



Handwritten text in Finnish, likely a letter or document. The text is written in a cursive script and includes a date at the top: "Torstai 17. 10. 1900".



Digital Library Update

December 16, 2011

Introductions & Acknowledgements

- ▶ **Digital Initiatives and Information Technology**
 - ▶ Digital Library team
 - ▶ Stephen Davison
 - ▶ Lisa McAulay
 - ▶ Jen Weintraub
 - ▶ Kristian Allen
 - ▶ Henry Chiong
 - ▶ Parinita Ghorpade
 - ▶ Islandora implementation
 - ▶ Gary Thompson
 - ▶ Charlie Castro
 - ▶ Gloria Rom
 - ▶ Andy Kohler

Agenda

- ▶ Welcome; Purpose; Introductions (Stephen)
- ▶ Where we came from; where we are now (Lisa)
- ▶ Introduction to the new Digital Collections portal (Lisa)
- ▶ Introduction to the Islandora framework (Kristian)
 - ▶ Drupal
 - ▶ Solr
 - ▶ Fedora
 - ▶ Demo (Los Angeles Times collection)
- ▶ Timeline and Services (Jen)
- ▶ Research support and collaboration (Stephen)
- ▶ Questions and Discussion

Our Current Infrastructure

- ▶ Assigned to LISA
- ▶ DLCS = Digital Library Collection System
 - ▶ Homegrown
 - ▶ Java
 - ▶ Oracle

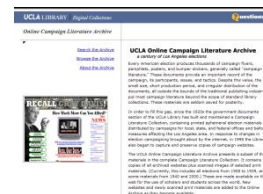
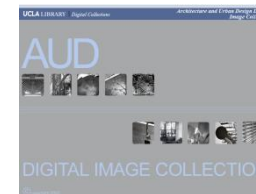
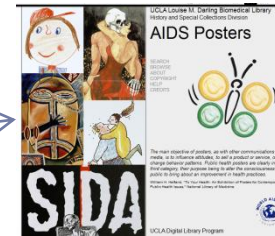
Screen shots here

Before DLCS

The screenshot shows the UCLA Digital Library Collections website. At the top, there is a navigation bar with "Questions" and "Libraries and Collections" dropdown menus. Below this, there are tabs for "About", "Collections", and "Contact". The main content area lists several digital collections:

- The AIDS Poster Collection**: The 625 posters in this collection, held by the Louise M. Darling Biomedical Library History and Special Collections, come from forty-four countries. They illustrate attempts by a variety of "institutions" to educate and warn their citizens about AIDS and offer advice and information in visual form about the still deadly autoimmune disease. These public health posters reach out to various at-risk groups as well as the general public. Some are more blunt and graphic than others, and they come in many styles.
- American Physical Society, Division of Particles and Fields**: Proceedings of the Los Angeles meeting held January 5-9, 1999.
- Architecture and Urban Design**: Images from the UCLA Department of Architecture and Urban Design's collections (UCLA only).
- Archive of Popular American Music**: Sheet music from the UCLA Music Library's Archive of Popular American Music.
- Campaign Literature Archive: A Century of Los Angeles Elections**: Ephemeral election materials such as posters, pamphlets, and copies of Web sites, from the Campaign Literature Collection in the Charles E. Young Research Library.
- Changing Times: Los Angeles in Photographs, 1920-90**: Featuring more than five thousand of the three million images contained in the *Los Angeles Times* and *Los Angeles Daily News* photographic archives housed in the Charles E. Young Research Library Department of Special Collections, this collection chronicles the history and growth of Los Angeles from the 1920s to 1990.
- Comitatus (Now published through the California Digital Library eScholarship Repository)**: Annual publication of the UCLA Center for Medieval and Renaissance Studies, devoted to publishing the work of young scholars.
- Corpus Juris Canonici (1582)**: A digital facsimile edition of the Decretals of Gregory IX as they appeared with marginal commentary in the 1582 edition.

Each project was a stand-alone website



With DLCS

BROWSE ALL COLLECTIONS

TITLE: [AIDS Poster Collection](#)

BROWSE BY: [Country of Origin](#) [Language](#) [Personal or Corporate Name](#) [Subject](#) [Type](#)

DESCRIPTION: Searchable database of digitized AIDS posters from 44 countries. Demonstrates the efforts taken by various countries to promote AIDS awareness and modify risk behaviors. The physical collection was originally acquired by the History and Special Collections division of the Louise M. Darling Biomedical Library in 2005, and has been expanded through later acquisitions.

TITLE: [Accion Socialista La](#)

BROWSE BY: [Language](#) [Name](#) [Subject](#) [Type](#)

TITLE: [Antiphonarium et hymnarium secundem morem Sanctae Romanae Ecclesiae](#)

BROWSE BY: [Name](#) [Subject](#) [Type](#)

TITLE: [Arkatov \(James\) Collection](#)

BROWSE BY: [Language](#) [Name](#) [Subject](#) [Type](#)

DESCRIPTION: Collection of hundreds of photographs of jazz and classical musicians, conductors and bandleaders performing in the Los Angeles area taken by photographer James Arkatov.

TITLE: [Armenian Manuscripts](#)

BROWSE BY: [Language](#) [Name](#) [Subject](#) [Type](#)

DESCRIPTION: Armenian manuscripts, gifted to the UCLA Library in 1968 by Dr. Garo Owen Minasian of Isfahan, Iran. The collection includes manuscripts of ecclesiastical character such as gospels, psalters, menologia, and ritual books, as well as theological and philosophical works, medical treatises, and anthologies of poetry.

TITLE: [Art Department Digital Image Collection](#)

BROWSE BY: [Name](#) [Type](#)

TITLE: [Bennett \(Walter E.\) Photographic Collection](#)

BROWSE BY: [Name](#) [Subject](#) [Type](#)

DESCRIPTION: Walter E. Bennett (1921-1995) was the first salaried photographer for Time, where he worked from 1952 to 1982. The collection consists of photographic materials such as prints, negatives and slides. It also includes miscellaneous manuscripts and ephemera related to Bennett's life and career.

TITLE: [Bound Manuscripts Collection, Collection 170](#)

BROWSE BY: [Coverage](#) [Language](#) [Name](#) [Subject](#) [Type](#)

DESCRIPTION: Collection consists of bound manuscripts of varying contents most notable for its early material which includes a papyrus fragment from the 7th century, a vellum Breviary (French, 15th century), a North French or Flemish Book of hours (late 15th or early 16th century), Persian and Arabic manuscripts, manuscript books from the 16th through 18th century, commonplace books, Friendship albums, and personal journals.

TITLE: [Bunche \(Ralph J.\) Papers, Collection 2051](#)

BROWSE BY: [Coverage](#) [Language](#) [Name](#) [Subject](#) [Type](#)

DESCRIPTION: Collection consists of correspondence, speeches, manuscripts, articles, publications, and photographs related to the life and career of Ralph J. Bunche. Includes materials related to his teaching and research, his affiliations with various organizations and conferences, and his participation in The Negro in America project spearheaded by Gunnar Myrdal. Includes materials on Africa: photographs, field notes, US Office of Strategic Services (OSS) reports, and Institute of Pacific Relations documents. Also materials related to Bunche's involvement with the United Nations, including UN records and materials dealing with his appointment as undersecretary of the UN in 1954. There are two boxes of materials on Bunche's receipt of the Nobel Peace Prize in 1950, and material from Joan Bunche and Brian Urquhart's notes from the biography he wrote, Ralph Bunche: an American life.

- ▶ Cross-Collections Searching and Browsing
- ▶ All Projects could be turned on and off easily
- ▶ Additional Features could still be designed

Revised Interface – the New Design

Growing Pains

- ▶ **35 Terabytes of Data**
- ▶ **250 Projects**
- ▶ **800,000 Digital Items**
- ▶ **1,600,000 Digital Content Files**
- ▶ **5,000,000 Pieces of Metadata**

Challenges

- ▶ **Scalability**
- ▶ **Performance**
- ▶ **Maintenance**
- ▶ **Feature Enhancements**
- ▶ **Development Time and Resources**

New Digital Library Architecture

▶ Islandora

- ▶ Packaged suite of open source tools

▶ See

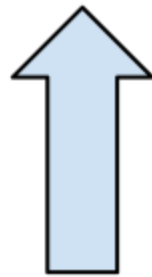
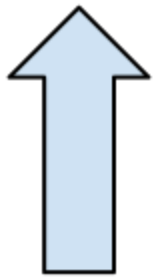
- ▶ Drupal delivers multiple user interface display options

▶ Search

- ▶ Solr offers a myriad of ways to explore and discover content

▶ Store

- ▶ Fedora delivers robust storage and repository features

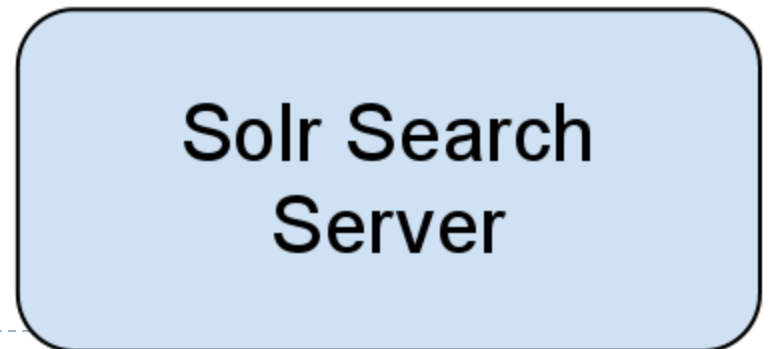
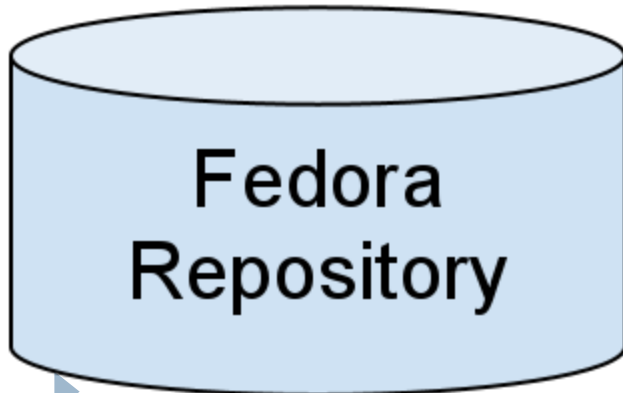
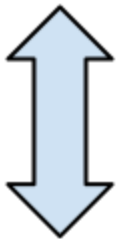


Web Site Delivery

Web Services API

Mobile Web Delivery

Drupal Platform



Drupal Overview

- ▶ Web-based Content Management System
- ▶ Robust, highly customizable and extendable
- ▶ Active community of developers
- ▶ Low learning curve to become productive
- ▶ Consolidation of technology within the Library
 - ▶ Basis of new library website (early 2012)

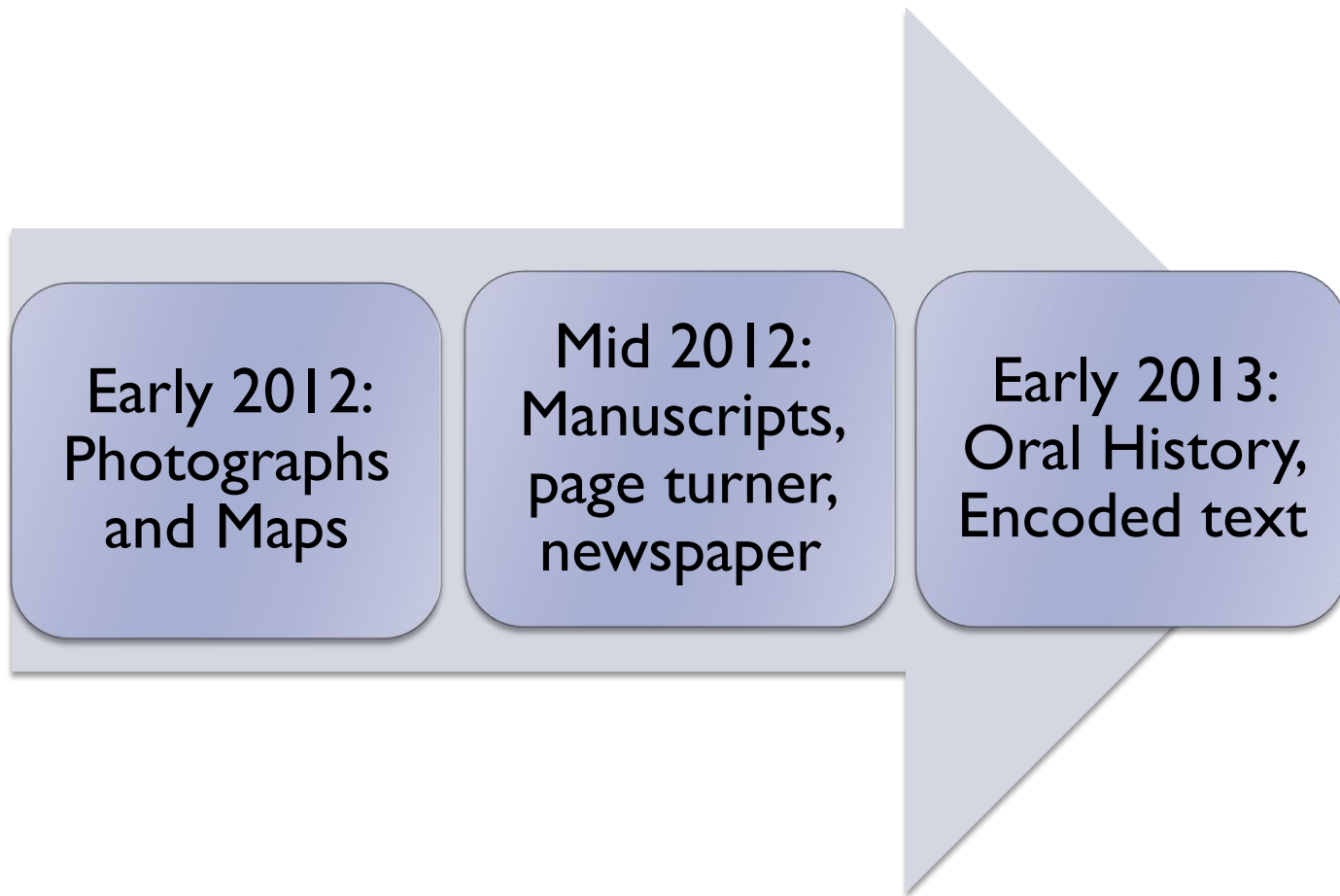
Solr Overview

- ▶ High performance enterprise search server
- ▶ Provides popular search options
 - ▶ Type-ahead, Suggested searches, Facets, Geo Spatial
- ▶ Scalable, mature software, used in high profile projects
 - ▶ Over 400TB, almost 3.5 Billion pages indexed by Hathi Trust
- ▶ Emphasis on speed of retrieval of data

Fedora Commons Overview

- ▶ Ability to store any type of content
- ▶ Proven scalability to millions of stored objects
- ▶ Provides critical needs of digital repositories
 - ▶ Content versioning, checksum integrity, replication option
- ▶ Customizable to any data relationship or model type

Projected Migration Timeline (while we also do new projects)



Expanding Digital Library Services

- ▶ **Born-Digital Archiving**
 - ▶ Social Media Archiving
 - ▶ Special Collections
- ▶ **Data Curation**
- ▶ **Robust Video Services**
 - ▶ Communication Studies Archive

Other proposed projects

- ▶ **Scalable Digital Content Management**
 - ▶ Ephemera Project
- ▶ **Institutional Repository**

Islandora as a collaborative tool and research platform

- ▶ “Solution packs”
- ▶ “Virtual Research Environments”

Islandora as a collaborative tool and research platform

- ▶ “Solution packs”
 - ▶ Software add-ons that address institutional and content specific needs
 - ▶ Examples:
 - ▶ Institutional Repository
 - ▶ Books
 - ▶ Magazines and Newspapers
 - ▶ Video and Audio
 - ▶ Maps
 - ▶ Oral History
- ▶ “Virtual Research Environments”

Islandora as a collaborative tool and research platform

- ▶ “Solution packs”
- ▶ “Virtual Research Environments”

Islandora as a collaborative tool and research platform

- ▶ “Solution packs”
- ▶ “Virtual Research Environments”
 - ▶ Custom data models and software components that provide best-practice stewardship of data
 - ▶ Examples:
 - ▶ Life Sciences
 - ▶ Epidemiology
 - ▶ Virtual Microscopy
 - ▶ Laboratory Management

Islandora as a collaborative tool and research platform: Metadata support

- ▶ Evolving, depending on needs of user community
- ▶ Strongest in support of Library community
 - ▶ MODS
 - ▶ MARCXML
 - ▶ Dublin Core (simple, qualified)
 - ▶ MIX
 - ▶ PREMIS

Islandora as a collaborative tool and research platform: Metadata support

- ▶ **Evolving, depending on needs of user community**
 - ▶ MODS, MARCXML, MIX, PREMIS
 - ▶ VRA (Visual Resources Association) Core
 - ▶ Dublin Core (simple, qualified)
 - ▶ PB Core (DC extension for Public Broadcasting metadata)
 - ▶ Darwin Core (DC extension for biodiversity informatics)
 - ▶ RefWorks XML (ProQuest, RefWorks Tagged Format in XML)
 - ▶ mzXML (file format for proteomics mass spectrometric data)
 - ▶ EMBL XML (Molecular Biology, nucleotide information)

Virtual Research Environments: 1

- ▶ UPEI Herbarium Specimen Collection:

 - <http://vre2.upei.ca/herbarium/>

- ▶ Darwin Core metadata: <http://rs.tdwg.org/dwc/index.htm>

 - ▶ The Darwin Core is body of standards. It includes a glossary of terms ... intended to facilitate the sharing of information about biological diversity by providing reference definitions, examples, and commentaries.

Virtual Research Environments:2

- ▶ **University of North Texas Apiary Project:**
<http://demo.apiaryproject.org:8019/drupal/>
- ▶ research to identify how human intelligence can be combined with machine processes for effective and efficient transformation of textual museum specimen label information into high-quality machine-processible parsed data.



ROI Type: Primary Label [Edit](#)
 Transcription: Undefined
 Parsing: Undefined
[Delete ROI](#)

ROI Type: Barcode [Edit](#)
 Transcription: Undefined
 Parsing: Undefined
[Delete ROI](#)



My Queue



My Queue Summary + ?

1 specimens
 2 ROIs
[open queue](#)

ap-specimen:Specimen-2 - Analyzing ?

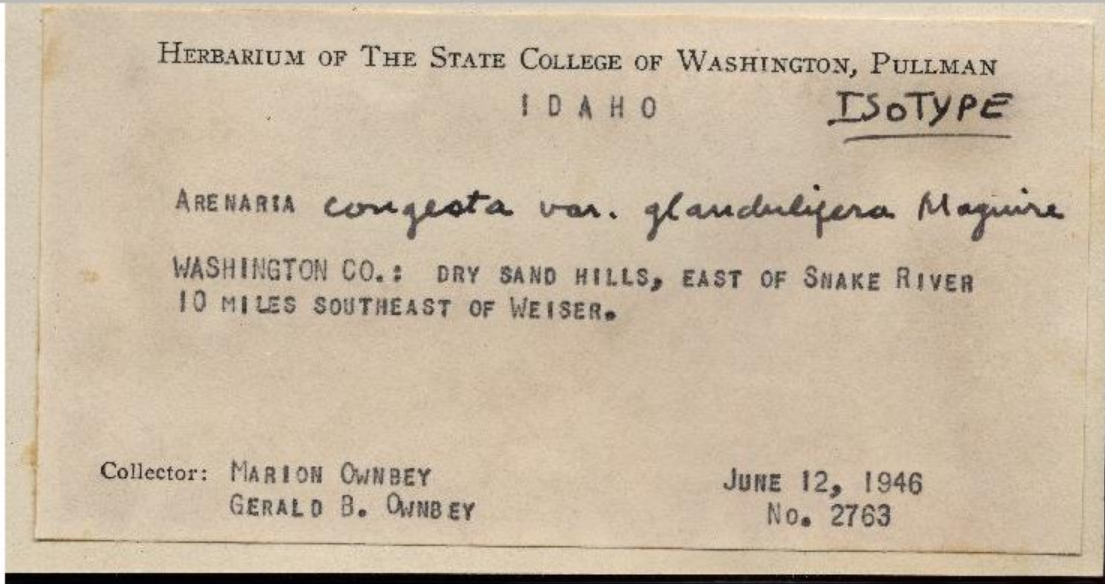
No roi currently selected
 Status: [▼](#)
[Add note](#)

ROI Legend ?

- Primary Label
- Annotation/Other
- Barcode
- Undefined

[Session Statistics](#)

[End Apiary Session](#)



Transcribe the text from the image to the left into the text area below.

Path:

Save text

My Queue



My Queue Summary + ?

- 1 specimens
- 2 ROIs
- [open queue](#)

ap-specimen:Specimen-2 - Transcribing ?

No roi currently selected

Status:

[Add note](#)

Session Statistics

End Apiary Session

Apiary Workflow

demo.apiaryproject.org/drupal/modules/apiary_project/workflow/index.php?workflow_id=1#

ANALYZE SPECIMEN TRANSCRIBE TEXT **PARSE TEXT** GETTING STARTED ?

TEXT IMAGE ? Parse ?

Who/When What Where HTML

Dry sand hills, east of Snake River 10 miles southeast of Weiser

Path: p

Who and When

- Verbatim Collector(s):
- Collector Number:
- Additional Collector Numbers:
- Verbatim Collection Date:
- Verbatim Determiner(s):
- Event Remarks:
- Verbatim Identification Date:
- Verbatim Annotation Date:
- Verbatim Annotator(s):

What

- Common Name:
- Taxon Remarks:
- Special Collection/Project Name:
- Other Barcode/Accession Numbers:
- Barcode Number:
- Herbarium Code:
- Previous Owing Institutions:
- Verbatim Field Notes:
- Identification Remarks:

Save Parsed Text Find taxa via uBio Parse using HERBIS

My Queue

ap-specimen:Specimen-2 - Parsing ?

No roi currently selected

Status: select a status

Add note

My Queue Summary + ?

1 specimens

2 ROIs

open queue

Session Statistics

End Apiary Session

start

Apiary Workflow - Go... Microsoft PowerPoint ...

EN 10:28 AM

Integration & Collaboration

▶ **Drupal:**

- ▶ ... is the Library standard for building presentation and interaction layers
- ▶ ... enables rapid development of Web sites and dynamic interfaces
- ▶ ... will replace Red Dot as the CMS driving the Library Web site

▶ **Solr:**

- ▶ ... can index any textual content, both metadata and text content
- ▶ ... can index across collections, across repositories, across systems

Next Digital Library “event”

- ▶ **Project oriented**
 - ▶ presentations and demos by project managers/participants
- ▶ **In the Research Commons**
- ▶ **Some examples**
 - ▶ Communications Studies News Archive
 - ▶ Social Media archiving
 - ▶ Digital preservation
 - ▶ Metadata
 - ▶ Islandora developments
 - ▶ Content oriented projects (maps, data, national and international collaborations)
- ▶ **Friday April 13, 2012**

Questions & Discussion



Los Angeles Actors Theater rehearse a scene from "Waiting for Godot" in Palmdale (1977)

- ▶ To remain informed:
 - ▶ Web site: <http://digital.library.ucla.edu>
 - ▶ Blog: <http://blogs.library.ucla.edu/digitallibraryprogram>
 - ▶ Twitter: <http://www.twitter.com/ucladiglib>
 - ▶ Facebook: <https://www.facebook.com/pages/UCLA-Digital-Library-Program/193496400490>