Music Documentation in Libraries, Scholarship, and Practice



June 4-6, 2012

RISM Data as Metadata for Digital Collections

Anne Graham and Deborah Pierce (University of Washington, Seattle)

English abstract:

The University of Washington Music Library holds a collection of rare music manuscripts and early editions including a portion dating from the seventeenth through nineteenth centuries that are indexed in RISM's series A/II project. To meet the needs of off-site researchers and reduce handling of the rare and fragile originals, the University of Washington Libraries digitized this collection between 2007 and 2010 with funding from an internal Libraries grant. A central enabling component of the digitization effort was the availability of RISM's high-quality metadata provided in electronic form by the RISM Zentralredaktion. The resulting online collection is freely available to any internet user and can be found at http://content.lib.washington.edu/mmweb.

This presentation will describe the methods that were used to transform the raw RISM metadata into standardized, searchable online metadata to accompany the visually rich display of navigable, printable, and zoomable images along with a description of the University of Washington Music Library Digital Scores Collection website, its functionality, and features. Others who are interested in digitizing and displaying their RISM collections online may find our experience of manipulating the raw data and harnessing it to describe the digital images to be helpful. Copies of the data dictionary used to describe the collection's metadata, including field origins, usage, format, and meanings, as well as mappings to Dublin Core, will be available to attendees.

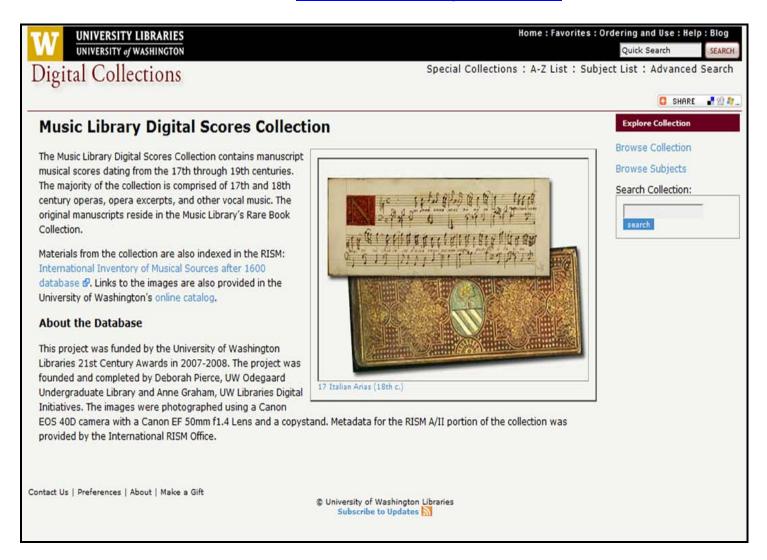
RISM Data as Metadata for Digital Collections

By

Anne Graham and Deborah Pierce University of Washington, Seattle

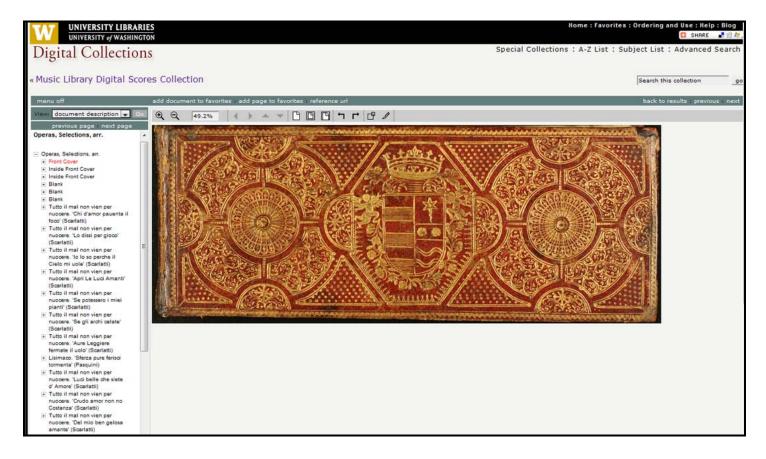
Presented at RISM's conference, *Music Documentation in Libraries, Scholarship, and Practice* in Mainz, Germany on June 4, 2012

The University of Washington Libraries began a project in 2007 to digitize and display online the portion of our rare music materials which are indexed in RISM's series A/II project -- music manuscripts representing composers born after 1570 and before 1770. After completion of the project, links to digital images in UW's digital collections database were added to both WorldCat and RISM's OPAC, thus bringing our digitized RISM materials directly to the world of music researchers via their own computers. The collection is served using CONTENTdm software and is available at http://content.lib.washington.edu/mmweb.



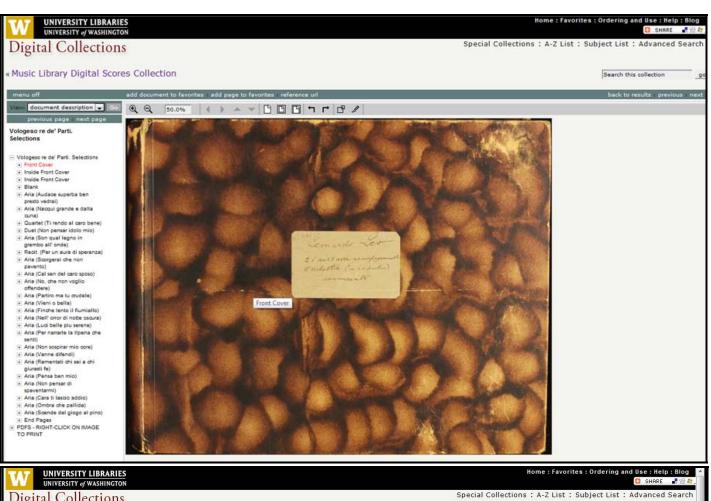
In the process, we also became the first U.S. library to utilize RISM cataloging in a digital collection.

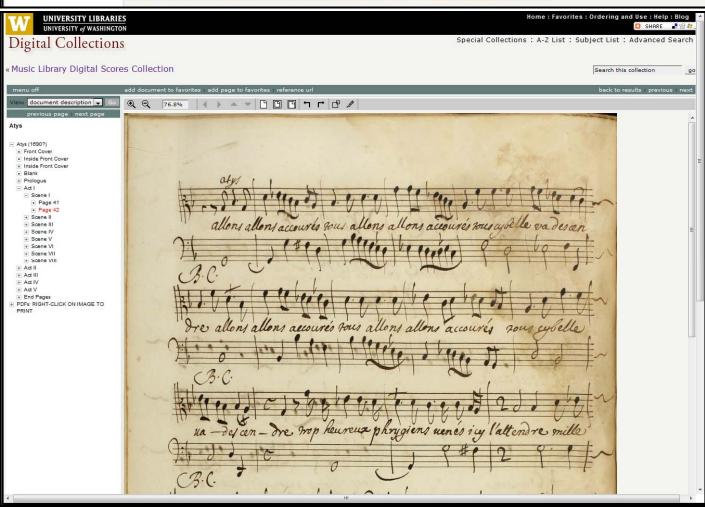
The manuscripts in the UW's Music Library that were indexed as part of the RISM series A/II consisted of 17th and 18th century operas, opera excerpts, and other vocal music including works by composers such as Christoph Willibald, Ritter von Gluck; Jean Baptiste Lully; Bernardo Pasquini; Alesandro Scarletti, and others.

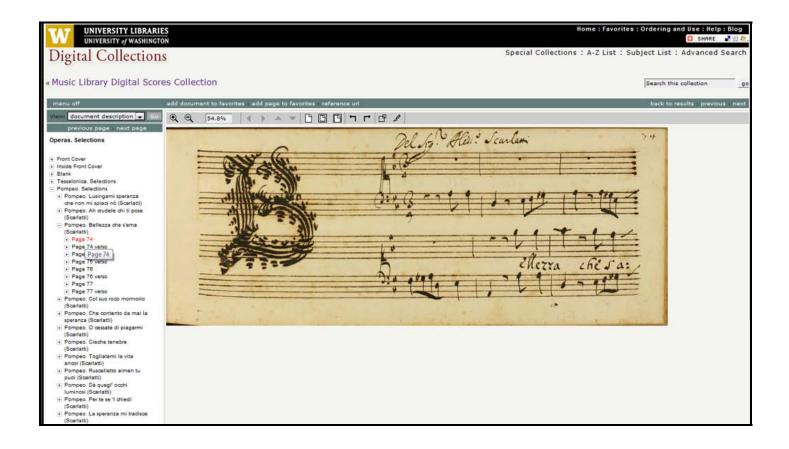


For quite some time now, librarians who shepherd rare musical materials have been living with the quandary of how to make copies of these materials easily available to researchers, while at the same time seeking to preserve them by handling the originals as little as possible. By 2007, other libraries had begun to digitize their rare music, and we started to think about digitizing ours too. We believed that doing so would address the preservation and access quandary, <u>and</u> we had some experience in this area from a previous project where we digitized our public-domain sheet music from the Pacific Northwest.

For this project we chose to digitize only our RISM artifacts which includes 30 artifacts or volumes, and represents 308 pieces of music. This set of manuscripts was chosen for two reasons: its popularity with patrons, and the fact that we were able to utilize existing RISM cataloging at no cost. Rather than digitizing our microfilm for this project, we chose to use copy photography techniques, because the end-user would have a more pleasant and realistic experience when viewing large, zoomable, crisp, full-color, digital images of the manuscripts.







Since our RISM series A/II materials had already been meticulously cataloged in the RISM series A/II project, we contacted RISM to see if they might provide us with that cataloging in digital format. We were delighted when Klaus Keil and Stephan Hirsh were willing to provide us with an export from their Kallisto database in Microsoft Access format. Note: other formats are available, including XML. This data gave us not only cataloging data for the artifacts, but also full analytic data for those artifacts that included multiple pieces of music in a single collection.

Metadata creation always begins with a lot of planning. We began our metadata creation by formulating a data dictionary which included all of the fields we wished to display to the end-user, as well as elements necessary to optimize searching, discovery, harvesting, and interoperability between our local instance of the CONTENTdm database and the internet at large CONTENTdm uses the Dublin Core metadata standard, so all fields were mapped to the appropriate Dublin Core element. This data dictionary specified two sets of metadata, one that described each artifact or volume as a whole and one that described each separate musical piece within the collected volumes. The data dictionary also detailed exactly what notation and style the contents of each field would use, as well as database-specific details, such as whether a field would be hidden, searchable, and/or required. Copies of the data dictionary are provided as an appendix.

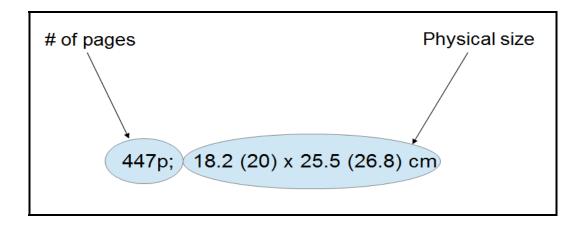
Music Library Manuscripts Collection

Data Dictionary date: July, 2008

Complete Volume Record:

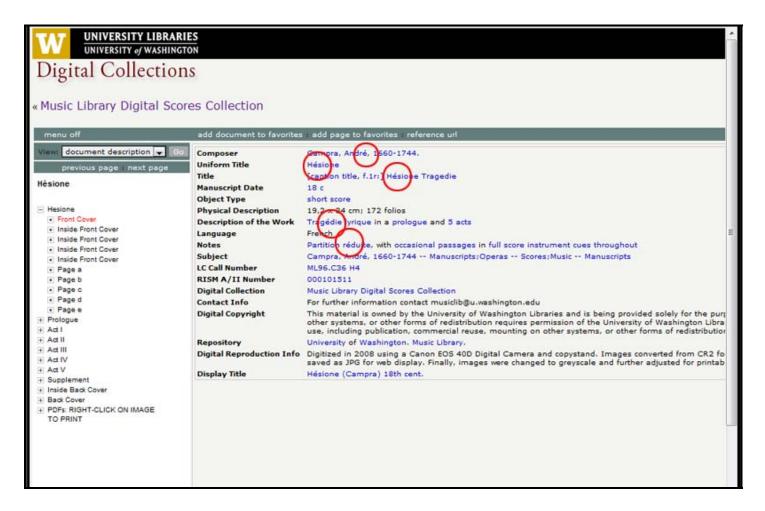
Project element	DC mapping	Comments
Composer	Creator: Search=yes, hidden=yes	Data taken from PEPRD field "rismhs". Example: Campra, Andre, 1660-1744. Note: remove diacritics from data files.
Composer (Unicode)	Creator: Search=yes, hidden=no	Data taken from PEPRD field "rismhs". Example: Campra, André, 1660-1744. Note: retain diacritics.
Uniform Title	Title: Search=yes, hidden=yes	Data taken from HSPR01 field "titreg" Note: remove diacritics from data files.
Uniform Title (Unicode)	Title: Search=yes, hidden=no	Data taken from HSPR01 field "titreg" Note: retain diacritics
Title	Title: Search=yes, hidden=yes	Data from HSPR01 field "diptit" Note: There will be a few entries which are blank in both Uniform title and title. In this case choose from fields: h3; h41830; kutit; or gnsart from the HSPR01 table. Note: remove diacritics from data files
Title (Unicode)	Title: Search=yes, hidden=no	Data from HSPR01 field "diptit" Note: There will be a few entries which are blank in both Uniform title and title. In this case choose from fields: h3; h41830; kutit; or gnsart from the HSPR01 table. Note: Retain diacritics
Manuscript date	Date: Search=no, hidden=no	Date that displays to public Data is found in various forms and fields in various formats, some encoded, we may input by hand from the coded information in the thematic catalog (ML1)

We then identified which RISM database fields contained the data that would match the requirements of the data dictionary. Kallisto is a relational database and CONTENTdm stores its data non-relationally as flat data files. Using a little reverse-engineering, we figured out which data fields from RISM's database mapped to the fields called for by our data dictionary. For example, two separate fields in RISM's relational database needed to be extracted and combined to create our physical description field.



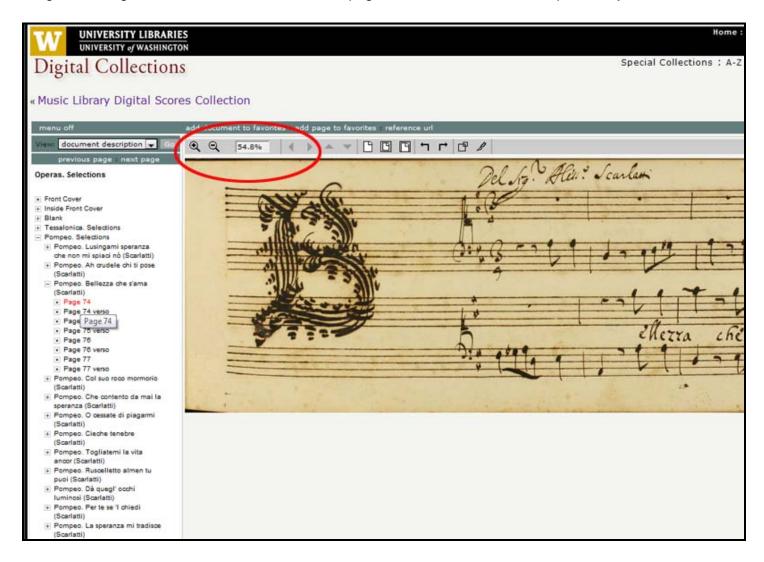
When all of the data had been extracted and placed into Excel, we performed some amount of data normalization and alterations to fit our goals. We converted some of the abbreviations to written-out text for our patrons' ease of use.

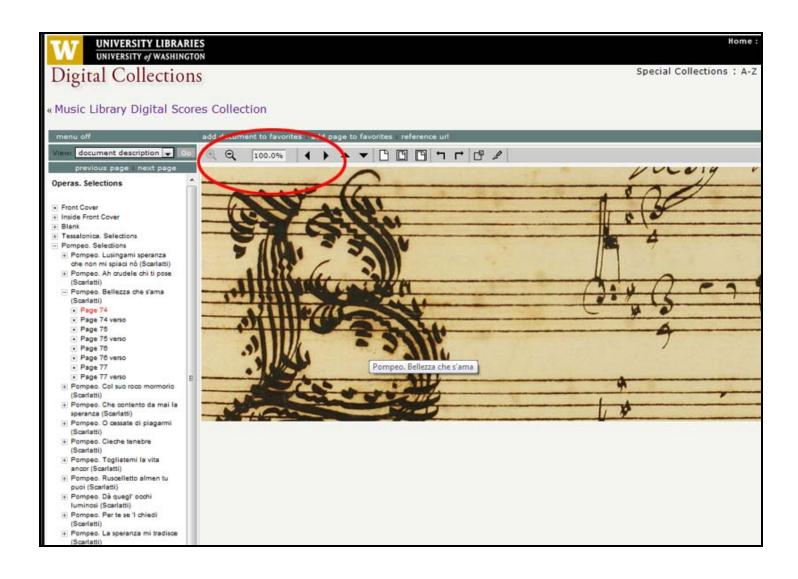
In addition, we needed to address the presence of diacritics in the data in order to allow users to search the database and retrieve results whether or not they typed "Leyermädchen" or "Leyermadchen." To accomplish this, we created two iterations for all fields containing diacritics. One field included diacritics and another field held the data without diacritics. The field without diacritics was searchable but was hidden from view.



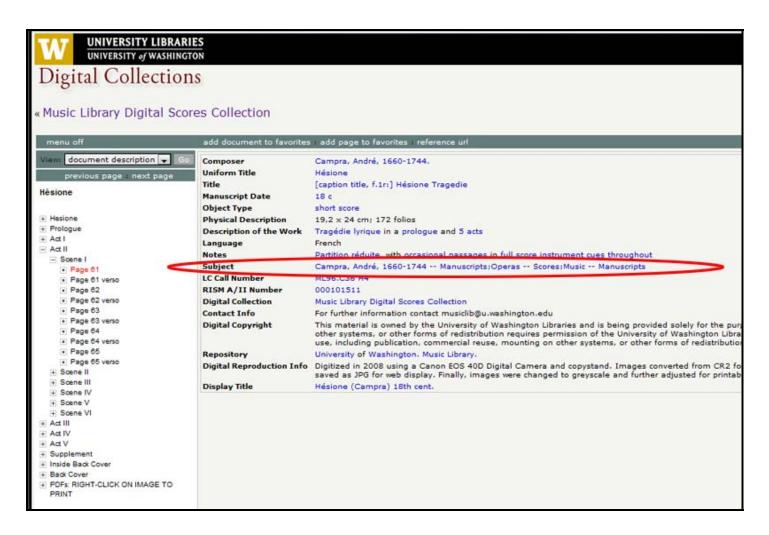
This allowed the data to be displayed correctly but still able to successfully retrieve results whether the user typed in diacritics or not.

The next step was to import the metadata into CONTENTdm along with the high-quality optimized JPG images from the camera. CONTENTdm's import function is capable of automatically creating zoomable JPG2000 images from regular JPGs and also assembles all pages of a volume into one compound object.

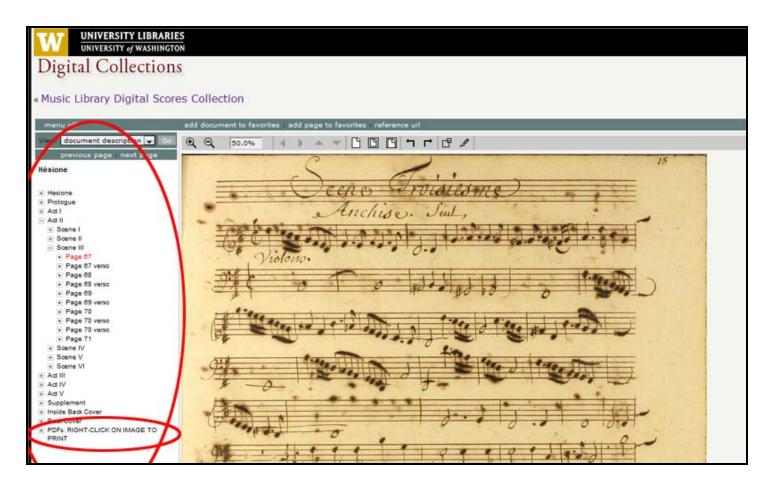




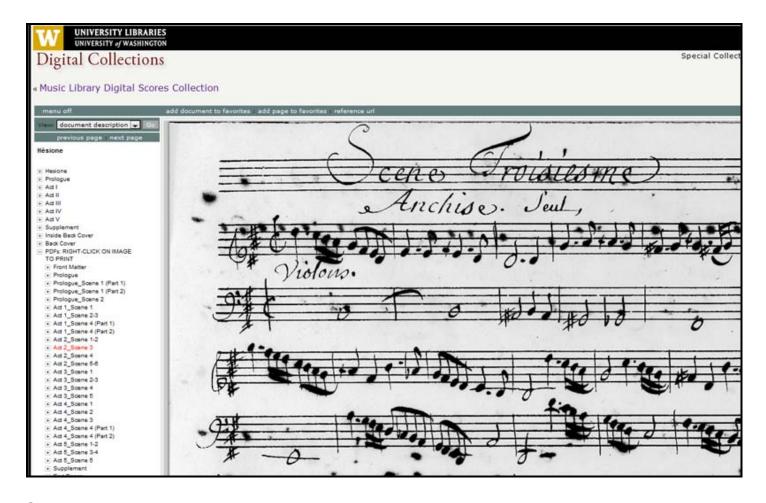
We added Library of Congress Subject headings for each of the artifacts and each analytical entry. And last, we added metadata which is routinely included in all of our digital objects, such as the name of the digital collection, contact information, and other administrativia.



The metadata from the RISM database gives detailed information which allows researchers to find a particular piece of music by searching the metadata. We needed to provide additional metadata elements for two reasons. First, so that the visual objects themselves could be easily navigated, which would in turn enhance search and discovery.



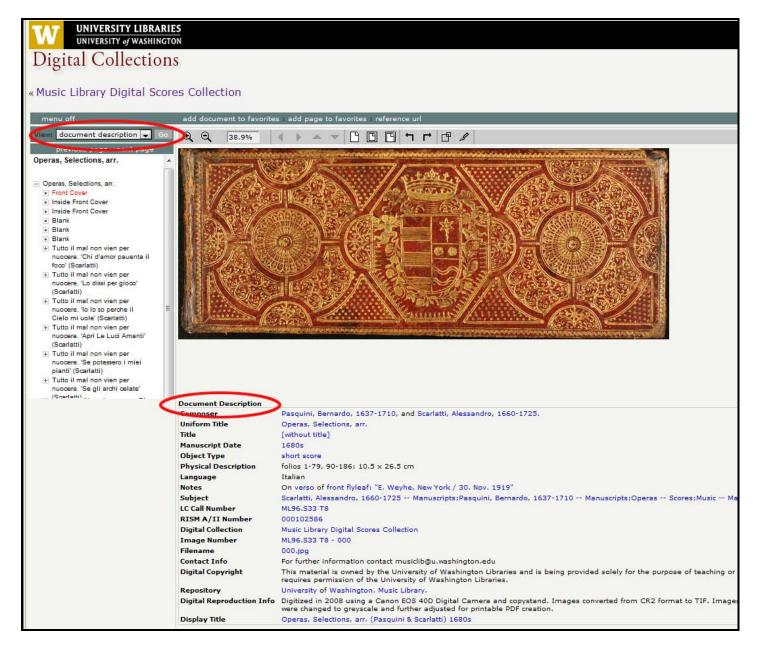
Second, in addition to the color JPG images, we provided black and white pdfs for printing, and needed to divide each volume into sections that allowed for reasonable download times of the pdfs and to help with onscreen navigation.



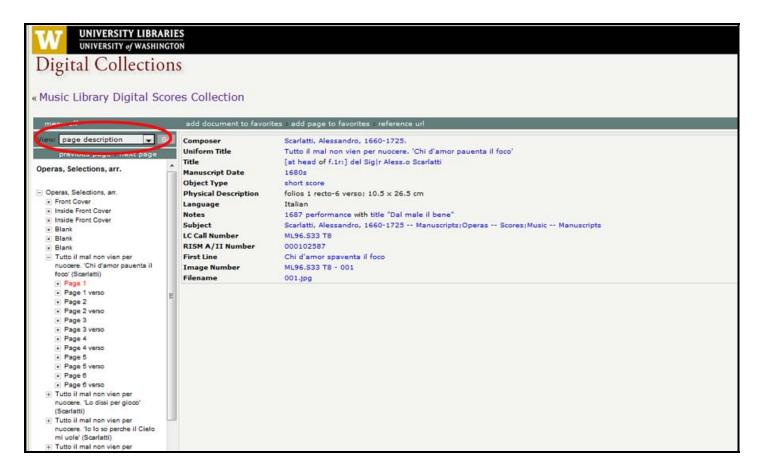
So that these divisions were intelligible to music researchers, we used Deborah Pierce's expertise as someone with much deeper music knowledge than I have. Deborah, the co-PI for this project, trained as a musicologist and has worked as a music reference librarian and cataloged music materials for over twenty-five years. Deborah created what we called "data maps" to guide Digital Initiatives staff so they would know where to create breaks and how to label portions of each volume.

Operas were divided into acts and scenes whenever possible and other document hierarchies were followed. Volumes containing collected works were divided by each piece to correspond to the RISM metadata. If a section was very long, we sub-divided it into 25 page chunks to control PDF printing.

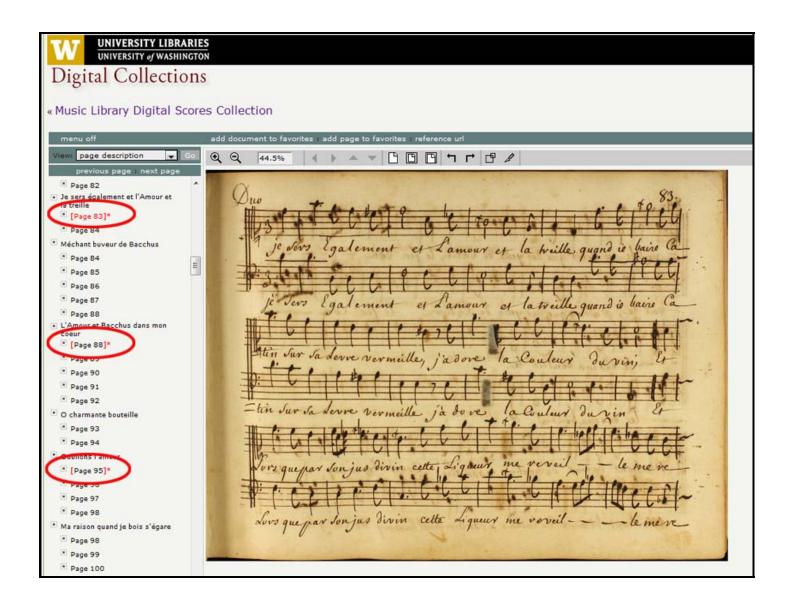
Metadata for the volume/artifact is displayed in CONTENTdm's "document description" area.



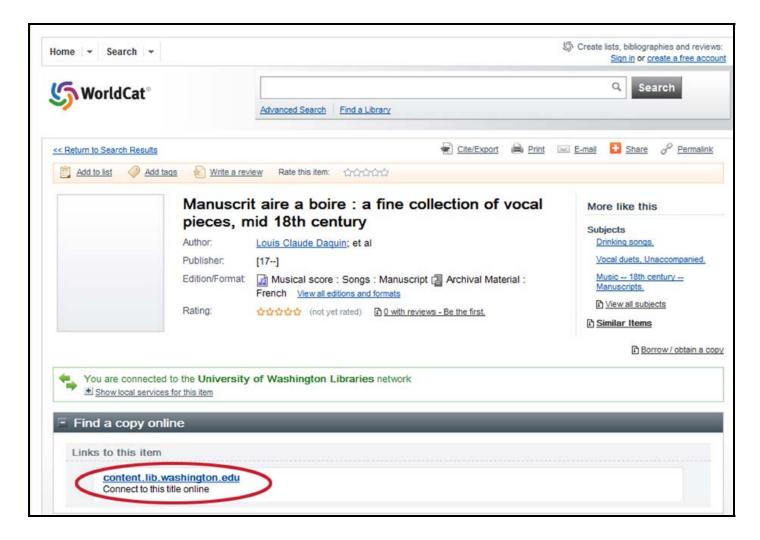
Detailed metadata for each individual piece in a volume appears with the image of the first page of each piece as a "page description."



All metadata is searchable; if a search result is found in the page description rather than the document description, then the page within the volume is highlighted.



When the CONTENTdm collection was completed, we added live links to the individual artifacts in the bibliographic records both in our local catalog and in WorldCat records.



RISM added links to our images from their OPAC and wrote a short article about the collaboration on their website.

To summarize, the project was a resounding success. The manuscripts are now freely available online to anyone in the world who wishes to view, print, or perform them. The originals do not need to be handled by patrons which promotes preservation of these unique and valuable materials. RISM's generous sharing of their cataloging for these volumes ensured that the metadata for the online objects is excellent, which enhances discoverability by end-users. In addition, the digital collection is harvestable via OAI-PMH, which could further enhance its discoverability.

Web statistics for the first ten months the collection was live show that the Music Library Digital Scores Collection received over 9,000 page visits. Currently, we are digitizing and adding <u>non-RISM</u> manuscripts and musical scores to the collection. We are very pleased with the results of the project and encourage other libraries to digitize their own collections – especially the RISM-cataloged collections -- and make them available to music researchers around the globe.

Data Dictionary for Music Library Digital Scores Collection

University of Washington July 2008

Complete Volume Record:

Project element	DC mapping	Comments
Composer	Creator: Search=yes, hidden=yes	Data taken from PEPRD field "rismhs" . Example: Campra, Andre, 1660-1744. Note: remove diacritics from data files.
Composer (Unicode)	Creator: Search=yes, hidden=no	Data taken from PEPRD field "rismhs". Example: Campra, André, 1660-1744. Note: retain diacritics.
Uniform Title	Title: Search=yes, hidden=yes	Data taken from HSPR01 field "titreg" Note: remove diacritics from data files.
Uniform Title (Unicode)	Title: Search=yes, hidden=no	Data taken from HSPR01 field "titreg" Note: retain diacritics
Title	Title: Search=yes, hidden=yes	Data from HSPR01 field "diptit" Note: There will be a few entries which are blank in both Uniform title and title. In this case use h3 from the HSPR01 table. Note: remove diacritics from data files
Title (Unicode)	Title: Search=yes, hidden=no	Data from HSPR01 field "diptit" Note: There will be a few entries which are blank in both Uniform title and title. In this case use h3 from the HSPR01 table. Note: Retain diacritics
Manuscript date	Date: Search=no, hidden=no	Date that displays to public Input date by hand from the coded information in the thematic catalog (ML113 I6 ser.AS/II pt. US Su).
Dates	Date: search=yes	Date field for searching, including enumerated date ranges

	hidden=yes	
Earliest Date	None: search=no hidden=yes	Earliest date in range or single date
Latest Date	None: search=no hidden=yes	Latest date in range or single date
Object Type	Type: search=yes, hidden=no	Data from HSPR01 field "ah400". Example: 1 short score
Physical Description	Format: search=no, hidden=no	Data from HSPR01 from fields "h02050" (pagination: e.g., 117f.) and HSS001 from field "format" (size: e.g., 22 x 29 cm) Example: 117 folios; 22 x 29 cm.
Description of the work	Type: search=yes hidden=no	Data from HSPR01 field "h41820" Example: Opera in 2 acts
Language	Language: search=yes, hidden=no	Language of written text, e.g., opera in French=French. Data from HSS001 field "h0203m" Note: This field is coded and needs to be fleshed out in transfer of data
Notes	Description: Search=yes, hidden=yes	Data from HSS001 field "bemueb" Note: Remove diacritics from data files
Notes (Unicode)	Description: Search=yes, hidden=no	Data from HSS001 field "bemueb" Note: Retain diacritics
Subject	Subject: Search=yes, hidden=no	LCSH. Take from the headings on the bibliographic records in the University of Washington Libraries' catalog.
LC call number	Identifier: Search=yes, hidden=no	University of Washington Libraries call number. Data from HSPR01 field "hsexsg"
RISM A/II number	Identifier: Search=yes, hidden=no	Data from HSPR01 field "rismnr"

First line	Title: Search=yes, hidden=yes	Data found in HSS001_i field "txtin1" Note: Remove diacritics from data files
First line (Unicode)	Title: Search=yes, hidden=no	Data found in HSS001_i field "txtin1" Note: Retain diacritics
Digital Collection	Relation-Is Part of: search=no, hidden=no	Music Library Digital Scores Collection
Image number	None: Search=yes Hidden=no	Number that users can reference which indicates the physical item
Contact info	Description: search=no, hidden=no, required	For further information contact musiclib@u.washington.edu
Digital Copyright	Rights: search=no, hidden=no; required	This material is owned by the University of Washington Libraries and is being provided solely for the purpose of teaching or individual research. Any other use, including publication, commercial reuse, mounting on other systems, or other forms of redistribution requires permission of the University of Washington Libraries.
Repository	Source: Search=no Hidden=no	The institution where the item is physically located. University of Washington. Music Library.
Digital reproduction information	Format: search=no, hidden=no	Describes the digital conversion process, date scanned, etc.
Display Title	None: Search=no,	Used as field which is visible in Browse mode, contains Title, Composer, Manuscript date. Example: Alceste (Lully) 1674
	Hidden=no	