Topical Scopes of Digital and Data Curation Curricula

Seungwon Yang
School of Library and Information Science
Center for Computation and Technology
Louisiana State University
Baton Rouge, LA
seungwonyang@lsu.edu

ABSTRACT
We analyzed topical scopes of two digital and two data curation curricula offered as (post)graduate certificates or specializations. We investigated a question of what the topical similarities and differences are between the curricula. To address this question: first, we identified topics from individual course and curriculum descriptions; second, we compared the topics from curriculum descriptions with each other; and third, we compared the topics from course descriptions that are part of a curriculum. We found several topics which were acting as least common denominators between two digital curation and between two data curation curricula. Such topics illustrated the fundamental components within the same type of curricula. The two common topics among all curricula were ‘standards’ and ‘management.’ This may indicate the importance of following the standards in preparation of digital (data) objects and in curatorial processes, as well as the importance of continued management of such resources in a repository. Our approach may be applicable to other types of educational opportunities such as workshops, MOOCs, and webinars. As we analyze more curation curricula, we may be able to develop a more accurate and comprehensive topical framework for the domain of digital and data curation, further enhancing the education for future digital and data curators.

Keywords
Digital/data curation, curriculum, topic, teaching/learning.

INTRODUCTION
Digital curation conferences and courses have been around for a couple of decades. However, “the need for more and better organized and networked continuing education offerings for digital and data curation (Tibbo, 2015)” is ever growing due to the exponential increase of digital content and data. Several education programs for digital and data curation such as specializations in the Master’s program, certificates, or workshops exist to address such needs. In addition, Massive Open Online Courses (MOOCs) relevant to digital and data curation were offered1 or are in development2. Although multiple curricula for digital and data curation exist, each seems to have its own unique portion as well as a shared portion.

In this study, we aimed to analyze topical similarities and differences in four curricula – two digital (“Certificate in digital”, 2006; “Digital curation at”, 2012) and two data curation (“Post-masters certificate”, 2006; “Specialization in data”, 2007). We approached this in three phases:

• Identify topics from curriculum descriptions as well as from course descriptions
• Compare topics from curriculum descriptions
• Compare topics from course descriptions that are part of each curriculum

This study contributes to the digital and data curation community by examining the topical coverage of curation curricula. As we analyze more curricula of digital and data curation, we may be able to have a comprehensive picture of such curricula allowing effective teaching/learning of future digital and data curators.

CURRICULA FOR DIGITAL AND DATA CURATION
Table 1 presents the curriculum type, curation type, institution names that provide the curriculum, and individual IDs. We use an ID to represent a specific curriculum throughout this study.

The University of North Carolina at Chapel Hill (UNC-CH) offers DI_U. This graduate certificate program requires 15

1 Digital curation at the University College London
2 Data management at the University of North Carolina, Chapel Hill
credits (5 courses) (“Certificate in digital”, 2006). The University of Maine (UMAINE) has a similar certificate program, DI_M, which requires 18 credits (6 courses) (“Digital curation at”, 2012). UNC-CH also provides a post-masters data curation certificate program, DA_U, which requires 30 credits (10 courses) (“Post-masters certificate”, 2006). DA_U is positioned between Master’s and Ph.D. and focused on practical applications of the knowledge. DA_I is a data curation specialization provided by the iSchool at Illinois. It requires a minimum of 15 credits (5 courses) and recommends taking 2-4 additional courses from their list of elective courses. It is a degree specialization for their Master’s program, in which students can customize their degree by taking a list of pre-selected courses (“Specialization in data”, 2007).

The full lists of digital and data curriculum courses can be found at http://bit.ly/1IkiliM.

<table>
<thead>
<tr>
<th>ID</th>
<th>Curation type</th>
<th>Curriculum type</th>
<th>Institution</th>
</tr>
</thead>
<tbody>
<tr>
<td>DI_U</td>
<td>Digital</td>
<td>Certificate (15 credits)</td>
<td>UNC, Chapel Hill</td>
</tr>
<tr>
<td>DI_M</td>
<td>Digital</td>
<td>Certificate (18 credits)</td>
<td>University of Maine</td>
</tr>
<tr>
<td>DA_U</td>
<td>Data</td>
<td>Post-masters certificate (30 credits)</td>
<td>UNC, Chapel Hill</td>
</tr>
<tr>
<td>DA_I</td>
<td>Data</td>
<td>Specialization (15-27 credits)</td>
<td>The iSchool at Illinois</td>
</tr>
</tbody>
</table>

Table 1. Selected digital and data curation curricula.

**COMMON COURSES**

To examine the macroscopic similarities between curricula, we compared course titles between DI_U and DI_M (Figure 1 (a)), between DA_U and DA_I (Figure 1 (b)), and between all the curricula (Figure 1 (c)).

![Figure 1. Course overlap: (a) DI_U and DI_M, (b) DA_U and DA_I, and (c) DI_U, DI_M, DA_U, and DA_I.](image)

The fundamental process of digital curation, which is to create collections of digital objects, to preserve them for long-term storage and access, and to manage such objects using information systems/digital libraries, was well represented in Figure 1 (a). For data curation, a means to represent data objects and to effectively organize such objects were reflected in ‘Metadata’ and ‘Databases’ courses in Figure 1 (b). The two courses ‘Digital libraries’ and ‘Digital preservation’ were common among all curricula analyzed in this study (Figure 1 (c)). It may indicate that these two were foundational courses in any curation-type curriculum.

**TOPICS FROM CURRICULUM DESCRIPTIONS**

As another macroscopic topic analysis, we compared topics from the curriculum descriptions and course titles. Figure 2 shows identified topics. Common topics between curriculum are marked with colors.

**TOPICS FROM COURSE DESCRIPTIONS**

As a microscopic analysis of topics, we compared topics from every course description and marked common ones with the same color (blue: common in digital curation curriculum, green: common in digital curation curriculum, and orange: common in all curricula) (Figure 3). We focused on the required courses, and thus excluded elective courses. In case students were allowed to choose from more than one course, we included all the courses and numbered them with additional alphabets (e.g., 2-a and 2-b in Figure 3 (a)).

![Figure 2. Curricular topics. Blue: common topics DI_U and DI_M, Green: common topics DA_U and DA_I, Orange: common topics in DI_U, DI_M, DA_U, and DA_I. 'Preservation' was a common topic across the four curricula. Between two digital curation curricula, DI_U and DI_M, 'access' was another common topic. 'Appraisal (of digital content)' is part of the 'acquisition' process, so we may consider 'acquisition' as another common topic. Between the data curation curriculum DA_U and DA_I, there were three more common topics: 'management', 'policy', and 'standards' in addition to 'preservation.'](image)
<table>
<thead>
<tr>
<th>No.</th>
<th>Topics</th>
<th>(a) DI_U</th>
<th>No.</th>
<th>Topics</th>
<th>(c) DA_U</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>IT, system, collection, policy, procedure</td>
<td></td>
<td>1</td>
<td>IT, system, collection, policy, procedure</td>
<td></td>
</tr>
<tr>
<td>2-a</td>
<td>ICT, management, preservation, records, organizational factor, strategy</td>
<td></td>
<td>2</td>
<td>Unstructured data, analytical paradigm, predictive modeling, data mining, text analytics, web analytics, Map Reduce, Storm, crowdsourcing, workflow, programming</td>
<td></td>
</tr>
<tr>
<td>2-b</td>
<td>Acquisition, development, digitization, mixed-mode holding, strategies, interface, metadata, interoperability, management, policy, evaluation</td>
<td></td>
<td>3</td>
<td>Design, implementation, semantic modeling, relational database, theory, normalization, query construction, SQL</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Acquisition, administration, management, record, manuscript, archive, resource type, format</td>
<td></td>
<td>4</td>
<td>Policy, management, repository, archive, integrated rule-oriented data system (iRODS)</td>
<td></td>
</tr>
<tr>
<td>4-a</td>
<td>Design, information system, analysis, modeling, functionality, data representation</td>
<td></td>
<td>5,6</td>
<td>Decided with a faculty supervisor</td>
<td></td>
</tr>
<tr>
<td>4-b</td>
<td>Identification, provision, evaluation, user needs, institutional environment</td>
<td></td>
<td>7</td>
<td>Metadata, schema, standards, application, information community</td>
<td></td>
</tr>
<tr>
<td>4-c</td>
<td>History, theory, method, archivist, preservation, enduring value</td>
<td></td>
<td>8</td>
<td>Collection development, digitization, mixed-mode holding, strategies, interface, metadata, interoperability, management, policy, evaluation</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Digitization, standards, quality control, digital asset, management, grant writing, metadata</td>
<td>(e) Format, preservation, management, standards</td>
<td>9</td>
<td>Digitization, standards, quality control, digital asset, management, grant writing, metadata</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Acquisition, collection, digitization, (e) Format, preservation, management, standards</td>
<td>(g) Management, standards</td>
<td>10</td>
<td>Data curation lifecycle, metadata, management, ingest, repository, access, policy, data reuse, implementation</td>
<td></td>
</tr>
</tbody>
</table>

Figure 3. Topics from course descriptions: (a) DI_U; (b) DI_M; (c) DA_U; (d) DA_I; (e) common topics of (a) and (b); (f) common topics of (c) and (d); (g) common topics of (a), (b), (c), and (d).

Figure 3 (a): DI_U
Students are allowed to choose either ‘2-a: Electronic records management’ or ‘2-b: Digital libraries.’ For 4-a to 4-c, students pursuing the Master’s in library science choose ‘4-a: Systems analysis.’ Students pursuing the Master’s in information science choose either ‘4-b:
Resource selection and evaluation’ or ‘4-c: Archival appraisal.’

**Figure 3 (b): DI_M**
DI_M is structured based on the process of acquisition \(\rightarrow\) representation \(\rightarrow\) access \(\rightarrow\) preservation, which is referred to as the ‘lifecycle of digital curation’ in their curriculum. These four steps correspond with courses 1-4 in Figure 3 (b). From 5-a to 5-e are internships, independent studies, or a fieldwork related to information systems, museum studies, and historical institutions.

**Figure 3 (c): DA_U**
This curriculum requires the highest number of courses (10) among the curricula in this study. Two classes, 5 and 6 in Figure 3 (c), are independent studies part I and II, where students can decide on a research topic based on consultations with a faculty member.

**Figure 3 (d): DA_I**
Compared to Figure 3 (c), where shared topics appear across all courses, course numbers 1-2 in Figure 3 (d) do not have any shared topics, therefore, there are no marked topics. The course titles are ‘1: Information organization and access’ and ‘2: Libraries, information and society.’ Students are required to take only 2 hours of course 2.

**Figures 3 (e) – (g)**
We discussed them in the next section in detail.

**DISCUSSION AND LIMITATIONS**
We found that our approach – identifying topics from each curriculum and course descriptions, and comparing the topics – was effective in analyzing the topical scopes of two digital and two data curation curricula:

- Macroscopic-level: topics from individual curriculum descriptions and course titles (Figures 1 and 2).
- Microscopic-level: topics from courses that are part of each curriculum (Figure 3).

We may consider topics appearing in more than one curriculum as the least common denominators (LCD) of such curricula. These LCD topics indicate the fundamental similarities between curricula (Figure 3 (e)-(g)).

The topics in Figure 3 (e) reflected an overview of digital curation where digital objects are prepared (digitization, format, standards), acquired (acquisition, collection), preserved, and managed for long-term access. The topics in Figure 3 (f) illustrated the fundamental components of data curation in two categories: (1) preparing data for sharing and reuse (metadata, policy, quality control, and standards), and (2) providing effective and reliable access to the data (theory, workflow, repository, and management). Finally, Figure 3 (g) showed the two topics, ‘management’ and ‘standards,’ which were common in all curricula. This fact may reflect the importance of preparing digital and data objects based on existing standards for effective sharing and reuse, as well as the importance of continuously managing the resources in a repository for any curatorial tasks.

**Limitations**
We analyzed topics from only two digital and two data curation curricula without considering workshops, MOOCs, and other opportunities. In addition, we extracted topics from the course descriptions instead of the course syllabi. This may have limited our scope of analysis to focus on a small number of academic programs.

**CONCLUSION**
In this study, we presented an analysis of topical scopes for two digital and two data curation curricula. After identifying topics from the description of individual courses and curriculum, we compared the topics to find the similarities and differences between curricula. Our approach was meaningful in that the identified common topics between the curricula – two digital and two data curation – uncovered fundamental components of the curatorial tasks in each type of curriculum. This approach may also be applicable to other types of educational opportunities for digital/data curation such as workshops, MOOCs, or webinars.

We aim to extend this study by including analyses for more curricula and other types of opportunities. It may lead to a more accurate and comprehensive picture of topical scopes for digital and data curation, as well as their topical relationships.

**ACKNOWLEDGMENTS**
We thank all the people who have taken the leadership of creating digital / data curation curricula used in this study.

**REFERENCES**


