (Harvard University Library) ©1998-2004 Altova GmbH & Altova, Inc.
Ln 13, Col 12
NUM


Altova XMLSPY - [MGISRAILTRA1_ss.xml]
File Edit Project XML DTD/Schema Schema design XSL Authentic Convert View Browser Tools Window Help
1 <MODS:mods xmlns:MODS="http://www.loc.gov/mods/v3">
2 <MODS:titleInfo >
3 <MODS:title >Rail Trails Datalayer</MODS:title>
4 </MODS:titleInfo>
5 <MODS:name type="personal" >
6 <MODS:namePart >Massachusetts Executive Office of Environmental Affairs</MODS:namePart>
7 <MODS:role >
8 <MODS:roleTerm authority="marcrelator" type="text" >creator</MODS:roleTerm>
9 </MODS:role>
10 </MODS:name>
11 <MODS:typeOfResource >cartographic</MODS:typeOfResource>
12 ...
13 <MODS:abstract >The Rail Trails Datalayer is a line coverage representing abandoned railroad rights-of-way and public bike trails that use the rights-of-way. The dataset was created by the Massachusetts Department of Environmental Management for the purpose of regional planning and mapping. The data was originally modified by DEM from the MassGIS Railroads Datalayer; it has since been modified for DEM by the University of Massachusetts in 1997. All modifications were made using information from various paper maps, and much of the data came directly from USGS Topographic Quadrangles.
</MODS:abstract>
14 ...
15 <MODS:subject authority="LCSH" >
16 <MODS:topic >Bicycle trails</MODS:topic>
17 </MODS:subject>
18 <MODS:recordInfo >
19 <MODS:recordCreationDate encoding="marc" >999999</MODS:recordCreationDate>
20 <MODS:recordIdentifier source="HGL" >MGISRAILTRA1</MODS:recordIdentifier>
21 </MODS:recordInfo>
22 </MODS:mods>
23
MODS (from FGDC)
Text Grid Schema/WSDL Authentic Browser
aleph_oai_ss.xml MGISRAILTRA1_ss.xml
XMLSPY v2004 rel. 4 U Registered to Staff (Harvard University Library) ©1998-2004 Altova GmbH & Altova, Inc.
Ln 14, Col 16
NUM

Metadata Mapping

MARC and VIA to MODS

MARC21

100 1 |a Picasso, Pablo |d 1881-1973.

VC MODS xml

```
<name type="personal">
<namePart>Picasso, Pablo</namePart>
<namePart type="date">1881-1973</namePart>
<role><roleTerm authority="..."  type="..."> creator</roleTerm></role>
</name>
```

VIA xml

```
<creator>
<nameElement>Picasso, Pablo</nameElement>
<dates>1881-1973</dates>
</creator>
```

Metadata Mapping

VC converts differing data sources into a standard format with a predefined set of elements.

In HOLLIS	In MODS	In VIA
title (24x)	titleInfo	title
author (1xx/7xx)	name	creator
published (260)	originInfo	production
genre (007/008)	genre	---
physical description (008)	physicalDescription/form	---
	typeOfResource	workType
subject (6xx)	subject/topic	location/*subject* topics

Metadata Mapping: MODS

- Why MODS
  - Richer metadata than qualified Dublin Core
  - Clean mapping from both MARC and VIA to MODS
  - Supports hierarchical VIA (and MARC) records

VIA

```
<surrogate componentID="olvsurrogate299383">
 <title><textElement>Total</textElement></title>
</surrogate>
```

MODS

```
<relatedItem type="constituent">
 <titleInfo><title>Total</title></titleInfo>
 <recordInfo><recordIdentifier>olvsurrogate299383
 </recordIdentifier>
 </recordInfo>
</relatedItem>
```

Virtual Collections - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Search Favorites Media

Address http://vctest.hul.harvard.edu:11080/vc/deliver/vcFullRecordInt?collection=testCollection&inoId=34&recordNumber=1&method=view Go

Google Go Bookmarks Popups okay Check AutoLink AutoFill Settings

Title	Collective farm girl on a bicycle
Name/Creator	Deineka, Aleksandr, 1899-1969, Russia, creator artist
VIA ID	olvwork125041
Location	State Russian Museum, Leningrad, Russia (Federation)
Creation Date	1935
Description	120 x 220 cm.
Form/Genre	still image paintings
Classification	CA D 3666 1935 Co
Culture	Russian
Materials Type	oil on canvas
Subject	landscapes ; bicycles ; farms ; girls

Hierarchy in VIA record

- - -

Title	Total
Digital Object	
Location	Harvard Fine Arts Library, Visual Collections - Slides and Digital Images
Form/Genre	still image color slide
Classification	CA D 3666 1935 Co

Done Internet

Metadata Mapping: Schema Definition

Virtual Collections Schema

- One primary schema for all collections

```
<virtualCollection>
<xs:import schemaLocation = "mods.xsd“/>
<xs:import schemaLocation=“VC_subject.xsd“/>
 <xs:element name = "virtualCollection">
   <xs:complexType>
    <xs:sequence>
       <xs:element name="modsGroup" ref="MODS:mods" />
      <xs:element name="vcSubj" ref="VC:subject" …/>
       <xs:element name="admSection" ref="admin" />
    </xs:sequence>
  </xs:complexType>
```

Common to all collections

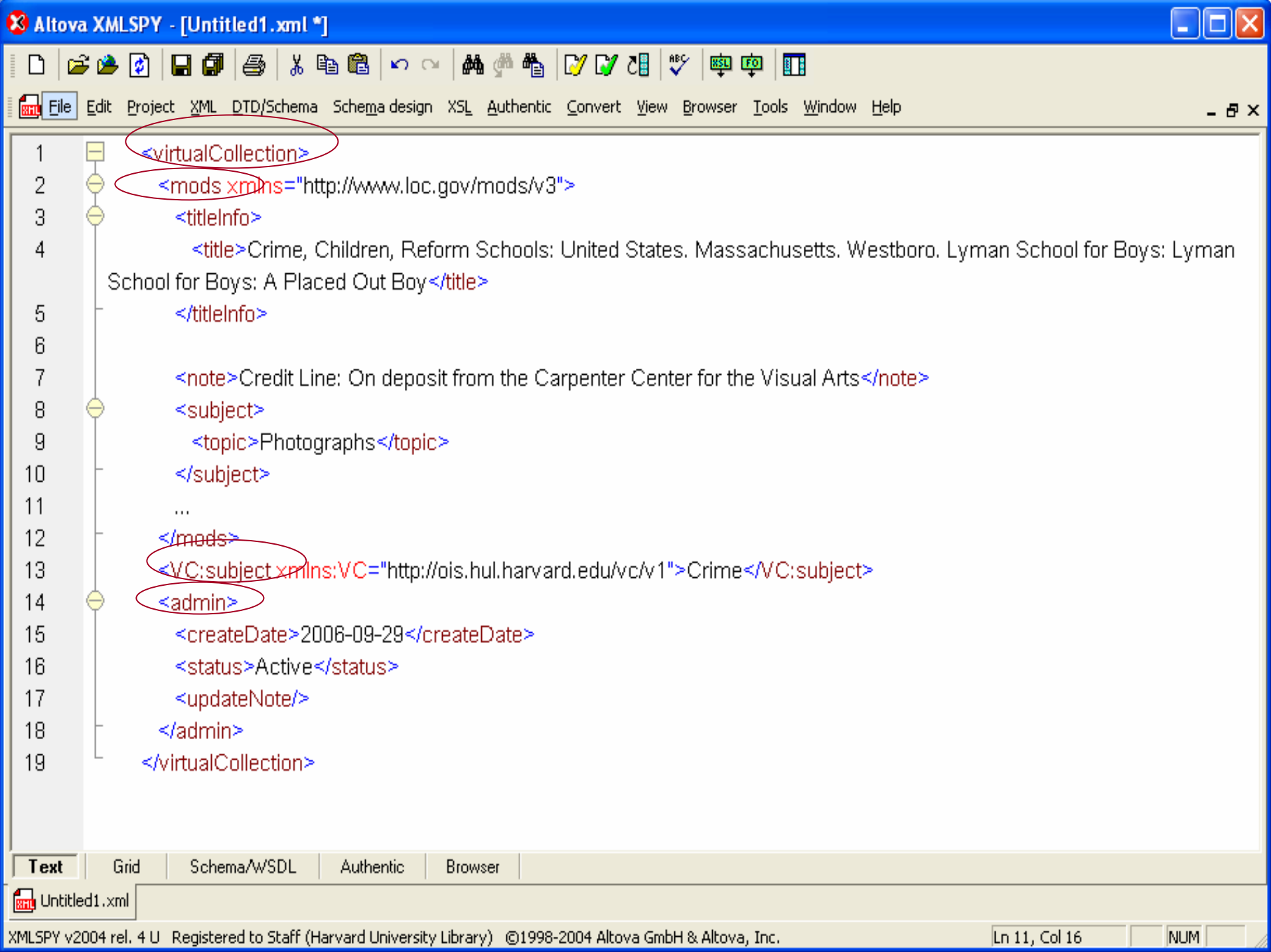
Distinct to each collection

Metadata Mapping: Schema Definition

- Local Subject schema for each collection

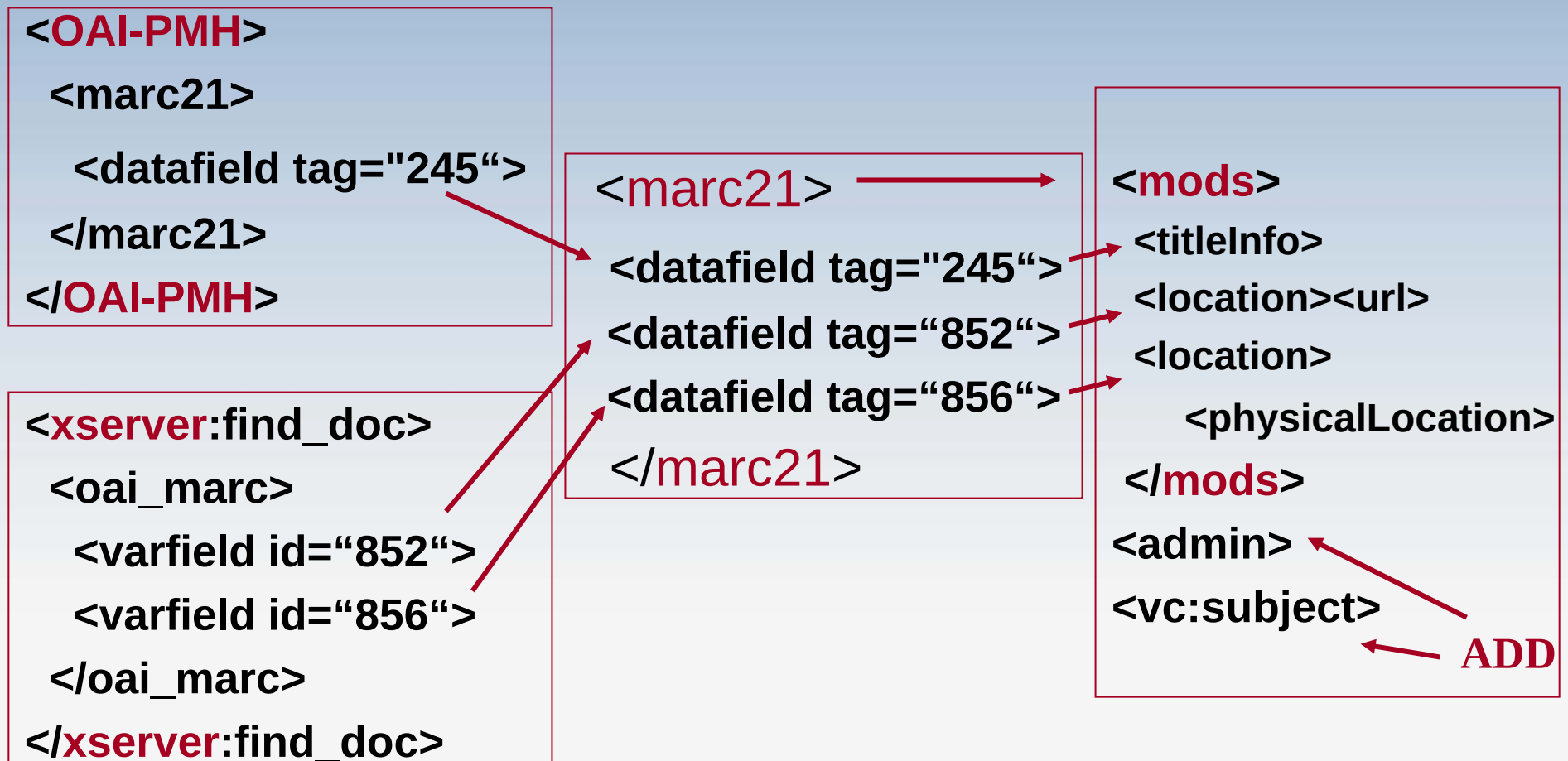
```
<xs:element name="subject">
    <xs:simpleType>
        <xs:restriction base="xs:string">
          <xs:enumeration value="African American"/>
          <xs:enumeration value="Aged women"/>
          <xs:enumeration value="Agricultural work"/>
          <xs:enumeration value="Anarchism"/>
          <xs:enumeration value="Artists"/>
                ...
        </xs:restriction>
    </xs:simpleType>
</xs:element>
```

Controlled list of Local Subjects


Altova XMLSPY - [Untitled1.xml *]
File Edit Project XML DTD/Schema Schema design XSL Authentic Convert View Browser Tools Window Help
<virtualCollection>
<mods xmlns="http://www.loc.gov/mods/v3">
<titleInfo>
<title>Crime, Children, Reform Schools: United States. Massachusetts. Westboro. Lyman School for Boys: Lyman School for Boys: A Placed Out Boy</title>
</titleInfo>
<note>Credit Line: On deposit from the Carpenter Center for the Visual Arts</note>
<subject>
<topic>Photographs</topic>
</subject>
...
</mods>
<VC:subject xmlns:VC="http://ois.hul.harvard.edu/vc/v1">Crime</VC:subject>
<admin>
<createDate>2006-09-29</createDate>
<status>Active</status>
<updateNote/>
</admin>
</virtualCollection>
Text Grid Schema/WSDL Authentic Browser
Untitled1.xml
XMLSPY v2004 rel. 4 U Registered to Staff (Harvard University Library) ©1998-2004 Altova GmbH & Altova, Inc.
Ln 11, Col 16
NUM

Metadata Mapping: Challenges

- Multiple transforms
  - **Original form MARC or MODS wrapped in OAI elements**
  - **Final XML record form is MODS + ADMIN + VC Subjects**



Storage of Metadata

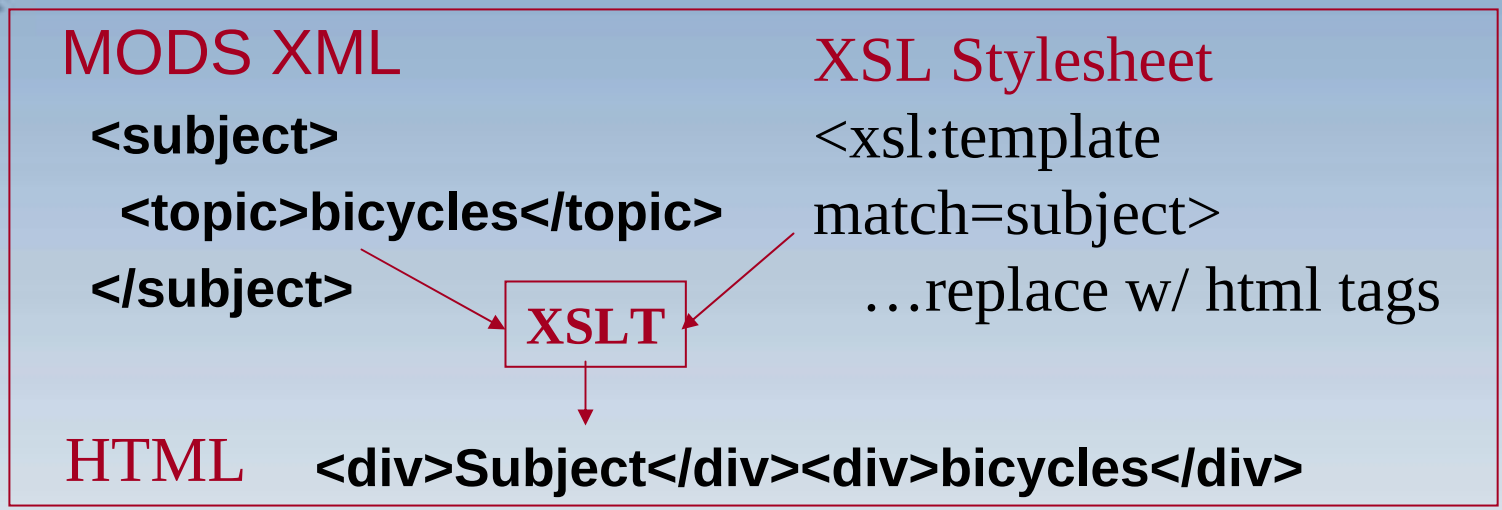
- Tamino native XML database
- Schema(s) defined within Tamino
- XQuery (W3C standard) used to query database for results
- Challenges
  - Native XML DBs not as mature as relational
    - **Tamino just now catching up with full implementation of XQuery**
  - Scaling as number and size of collections grow

User Interface and Customization

- Challenge: provide an out-of-the-box option and customizable option for collections
  - Different needs for different collections
- Solution: Standalone and Integrated
  - Default XSL and CSS templates for standalone
  - Override default templates with custom XSL and/or CSS for integrated collections

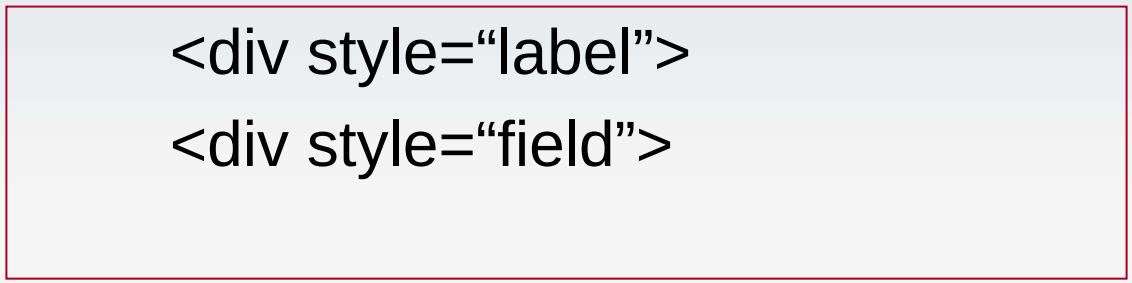
User Interface: Stylesheets

- XSL used to render XML to HTML



- CSS used to render “look and feel”

(true of Standalone as well as Integrated)



User Interface: Stylesheets

- Custom XSL (and CSS) passed as url parameters to overwrite default templates
- XSL
- **&baseUrl=http://defaultUrl&resultSetXsl=defaultRS.xsl**
- **&baseUrl=http://customUrl&resultSetXsl=customRS.xsl**
- CSS
- **&baseUrl=http://customUrl&css=custom.css**

Virtual Collections - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Search Favorites Media

Address http://vc.hul.harvard.edu:11080/vc/deliver/vcFullRecordInt?collection=immigration&inoId=6493&recordNumber=3&method=view Go

Google Go Bookmarks No popups Check AutoLink AutoFill Send to Settings

Full Record Display

Record 3 of 8 for search: **holland** . Jump to record # Go

Previous Record | Next Record | Back To Results Display

Title Anarchism: Holland. Portraits of Salomonson: Anarchism: Salomonson - a Dutch apostle of the return to nature.: Monsieur Joseph Salomonson d'Amsterdam. Ex-consul Meva, l'apôtre de la vie naturelle Social Museum Collection

Name/Creator Unidentified Artist, creator Artist

VIA ID HUAM20778soc

Digital Object

Location Fogg Art Museum, Harvard University Art Museums

Description 14 x 9 cm (5.51 x 3.54 in)

Form/Genre still image

photograph

Culture European ?

Materials Type Gelatin silver print

Subject Photographs

Default XSL

Internet

Virtual Collections - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Search Favorites Media

Address G:\Share\michaelv\dlf\FullRecord4.htm Go

Google Go Bookmarks Popups okay Check AutoLink AutoFill Settings

Full Record Display

Record 3 of 8 for search: **holland** . Jump to record # Go

Previous Record | Next Record | Back To Results Display

Digital Object

Custom XSL (change order)

Name/Creator	Unidentified Artist, creator Artist
Title	Anarchism: Holland. Portraits of Salomonson: Anarchism: Salomonson - a Dutch apostle of the return to nature.: Monsieur Joseph Salomonson d'Amsterdam. Ex-consul Meva, l'apôtre de la vie naturelle Social Museum Collection
Location	Fogg Art Museum, Harvard University Art Museums
VIA ID	HUAM20778soc
Description	14 x 9 cm (5.51 x 3.54 in)
Form/Genre	still image photograph
Culture	European ?
Materials Type	Gelatin silver print
Subject	Photographs

Done Local intranet

Virtual Collections - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Search Favorites Media

Address G:\Share\michaelv\dlf\FullRecord5.htm Go

Google Go Bookmarks Popups okay Check AutoLink AutoFill Settings

Full Record Display

Record 3 of 8 for search: **holland** . Jump to record # Go

Previous Record | Next Record | Back To Results Display

Digital Object

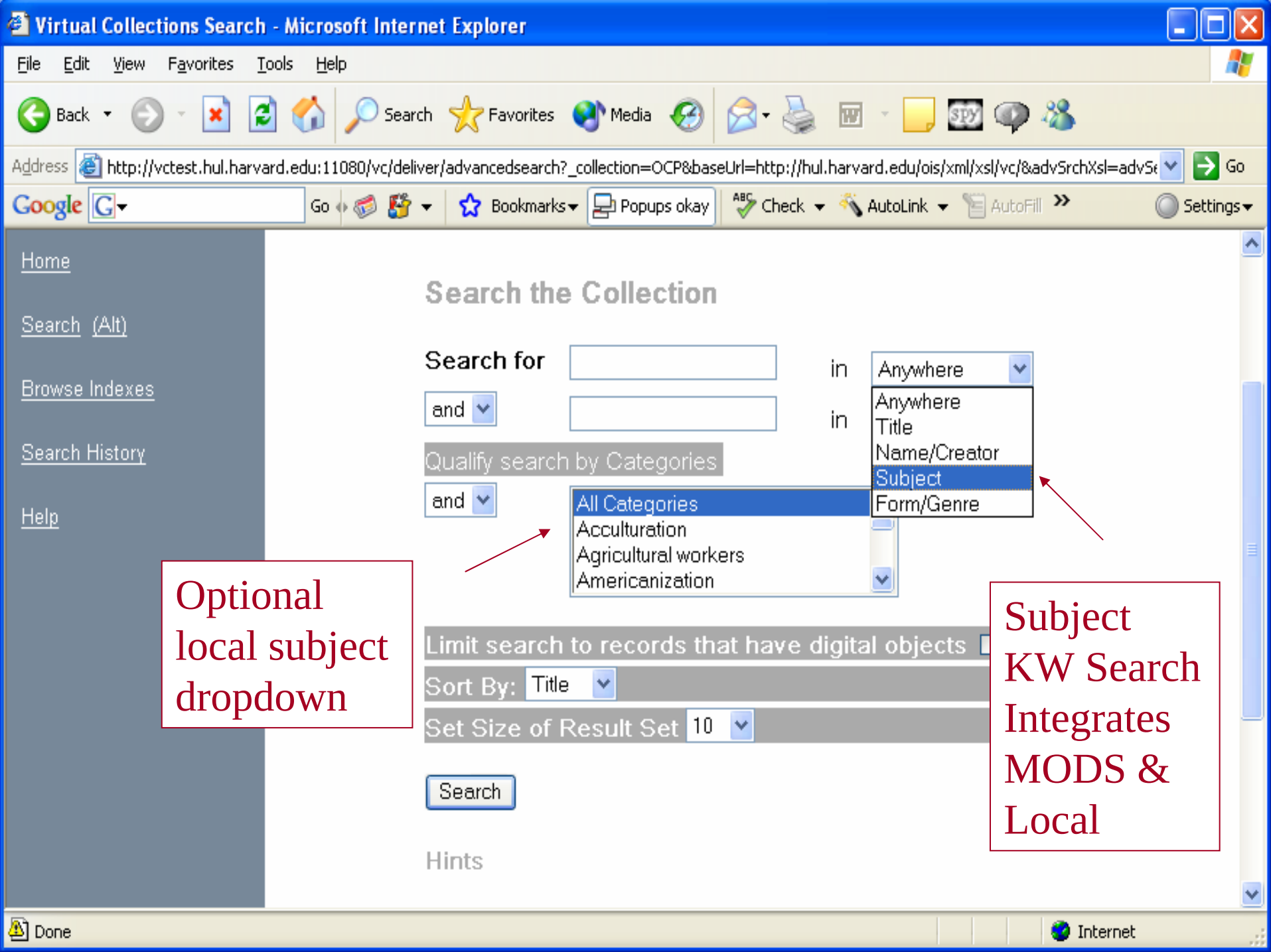
Custom CSS
(change look and feel)

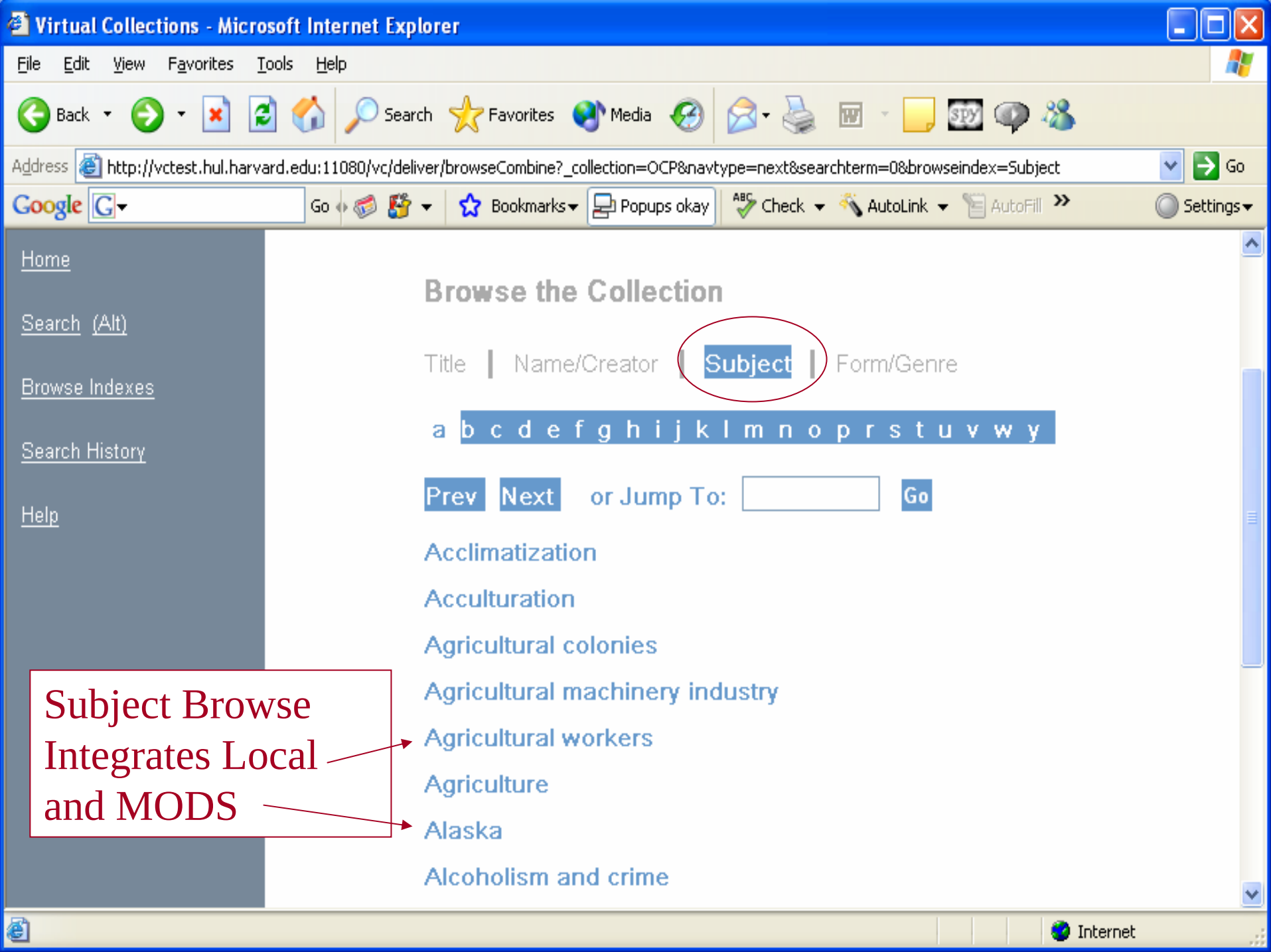
Name/Creator	Unidentified Artist, creator Artist
Title	Anarchism: Holland. Portraits of Salomonson: Anarchism: Salomonson - a Dutch apostle of the return to nature.: Monsieur Joseph Salomonson d'Amsterdam. Ex-consul Meva, l'apôtre de la vie naturelle Social Museum Collection
Location	Fogg Art Museum, Harvard University Art Museums
VIA ID	HUAM20778soc
Description	14 x 9 cm (5.51 x 3.54 in)
Form/Genre	still image photograph
Culture	European ?
Materials Type	Gelatin silver print
Subject	Photographs

Done Local intranet

User Interface: Local Subjects

- Challenge: how to separate/integrate Local and MODS subjects
  - Hardest issue to reach concensus on
- Solution
  - Local subjects optional
  - Local and MODS subjects integrated for KW search as well as browse
  - Optional local subject dropdown on search screen
  - MODS and Local subjects separated in Full Record display


Virtual Collections Search - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Address http://vctest.hul.harvard.edu:11080/vc/deliver/advancedsearch?_collection=OCP&baseUrl=http://hul.harvard.edu/ois/xml/xsl/vc/&advSrchXsl=advSe
Go
Google Go Bookmarks Popups okay Check AutoLink AutoFill Settings
Home
Search (Alt)
Browse Indexes
Search History
Help
Search the Collection
Search for
in
Anywhere
Anywhere
Title
Name/Creator
Subject
Form/Genre
and
in
Qualify search by Categories
and
All Categories
Acculturation
Agricultural workers
Americanization
Limit search to records that have digital objects
Sort By: Title
Set Size of Result Set 10
Search
Hints
Optional local subject dropdown
Subject KW Search Integrates MODS & Local
Done
Internet


Virtual Collections - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Address http://vctest.hul.harvard.edu:11080/vc/deliver/browseCombine?_collection=OCP&navtype=next&searchterm=0&browseindex=Subject
Go
Google Go Bookmarks Popups okay Check AutoLink AutoFill Settings
Home
Search (Alt)
Browse Indexes
Search History
Help
Browse the Collection
Title | Name/Creator | Subject | Form/Genre
a b c d e f g h i j k l m n o p r s t u v w y
Prev Next or Jump To: Go
Acclimatization
Acculturation
Agricultural colonies
Agricultural machinery industry
Agricultural workers
Agriculture
Alaska
Alcoholism and crime
Subject Browse Integrates Local and MODS
Internet

User Interface: Mapping

- Challenge: consensus on fields and labeling

In HOLLIS	Label in VC	In VIA
Title (24x)	Title	Title
Author (1xx/7xx)	Name/Creator	Creator
Published (260)	Publisher Place of Origin	Production
Genre Physical Description	Form/Genre	Work Type

VC Future Developments

- OAI Data Provider for VCs (this year)
- Portfolio function
- Ability to harvest from other Harvard catalogs
- Ability to search multiple virtual collections

Questions?