

# The Sheet Music Consortium: A Specialized Open Archives Initiative harvester project

Stephen Davison

University of California, Los Angeles

ISMIR 2003, October 28

# Outline

- ◆ The Open Archives Initiative (OAI) Protocol
- ◆ Applying OAI to Sheet Music Collections Online: Benefits
- ◆ Short History of the OAI Sheet Music Project
- ◆ Data Mapping Issues
- ◆ Demonstration of the OAI Sheet Music Service Provider
- ◆ Service Provider Usability Study
- ◆ Future Plans: Expanding the Service to Other Collections

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

- ◆ Defines format for metadata exchange
- ◆ Like Web search engines
  - Obtain metadata from servers worldwide
  - Stores data in a database
  - Responds to queries
- ◆ Like Union Catalogs
  - Stores holdings of specific collections
  - Can provide value-added services
  - Cooperative venture between institutions
- ◆ Centralized data repository: not distributed searching

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

- ◆ Based on concept of metadata harvesting
- ◆ Two types of participants:
  - Data providers
  - Service providers
- ◆ Two types of software applications:
  - Repositories
  - Harvesters
- ◆ Data providers operate OAI-compliant repositories that expose their metadata to harvesters
- ◆ Service providers operate metadata harvesters that issue OAI requests as a means of collecting metadata from repositories
- ◆ Data is exchanged as a stream of XML data

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

## ◆ An OAI Data Provider:

- Creates and updates data
- Date stamps data records so a harvester can recognize new and/or updated data
- Responds to harvester requests for data updates
- Can provide data to any number of different harvesters

## ◆ An OAI Service Provider:

- Harvests metadata on a regular basis
- Provides searching and other services
- Provides links back to the items in repositories

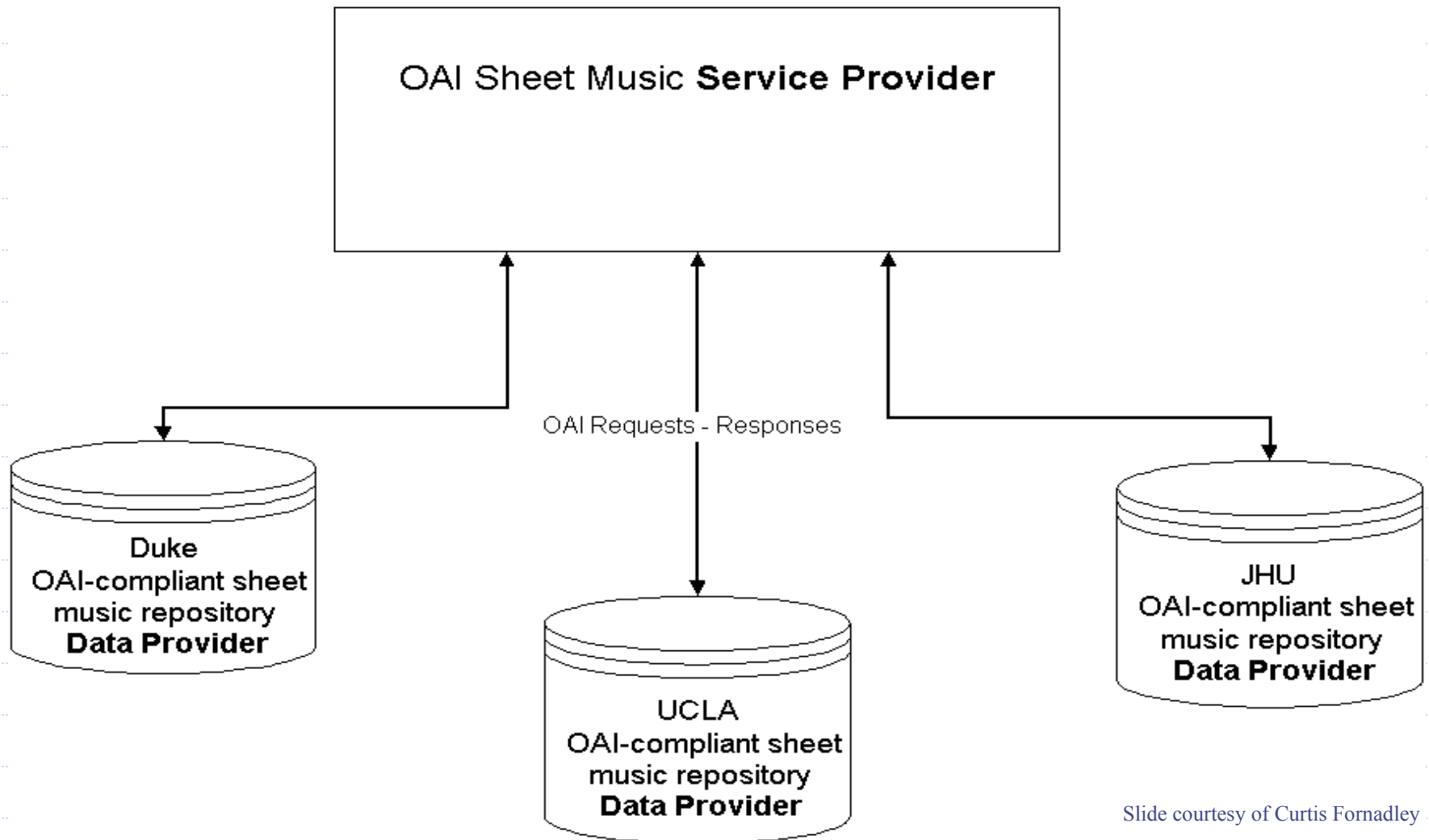
# Important OAI Verbs

- ◆ Identify
  - Returns information about a repository
- ◆ ListMetadataFormats
  - Returns available metadata formats
- ◆ ListSets
  - Returns set structure within repository
- ◆ GetRecord
  - Arguments: record identifier; metadata format
- ◆ ListIdentifiers
  - Returns list of record identifiers within (optional) selection criteria
- ◆ ListRecords
  - Returns records within (optional) selection criteria



<http://digidev.library.ucla.edu/oai>

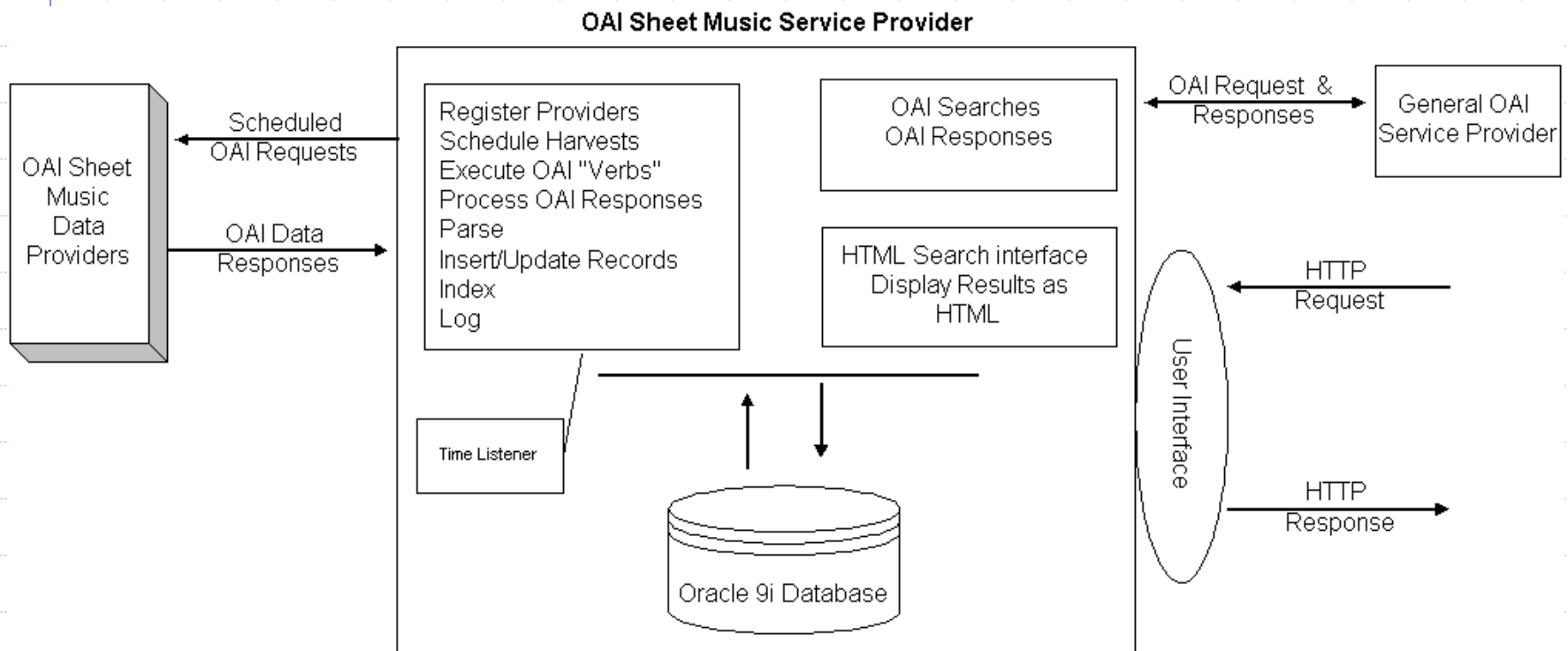
# OAI Sheet Music Data and Service Providers



Slide courtesy of Curtis Fornadley



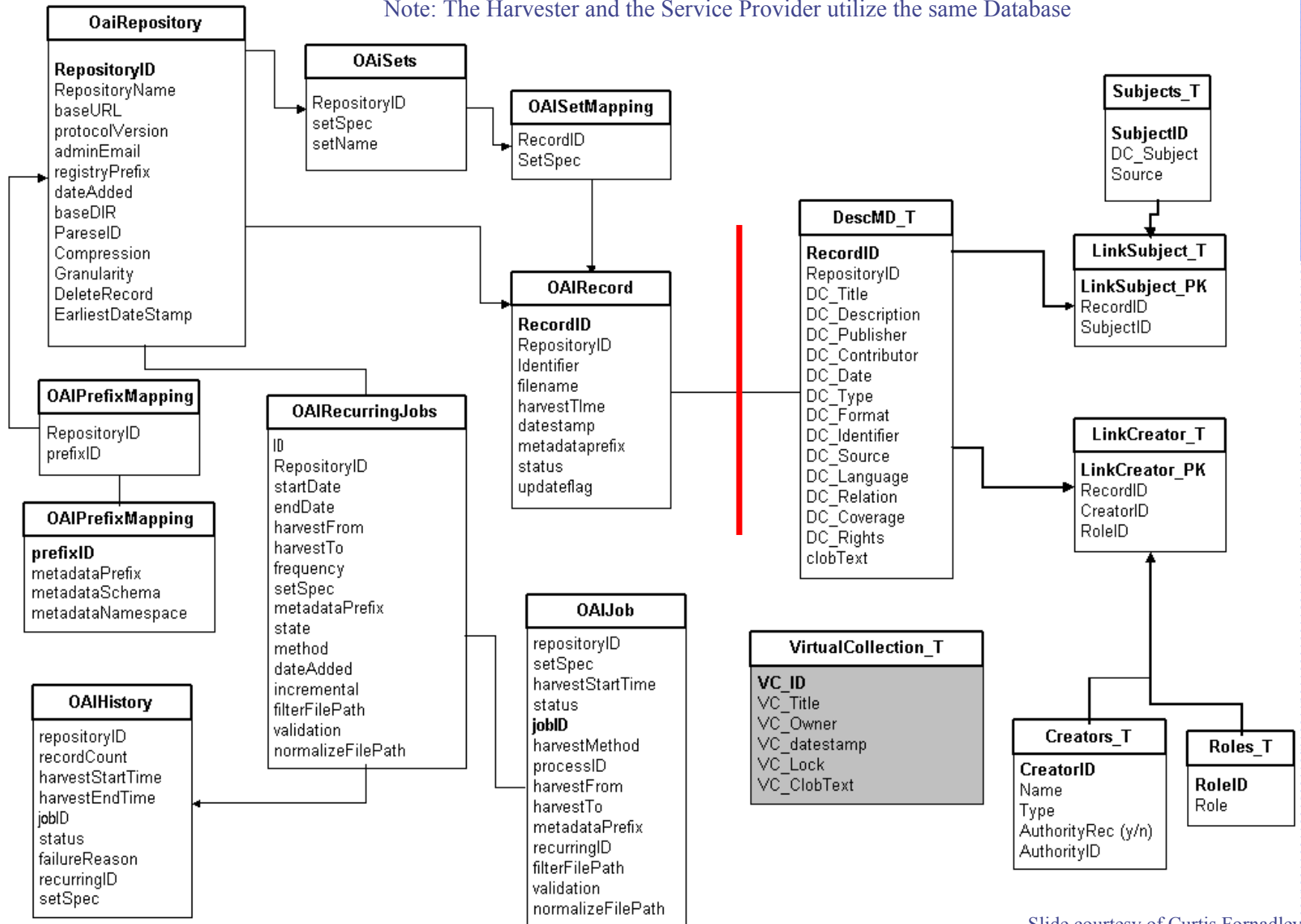
# Design of OAI Sheet Music Service Provider



Slide courtesy of Curtis Fornadley

# UCLA OAI Sheet Music Harvester Database Schema

Note: The Harvester and the Service Provider utilize the same Database



# OAI and Sheet Music: Benefits

- ◆ “Union catalog” of sheet music; collections cross-searchable
- ◆ “One-stop shopping”: not necessary to know where the different collections reside
- ◆ Can easily find variant versions of the same sheet in the collections
- ◆ Items become “known” by a variety of harvesters; digital items gain from being used in a variety of contexts, leading to ...
- ◆ A broader audience for sheet music collections, including non-musicians (sociologists, cultural historians, art historians, etc.)

# Short History of the Project

- ◆ Fall 2001
  - UCLA, Indiana, Johns Hopkins form steering team
- ◆ March 2002
  - Planning meeting at IU Bloomington. Included representatives from a dozen institutions, including Duke, Brown, Chicago, NYU, UC Berkeley
  - Planning groups formed:
    - ◆ Technical design
    - ◆ Data mapping
    - ◆ Interface design/Usability

## Short History of the Project (cont.)

- ◆ March-August 2002
  - Harvester and data provider development
  - Harvester upgrade to OAI 2.0
- ◆ September-October 2002
  - Test load of data, debugging
- ◆ November 2002 – February 2003
  - Usability testing at UCLA, Indiana, Johns Hopkins, Duke, Brown
- ◆ March-August 2003
  - Redesign of service provider
  - Duke data provider development
  - Data reharvested
- ◆ September 2003
  - Service provider launched

# Data Mapping Guidelines

- ◆ Mapping to Dublin Core from existing metadata
- ◆ Creation of metadata for new collections
- ◆ LC mapping existed already
  - MARC to DC
- ◆ Simple Dublin Core to start
- ◆ Qualified Dublin Core planned

# Unqualified Dublin Core

<b>Identifier</b> Unique reference (URL)	<b>Title</b>	<b>Creator</b>
<del><b>Contributor</b> Not used. Use Creator instead.</del>	<b>Publisher</b>	<b>Date</b>
<b>Description</b>	<b>Format</b>	<b>Source</b>
<b>Type</b>	<b>Language</b>	<b>Subject</b>
<b>Relation</b>	<b>Coverage</b>	<b>Rights</b>

# DC\_Title

- ◆ “A name given to the resource. Typically, a Title will be a name by which the resource is formally known.”
- ◆ **Title (and variants), first line, first line of refrain/chorus, and other forms in which the work might be known. The definitive title should be first listed title.**
- ◆ **Recommendation: Following AACR2 capitalization rules. Omit the statement of responsibility.**
- ◆ April fool [song title]
- ◆ Martini [show title]
- ◆ The Cornell Masque presents Martini [alternative title]
- ◆ Poets say, that an April day is much like the ways of lovers [first line]
- ◆ Gloomy season, without reason, makes our hearts seem far apart [first line of chorus]



# DC\_Creator

- ◆ “An entity primarily responsible for making the content of the resource. Examples of a Creator include a person, an organization, or a service. Typically, the name of a Creator should be used to indicate the entity.”
- ◆ **Composer, lyricist, arranger. Other contributors, including illustrator, artist, photographer, etc. Recommend last name first.**
- ◆ **Recommendation: Invert name. Use the authorized form of name where possible. If needed (e.g. for an alias) repeat the field for the alternative form.**
- ◆ Shaw, J. B., Jr. [lyricist]
- ◆ Stanley, J. Selwyn [composer]
- ◆ Lyrics and Music By Irving Berlin. Book by Harry B. Smith.

# DC\_Publisher

- ◆ “An entity responsible for making the resource available. Examples of a Publisher include a person, an organization, or a service. Typically, the name of a Publisher should be used to indicate the entity.”
- ◆ **Place of publication and name of the publishing person or corporate entity. Use AACR2 formatting if possible.**
- ◆ **Follow standard AACR2 formatting if possible (Place followed by Publisher). It is likely that the place will be split off into a separate field in the DCQ phase.**
- ◆ New York : Cornell Masque Ithaca
- ◆ New York : M. Witmark & Sons

# DC\_Date

- ◆ “A date associated with an event in the life cycle of the resource. Typically, Date will be associated with the creation or availability of the resource. Recommended best practice for encoding the date value is defined in a profile of ISO 8601 [[W3CDTF](#)] and follows the YYYY-MM-DD format.”
- ◆ **Date of publication. The most recent date to appear on the music, or, the actual date of publication if not present but known. Include other dates (e.g. date of composition) if known. Codes “c” for copyright and “ca.” for circa in front of the date is allowed for now. Use repeated DC fields for each date if needed.**
- ◆ 1920 [date of publication]
- ◆ 2002-07-17 [date of digitization]
- ◆ c1902
- ◆ [ca. 1800]
- ◆ 18--

# DC\_Subject

- ◆ “The topic of the content of the resource. Typically, a Subject will be expressed as keywords, Key phrases or classification codes that describe a topic of the resource. Recommended best practice is to select a value from a controlled vocabulary or formal classification scheme.”
- ◆ **Subject headings, usually from controlled vocabularies such as LCSH, LCTGM, AAT, etc. Also local vocabularies.**
- ◆ **Recommendation: Attempt to map to existing vocabularies wherever possible.**
- ◆ Songs with piano [LCSH]
- ◆ Dance [TGM]
- ◆ Music boxes [TGM]
- ◆ Musical revues & comedies [TGM]
- ◆ Women [TGM]

# Sheet Music OAI Service Provider

- ◆ Currently provides searching and access to collections at:
  - Indiana University
  - Johns Hopkins University
  - Library of Congress
  - Duke University
  - UCLA
- ◆ Future additions to service include:
  - Brown University
  - National Library of Australia

# OAI Sheet Music Service Provider

(October 2003)

- ◆ Library of Congress, American Memory
  - 47,528 records (data & images)
- ◆ Duke University
  - 17,698 records (3,042 with images)
- ◆ Indiana, Lilly Collection
  - 17,417 records (data only)
- ◆ Johns Hopkins, Levy Collection
  - 11,590 records (data & images)
- ◆ UCLA, Archive of Popular American Music
  - 2,173 records (data & images)

# OAI Sheet Music Service Provider

<http://digital.library.ucla.edu/sheetmusic/>

# Usability Testing

- ◆ Would a community specific OAI harvester be useful for sheet music users?
- ◆ How might we improve the prototype for intended users?
- ◆ Usability testing design: focus groups and scenario-based testing
- ◆ Mellon Foundation support
- ◆ Profile of participants: performers and historians; undergraduate, graduate, and faculty; some from non-music disciplines



# Usability Testing

## ◆ Focus groups

- 6 researchers per group per site (total ca. 30)
- Discussion of information seeking needs of participants

## ◆ Scenario-based testing

- 5 one-on-one scenarios site (total 25)
- Scenario using current service provider
- Comments of prototype service provider

# Usability Testing Results

- ◆ Focus groups
  1. Research needs
  2. Search strategies
  3. Desired content
  
- ◆ Scenario-based tests
  1. Inconsistent metadata
  2. Terminology
  3. Navigation
  4. Searching

# Issues

- ◆ Creating a metadata harvesting service vs. an institutional resource
  - Some of the functionality that focus group participants identified is simply beyond the limits of what the metadata can support.
  - What is appropriate functionality for a harvesting service?
  - What functionality should be addressed at the institutional level?
    - ◆ Example: Links to sound files

## Where to from here?

- ◆ Additional collections
- ◆ Enhanced services
- ◆ Cooperation/interoperability with OAI harvester for musical scores
- ◆ Repurposing metadata for other projects

# Sheet Music Consortium Steering Team

- ◆ Stephen Schwartz, Head of Library Information Systems Development, UCLA (Convener)
- ◆ Rebecca Graham, formerly Head of Library Systems, Johns Hopkins University
- ◆ Kristine Brancolini, Director, Digital Library Program, Indiana University
- ◆ Stephen Davison, Music Librarian for Special Collections & Digital Projects Consultant, UCLA
- ◆ Cynthia Requardt, Head of Special Collections, Johns Hopkins University

# Other Contributors

- ◆ Curtis Fornadley, Digital Library Lead Programmer, UCLA
- ◆ Brian Harrington, formerly Curator, Digital Collections, Johns Hopkins University
- ◆ John Dun, Assistant Director for Technology, Digital Library Program, Indiana University
- ◆ Teal Anderson, Usability Specialist, Digital Knowledge Center, Johns Hopkins University

# Resources & Contacts

- ◆ OAI Website
  - <http://www.openarchives.org>
- ◆ OAI Sheet Music Service Provider
  - <http://digital.library.ucla.edu/sheetmusic/>
- ◆ Sheet Music Consortium contacts:
  - Stephen Schwartz, [shs@library.ucla.edu](mailto:shs@library.ucla.edu)
  - Kristine Brancolini, [brancoli@indiana.edu](mailto:brancoli@indiana.edu)
  - Cynthia Requardt, [Cynthia.Requardt@jhu.edu](mailto:Cynthia.Requardt@jhu.edu)
  - Stephen Davison, [sdavison@library.ucla.edu](mailto:sdavison@library.ucla.edu)