Next Generation Sheet Music Consortium

Michelle Dalmau, Digital Projects & Usability Librarian
IU Digital Library Program
Digital Library Brown Bag Series, 30 November 2011
SMC Hot Topics

- Motivations
- Historical Overview
- Next Generation
  - Metadata Harvesting Services
  - Data Provider Services
  - End-User Services
- Sustainability

Parts of this talk have been borrowed as-is or adapted from previous presentations given by Stephen Davison and Jenn Riley.
To build a *community-based* shared metadata resource for sheet music held in repositories worldwide

- Address bibliographic challenges surrounding sheet music (description either lacking or not standardized)

- Accommodate significant interest in and high usage of sheet music; provide access to a wide-ranging audience from enthusiasts to performers, from genealogists to researchers

- Preserve the ephemeral nature of the publication: covers, advertisements, different arrangements
Historical Overview

- **2001**
  - Initial discussions at CNI and DLF
  - Established initial partnership: UCLA, IU, JHU, Duke

- **2002**
  - Planning meeting in Bloomington

- **2003**
  - Launch of SMC Web site: [http://digital.library.ucla.edu/smc](http://digital.library.ucla.edu/smc)

- **2003-2010**
  - Additional collections harvested, but no new functionality
    - Library of Congress, Maine Music Box, National Library of Australia
Historical Overview: OAI-PMH Segue

- Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) facilitates the “harvesting” or collecting of metadata from various repositories that can be aggregated for “federated” searching.
  - Data Providers expose structured metadata (Dublin Core and MODS) for harvesting
  - Service Providers harvest the metadata for various purposes, most commonly to provide searching across multiple repositories (e.g., OAIster)

- The initial aim of the Consortium was to demonstrate the application of the OAIPMH protocol to a specialized set of digital collections, and, more specifically, to build a sheet music service that would provide unified access to those collections.

See Fall 2003 brown bag presentation “OAI and the Sheet Music Consortium” by Jon Dunn and Jenn Riley for technical details.
Pause for Questions
Next Generation SMC

- **2007-2008: IMLS Planning Grant**
  - January 2008: Planning meeting, Bloomington, IN
  - February 2008: Public meeting @ Music Library Association Annual Meeting, Newport, RI
  - Ongoing needs assessment of data providers and end-users

- **2009-2012: IMLS Leadership Grant**
  - Community-building: convene advisory board, outreach
  - New data harvester: more robust data harvesting
  - New service provider: better searching, support user-contributed content
  - Metadata standards: greater uniformity
  - Metadata creation tools: greater participation
  - mapping tool, static repository & repository gateway
Leverage jOAI software package (open source, well supported) for the harvesting and service provider infrastructure

Automate the harvesting workflow to more efficiently support the addition of new data providers and regular updates for existing providers (records refresh)

Broaden support beyond Dublin Core to include QDC (ContentDM users) and MODS; records stored in MODS

Improve automated metadata remediation/normalization (dates, de-duplication, etc.)

Harvest thumbnails based on UIUC’s Thumbgrabber

Setup Static Repository Gateway for those data providers who do not have the tech support to setup their own data provider service (or who do not have built in provider services found in ContentDM, Dspace, etc.)
Next Gen SMC: Data Provider Services

- SMC Metadata Mapping Tool: http://www.dlib.indiana.edu/smcmigrator/
- Sheet Music Metadata Guidelines that are agnostic in terms of format, promote share-ability, and provide guidance as opposed to prescriptive rules for description http://digital2.library.ucla.edu/sheetmusic/aboutProject.html
Next Gen SMC: Metadata Mapping Tool
Next Gen SMC: Metadata Mapping Tool
Next Gen SMC: Metadata Mapping Tool

Upload a File

Choose a file to upload and map: [Choose File] Pitt Choral Collection.xls

Administrator's email: mdalmau@indiana.edu

Where are you planning on hosting the mapped XML file?

URL for XML file location: http://www.dlib.indiana.edu/sheetmusic/

Choose an existing mapping: [PittChoral]

Or enter a new collection/mapping name: PittChoral

NOTE: The XML filename will be the collection/mapping name you provide.

Your URL will be:

http://www.dlib.indiana.edu/sheetmusic/PittChoral.xml
Next Gen SMC: Metadata Mapping Tool

Map Metadata
Validate Metadata
About
Help
Using the Tool
Preparing Files
Glossary
Log In

Step 1: Upload  Step 2: Import  Step 3: Crosswalk  Step 4: Validate

The first row of the uploaded file is:

#
Title
Composer
Arr.
Status
Bib #
Notes

What type of information is this?

First Data Record  Field Names  Start Over

URL: http://www.dlib.indiana.edu/smcmigrator/uploadproc.php
Comments: diglib@indiana.edu
Metadata Mapping Tool brought to you by the IU Digital Library Program and the UCLA Digital Library Program
IU Libraries Privacy Policy
Copyright 2011, The Trustees of Indiana University
### Next Gen SMC: Metadata Mapping Tool

#### Step 1: Upload

<table>
<thead>
<tr>
<th>Row</th>
<th>Field Name</th>
<th>Source</th>
<th>Record 0</th>
<th>Record 1</th>
<th>Record 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Discard</td>
<td>#</td>
<td>1</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>Title</td>
<td>none</td>
<td>Title</td>
<td>Four Winds</td>
<td>Loon</td>
</tr>
<tr>
<td>3</td>
<td>Composer</td>
<td>lcmaf</td>
<td>Composer</td>
<td>McKay, George Frederick</td>
<td>Strom, Ina L.</td>
</tr>
<tr>
<td>4</td>
<td>Discard</td>
<td>Arr.</td>
<td>SATB</td>
<td>SATB</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Discard</td>
<td>Status</td>
<td>Imported to Voyager</td>
<td>No rec. in OCLC</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Call Number</td>
<td>n/a</td>
<td>Bib #</td>
<td>5430062</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Notes/Description</td>
<td>n/a</td>
<td>Notes</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Additional Static Fields

<table>
<thead>
<tr>
<th>Language</th>
<th>Source</th>
<th>English</th>
<th></th>
</tr>
</thead>
</table>
DC/QDC/MODS metadata validation automatically generates a report as part of the Mapping Tool

Static repository XML files generated independent of the Mapping Tool can also be validated

Validation report or “report card” is organized by:

- Gateway Harvesting, Browsing Support, Search Refinements, Record Context (i.e., language or identification numbers) and Data Integrity (i.e., CVs)
- Recommendations based on the Sheet Music Metadata Guidelines and OAI-PMH requirements; no bad grades are ever given; just incentives to attain A+ in all categories
- Only two possible errors could result: nonexistent titles (minimum metadata requirement) or MODS/DC/QDC schema validation problem
Next Gen SMC: Metadata Mapping Tool

Congratulations!

Your collection meets the minimum requirements for submission to the Sheet Music Consortium static repository. To see how your data will be used, or to identify ways in which you might make your data richer or support more end-user functionality, please review the report categories below.

- Gateway Harvesting - 3 messages
- Browsing Support - 2 messages
- Search Refinements - 2 messages
- Record Context - 2 messages
- Data integrity - 2 messages

Retrieve XML File

- `<oai:repositoryName>` - Make sure the name inside this tag correctly identifies the collection you are mapping.
- `<oai:baseURL>` - The base URL in the generated file will be a creation of the following parameters:
  "http://oailgateway.library.ucla.edu/gatewaynet/oai.aspx" plus (Step 1: Upload's URL for data harvesting minus "http://")

So the `<oai:baseURL>` in the .xml file generated from the Map Metadata - Step 1: Upload help example would be the following:
http://oailgateway.library.ucla.edu/gatewaynet/oai.aspx/www.example.edu/webDirectory/SampleFile.xml

Make sure the URL in your generated file reflects where the file will be located and accessible to the OAI-PMH harvester.
**Next Gen SMC: Metadata Mapping Tool**

---

**Metadata Report for pittchoralmichelle**

**Congratulations!**

Your collection meets the minimum requirements for submission to the Sheet Music Consortium static repository. To see how your data will be used, or to identify ways in which you might make your data richer or support more end-user functionality, please review the report categories below.

---

**Gateway Harvesting - 3 messages**

**Requirements for Submission to Static Repository Gateway**

We encourage institutions to share metadata records with the Sheet Music Consortium via the OAI Static Repository Gateway. In order for a file to be submitted to the OAI Static Repository Gateway, it needs to contain the following information: a repository name that correctly identifies the collection, an e-mail address that can be used as a point of contact for the collection, and a base URL that reflects where the file will be located and accessible to the OAI-PMH harvester. For more information on ensuring that a file meets the minimum requirements for submission to the Sheet Music Consortium static repository, see [Map Metadata - Step 4: Validate](#).

<table>
<thead>
<tr>
<th>Check</th>
<th>SMC Recommendation</th>
<th>Feedback</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base URL</td>
<td>In order to submit your file to the OAI Static Repository Gateway, the URL you provide must begin with the path to the OAI Static Repository Gateway - <a href="http://oaigateway.library.ucla.edu/gatewaynet/oai.aspx/">http://oaigateway.library.ucla.edu/gatewaynet/oai.aspx/</a> and end with the address of the XML file, without the opening &quot;http://.&quot;</td>
<td>Congratulations! The URL you have provided begins with the correct path to the OAI Static Repository Gateway. Please make sure that <a href="http://oaigateway.library.ucla.edu/gatewaynet/oai.aspx/">http://oaigateway.library.ucla.edu/gatewaynet/oai.aspx/</a> <a href="http://www.dlib.indiana.edu/PittChoralMichelle.xml">http://www.dlib.indiana.edu/PittChoralMichelle.xml</a> is the correct value for the full URL.</td>
</tr>
<tr>
<td>Correct</td>
<td>The email address provided will be used for communication regarding the collection.</td>
<td>Please make sure that '<a href="mailto:mdalmau@indiana.edu">mdalmau@indiana.edu</a>' is the correct email contact for the repository.</td>
</tr>
<tr>
<td>Repository Name Correct</td>
<td>Since the repository name allows users to locate records in your collection by browsing, it is important for the repository name to be correct.</td>
<td>Please make sure that 'PittChoralMichelle' is the correct name for the repository.</td>
</tr>
</tbody>
</table>

**Browsing Support - 2 messages**
Next Gen SMC: Metadata Mapping Tool

Data that will support browsing functionality in the discovery interface

Titles, names, subjects, and dates are currently used in the discovery system as key access points for browsing. Including this information will enable users to locate your records by browsing. For example, users interested in sheet music published in the 1940s might browse for these records. Providing a copyright date for each record will help users to locate resources, especially your records, when browsing by date.

<table>
<thead>
<tr>
<th>Check</th>
<th>SMC Recommendation</th>
<th>Feedback</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject/Topic</td>
<td>Subject metadata is an important access point for end users to find all the songs about a given subject. This is a common way end users want to access the Sheet music collection.</td>
<td>32/32 of your resources do not contain subject information. Consider including this information, as it will allow users to locate your records through browsing by subject. [See records]</td>
</tr>
</tbody>
</table>

| Example Records |
|-----------------|-----------------|-----------------|
| Standard        | OAI identifier  | Record Title    |
| MODS            | PittChoralMichelle0 | Four Winds      |
| MODS            | PittChoralMichelle1 | Loon            |
| MODS            | PittChoralMichelle2 | Conchita         |
| MODS            | PittChoralMichelle3 | Village Bells   |
| MODS            | PittChoralMichelle4 | O Watchers of the Stars |
| QDC             | PittChoralMichelle0 | Four Winds      |
| QDC             | PittChoralMichelle1 | Loon            |
| QDC             | PittChoralMichelle2 | Conchita         |
| QDC             | PittChoralMichelle3 | Village Bells   |
| QDC             | PittChoralMichelle4 | O Watchers of the Stars |
| DC              | PittChoralMichelle0 | Four Winds      |
| DC              | PittChoralMichelle1 | Loon            |
OAI Static Repository Gateway

Welcome, to our Static Repository Gateway. This service transforms Static Repository files (read the specs [here](#)) into OAI-PMH harvestable repositories.

Please be aware of the following restrictions we set:

- The OAI Static Repository **must** conform to the [OAI Static Repository specification](#).
  
  Specifically:
  - The OAI Static Repository **must** validate against the [XML Schema](#) that defines the format for OAI Static Repositories.
  - The network-location of the OAI Static Repository **must** be a HTTP address, i.e. the network-location **must** start with `http://` or `https://`.

- In addition, our Gateway imposes the following restrictions:
  - The network-location of the OAI Static Repository **must** have a suffix `.xml`. E.g. [http://digidev.library.ucla.edu/static-repository/sample.xml](http://digidev.library.ucla.edu/static-repository/sample.xml)
  - The network-location of the OAI Static Repository **must** contain characters from the following set: `[a-z] [A-Z] [0-9] / ' + * ' % ' ! ' ' ' ' ' ' ' ' ~


Sheet Music Cataloging Guidelines serve both collection managers who have yet to describe their sheet music collection and those who have described their sheet music but see opportunities for improvement.

- Metadata descriptions of fields and how the fields/formatting impact discovery (what and why)
  - Title, Name, Publisher, Subject, ID Numbers, Relational Identifiers, Notes, Dates and Language

- The validation tool provides recommendations based on the Sheet Music Metadata Guidelines in support of interoperable metadata that the SMC portal web site can then leverage to provide more advanced discovery and results manipulation options.
Pause for Questions
“New and Improved” web site with more robust and intuitive searching and browsing capabilities

More content (because of the low-barrier methods for sharing sheet music metadata described earlier); 16 institutions (up from 7)

Prominent and intuitive Virtual Collections feature for compiling and annotating records that can be public, private or shared with students or colleagues

User-contributed tagging (semi-structured: genre and instrumentation) and annotations (unstructured for evaluative content, links to related resources, etc.)

Zotero integration/export of citations; RSS feeds (new collections, re-harvested collections, etc.); and OAI-ORE or Linked Data implementation (experimental)

See Fall 2008 brown bag presentation “An Introduction to the Open Archives Initiative Object Reuse and Exchange (OAI-ORE)” by Jenn Riley for more information.
Next Gen SMC: Have a Look!

http://digital2.library.ucla.edu/sheetmusic/
UCLA has committed to indefinitely support harvesting and service-provider services (at least as-is by the grant’s end)

IU has committed to indefinitely support the metadata mapping tool and validation service

However, technology evolves; user needs shift, etc. so this is the clincher if the SMC is to remain a relevant organization/resource (indefinitely)

- Tiered membership model (based on cost-share or funds)
- Sponsor/Support “collection grants” to help smaller organization describe and digitize sheet music collections
- Help solicit institutional support by providing evidence of usage (e.g., statistics, user stories, etc.)
- Do you have ideas?
Shout Out to SMC Project Team

**UCLA**
- Stephen Davison, PI, Head of the Digital Library Program
- Henry Chiong, Digital Library Architect
- Parinita Ghorpade, Programmer
- Claudia Horning, Metadata Specialist
- Elizabeth ("Lisa") McAulay, Usability Lead
- Erika Troll, Interface Design

**IU**
- Jenn Riley/Michelle Dalmau/Jon Dun, co-PI
- Brian Keese, Programmer
- Julie Hardesty, Interface Design and Usability Lead
- Elizabeth Munson, Metadata Specialist
Questions?