Metadata and Greenstone in Chopin Early Editions

Tod A. Olson <tod@uchicago.edu>
The University of Chicago Library
November 12, 2004

1 Introduction

This paper discusses the use of metadata in the Chopin Early Editions (CEE) collection of online piano scores, the use of Greenstone Digital Library Software in delivering this collection, and highlights issues applicable to other similar collections, such at the Online Chopin Variorum Editions (OCVE). The University of Chicago Library is home to a significant collection of Chopin scores published between 1830 and 1880 (Platzman, 2003). CEE is composed of scanned images of scores from this collection, and is made freely available on the Web to musicians, music scholars, and the general public.

2 Metadata

This section discusses the role of descriptive, administrative, and structural metadata in CEE, and the use of METS and MODS as containers for this metadata. Particular issues for descriptive metadata in providing access for CEE are discussed, as are issues with structural metadata in modeling the scores.

2.1 Descriptive metadata

Descriptive metadata for CEE is maintained in the form of detailed library cataloging records, recorded in Machine Readable Cataloging (MARC) format according to the Anglo-American Cataloging Rules, 2nd ed. (AACR2). MARC and AACR2 are not optimal for musical scores, but they do support fine granularity, are used in our regular cataloging processes, and allow us to share these records widely. MARC records are mapped to the Metadata Object Description Schema (MODS). MODS is maintained by the Library of Congress, preserves most of the granularity of the MARC record in an XML format, and is friendly to electronic resources.

MODS is an XML schema, with data arranged in a hierarchy. It is designed to offer more flexibility than MARC as a container for bibliographic information. MODS is meant to be a metadata container capable of supporting AACR2 cataloging rules, but not limited by AACR2.

The MARC records for the scores are cross-walked to MODS (Metadata Object Description Schema) records according to the *MARC Mapping to MODS Version 3.0* (2003). The MODS data is mapped to the metadata fields used in CEE. The MODS data may also be transformed in the future for other purposes, such as sharing through the Open Archives Initiative (OAI). By retaining the granularity of the MODS data, which is not needed immediately by CEE, we have greater flexibility in future reuse of the records.

Below is examined the use of uniform tiles for bringing together, or collocating, various editions of the same score, the use of genre for access to CEE, and problems encountered with the opus number.

2.1.1 Uniform titles and collocation

It is well established that music has an extremely high rate of republication under varying names and titles (Smiraglia, 1989b). Chopin is no exception. For example, the Chopin's opus 15 was published under such diverse titles as *Trois nocturnes pour le pianoforte, oeuv. 15* and *"Les zephyrs": 4me., 5me., et 6me. nocturnes*. Displaying these titles in a list gives the user no indication they are the same work. Because of these varied titles, a device is needed to collocate the scores.

The bringing together of multiple editions of the same work, regardless of title changes, is a fundamental premise in the organization of knowledge in libraries. Uniform titles are a syndetic device which "draw[s] together all variant manifestations of a work, and distinguish[es] similar titles but different works". (Smiraglia 1989a, 54) The titles above are collocated under the uniform title *Nocturnes, piano, op. 15*. This organization allows the reader to easily discover all versions of a work, despite widely varying titles, and gives the reader confidence that manifestations with different titles really are the same work.

CEE provides this collocation by uniform title, which is presented to the user first as a list of all uniform titles in the collection. Clicking on the uniform title reveals all scores available that match that uniform title, identified by title proper and imprint information. The uniform title listing in CEE provides two functions: it gives the user a concise view of the entire holdings of the collection, and it allows the user to quickly drill down to specific scores of interest.

2.1.2 Genre

Genre was considered one of the most important forms of access by the CEE project team. Early usage statistics support this notion (Olson & Downie, 2003). CEE provides a browse list of genres. Some genres (e.g. Mazurkas, Waltzes) have a large number of scores, and it can be difficult to determine manifestations of the same work. To make the available works more evident, regardless of number of manifestations, the genre list shows all uniform titles, which in turn expand to lists of scores. This gives the user an accurate view of what is held under a particular genre, and still allows quick access to desired scores.

Access by musical genre is very useful for a musician or music scholar; this is reinforced by access patterns observed in Chopin Early Editions (Olson & Downie, 2003). Classical piano genre terms, as used in Chopin Early Editions, are difficult to provide through the MARC/AACR2 bibliographic record. The vocabulary for genre is generally agreed upon in the music field and the terms can be found in standard works such as *The New Grove Dictionary of Music and Musicians* (Sadie, 2001). However, there is no formal controlled vocabulary, in thesaurus or other form, for cataloging.

2.1.3 Opus

The CEE metadata currently does not support access by opus number well, and several users have requested improvements. The problem lies in the metadata, as the opus number is not coded consistently. The opus number may not be present in the record if it does not appear on the printed work and if it is not in the standardized title, for example, when the composer has named the piece. If present, the opus number may appear with

other enumeration in the same field. For example, in *Ballades, piano, no. 2, op. 38*, the values “no. 2, op. 38” are coded in the same field. As the entire field is indexed, a search for opus 2 that tries to simple match on the integer value will return *Ballades, piano, no. 2, op. 38* as a false hit.

To solve this problem, a local field containing only the opus number will be added to the catalog records. Traditional publisher-assigned opus numbers will be prefixed with “op.”; for pieces lacking an opus number, the number assigned by Brown (1960) will be used, prefixed with a “B.”. MODS does not have an element restricted to opus number. The opus number from the catalog record will be mapped into an opus number element in a locally-defined schema and added to the MODS record under the *extension* element, which can hold elements from any schema. Once the improved metadata is available, CEE will likely offer a browse listing by opus number.

2.2 Administrative metadata for images

Technical metadata for each scanned image is recorded at scanning time, in the same database as the structural metadata. This information includes the software and hardware used to capture the image, software settings, scanner operator, and the like. Though recorded, this information is not yet used or available in CEE. This metadata would be mapped to a suitable XML schema, such as the upcoming NISO Metadata for Images Standard (NISO MIX), for use in CEE.

CEE currently records no rights management metadata, though copyright of the images is asserted through a statement on the site.

2.3 Structural metadata

Structural metadata for CEE includes a document number for each score, the sequence of images in each score, page numbers as they appear on the scores, and milestones within scores. Milestones are significant markers in the score external to the music itself, such as cover, title page, and the beginning page of any work, such as “Nocturne no. 4” and “Nocturne no. 5” within *Nocturnes, piano, op. 15*. Milestones are used in CEE to provide a convenient index into the score as printed.

One issue with structural metadata is the degree to which we mimic the physical score. Some scores may have been altered significantly since their initial purchase. Sometimes collectors will bind several separate scores into one cover; or a library will give a score a hard binding to make it sturdier on the shelf. In one unusual case in CEE, a two-volume set of scores was bound together with volume 2 in front of volume 1. In these cases, a decision must be made about how faithfully the digitized score should reproduce the physical score, and the structural metadata must be designed to support these decisions. For example, it may be desirable to scan and record added covers to have a complete representation of the physical score (perhaps to aid in identifying the score in case of loss) but only to display them if the covers are particularly interesting. To support this, the structural metadata for each image may have an optional flag to suppress the undesired image from the user display, while allowing a complete set of images for internal purposes.

2.4 MODS

The Metadata Object Description Schema (MODS) is maintained by the Library of Congress, and is intended to preserve most of the granularity of library bibliographic records in an XML format. For example, separate versions of the title can be coded as appearing on the title page or on the spine, or assigned as the uniform title. Components within collected works can be encoded. Persons associated with a published score are coded in the *name* element, along with their role (e.g., composer, editor, dedicatee).

MODS is also flexible, and extensible. The *extension* element accepts child elements from any schema. For example, as MODS has no element restricted to opus number (*partNumber* may contain other enumerations), that specific value will be coded according to another schema and added under *extension*.

2.5 METS

The Metadata Encoding and Transmission Standard (METS) is an XML schema which allows a single file to encapsulate both the data files and metadata for an electronic object or collection, and the relationships between them. METS defines a mechanism for encoding the structural hierarchy and files for the object or collection, and allows the use of other namespaces to carry descriptive and administrative metadata. METS is being developed by the digital library community, drawing on previous experiences with large digital library projects, and is used to model and archive both digital objects and collections of objects. METS is used for both the transmission and archiving of digital objects. Cundiff (2004) provides an excellent introduction to METS.

METS is used by CEE to provide reference versions of Chopin scores. These METS objects are automatically generated from the recorded structural and descriptive metadata, and then transformed for loading into Greenstone.

METS represents objects as a hierarchical structure of nodes. Any node can have associated descriptive metadata, administrative (technical or rights management) metadata, and files. In CEE, the top-level node represents the score as a whole and descriptive metadata for the score is associated with it. Nodes representing individual pages are underneath this node, and are associated with page images.

METS objects are generated by program for each score in CEE. These objects contain the structural and descriptive metadata for the scores, and pointers to the constituent image files. Though the structural and descriptive metadata are maintained in other databases, these files are considered the reference form for the digital scores. If changes are made to any of the metadata, the METS objects can simply be regenerated.

There are two key benefits to using METS for digital scores in CEE. METS allows the flexibility to integrate new forms of data and metadata (e.g., technical) into the document structure. Currently, we plan to add a new image format to CEE, but METS would also allow the association of the score images, or bars on those images, with the corresponding portions of encoded scores files or recorded performances. Second, by applying XSL Transformations, the METS documents can be reused for different purposes. In CEE, the METS objects are transformed for loading into Greenstone. METS objects are also transformed to provide records for OAI harvesting. The CEE scores could also be delivered through different software by applying a different transformation.

Since the relationships between all of the score components are encoded in the METS document, those relationships are available in any downstream processing.

The generality and extensibility of METS makes it difficult the same user interface to completely support the full range of possible METS documents. METS projects, and communities sharing similar documents, tend to establish well-defined constraints that their METS objects adhere to. Formalizing these constraints in a METS Application Profile allows the assumptions about adhering documents to be shared. Profiles offer the prospect of greater interoperability between METS-base digital libraries, and offer guidance to projects deciding on document models.

3 Greenstone in CEE

This section discusses the use of Greenstone Digital Library Software in delivering CEE to the public. Greenstone is robust, flexible software, used for a wide variety of digital library projects. Greenstone provides all indexing and user interface features in CEE.

3.1 Deciding on Greenstone

Originally, the CEE project plan was to load the METS records in a native XML database, and construct an entirely custom user interface. After several weeks, it was determined that the database could not accept the METS documents due to immature support for XML schema and no support for namespaces. We switched to Greenstone, implemented the previously specified interface features, and still made our project milestones.

3.2 Metadata and document structure in Greenstone

Greenstone converts documents, on import, into the Greenstone Archive Format (GSAF) for internal storage, indexing, and access. GSAF is an XML: format which represents documents as a hierarchical structure of nodes. Metadata and displayable content may be contained in any node.

3.2.1 Metadata in Greenstone

Structural metadata in Greenstone is reflected as nested *Section* elements in the hierarchical structure of the document. Metadata for access and organization is always in the form of field-value pairs, encoded as *Metadata* elements, and can be attached to any node in a document's hierarchy. In the case of a score, metadata can be associated with any page node. Making up some plausible field names, one may attach a PAGE with value"Page 1" and MILESTONE with value "First movement" to a node.

3.2.2 Document identifiers and persistent URLs

Every document in Greenstone has a unique identifier, stored as the *Identifier* metadata field in the top-level *Section*, and is used in constructing URL references to the document. By default, this identifier is created by computing a hash key on the original document. If the document changes, so does its hash, and therefore the *Identifier*. This makes it difficult to persistently identify a document should it undergo changes, such as revising the content.

In CEE, every score has a unique project-assigned document number. This document number, with the prefix “CHOP”, is used as the Identifier value. We can therefore compute the URL for any CEE score, given its assigned document number. This makes it easy to provide a persistent identifier (PI) for each score. For example, the PI for CEE score number 088 is http://pi.lib.uchicago.edu/1001/dig/chopin/088, and is redirected to Greenstone by a simple URL rewrite. Public references to score 088, as in our online catalog, use this PI. If CEE ever moves to a different software platform, updating the published URLs to the scores is a simple matter of changing the redirect for the PI.

3.2.3 Document structure in Greenstone

Documents in Greenstone are stored internally as XML documents in Greenstone Archive Format (GSAF). Documents are modeled as a hierarchy of *Section* elements. A *Section* can contain a *Description* element, a *Content* element, and any number of child *Section* elements. *Description* carries *Metadata* elements, which represent metadata field-value pairs. The *Content* element for any *Section* is formatted for display to the user, and used in full-text indexing.

```
<Section>
      <Description>
            <Metadata name='fieldname'>metadata value</Metadata>
            …
      </Description>
      <Content>[Encoded HTML, image links, etc. goes here]</Content>
      <Section>
            <Description>
                  <Metadata name='fieldname'>metadata value</Metadata>
                  …
            </Description>
            <Content>[Encoded HTML, image links, etc. go here]
            </Content>
            …
      </Section>
      <Section>
            …
      </Section>
</Section>
```

GSAF structure

Additionally, the experimental version of Greenstone replaces GSAF with METS for its internal storage format. METS has been added as an optional storage format in Greenstone 2.52, the most recent production release.

3.3 Document structure in CEE

Currently, documents in CEE are structured with the top-level *Section* corresponding to the score as a whole. The *Description* contains document-level *Metadata* (Title, Publisher, etc.) while its *Content* contains the formatted bibliographic view. Each page is a child Section where its *Description* contains page-level *Metadata* (e.g., a page label of “Page 1: Nocturne no. 15”) and *Content* is an (HTML reference to a) lower-quality JPEG

image with a text link to a higher-quality JPEG image. Navigationally, this means that the user can see the score page-by-page at a lower resolution, but must take an extra action to see the higher-resolution image for each page.

There are plans to revise this document model, based on web logs, observed user behavior, and an additional document format. Web logs show that users looking at many pages of a score will often follow through from the default page image to the higher–resolution image. We infer from this use pattern that there is some desire to work only with the higher-resolution images, and we wish to make this a possibility. From observed user behavior, we see some problems with linking to a higher-resolution image. The higher-resolution images are 2000 pixels wide, certainly much larger than most computer displays. Some common web browsers "helpfully" shrink the image to fit into the viewable area, and require a user action to show the image at its original size. During user observations, users unaware of this browser feature expressed irritation that the supposedly higher-quality image was actually smaller than the original page image.

Score images are now being produced also in the DjVu format, which allows zooming, printing, and page-turning features within the browser plugin. These are to be added to the CEE interface, though the details have not yet been decided.

A revised document model for CEE would allow the user to select any of three image options: 700-pixel wide JPEG, 2000-pixel wide JPEG, or DjVu. All images would appear as images embedded in HTML to avoid browser resizing (though extreme width may pose other usability problems). Ideally, the user would be able to page turn through any specific image format and, from any page, be able to switch to that page in another image format and continue browsing in that format. Greenstone should allow modeling of most, and perhaps all, of these features. (Interaction with the native page turning in DjVu has not yet been explored and may be the limiting factor.)

3.4 Customizing Greenstone

Greenstone can be customized through two types of mechanisms, the dynamic page generation features, and custom metadata and content in the GSAF files. These features can interact.

Greenstone has a macro system, which generates many of the page features such as the headers and footers, and implements the multi-lingual interface. Macros can be customized to make global or collection-specific changes. Greenstone also has a collection-specific format string facility, which formats the browse lists, search results lists and document displays. Format strings can mix HTML markup with the dynamically expanded macros and metadata values. Both the macros and the format strings are evaluated at runtime. Any changes made are immediately available and have a rapid debug cycle. Changes to format strings require a page reload in the web browser; changes to macros may require a new Greenstone session. Customizations that rely only on the macro or format strings can be very quick to implement.

Custom metadata can be added to Greenstone documents. Document content can include Greenstone macros, which will be expanded at runtime. Changes to the content and metadata of the documents require rebuilding the collection to take effect.

The combination of these features is very powerful. CEE uses them to provide unique document navigation features in the form of a custom page-turner and a custom menu of document pages.

3.5 Custom document naviagation in CEE

CEE implements three custom navigation features without modifying the underlying Greenstone architecture: tabs to switch between the bibliographic description and the score images, a custom page-turner, and a selection menu of all pages in the score. These rely on (1) the fact that Greenstone allows every *Section* of a document to be addressed by a computable identifier, (2) dynamically expanding macros and format strings, and (3) the ability to custom-generate document metadata and content. Macros can generate the bulk of the URL, including all session state parameters. The document *Identifer* value, suffixed by coordinates within the hierarchy, uniquely identify a section. For example, CHOP008.2.1 would identify document CHOP008, second *Section*, first sub-*Section*, if it existed. CEE implants it custom page turner by seeding each *Section* with metadata containing the identifier for the previous and next pages, and referring to them in the *DocumentText* format string (underscores denote macros, brackets denote metadata):

```
<table border=0 align=right cellpadding=0 cellspacing=0>
    <tr>
     <td width=45%>{If}{[PrevPgId],
         <a href=\"_httpdocument_&cl=_cgiargcl_&d=[PrevPgId]\">
                                    _iconprev_prev</a>, }</td>
     <td width=10%> </td>
     <td width=45%>{If}{[NextPgId],
         <a href=\"_httpdocument_&cl=_cgiargcl_&d=[NextPgId]\">
                                    next_iconnext_</a>, }</td>
    </tr>
</table>
```

Similarly, macros and static section identifiers can be added to any of the *Content* elements to link directly to another *Section*.

If any document navigation feature proves too difficult to implement natively in Greenstone, it is possible to link out to an external document-viewing interface. There is precedent for this approach: the Washington Research Library Consortium (WRLC) implementation of Greenstone uses a page turner developed in-house before Greenstone included such a feature natively.

3.6 Search and retrieval

Greenstone provides full text and metadata searching at the document and section levels, and full text searching at the paragraph level. These are determined at indexing time by the collect.cfg configuration file. Therefore CEE could provide searching of the milestones by simpling adding an index line to collect.cfg and reindexing.

It may be possible to leverage text-based indexing for music content retrieval, though this is a research area (Downie & Olson, 2004).

One issue is in searching non-ASCII text. The indexing and querying facilities do not normalize their input to remove accents (though they are case-insensitive). This means a

search for “etudes” will not match the indexed term “études” because of the accent. This is an issue for CEE because many users will not expect the accent over the “e” to be required in their query. To solve this in CEE, we have parallel metadata fields for most metadata, one for indexing and one for display. For example:

```
<Metadata name=“Title”>Douze grandes études </ Metadata >
<Metadata name=“TitleIdx”>Douze grandes etudes</ Metadata >
```

TitleIdx will be used for searching and ordering browse lists, but *Title* will be displayed to the user. This solution involves no modifications to the underlying Greenstone source code.

3.7 Hierarchical browse lists

The ability to browse CEE in hierarchical lists organized by genre and uniform title was discussed above. Greenstone’s *hierarchy classifier* is used to implement these lists. The hierarchy for any list is defined by a file which associates a metadata value with a node in the hierarchy and display label for that node. This is a very flexible mechanism, as the collection designer has full control of the hierarchy. There are no limits on depth or breadth of the hierarchy, and metadata and display values can be in any language or symbol set represented in Unicode. The hierarchy file itself uses UTF-8 encoding.

There is a maintenance issue here: it is possible that a score will be added with a genre or uniform title that does not already appear in the hierarchy file for one or the other browse lists. Rather than trying to predict all possible values of genre and uniform title, or relying on staff to manually update the file when needed, we automated the process.

A program was written to scan all of the scores in their Greenstone format and generate a hierarchy file organized by metadata fields specified at runtime. Thus the file describing the genre browse is automatically generated from values in our generated GSAF files, and is organized by genre and uniform title. A file for the uniform title browse is also automatically generated. As an added benefit, typographical errors in these metadata are easily detected by scanning the browse list.

3.8 Graphical Design and Greenstone

CEE has done little to change the graphical flavor of Greenstone, focusing on access and navigation features in the initial phase. Other collections have made significant changes to the look of Greenstone, particularly the Digital Bridges collection at Lehigh University (2002) and the digital collections at the Washington Research Library Consortium (WRLC) (2004). The Greenstone macro system allows customization of many of the graphical display elements; the methods used by the WRLC are documented in Zhang (2003).

3.9 Document Workflow

Greenstone provides two ways to create and build collections: a set of command line utilities, and a graphical tool call the Greenstone Librarian Interface (GLI). In either case, a collection is created, documents are selected and imported, indexes are specified, and the collection is indexed. This workflow has been adapted to different environments. For example, WRLC has built a web-based interface for assigning Dublin Core metadata to

their digitized objects. The Mercy Corps, a relief agency, has developed a workflow that allow field offices to submit documents and metadata over the web, to be vetted and added to their collections (Witten, 2003).

In CEE, library staff create and maintain descriptive and structural metadata in tools that are familiar to the (library cataloging software and MS Access), and page images are stored on the Library's web server. A software program knits the pieces of each score together into a METS object, which is transformed into GSAF via XSLT, and loaded into Greenstone using the command line tools.

3.10 Greenstone 3 and the future

The current Greenstone codebase is solid and will be supported for the foreseeable future. There is also a new version of Greenstone, Greenstone 3, being developed to take advantage of more recent technological advances. Greenstone 3 is built on a service-based, modular architecture, is capable of running in a distributed environment, and provides backward compatibility with the current version of Greenstone (Bainbridge, et al. 2004). This version is written in Java (the current version is written in C++) and stores documents internally in METS. Components communicate via SOAP, and XSLT and CSS are used to control document display. METS is a digital library standard, and Greenstone's adoption of METS may result in greater interoperability with other digital library software. The use of XSLT and CSS may result in greater control of the look and feel without altering the Greenstone code. The modular architecture allows components to be replaced or new components to be added, while retaining access to the rest of Greenstone as a whole. For example, one can now choose Lucene for indexing, while retaining all of the document importing and user interface features in Greenstone 3. Similarly, the architecture provides a framework wherein one may add new interface features or query modules. Bainbridge, et al. report as examples a geographical map based interface, and a cluster visualization interface. See Carey, Heesh, and Rüger (2003) for more information on the cluster visualization interface.

Meanwhile, UNESCO's commitment to the current version of Greenstone guarantees it will be supported for the foreseeable future, nor does that version stand still. Certain features developed for Greenstone 3 have been integrated into the current production codebase. For example, Greenstone 2.52 allows the optional use of METS for internal storage, and allows the use of Lucene for indexing.

4 References

Anglo-American Cataloging Rules, Second Edition, 2002 revision. Chicago, American Library Association, 2002.

Bainbridge,D., Don, K.J., Buchanan, G.R., and Witten, I.H. (2004) "Dynamic digital library construction and configuration." *Proc European Digital Library Conference (ECDL 2004), Bath, England*. Available from http://www.sadl.uleth.ca/greenstone3/Dynamic_DL_Construction.pdf, accessed September 7, 2004.

Brown, Maurice John Edwin. (1960). *Chopin: an index of his works in chronological order*. London, Macmillan; New York, St. Martin's Press, 1960.

Carey, M., Heesch, D. & Rüger, S. (2003). Info Navigator: A visualization tool for document searching and browsing. *Proc of Intl Conf on Distributed Multimedia Systems, Florida, Sep 2003*. http://www.doc.ic.ac.uk/~srueger/www-pub/dms-2003.pdf, accessed November 21, 2004. Demo available at http://km.doc.ic.ac.uk/demos/vishnu.html.

Cundiff, Morgan V. (2004). An introduction to the Metadata Encoding and Transmission Standard (METS). *Library Hi Tech, 22*(1) 2004, pp. 52-64.

Chopin Early Editions. (2002). University of Chicago Library. http://chopin.lib.uchicago.edu,

Digital Bridges. (2002). Lehigh University. http://bridges.lib.lehigh.edu/, accessed October 19, 2004.

Downie, J. Stephen, & Olson, Tod. (2004). Toward melodic access and title uniformity in the Chopin Early Editions music digital library, in *Proceedings of the 2004 Joint ACM/IEEE Conference on Digital Libraries*, p. 406.

MARC Mapping to MODS Version 3.0. (2003). Library of Congress.

MARC 21 Concise Format for Biblographic Data. (2003). Library of Congress Network Development and MARC Standards Office. 2003 Concise edition. http://www.loc.gov/marc/bibliographic/ecbdhome.html.

Metadata Object Description Schema (MODS). (2004). Library of Congress Network Development and MARC Standards Office. http://www.loc.gov/standards/mods/.

Metadata Encoding and Transmission Standard (METS). (2004). Library of Congress Network Development and MARC Standards Office. http://www.loc.gov/standards/mets/.

NISO Metadata for Images in XML (NISO MIX). (2004). Library of Congress Network Development and MARC Standards Office. http://www.loc.gov/standards/mix/.

Olson, Tod, & Downie, J. Stephen. (2003). Chopin Early Editons: The Construction and Usage of a Collection of Digital Scores. In Holger H. Hoos & David Bainbridge (Editors), *Proceedings of the Fourth International Conference on Music Information Retrieval: ISMIR 2003*. (pp. 247-248).

Platzman, George W. (2003). A descriptive catalogue of early editions of the works of Frédéric Chopin in the University of Chicago Library. Chicago : University of Chicago Library : distributed by the University of Chicago Press, 2003.

Sadie, Stanley. (2001). *The New Grove Dictionary of Music and Musicians*. 2nd ed. New York : Grove's Dictionaries ; London : Macmillan Publishers Limited, 2001.

Smiraglia, Richard P. (1989a). *Music Cataloging: The Bibliographic Control of Printed and Recorded Music in Libraries*. Englewook, CO : Libraries Unlimited, 1989.

Smiraglia, Richard P. (1989b). Uniform Titles for Music: An Exercise in Collocating Works. *Cataloging & Classification Quarterly, 9*(3) 1989, pp. 97-114.

Smiraglia, Richard P. (2001). *The Nature of "A Work": Implications for the Organization of Knowledge*. Lanham, MD, Scarecrow Press, 2001.

Witten, Ian H. (2003) "Examples of Practical Digital Libraries: Collections built internationally using Greenstone" D-Lib Magazine,Vol 9, No. 3 March 2003.

Witten, Ian H, and Bainbridge, David. (2003). *How to Build a Digital Library*. Boston: Morgan Kaufmann, 2003. http://www.dlib.org/dlib/march03/witten/03witten.html .

Zhang, Allison. (2003). *Customizing the Greenstone User Interface*, Washington Research Library Consortium. http://www.wrlc.org/dcpc/UserInterface/interface.htm. accessed November 5, 2004.