

LBSC 671 Midterm (Spring 2014) [with answers]

Name: \_\_\_\_\_

You have 3 hours (180 minutes) to complete this exam (although it should not take you that long). Time starts when you proceed past the next page (i.e., you cannot read the exam without starting the 3 hours).

Record your start date and time here: \_\_\_\_\_

This is an open book, open notes, open Internet exam. You may use anything that existed on the Internet before you first read the exam, and you may use any program on any computer. Note in particular that you may look at any homework assignment from any student in this class (not just your own homework) if you find it helpful to do so.

You may not communicate in any way with any person about any topic during this exam. You may not discuss this exam or provide any information about this exam (such as whether you found it to be easy or hard) until the professor posts solutions, discusses the exam in class, or sends you your grade. The reason for this is that different people will take the exam at different times.

You may send your answers in any reasonable form (e.g., Word, text or PDF).

You must finish the exam within 3 hours of starting it, and you must finish it before 5:30 PM on Monday March 31, 2014.

Record your end date and time here: \_\_\_\_\_

Keep the following general test taking strategies in mind. If you find a question to be ambiguous, explain your confusion but provide an answer. This will make it possible to consider the cause of your confusion when grading your answer. Don't spend so much time on one question that you run out of time for other questions. Look over all the questions initially and answer the ones that will get you the most points (for the time invested) first!

A maximum of 100 points are possible on this exam. The questions are on the next two pages. There are four questions on this exam.

1. (60 points) Answer ONE of the following two questions (either a or b, but NOT BOTH). For either answer, the grade will be divided equally between 20 points for the application of the description standard (RDA or DACS), 20 points for proper use of the representation (MARC, MODS, or EAD), and 20 points for (the combination of) subject headings and name authority. If you answer both questions a and b, only the first one you wrote an answer for will be graded. There is no length limitation for your answer to this question.
  - a. [Descriptive cataloging] Download the scanned PDF document from <http://www2.gwu.edu/~nsarchiv/NSAEBB/NSAEBB453/docs/doc09.pdf> . If for any reason you are not able to get the document from that address, I have placed another copy temporarily at <http://www.umiacs.umd.edu/~oard/doc09.pdf> . Create a complete bibliographic metadata record for this document using RDA and render that metadata record using your choice of MODS or MARC 21. Your bibliographic record must include one or more Library of Congress Subject Headings. The Library of Congress Name Authority File should be used as the source for names; for names that are not contained in that authority file, you should render the name in the form specified by RDA. For full credit, every part of your answer should be included in the representation that you have chosen (MARC 21 or MODS), but if there is specific information required by this question that you do not know how to include in that representation you may (for partial credit) include it separately (with an explanatory note) immediately following your bibliographic record.

MODS:

```
<?xml version="1.0" encoding="UTF-8" ?>
<mods:mods xmlns:mods="http://www.loc.gov/mods/v3" xmlns:xlink="http://www.w3.org/1999/xlink"
xmlns="http://www.loc.gov/mods/v3" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.loc.gov/mods/v3 http://www.loc.gov/standards/mods/v3/mods-3-2.xsd"
ID="MODS" version="3.2">
<mods:titleInfo displayLabel="Pueblo Committee Report to President">
<mods:title>Pueblo Committee Report to President</mods:title>
<mods:subTitle>Sixth Draft</mods:subTitle>
<mods:partNumber>February 7th, 1968</mods:partNumber>
</mods:titleInfo>
<mods:name type="personal" authority="LCNAF">
<mods:namePart type="family">Mark</mods:namePart>
<mods:namePart type="given">Clark, Mark W. (Mark Wayne)</mods:namePart>
<mods:namePart type="date">1896-1984</mods:namePart>
<mods:role>
<mods:roleTerm authority="marcrelator" type="text">Author</mods:roleTerm>
</mods:role>
</mods:name>
<mods:name type="personal" authority="LCNAF">
<mods:namePart type="family">Kuter</mods:namePart>
<mods:namePart type="given">Kuter, Laurence Sherman</mods:namePart>
<mods:namePart type="date">1905-1979</mods:namePart>
<mods:role>
<mods:roleTerm authority="marcrelator" type="text">Author</mods:roleTerm>
</mods:role>
</mods:name>
<mods:name type="personal" authority="LCNAF">
<mods:namePart type="family">McDonald</mods:namePart>
<mods:namePart type="given">McDonald, David L. (David Lamar)</mods:namePart>
```

```
<mods:namePart type="date">1906-1997</mods:namePart>
<mods:role>
<mods:roleTerm authority="marcrelator" type="text">Author</mods:roleTerm>
</mods:role>
</mods:name>
<mods:name type="personal">
<mods:namePart type="family">Ball</mods:namePart>
<mods:namePart type="given">Ball, George W.</mods:namePart>
<mods:role>
<mods:roleTerm authority="marcrelator" type="text">Author</mods:roleTerm>
</mods:role>
</mods:name>
<mods:typeOfResource>text</mods:typeOfResource>
<mods:genre authority="marcgt">reporting</mods:genre>
<mods:originInfo>
<mods:dateCreated keyDate="yes" encoding="w3cdtf" point="start">1968-02-07</mods:dateCreated>
<mods:dateCreated encoding="w3cdtf" point="end">2005-07-22</mods:dateCreated>
<mods:place>
<mods:placeTerm>Pueblo Committee and National Archives and Records
Administration</mods:placeTerm>
</mods:place>
<mods:publisher>not identified</mods:publisher>
</mods:originInfo>
<mods:language>
<mods:languageTerm type="code" authority="iso639-2b">eng</mods:languageTerm>
<mods:languageTerm type="text" >English</mods:languageTerm>
</mods:language>
<mods:physicalDescription>
<mods:internetMediaType>image/gif</mods:internetMediaType>
<mods:extent>16 p.</mods:extent>
<mods:digitalOrigin>digitized other analog</mods:digitalOrigin>
</mods:physicalDescription>
<mods:abstract>This is the report of the Pueblo Committee to the President of the United States. This is
about the USS Pueblo which was captured by Korean forces. This report analyses the mission, the
events leading to the capture, and recommendations for the future. </mods:abstract>
<mods:subject authority="lcsch">
<mods:topic>Pueblo Incident, 1968</mods:topic>
<mods:topic>Korea</mods:topic>
<mods:genre authority="lcsch">Report</mods:genre>
</mods:subject>
<mods:identifier displayLabel="Pueblo Incident" type="lccn">sh 88006555</mods:identifier>
<mods:location>
<mods:url usage="primary display"
>http://www2.gwu.edu/~nsarchiv/NSAEBB/NSAEBB453/docs/doc09.pdf</mods:url>
</mods:location>
<mods:recordInfo>
<mods:recordCreationDate encoding="w3cdtf">2014-03-30</mods:recordCreationDate>
<mods:languageOfCataloging>
<mods:languageTerm type="code" authority="iso639-2b">eng</mods:languageTerm>
<mods:languageTerm type="text" >English</mods:languageTerm>
</mods:languageOfCataloging>
</mods:recordInfo>
</mods:mods>
```

MARC:

006 m\*\*\*\* \*\*d\* \*\*\*\*\*  
007 cr mn  
040 CSt\$ beng\$ erda\$ epn\$ cCSt  
110 \$aPueblo Committee  
245 \$aPueblo Committee Report to President /\$nSixth Draft /\$cPueblo Committee  
264 \$cFebruary 7, 1968.  
300 1 online resource (16 pages) :  
336 text \$2rdacontent  
337 computer \$2rdamedia  
338 online resource \$2rdacarrier  
520 A declassified Top Secret document, it was originally commissioned by Johnson, Lyndon B. (Lyndon Baines), 1908-1973 to address a significant compromise of the National Security Agency's code secrets when the American spy ship *USS Pueblo* was captured by communist forces off the coast of North Korea on January 23, 1968.  
588 Viewed March 30, 2014.  
600 \$xNational Security Archive (U.S.)  
650 United States. President (1963-1969 : Johnson).  
700 Clark, Mark W. (Mark Wayne), 1896-1984; Kuter, Laurence Sherman, 1905-1979; McDonald, David L. (David Lamar), 1906-1997; Ball, George W. /\$eauthor.  
776 \$iPrint version: \$aPueblo Committee. \$tPueblo Committee Report to President.  
\$h 16 pages ;  
856 40\$u<http://www2.gwu.edu/~nsarchiv/NSAEBB/NSAEBB453/docs/doc09.pdf>

- b. [Archival description] Use a Web browser to examine the collection of oral history interviews at [http://www.jsc.nasa.gov/history/oral\\_histories/oral\\_histories.htm](http://www.jsc.nasa.gov/history/oral_histories/oral_histories.htm) and then create a complete metadata record for that collection using DACS and render that record in EAD. Your EAD metadata must include one or more Library of Congress Subject Headings. The Library of Congress Name Authority File should be used as the source for names; for names that are not contained in that authority file, you should render the name in the form specified by RDA. For full credit, every part of your answer should be represented in EAD, but if there is specific information required by this question that you do not know how to include in EAD you may (for partial credit) include it separately (with an explanatory note) immediately following your EAD.

```
<ead>
<eadheader>
<eadid countrycode="US" mainagencycode="UMD" identifier="JSCOHP">usUMDjscohp</eadid>
<filedesc>
<titlestmt>
<titleproper>Johnson Space Center Oral History Project</titleproper>
</titlestmt>
</filedesc>
<profiledesc>
<creation>
Created by guest1 using the Archives Hub EAD Editor
<date>2014-03-31</date>
</creation>
</profiledesc>
</eadheader>
<archdesc level="file">
<did>
<unitid countrycode="US" repositorycode="UMD" identifier="JSCOHP" type="persistent" label="current
">JSCOHP</unitid>
<unittitle>Johnson Space Center Oral History Project</unittitle>
<unitdate normal="2014 ">1996</unitdate>
<physdesc>
<extent>
Website containing 675+ interviews, including mp3 and transcripts
</extent>
</physdesc>
<repository>
National Aeronautics and Space Administration Johnson Space Center
</repository>
<origination>Johnson Space Center Oral History Project</origination>
</did>
<bioghist>
<p>The Johnson Space Center Oral History Project was started in 1996 to collect and preserve the
stories of the individuals who pioneered space exploration.</p>
</bioghist>
<scopecontent>
```

<p>Over 675 personal accounts of individuals who worked on the Mercury, Gemini, Apollo, Skylab, and Shuttle missions. Interviewees include managers, engineers, technicians, doctors, astronauts, and other employees of NASA and aerospace contractors</p>

</scopecontent>

<accessrestrict>

<p>This collection is available for research.</p>

</accessrestrict>

<otherfindaid>

<p>Searchable by name of participants</p>

</otherfindaid>

<controlaccess>

<subject source="lcsh">

<emph altrender="a">

United States. National Aeronautics and Space Administration--History--20th century.

</emph>

</subject>

<subject source="lcsh">

<emph altrender="a">

Lyndon B. Johnson Space Center--History--21st century.

</emph>

</subject>

<subject source="lcsh">

<emph altrender="a">Oral History</emph>

</subject>

</controlaccess>

</archdesc>

</ead>

For the remaining questions (numbered 2-4), your answer for each question should be about half a page. Shorter answers that answer the question well would be fine. Longer answers, if you feel that a longer answer is needed, must be no longer than a page (and you should NOT make all your answers that long).

2. (20 points) Explain how the structure of the publishing and the distribution industries affect the ways in which public libraries acquire materials. A complete answer should address both manuscripts (e.g., books) and serials (e.g., journals), and it should address both physical and digital materials.

Publishers work with authors to create content, handling the copyright, editing, peer review, and actual publication of manuscripts and serials. The distributor, who may also be the publisher, will provide access, in the case of e-resources, as well as value-added services for both print and e-resources. For print books, there are stock profiles, where a prebuilt order can generate and ship, based on the monthly NY Times best sellers, for instance. Also, books can be shipped "shelf ready." E-Books also have a similar format to the stock profile, such as Demand Driven Acquisitions, whereby e-book titles are "rented" a certain number of times by patrons, before triggering a purchase automatically, in addition to providing access to the electronic content.

For digital serials, the distributor may have bundled packages, based on subject or publisher, offering a lower price for the total package, or offering titles as "comes with," which are additional titles included with a resource. Again, with e-serials, the distributor must also provide access; however, the e-journal may also have the ability to provide access to archival, or backfiles of the title; these may be free, included with the title purchase, or may be a separate purchase, purely for the back files. Also, the distributor may provide for perpetual access, where access will be granted for the subscription term even after cancellation. This often requires an annual maintenance fee. Public libraries also often maintain subscriptions to print serials such as newspapers and popular magazines. With the exception of local newspapers that may not be collected elsewhere, print serials are often discarded after a period of use (e.g., a month for newspapers, or a year for magazines) rather than bound and retained.

3. (10 points) Compare and contrast Portico and LOCKSS as means for long-term preservation of access to certain types of digital content. In your answer, you should briefly identify which broad classes of risks both services seek to address, and then you should focus for most of your answer on the differences in the types of risks that each service seeks to protect against. In each case (i.e., for each type of risk that is protected against by LOCKSS or Portico or both) you should briefly describe how that service mitigates that risk. For example, you might say that Portico, but not LOCKSS, seeks to protect against [insert type of risk here], and that it does so by [insert how it does it here]. Note that I am not just looking for a description of how each service operates, but rather for a description of what each service achieves by operating in the way it does.

Portico and LOCKSS both seek to help libraries preserve the electronic materials they subscribe to. In this case if one of the companies that provides access to digital content goes out of business the material will still be preserved. The two models also protect against other risks such as, natural disasters, cyber-attack, vandalism, technical failure, etc.

Portico is run by ITHAKA and is set up as a centralized preservation management system. The entire archive is preserved in a standard format, which migrates as technology changes. The material in Portico remains in a dark archive until a “trigger event” occurs and at that point only Portico subscribers can see the information. Portico is primarily structured so that it mitigates against:

- 1) The collapse of a publishing service
- 2) Technical failure (hence the data migration—although having everything in one format may not be a great idea)
- 3) Cyber-attack (It is not great that it is centralized, but as a dark archive it seems to be better protected than LOCKSS)
- 4) Vandalism (It is not great that it is centralized, but as a dark archive it seems to be better protected than LOCKSS)
- 5) Natural Disasters (Does not do a good job of mitigating for natural disasters because it is centralized and inaccessible by others)

LOCKSS (Lots of Copies Keeps Stuff Safe) is set up so individual libraries keep their materials (which are harvested from web crawlers) in “LOCKSS Boxes.” These participating institutions also have access to the materials that other LOCKSS participants are saving. Unlike Portico, the materials collected through LOCKSS remain in the same format as they were created—only migrating the content when necessary. The main focus of LOCKSS is the redundancy because it protects against loss. LOCKSS is primarily structured so that it mitigates against:

- 1) The collapse of a publishing service
- 2) Technical failure (having the materials in different formats might be a positive because if there is a problem with a certain format not everything will be effected)
- 3) Natural Disasters (the material is not centralized in one location and can be accessed by anyone who participates in LOCKSS)
- 4) Cyber-attack (The material is not centralized in one location. Although it does not seem as secure as Portico because LOCKSS is not a dark archive)
- 5) Vandalism (The material is not centralized in one location. Although it does not seem as secure as Portico because LOCKSS is not a dark archive)



(10 points) In our readings for Week 2, Thelwall and Vaughan analyze some differences in the coverage of the Internet Archive for Web pages from different countries. Ultimately, these differences arise from the way in which the Internet Archive acquired those pages. Use Broder's concept that the World Wide Web can be described by an analogy to a "bowtie" to explain why the Internet Archive might have more difficulty acquiring some Web pages than others.

The challenge in Web archiving rests with the fact that in order to archive the Web we must rely on the use of web-crawling technology. Web-crawling technology operates by following links on websites to discover other websites, but if the website has not been linked to by any other website, it cannot be crawled. This leads to websites that are disconnected from the rest of the Web. If we look at the web as a bowtie, all of the pages in the center "knot" of the bowtie link to each other making it possible to discover and access every page in that center knot when the crawler continuously follows the links provided by the center pages. On the left side of the bowtie are the pages that nobody links to, but they link to other pages in the center. In order to be discovered in a web crawl, the crawl must start on these pages, otherwise they will never be found as the crawler moves from the left side to the center of the bowtie. The pages on the right side of the bowtie are pages that are linked to from pages in the center, but don't link to any additional pages. These pages can be discovered during a crawl, but they don't add to the crawler's knowledge of new pages. Some foreign language websites are located on the left side of the bowtie and are not linked to by the English language pages in the center of the bowtie. Others are located in islands that link among themselves, but that are not linked from (and do not link to) the knot in the center of the bowtie. Unless such pages submit their URLs to the web crawler that contributed the crawl to the Internet Archive, they will more likely not be discovered and will not, therefore, be archived.

Hand type the following university honor pledge in your answer (cut and paste is not allowed):

“I pledge on my honor that I have not given or received any unauthorized assistance on this exam”

Don't forget to record your end time. Email your exam to [oard@umd.edu](mailto:oard@umd.edu) immediately after finishing with the subject line “671 Midterm”

----- End -----