Graduate School of Library and Information Studies, Queens College, CUNY

GSLIS 729 — Metadata for Digital Resources

Instructors: Kwong-Bor Ng (Ph.D.)   KwongborNg@gmail.com

Please send email to KwongborNg@gmail.com to reach the instructor for all 729 email communication, no exception. If you send emails to the instructor's other email addresses, they may not be read or answered in time. Please specify in the subject line of your email that it is about 729.

I.   Course Description

Metadata is fundamental in enabling access to resources. This course is an **advanced technology course** about the theories, development and encoding of metadata. Students will learn how to apply and encode metadata to describe and represent digital content objects (e.g., eBooks, web sites, etc.) in a distributed network environment (e.g., the Internet) to facilitate online access and retrieval.

We will cover three major metadata applications commonly used in digital libraries and cultural heritage institutions, with a focus on the meta mark-up language behind most of the metadata encoding schemes, i.e., The XML. The three metadata applications are: MARCXML, Dublin Core, and MODS (Metadata Object Description Schema). MARCXML and MODS are XML schema developed and maintained by The Library of Congress for digital libraries. The Dublin Core is one of the most popular metadata element sets for use with RDF (Resource Description Framework), more recently in the context of the Linked Data movement.

To understand how metadata encoding schemes work, you need to have some background in various XML technologies. Therefore, we will cover topics like DTD, W3C Scheme Language XSD, Namespace in XML, XPath, XSLT, etc., to help students see the whole picture and comprehend the rationale behind the encoding. Be prepared to do a lot of coding (in XML, DTD, XSL, XSD and PHP) in this class.

At the end of the semester, students will understand the critical roles that different types of metadata can play for digital resources, with the knowledge and skills relevant to entry-level metadata librarian and metadata specialist positions in digital libraries and cultural heritage organizations.

Bulletin Description

This is a course about encoding and applying metadata standards to describe and represent electronic information as objects (e.g., eBooks, web sites, etc.) in a distributed network

environment. It covers various applications in libraries and information institutions, with a focus on the meta mark-up language, XML.

Prerequisites:

- GSLIS 700: Technology of Information (or GSLIS 706: Advance Technology Concepts)
- GSLIS 701: Fundamentals of Library and Information Science
- GSLIS 702: Information Sources and Service: General
- GSLIS 703: Introduction to Technical Services.

II.   How the Course is conducted:

- This is an **asynchronous online course**. Within the semester, links of each class will be added and activated on the date specified in Blackboard. Students can read the materials and do the exercises anytime they like (except for the exams), provided that they can finish their assignments before the deadlines.
- Students must take the midterm exam and final exam between 10:00 am – 10:00 pm on the exam days. The tasks in the exams may be based on previous assignments, so make sure you have finished your assignments before the exam days. Students should be able to finish the exam tasks within 2 hours. However, take the exam as early as you can and allow yourself more than enough time. Exam answers submitted after 10:00 pm of the exam day will not be accepted.
- There are two main types of content for this course: conceptual (e.g., theories and standards of metadata) and procedural (e.g., technical specification of how to encode certain metadata elements in some XML schema). For the first type of content, the instructor will select papers for students to read and discuss. For the second type of content, the instructor will create step by step tutorials. In addition, students will need to read the various official recommendations (i.e., authoritative documents of standards) of the XML languages and some metadata schema in order to understand the relationship between the conceptual and procedural components.

III.   Software and Server

- You will encode metadata in XML, and you will also need to do some coding in languages like CSS, XSL, PHP, etc. You can use your own favorite code editor software (e.g., oXygen, XMLSpy, Notepad++, Dreamweaver, Aptana, etc.) The instructor will use the free version of Aptana, which is a very powerful open-source web development IDE, with Mac and PC version. More will be discussed in the lectures.
- Students will need to upload their encoded metadata files and other programs to a web server for public HTTP access, and for server-side processing (by PHP). If you have your

own server, feel free to use it. If you don't have your own server, use some free server service. We will introduce how to use free server service in the beginning of this course.
- Most of the code editor programs will have FTP function for files transfer. There are also FTP programs and utilities you can download and use for free, e.g., FireFox's FireFTP add-on, etc. More will be discussed in the lectures.

IV.   Course Objectives and GSLIS Program Objectives

Course Objectives Stated as Student Learning Outcomes	Assignments	GSLIS Program Objectives (as listed below this table)							
		A	B	C	D	E	F	G	H
Articulate important concepts, issues, and terminology related to metadata theory, standards, and applications relevant to cultural heritage institutions	Classification of descriptive metadata Reading Summaries	X							
Design and document metadata schemes to meet the functional requirements of specific collections and projects.	Application Profile Creation			X			X		
Rendering display of metadata and XML document through web browser	CSS assignment						X		
Encode metadata according to national and international metadata standards	DTD assignment XSD assignment (in MARCXML, DC, MODS)			X			X		
Deliver metadata records (or digital content objects with embedded metadata to client through server side processing)	XSL assignment PHP assignment						X		

GSLIS Program Objectives

Program and course requirements in the GSLIS are designed to ensure that graduates have met the following Student Learning Outcomes (SLOs). These SLOs (A-H) state that graduates will have the ability to:

A. Assist users in gaining access to information and knowledge, including its creation, acquisition, organization and management, storage and retrieval, by demonstrating that they can:
   a. Identify, acquire, create, organize, process, store and provide access to information in all its forms for libraries, cultural institutions and other information organizations in a global environment.
   b. Identify, retrieve, evaluate and use general and specialized resources to address current and future information needs and provide related services to diverse user communities.
B. Articulate the role and importance of ethics, values, and advocacy within the legal and historical frameworks underlying the practice of librarianship and the information professions
C. Apply the appropriate practices and policies of established Library and Information Science professional standards in various specializations
D. Find, analyze, assess, apply, and conduct research in Library and Information Science and other disciplines in response to gaps in knowledge and practice
E. Contribute to a diverse, global society—including the role of addressing the needs of underserved groups--through exemplary Library and Information Science practice and research
F. Identify, evaluate and implement current and emerging technologies and services to meet the evolving information needs of diverse user communities in an increasingly interconnected environment
G. Demonstrate understanding of the importance of continuing professional development in Library and Information Science; articulate and apply principles, theories and measures underlying the role of the library in supporting lifelong learning within the community
H. Explain and apply principles of effective management and leadership in the library and related information institutions

V.   Textbook and References

Students DO NOT need to purchase any textbooks for this course; all required readings are accessible online. The following is a list of primary references. Additional readings may be listed in the Blackboard.

- Baca, Murtha. (2008). (Ed.) *Introduction to Metadata*. Online edition version 3.0. Los Angeles, CA : Getty Research Institute. Retrieved March 18, 2014 from:

- http://www.getty.edu/research/publications/electronic_publications/intrometadata/index.html
- Fawcett, J., Quin, L.R.E. & Ayers, D. (2012). *Beginning XML*. 5th ed. Indianapolis: John Wiley & Sons. (Accessible through Queens College Library's online catalog. At the library's web site, search for the book, retrieve its record. On the catalog record of the book, there are multiple links, click on the link: View online ; access limited to Queens College.)
- *Understanding Metadata* (2004). Published by National Information Standards Organization. Retrieved March 18, 2014 from: http://www.niso.org/publications/press/UnderstandingMetadata.pdf
- Han, Myung-Ja (2012). Metadata with Levels of Description: New Challenges to Catalogers and Metadata Librarians. Retrieved March 18, 2014 from: http://conference.ifla.org/past-wlic/2012/80-han-en.pdf
- NISO Framework Advisory Group. (2007). "Metadata." In: *A Framework of Guidance for Building Good Digital Collections*. 3rd ed. Priscilla Caplan, Grace Agnew, Murtha Baca, Carl Fleischhauer, Tony Gill, Ingrid Hsieh-Yee, Jill Koelling, Christie Stephenson, and Karen A. Wetzel. Retrieved March 18, 2014 from: http://www.niso.org/publications/rp/framework3.pdf
- Cole, Timothy & Han, Myung-Ja (2013). *Third Millennium Cataloging : XML for Catalogers and Metadata Librarians.* Oxford: ABC-CLIO. (The book can be accessed through ebrary.com: http://site.ebrary.com/lib/qc/login.action. Use the bar code on your student ID card to logon.)

VI.   Methods of Assessment

- Attendance and Class Participation   10%
- Assignments   30%
- Mid Term Exam   30%
- Final Exam   30%

A+	A	A-	B+	B	B-	C+	C	C-	D+	D	F
97-100	93-96	90-92	87-89	83-86	80-82	77-79	73-76	70-72	67-69	60-66	0-59

Use of Student Work

All programs in New York State undergo periodic reviews by accreditation agencies. For these purposes, samples of students' work are made available to those professionals conducting the review. Candidate anonymity is assured under these circumstances. If you do not wish to have

your work made available for these purposes, please let the professor know before the start of the second class. Your cooperation is greatly appreciated.

CUNY Policy on Academic Integrity

The Policy on Academic Integrity, as adopted by the Board is available to all candidates. Academic Dishonesty is prohibited in The City University of New York and is punishable by penalties, including failing grades, suspension, and expulsion. This policy and others related to candidates' issues are available to you at: http://qcpages.qc.cuny.edu/provost/Policies/index.html.

Use of Candidate Work

All programs in New York State undergo periodic reviews by accreditation agencies. For these purposes, samples of candidates' work are made available to those professionals conducting the review. Candidate anonymity is assured under these circumstances. If you do not wish to have your work made available for these purposes, please let the professor know before the start of the second class. Your cooperation is greatly appreciated.

Reasonable Accommodations for Candidates with Disabilities

Candidates with disabilities needing academic accommodation should: 1) register with and provide documentation to the Special Services Office, Kiely 171; 2) bring a letter to me indicating the need for accommodation and what type. This should be done during the first week of class. For more information about services available to Queens College candidates, contact: Special Service Office; 171 Kiely Hall; 718-997-5870 (8:00 a.m. to 5:00 p.m.).

VII.   Class Schedule

The following schedule is based on regular semester with 15 classes. In Summer semester, the same coverage will be offered in the same sequential order, but topics in adjacent classes may be combined to fit a more intensive schedule.

Class	Coverage
1	• **Topics and Required Readings:** (Please read the following 3 papers. If there are links in the document, no need to follow the links to read other pages.) 1. NISO: Understanding Metadata 2. Getty Institute: Setting the Stage 3. JISC: An Introduction to Metadata • **Examples:** (After reading 3 papers above. Examine the following 2 examples.) 1. MARC XML Architecture and a Metadata Record in MARC as well as in Dublin Core Format. 2. A Metadata Record encoded in The Library of Congress' MODS schema (Depends on the type and version of the browser you use, you may need to use the view source function of your browser in order to see the XML code.) • **Reference:** 1. XML Standards created and maintained by the Library of Congress. • **Assignment:** **Task:** In example 1 (above, i.e., MARC XML Architecture and a Metadata Record in MARC as well as in Dublin Core Format), from page 4 to the first 3 lines of page 5, there is a Dublin Core Metadata Record encoded in XML/RDF format. Based on your understanding of the required reading, group the metadata elements (e.g., title, publisher, creator, etc.) into different categories (e.g., administrative metadata, technical metadata, etc.) Send me (Kwongbor.Ng@qc.cuny.edu) your answer. **Notes:** If the categories of metadata listed in the required reading are different among themselves, use the categories that make most sense to you.

Class	Coverage
2	• **Topic:** A Non Technical Introduction to Basic XML. 1. Most of the links on the left navigation panel point to examples of different XML documents. 2. Click on each link on the left navigation panel, examine each and every result page. 3. The link **Slide Presentation** points to a set of slides. Many slides have hyperlink (i.e., words with underline) on them for further elaboration. You can click on the hyperlink to learn more about the topics. Some of the links point to XML files. you may need to use the **view source** function of the browser to see the XML tags. 4. This is a non-technical introduction. So it is OK if there are terminologies that you don't understand for now. We will have a technical introduction to cover them later. 5. The last 3 links on the navigation panel are the **assignments**. For now, I don't need you to have a workable server, simply create the files (there should be 6 files: 1 html file, 4 xml file and 1 css file) and send them to me as email attachment (Kwongbor.Ng@qc.cuny.edu or Kwongbor.Ng@gmail.com). Caveat: For XML files, the first line must be <?xml version="1.0" encoding="UTF-8"?>, there must not be any empty line above it, nor any characters, not even empty space, in front of it.
3 & 4	• **Topics:** 1. Prepare the Tools ▪ Using code editor program (Aptana) ▪ Get a web server account ▪ Using FTP 2. A Technical Introduction to XML, Part 1 ▪ Prolog, XML Declaration ▪ Elements and Attributes ▪ PCDATA and CDATA ▪ Entities and Entity References ▪ Well-formedness • **Assignments:** 1. Send me the URL of exercise04.xml, when I click on it, I should see something like this. (Hints: follow the 3 tutorials in the Prepare the Tools section above.) 2. Send me the URL of class03.xml, when I click on it, I should see something like this. (Hints: See the bottom of A Technical Introduction to XML, Part 1). • **Reference:** 1. W3School: XML Tutorial. (From XML Home to XML CSS) 2. XML Step by Step: Chapter 3: Creating Well-Formed XML Documents 3. W3C: The W3C XML 1.1 Recommendation.

Class	Coverage
5	• **Topic:** A Technical Introduction to XML, Part 2 1. Document Type Declarations and Document Type Definition 2. Internal document type definition and External document type definition 3. Element declaration and Attribute list declaration • **Assignments:** 1. See the bottom part of A Technical Introduction to XML, Part 2. • **References:** 1. W3Schools: XML Document Types and XML DTD 2. W3Schools: DTD Tutorials (From DTD Home to DTD Summary) 3. W3C: The W3C XML 1.1 Recommendation.
6	• **Topics:** o CSS and XML o XML Namespace • **Assignments:** o See the bottom of CSS and XML and XML Namespace o The answer of the Namespace assignment (see page bottom) o Reading Summary assignment (see Blackboard for details.) • **Reference:** o W3schools' Displaying XML with CSS o W3C's CSS2 Specification o W3schools' XML Namespace Tutorial o Wikipedia: XML Namespace o Eric van der Vlist's A Ten-Minute Guide to XML Namespaces o W3C's Namespaces in XML 1.1 (Second Edition)
7	• **Mid Term Exam**

Class	Coverage
8	• **Topics:** o Server Side Validation using PHP o Introduction W3C Schema Language 1. What are the limitation of DTDs 2. Overview of W3C XML Schema 3. Schema Basics 4. Working with Namespaces 5. Complex Types 6. Empty Elements 7. Simple Content 8. Mixed Content 9. Allowing Any Content 10. Controlling Type Derivation • **Assignments:** o Create the PHP (i.e., validate01.php) and XML files we used today. • **References:** o XML Schema Primer (part 0) (Second Edition) o XML Schema Primer (part 1) (Second Edition) o XML Schema Primer (part 2) (Second Edition)
9 & 10	• **Topic:** 1. MARC XML. 2. Dublin Core 3. See my note 01 and my note 02 on the W3C XML Schema Language. • **Assignments:** 1. See bottom of the two topic pages. • **Reference for MARC XML:** 1. MARC21 SLIM schema 2. Expressing Dublin Core Description Sets using XML

Class	Coverage
11 & 12	• **Topic:** 1. Introduction to MODS (I am using the version MODS 3.4 in my class materials and examples. The latest version of MODS is 3.5. The only difference between the two versions is, there are more attributes added into the schema. 2. See here for details.) **Assignments:** 1. Pick one of your fairy tales or short stories, create an XML document based on MODS schema (e.g., like this). Check its validity. Upload your files (i.e., the document instance and the schema files) to your course web site. Send me the URL of your XML document. 2. You can use either MODS 3.4 or MODS 3.5 for the assignment. For pedagogical purpose, I recommend you use 3.4. You don't need to use any of the new attributes of version 3.5 anyway. In addition, you probably want to use my slightly edited/modified version (see comments at the beginning of the file for the modification.) 3. Application Profile Creation Assignment, see blackboard for details. • **Reference:** ○ MODS Official Web Site

Class	Coverage
13 & 14	**Topics: Introduction to XPath, XSLT, and Server Side XML Processing Using PHP** 1. [Introduction to XPath](#) 2. [Introduction to XSLT (Part I)](#), with a little bit of PHP at the end. 3. [Introduction to Server Side XML Processing Using PHP](#) 4. [Introduction to XSLT (Part II)](#) **Assignments:** 1. Same as the [product](#) in the Step by Step Introduction section of [Introduction to XSLT (Part I)](#) but instead of Ugly Duckling, you need to use your own story/fairytale. 2. Please note that there are multiple files involved: a php file (e.g., my [demoC01.php](#), code [here](#)), an xml file (e.g., my [Andersen-UglyDuckling02.xml](#), which has two groups of metadata in the beginning, one group follows MODS specification and the other group follows the Dublin Core specification -- Please note that if you directly open this document by http, you may get an error message because the value of the href attribute on the second line is empty), the associated xsd files (e.g., my [book-v02.xsd](#), which uses [mods-3-2.xsd](#) and [qualifieddc.xsd](#), and, mods-3-2.xsd uses [xml.xsd](#) and [xlink.xsd](#), and, qualifieddc.xsd uses [simpledc.xsd](#), [dcmitype.xsd](#), [dcterms.xsd](#) and [dc.xsd](#)), and the xsl file (e.g., my [fairytale01.s7.xsl](#)). The HTML file that you see at the client side is only the result of the execution of the php file on the xml file following the transformation instructions in the xsl file, and, the xml file is following the format specified by the xsd files. 3. Upload all files to your server, send me the URLs of your php file, your xml file, your xsd files and your xsl file. **References:** 1. W3C's XSL Recommendation: [Extensible Stylesheet Language (XSL)](#) 2. W3C's XSLT Recommendation: [XSL Transformations (XSLT)](#) 3. W3C's XPath Recommendation: [XML Path Language (XPath)](#) (Depends on the type and version of the browser you use, for the 3 links below, you may need to use the view source function in order to see the XML code.) 4. Library of Congress' [XSLT for converting MODS to MARCXM](#) 5. Library of Congress' [XSLT for converting Dublin Core to MODS](#) 6. Library of Congress' [XSLT for converting MODS to Dublin Core](#)
15	• [Final Exam](#) (Link will be activated at 10:00 am)