# Digital Preservation in LIS Education: A Content Analysis of Course Syllabi

Ayoung Yoon, Angela P. Murillo, and Paula Anders McNally

Following a revolution in digital information and technological development, both the need for and opportunities in digital preservation education have been growing over the last 30 years. This study aimed to examine digital preservation course content through an analysis of course syllabi to understand what is taught in library schools through the graduate level. Our analysis demonstrates the significant growth in digital preservation education during the last decade but also presents several areas that need to be improved, such as more integration of technology into the courses, providing advanced courses, developing core sets of literature, and developing more content for teaching issues related to ethics and diversity.

**Keywords:** course analysis, digital preservation, LIS education, syllabi analysis

Preservation education has a long history, beginning as early as the 1970s when several conferences and symposia were held to assess preservation education needs (Conway, 2014). Following a revolution in digital information and technological development, both the need for and opportunities in digital preservation education have grown over the last thirty years. Kim (2015) argued that, due to the explosion in the amount of digital information as well as the challenges presented by new types of digital technology, there is a great need to educate professionals for the long-term management and curation of digital information, especially concerning the value and long-term sustainability of such information. The production of professionals with relevant skills and knowledge has become one of the core educational goals for academic programs in library and information science (LIS) because "our increasingly digitally mediated societies face a loss of cultural and social heritage" (Shankar & Cushing, 2016, p. 13) when digital information is not properly managed and preserved. Many scholars have further underscored the need to educate and train professionals with these skills. As one scholar explains, "Not having solid understanding of digital preservation fundamentals could be a disadvantage" for information professionals, who then may be required to undergo additional training in the workplace (La Beaud, 2017, p. 8).

As the world shifts to a digital reality, many LIS schools in the US have launched,

**KEY POINTS:**

- The study findings demonstrate the need to integrate digital preservation tools and technologies into course content through class activities.
- While the course contents were commonly built around preservation actions, less emphasis was placed on topics such as policy and ethics, which need to be strengthened in preservation education.
- Digital preservation education lacks topics on diversity and inclusion, despite the fact that library professionals are committed to diversity and inclusion as a foundation of the profession.

or have continued to provide, digital preservation courses or programs at the graduate level (as well as continuing education courses for professionals) in order to meet the need for updated practices and skills within established professional fields, such as archives and record management, library and information services, and scholarly publishing (Botticelli, Fulton, Pearce-Moses, Szuter, & Watters, 2011). Several institutions in the United States have shared their experiences in digital preservation education—for instance, Simmons College (Bastian, Cloonan, & Harvey, 2010; Mahard & Harvey, 2013), the University of North Carolina at Chapel Hill (Lee, Tibbo, & Schaefer, 2007; Lee & Tibbo, 2011), the University of North Texas (Kim, 2015; Moen, Kim, Warga, Wakefield, & Halbert, 2012), the University of Arizona (Fulton, Botticelli, & Bradley, 2011), and Clayton State University (Botticelli et al., 2011). Other countries, such as the United Kingdom, Australia, New Zealand, Germany, and Switzerland, have also reported a focus on digital preservation education (Hedges, 2014; La Beaud, 2017; Recker, Engelhardt, Oßwald, & Strathmann, 2014), which indicates the level of international interest in this expanding discipline.

Despite these efforts and continued progress in the field of digital preservation education, studies (e.g., Bastian, Harvey, Mahard, & Plum, 2010; La Beaud, 2017) have found that there is little consensus among LIS courses and programs in digital preservation education, and they have noted the lack of a "standard, optimized model of best practice in digital preservation education" (La Beaud, 2017, p. 4). This suggests the need to closely investigate the current content of digital preservation education in order to propose core content, describe new and necessary skills, and integrate new technology into educational modules.

This study aimed to examine digital preservation course content using the analysis of course syllabi to understand what is being taught in library schools at the graduate level—a question that will be critical to the next generation of digital professionals. Syllabi analysis is one way of examining digital preservation education—both its content and its methods—in order to provide a systematic understanding of core course content, major topics, skill development, and integrated technology, as well as core literature in the field. While past studies (e.g. La Beaud, 2017) have reported some findings from course analysis, we suggest two ways in which these studies were deficient: either the findings were outdated, since the digital preservation field continues to change rapidly with the advent of new digital technologies (e.g., Bastian et al., 2010), or the studies themselves took a different methodological approach (e.g., La Beaud, 2017). While this study will update the findings of Bastian et al. (2010), this paper examines digital preservation syllabi from ALA-accredited MLIS programs in order to examine the current course content in digital preservation courses and to provide areas of improvement for digital preservation courses. We will also suggest several implications for the design of future digital preservation courses, with common topics covered across different institutions, as well as common skills that can be acquired through various learning activities.

## Literature review

Digital preservation education efforts have evolved significantly since the 1990s, when the Northeast Document Conservation Center hosted its first conference, "School for Scanning: A Conference on Digitization, Microfilm, and Preservation" (Tibbo, 2015). The field

of archival science contributed a great deal in paving the way for preservation education with its early interest in preservation, which began in the late 1980s (Conway, 2014). It also added value to the design of related courses, such as those regarding digital libraries (Ross, 2012). Despite the past achievements of archival education and its contribution to digital preservation education, scholars cite the continuing challenges presented by digital preservation across different types of cultural institutions, including libraries, museums, archives, and other institutions of cultural memory (Constantinescu, 2010). The need to educate professionals in this area is more pressing than ever (Tibbo, 2015) due to the rapid, massive increase in digital information. This need is also demonstrated by market demand. A recent report by the National Research Council (2015) illustrates the growth in jobs that require skills relevant to digital preservation, curation, and management.

Mahard and Harvey (2013) have argued the need to provide educational and training opportunities for digital preservation at three levels: (1) graduate programs (e.g., master's and Ph.D.), (2) post-graduate or continuing education (e.g., short courses, certificates), and (3) informal learning opportunities (e.g., job training, workshops). While all levels of education are necessary, integration into graduate education is certainly an important starting point; training those who are responsible for preserving digital information needs to be an ongoing commitment, as digital preservation strategies evolve continuously (Duff, Marshall, Limkilde, & van Ballegooie, 2006). Furthermore, opportunities in post-graduate training begin where library school teachings leave off, and the specializations and certificate programs from the American Library Association (ALA) do not currently provide a "distinct digital preservation focus" (La Beaud, 2017, p. 2).

Along with international efforts to provide digital preservation education through graduate programs (e.g., Hedges, 2014; La Beaud, 2017; Recker et al., 2014), a number of ALA-accredited US institutions have been providing courses that focus on digital preservation or digital curation, which is a way of conceptualizing preservation more broadly. These institutions include Simmons College, the University of North Carolina at Chapel Hill, the University of North Texas, the University of Arizona, and the University of Michigan. More institutions are currently integrating digital preservation–related topics into their curricula through courses on topics such as digital libraries, cataloging, or reference work (La Beaud, 2017).

These institutions have also reported challenges in teaching digital preservation or curation. One challenge is related to the field itself, as digital preservation or curation is "a rapidly developing field with an abundance of literature" (Botticelli et al., 2011, p. 151); it is difficult to keep up with these changes when selecting course content. Further, Yakel, Conway, Hedstrom, and Wallace (2011) have noted that digital curation, in particular, is "one of the hardest topics to teach precisely because ubiquitous computing dulls the sense of urgency and reinforces a sense of complacency that only those deeply immersed in the technical challenges of digital curation understand to be a chimera" (p. 23). Perhaps due to this characteristic of the field, La Beaud (2017) has noted that there is currently no "standard, optimized model of best practice" in digital preservation among digital preservation and/or curation courses across different institutions and countries (p. 4). Other common challenges include the integration of technology through laboratory or

field experiences (Bastian et al., 2010), teaching the implications of technological systems in a preservation setting beyond the understanding of what technology does (Galloway, 2014), and helping students acquire the practical skills for day-to-day work in curation (Botticelli et al., 2011).

Several studies have tried to systematically understand digital preservation and/or curation courses through the content analysis of course syllabi (e.g., Bastian et al., 2010, 2011; Costello, 2010). These studies reported that, although some consensus may have been reached on some major topics, there is still no agreement regarding the core literature within the field, since curriculum development is still in a relatively early stage. While the International Digital Curation Education and Action Working Group has been developing a shared digital curation curriculum from the analysis of existing curricula on digital preservation, curation, and stewardship, how these topics influence existing courses or curricula, or how digital preservation education has changed since the beginning of that project, is still unknown. Most recently, La Beaud (2017) conducted a syllabi analysis on those courses with a component relevant to digital preservation, but that study does not investigate courses with a sole focus on digital preservation. This study aims to investigate current digital preservation courses through the content analysis of course syllabi. Identifying common topics and skills taught within the courses, as well as the educational methodologies they represent, will not only help us to understand the current educational market but will also identify what is missing in current courses and how to improve those courses.

## Methods

This study analyzed digital preservation course syllabi using content analysis as the main method. Content analysis is "an observational research method that is used to systematically evaluate the symbolic content of all forms of recorded communication" (Kolbe & Burnett, 1991, p. 243) and is thus appropriate for our research purpose, which is to examine texts that are not subject to the influence of interests.

### Sample and data collection

In order to identify digital preservation courses taught at LIS schools, the project team examined 60 institutions from the 2017 list of ALA-approved MLS programs in the United States and Canada. Each school was given a unique number, and course catalogs and directories were examined for the presence of digital preservation courses in the brief descriptions of the courses. Our intention was to choose courses that focused solely on the topic of digital preservation, and therefore we excluded courses on digitization, digital archives, or other courses related to digital preservation issues that dealt with the topic as only a small portion of the entire course content (e.g., references or description). We included courses on digital curation because the terms *preservation* and *curation* are often used interchangeably. When an institution offered more than one course on digital preservation, we included all courses. We selected the most recent versions of the syllabi when possible, though some syllabi dated from two years back, especially in courses offered every other year. Note that we collected only those syllabi that were available online and did not make syllabi requests directly to instructors.

### Protocol development

Once we identified the digital preservation courses, we developed a protocol for analysis based on the sample set of syllabi randomly chosen from our collection. Once the protocol was finalized from the sample set, we applied the protocol to all syllabi. Our protocol consisted of two parts: the course basics and the course content. The course basics aimed to understand the course context and included course title, course-offering frequency, course level (e.g., introductory or advanced), mode of offering (e.g., online versus in-person), prerequisites, and approach to digital preservation. Regarding approach, we were particularly interested in whether the course considered (or acknowledged) archival theories or practices, considering the contribution of archival science in preservation education, as past literature has pointed out. In order to identify the course approach, we checked if the course was part of archival science education (e.g., archives specialization), if the course stated the approach in the syllabus, or if the course included archival concepts or theories as a course topic (e.g., concept of integrity, authenticity, and provenance).

The second part, the course content, aimed to understand the learning goals for students as well as the skills and knowledge instructors intended to teach. We investigated other course content in addition to course syllabi in order to understand other courses' learning goals because only a few syllabi stated the intended learning outcomes for students. The course content that we investigated included topics covered in the course, technology or tools utilized or taught in the class, textbooks used in the course, assignments (e.g., types of assignments, whether the assignments were solo or group work), and literature or other reading materials that were provided to students as assigned class readings.

### Data analysis

The identified syllabi were coded from April to August 2018. The inter-rater reliability among three project team members was 95.8%. We employed different methods for analysis. First, we used SPSS for a descriptive analysis of the data that three team members coded based on the protocol. Textual data were analyzed using Excel.

Second, we used a topic-modeling technique to identify the key themes found in the syllabi. Course topics were gathered from the course schedule sections of the syllabi. The course topics were typically located on the schedule portion of the syllabi through weekly headings and indicated the topics and subjects that would be covered during that week of the course. These topics were gathered and placed manually into a Microsoft Excel spreadsheet. They represent the overall topics or subjects examined throughout the course.

Since we were interested in gaining a better understanding of the topics covered in digital preservation curriculum, topic modeling was a good choice for analysis. Topic modeling has "proved useful for analyzing and summarizing large-scale textual data" and is a "well received, unsupervised method" (Song & Ding, 2014, p. 235). The specific algorithm we used was Latent Dirichlet Allocation (LDA). LDA is "a generative probabilistic model for collections of discrete data such as text corpora" (Blei, Ng, & Jordan, 2003, p. 993). The concept behind LDA is that "one document contains multiple topics, and each topic requires

specific words to describe it" (Song & Ding, 2014, p. 235); therefore, the observed variables are words in the documents, and the hidden variables are topics. Additionally, LDA documents have the same set of topics, but each document contains different proportions of those topics.

The probability of generating a word $w$ from a document $d$ is

$$P(w|d,\theta,\phi) = \sum_{z\in T} P(w|z,\phi_2)P(z|d,\theta_d)$$

The likelihood of a document $D$ is defined as

$$P(Z,W \mid \Theta,\Phi) = \prod_{d\in D} \prod_{z\in T} \theta_{dz}^{ndz} \times \prod_{z\in T} \prod_{v\in V} \phi_{zv}^{nzv}$$

To create the topic models, we employed the Stanford Topic Modeling Toolbox (STMT), which is a Java-based, open source tool developed by the Stanford Natural Language Processing Group (n.d.). This tool provides the ability to explore several topic model methods. A Collapsed Variation Bayes Approximation and a Collapsed Gibbs Sampler were both tested, and the Collapsed Variation Bayes Approximation was ultimately used to create the topics maps. Additionally, the STMT provided the ability to calculate the optimal number of topics for the data with perplexity calculations. The perplexity calculations indicated that five topics were the optimal number of topics for our data. Therefore, five topics were created, and the LDA model parameters were employed. The output of the topic models included the five topics, the top 20 terms per topic, and the probability of each topic and term.

LDA uses soft clustering, so each term can appear in more than one topic. Therefore, it is recommended to incorporate human judgment to increase thematic meanings of topics (Song & Ding, 2014, p. 253). We employed human interpretation of the meaning of the topics through considering the prevalent terms in each topic, the probability of each term, and the themes seen in the terms for each topic. Through this analysis, the thematic meaning of each topic was determined.

Lastly, we conducted a citation analysis to explore the nature of reading materials that were used in the classes. A reading list from the syllabi was collected manually and placed into an Excel spreadsheet. The reading list items were cleaned for consistency and accuracy using Open Refine and Excel tools. For each item, the following data were gathered and organized: author(s), year, title, journal, book chapter/tutorial chapter, volume (issue number), editor, conference, publisher, pages, link or other identifier, and type. With regard to type, each item was placed into the following categories: journal articles, books, book chapters, conference papers, master's theses, news articles, newsletters, presentations, reports, tutorials, tutorial chapters, videos, websites, websites (blogs), websites (tools), and unknown if the type of resource was unclear.

Using pivot tables in Excel, we quantitatively analyzed the reading list and created tables for the following items: publication year, publication type, top titles, top authors, top journals, top conferences, top websites, top reports, top books, top publishers for reports, top publishers for websites, and top websites (tools).

## Findings

Among the 60 ALA-accredited institutions we examined for digital preservation courses, more than half (58.3%) offered courses focusing on the topic of digital preservation or curation. These 35 institutions offered a total of 59 courses on digital preservation and/or curation, and several institutions offered more than one course. One institution offered nine courses. Some courses did not have their syllabi available online, so we further analyzed only the 36 courses that had syllabi, which were from 22 institutions.

### Course basics

All courses had either the term *preservation* or *curation* in the title. The most common title across different institutions was simply "digital preservation" or "digital curation," especially in institutions offering only one course on digital preservation. Several variations included "Creating, Managing, and Preserving Digital Assets" and "Archiving and Preserving Digital Media." When more than one course was offered, the course titles became more specific, such as "Policy Issues in Digital curation" and "Digital Forensics for Curation of Digital Collections."

A majority of the courses (86.1%) were at an introductory level, as stated on the syllabus. This was not surprising because only five institutions (22.7%) offered more than one course and included advanced-level courses. As many of them were introductory, only just over one-third (36.1%) required prerequisites before taking digital preservation courses. Many prerequisites were degree-related requirements. Other prerequisites were required in the specific context of the courses; for instance, when the course was more advanced, students were required to take introductory-level digital preservation. Three courses required archive foundational courses as prerequisites, such as "Introduction to Archives," "Electronic Recordkeeping," or "Records Access, Storage, and Retrieval." It comes as no surprise that universities required the completion of courses in archival science, considering its integral contributions to the field. Further, when we examined the course approaches, almost two-thirds (63.9%) adopted an archival approach when teaching digital preservation. However, over one-third (36.1%) also implemented a combination of library and archival techniques, while 33.3% of courses did not specify their approach to digital preservation.

The frequency of course offerings varied: just over half (52.8%) of the courses were offered annually, 27.8% were offered every semester, and 2.8% were offered every other year. The frequency for six courses (16.7%) was also indicated. Face-to-face class meetings still accounted for the majority (52.8%), but just under half of the courses were taught online (41.7%), indicating the growth of online education in LIS. Two courses (5.6%) were taught both online and face-to-face.

### Course content

#### *Textbooks*

Only eleven courses (30.6%) required one or more textbooks for the class. While some textbooks were specific to a certain context in advanced classes (e.g., film preservation), there were books commonly used in general digital preservation courses. Four different courses used *Digital Curation: A How-To-Do-It Manual* (Harvey, 2010), and other textbooks

were used in only one course each. Appendix A presents the full list of textbooks and the frequencies at which they appeared in the course syllabi.

### *Course topics*

From our topic-modeling analysis, we identified 58 common topical terms that appeared across the syllabi (Appendix B). We identified five groups of high-level topics and the top 20 terms per topic. Table 1 presents a summary of the five topics with their top 20 terms.

The probability of each topic and top 20 terms was provided through the Stanford Topic Model Toolbox (STMT). Topic 1, *preservation actions throughout digital information lifecycle*, was the most prevalent topic with 30%. Topic 2, *preservation methods*, followed as the second most prevalent with 22%, followed by Topic 3, *digital curation* (20%), Topic 4, *archival approaches to preservations* (16%), and Topic 5, *repositories* (13%). The STMT output provided the probability of each term in each topic, the probabilities were recalculated as percentages for ease of reading. Additionally, visual representations of the topics and terms are presented in Appendix C.

These findings indicate that the majority of the courses focused their content on *preservation actions throughout the digital information lifecycle* and *preservation methods*, while *archival approaches to preservation* and *repositories* were not given the same amount of focus. As indicated in Table 1 and Appendix C, Topic 1, *preservation actions throughout digital information lifecycle*, focused on broad preservation actions and functions. Higher term probabilities included *selection*, *access*, *appraisal*, and *policy*, and the two highest term probabilities were *digital* and *preservation* (over 60%). For Topic 2, *preservation methods*, the spread of terms was distributed quite evenly and focused on the methods, tools, and techniques of digital preservation. These include *metadata*, *emulation*, *repositories*, *models*, and *migration*. This group of terms shows the vast number of skills needed to work with digital preservation. Topic 3 is the only topic where we see *curation* as a dominant term in the topic, as it is the second most common term in the model. Additionally, the term *lifecycle* is found in this topic. With this, we can see the relationship between digital curation and digital preservation. Topic 4 differed from the other topics in that it thematically took on a greater archival theme, as nearly 50% of the topic focused on *digital curation* and *archival approaches to preservation*. Finally, Topic 5 reveals digital preservation as a core function at repositories, which also demonstrates that digital preservation is also taught in the context of repository functions. The terms related to preservation functions and repositories make up nearly 40% of Topic 5, and the terms related to sustainability (e.g., *planning and sustainability*, *disaster, risk*) make up nearly 30%. Thus, it is shown that Topic 5 focuses more on the granular details of digital preservation functions at repositories.

### *Assignments*

Examining assignments can often indicate the kinds of skills and knowledge that instructors want students to learn. A total of 140 assignments were analyzed from the syllabi, though the level of description for each assignment varied. Project-based assignments were most common (41.4%). The nature of projects varied, but usually projects seemed to be designed

**Table 1: Topic probability with top 20 terms for each topic**

| Topic 1 | Topic 2 | Topic 3 | Topic 4 | Topic 5 |
|---|---|---|---|---|
| Preservation actions throughout digital information lifecycle (30%) | Preservation methods (22%) | Digital curation (20%) | Archival approaches to preservation (16%) | Repositories (13%) |
| Digital (29.90%) | Management (8.44%) | Digital (15.83%) | Digital (23.42%) | Preservation (27.35%) |
| Preservation (29.08%) | Metadata (7.88%) | Curation (12.19%) | Archives (10.96%) | Repositories (13. 04%) |
| Selection (4.59%) | Authenticity (7.34%) | Data (12.16%) | Digitization (9.23%) | Materials (8.96%) |
| Access (3.82%) | Repository (7.26%) | Preservation (7.15%) | Archival (7.87%) | Planning (8.14%) |
| Appraisal (3.72%) | Emulation (6.44%) | Content (5.73%) | Management (6.99%) | Institutional (7.09%) |
| Policy (3.02%) | Digital (6.41%) | Information (5.46%) | Tools (5.55%) | Sustainability (6.95%) |
| Objects (2.98%) | Formats (5.76%) | Standards (4.69%) | Ethics (5.39%) | Forensics (5.13%) |
| Collection (2.89%) | Models (5.74%) | Management (4.61%) | Systems (4.09%) | Archival (4.23%) |
| Strategies (2.81%) | Migration (5.37%) | Records (4.58%) | Plans (4.06%) | Collection (3.86%) |
| Metadata (2.70%) | Preservation (5.14%) | Research (4.57%) | Formats (3.65%) | Disaster (3.57%) |
| Curation (2.52%) | OAIS (4.88%) | Trusted (4.26%) | Concepts (2.77%) | Web (3.53%) |
| Use (2.51%) | Data (4.70%) | Repositories (3.89%) | Media (2.74%) | Selection (3.45%) |
| Authenticity (2.31%) | Storage (4.26%) | Users (3.37%) | Representation (2.71%) | Standards (1.58%) |
| Challenges (2.17%) | File (3.38%) | Workflows (3.24%) | File (2.42%) | Models (0.93%) |
| Trusted (1.70%) | Legal (3.22%) | Access (3.24%) | OAIS (1.86%) | Risk (0.81%) |
| Planning (1.07%) | Concepts (3.17%) | Lifecycle (2.24%) | Collections (1.80%) | Reuse (0.44%) |
| Digitization (0.99%) | Project (2.93%) | Use (1.59%) | Strategies (1.71%) | Issues (0.27%) |

(*Continued*)

| Topic 1 | Topic 2 | Topic 3 | Topic 4 | Topic 5 |
|---|---|---|---|---|
| Models (0.86%) | Archival (2.83%) | Reuse (0.58%) | Project (1.69%) | Workflows (0.24%) |
| Reuse (0.19%) | Collections (2.72%) | Archival (0.42%) | Preservation (0.65%) | Archives (0.22%) |
| Information (0.18%) | Properties (2.12%) | Collections (0.18%) | Use (0.43%) | Challenges (0.20%) |

to give students a real-life context of preservation work through existing cases or real-life examples. Several common examples involved working with local partner organizations, either to analyze their preservation needs or to develop preservation strategies. Sometimes, students were given a scenario in which they had to identify solutions for various preservation issues. Several assignments prompted students to work with preservation tools and technology (e.g., DSpace, BitCurator) either to suggest a solution for identified preservation issues or to implement preservation strategies. Presentations (7.8%) were usually associated with the project assignments as a way for students to share what they learned from their projects. The second common type of assignment was writing assignments (33.5%). Short essays (16.4%), including written discussions and journals, were required of many students across online and on-campus classes. Usually, these short essays involved assigned weekly readings or weekly topics, which also appeared to be part of in-class participation for on-campus classes. Other writing assignments (17.1%) included research papers on the topic of digital preservation, literature reviews, or other types of analysis, such as preservation policy comparison. Quizzes were rarely used as an assignment, and when they were used, they usually concerned the core concepts of digital preservation and models (e.g., Open Archival Information System Reference Model, Digital Curation Center's Digital Curation Lifecycle Model). Among different types of assignments, project-based assignments had a high percentage of collaborative work, with two to five students (40%). Still, a majority of assignments (75%) required individual work, while only 13.5% of all the assignments involved group work.

We looked into the details of assignments and identified the top 10 topics covered in the assignments. The most common focus of the assignments was the investigation of digital preservation and curation issues through analysis of case studies or scenarios in order to suggest strategies or solutions (22.8%). Other assignments focused on preservation assessment or metadata. This was usually done as a group project and required a holistic approach that encompassed the many topics that the class had been exploring (e.g., policy, workflow). While a number of classes assigned different forms of short essays, such as written discussions, many assignments (15%) also stated that these assignments should address the assigned weekly topic. Only 7.9% of assignments intended to give students the opportunity to either explore or work with digital preservation tools and technology. Topics relating to digital repository practice appeared to be the fourth most common among

**Table 2: Types of assignments (*N* = 140)**

| | | Collaboration | | | Total (with percentage) |
|---|---|---|---|---|---|
| | | Individual | Group | Not specified | |
| Type | Short essay (e.g., written discussion, journal) | 23 | 0 | 0 | 23 (16.4%) |
| | Paper (e.g., research, literature review, other types of analysis) | 20 | 2 | 2 | 24 (17.1%) |
| | Project (e.g., hands-on project, project with partners) | 32 | 13 | 13 | 58 (41.4%) |
| | Quiz | 4 | 0 | 0 | 4 (2.9%) |
| | Presentation | 7 | 3 | 1 | 11 (7.9%) |
| | In-class participation | 14 | 0 | 0 | 14 (10%) |
| | Other | 5 | 1 | 0 | 6 (4.3%) |
| Total (with percentage) | | 105 (75%) | 19 (13.5%) | 16 (11.5%) | 140 (100%) |

the assignments, followed by preservation policy (5.7%) and file format (4.3%). Although we excluded courses focused solely on data curation and data management, some digital preservation or curation courses guided students in completing at least one assignment on data management (2.1%). Other topics included outreach and programming (1.4%), grant writing (1.4%), and disaster planning (1.4%).

### *Technology*

When we examined the digital preservation–related tools and technologies integrated throughout the courses, either through the assignments or other types of class learning activities, only 12 courses (33.3%) specified the tools and technology that the course aimed to teach to the students. We defined the use of technology as an "active learning activity" such as hands-on laboratories using digital preservation infrastructure (e.g., DSpace), a forensic tool (e.g., BigCurator), file identification (e.g., JOVE), or utilizing these technologies in students' assignments (e.g., review of tools, comparison of existing tools). Since we excluded the courses that only briefly mentioned the technology and included only the courses that required students to directly work with the tools and technology, it is possible that more courses have introduced these resources. Among the tools and technology utilized in the classes, DSpace was the most highly used (four courses, 33.3%), followed by BitCurator (three courses, 25%). Other tools included iRods, DRAMBORA, and Fedora Commons. Some archival management tools were also mentioned, such as ArchivesSpace and Archivematica.

**Table 3: Assignment topics (*N* = 140)**

| | Topics | Frequency with percentage | Sub-topics | Frequency |
|---|---|---|---|---|
| **1** | Digital preservation & curation (general) | 32 (22.8%) | Case study | 12 |
| | | | Assessment & planning | 10 |
| | | | Metadata | 6 |
| | | | Issue searching & analysis | 3 |
| | | | Guide comparison | 1 |
| **2** | Weekly readings or assigned weekly topics | 21 (15%) | | |
| **3** | Tools & Technology | 11 (7.9%) | Review or evaluation | 7 |
| | | | Working with DSpace, DRAMBORA, etc. | 4 |
| **4** | Digital repository | 10 (7.1%) | Trusted digital repository review & analysis | 3 |
| | | | Repository evaluation | 3 |
| | | | Implementation | 2 |
| | | | Not specified | 2 |
| **5** | Preservation policy | 8 (5.7%) | Policy development | 5 |
| | | | Policy analysis & comparison | 3 |
| **6** | File format | 6 (4.3%) | Assessment, analysis, & strategies | 6 |
| **7** | Data management plan | 3 (2.1%) | Analysis, evaluation, plan | 3 |
| **8** | Outreach/Programming | 2 (1.4%) | | |
| **9** | Grant writing | 2 (1.4%) | | |
| **10** | Disaster planning | 2 (1.4%) | | |
| **Total** | | **97 (69.2%)** | | |

### *Reading resources*

Literature or other reading materials required for students are an integral part of the learning process, since they not only help students explore the topics but also introduce different approaches or perspectives on digital preservation issues. We analyzed a total of 1,919 reading resources from the syllabi to understand what types of content and materials were introduced to the students. In terms of publication types, journal articles were the most common type of reading resources that were required for the students (32.1%), followed by reports (19.6%) and websites (19.0%). Figure 1 illustrates all resource types by count. A few resources dated from the 1980s and early 1990s, but most resources were published sometime during the 2000s, and 44.4% were from after 2010 (Figure 2).

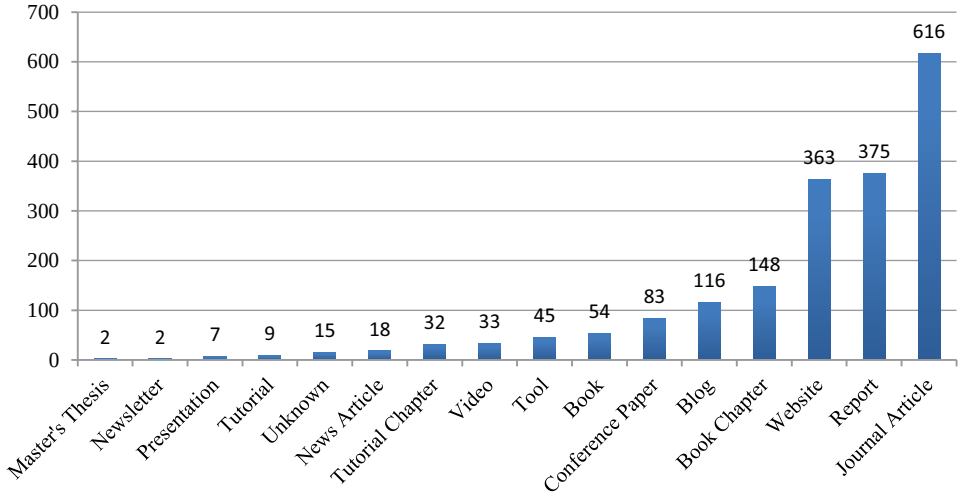


**Figure 1:** Reading resource types by count (*N* = 1,919)

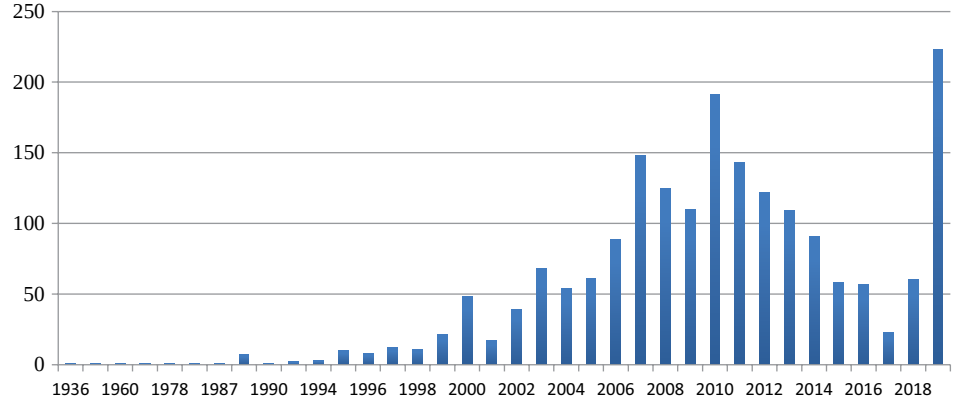


**Figure 2:** Publication years by count (*N* = 1,919)

While the materials provided for students to read were diverse in terms of types and titles—and no group of materials was dominant among all reading resources—we were able to identify the most commonly assigned reading materials across different source types. Table 3 presents the top 10 reading materials for students by both frequency and percentage. Note that *Digital Curation: A How-To-Do-It Manual* (Harvey, 2010) was the most commonly used textbook as well, and when it was not used as a textbook, the manual was still heavily utilized as a class reading resource.

**Table 4: Top ten reading materials (*N* = 1,919)**

| | Title | Frequency | Percentage |
|---|---|---|---|
| 1 | Cornell University Library. (2007). *Digital Preservation Management Tutorial: Implementing Short-Term Strategies for Long-Term Problems*. http://www.dpworkshop.org/dpm-eng/eng_index.html. | 28 | 1.5% |
| 2 | Harvey, R. 2010. *Digital Curation: A How-To-Do-It Manual*. New York: Neal-Schuman Publishers, Inc. | 23 | 1.2% |
| 3 | The DCC Curation Lifecycle Model. http://www.dcc.ac.uk/resources/curation-lifecycle-model; Higgins, S. (2008). The DCC Curation Lifecycle Model. *International Journal of Digital Curation*, *3*(1). | 15 | 0.8% |
| 4 | Harvey, R. (2011). *Preserving Digital Materials*. Berlin, Boston: De Gruyter Saur. | 13 | 0.7% |
| 5 | Cornell University Library. (2003). *Moving Theory into Practice: Digital Imaging Tutorial*. http://preservationtutorial.library.cornell.edu/contents.html. | 13 | 0.7% |
| 6 | Kirschenbaum, M.G., Ovenden, R., and Redwine, R. (2010). *Digital Forensics and Born-Digital Content in Cultural Heritage Collections*. Council on Library and Information Resources (CLIR). http://www.clir.org/pubs/reports/pub149/reports/pub149/pub149.pdf | 11 | 0.6% |
| 7 | Hughes, L. (2004). *Digitizing Collections: Strategic Issues for the Information Manager*. New York: Neal-Schuman Publishers, Inc. | 10 | 0.5% |
| 8 | CCSDS. 2001. *Audit and Certification of Trustworthy Digital Repositories, Magenta Book*. Washington DC: CCSDS. http://public.ccsds.org/publications/archive/652x0m1.pdf. | 9 | 0.5% |
| 9 | Blue Ribbon Task Force on Sustainable Digital Preservation and Access. (2010). *Sustainable Economics for a Digital Planet: Ensuring Long Term Access to Digital Information*. http://brtf.sdsc.edu/biblio/BRTF_Final_Report.pdf. | 8 | 0.4% |
| 10= | Library of Congress. (2003). *Sustainability of Digital Formats Planning for Library of Congress Collections*. http://www.digitalpreservation.gov/formats/. | 7 | 0.4% |
| 10= | Ball, A. (2010). *Preservation and Curation in Institutional Repositories*. Digital Curation Centre. http://www.dcc.ac.uk/sites/default/files/documents/reports/irpc-report-v1.3.pdf. | 7 | 0.4% |
| 10= | Digital Curation Centre & Digital Preservation Europe. (2007). *Digital Repository Audit Method Based on Risk Assessment* (*DRAMBORA*) *Toolkit. 1.0*. http://www.repositoryaudit.eu/. | 7 | 0.4% |
| Total | | 151 | 8.1% |

We further examined several types of resources that were commonly utilized—including journal articles, conference proceedings, web resources (including general websites, blogs, and websites that introduced the preservation tools), and reports—to identify the parent sources, publishers, or organizations they had in common. From the analysis, we identified 150 different journal names from 616 journal articles, 52 conference names from 83 conference papers, 523 organization or publisher names from 524 web resources, and 374 names of organizations from 376 reports. Table 5 presents the five most common parent source names from each category.

## Discussion

Our syllabi analysis suggests a noticeable improvement in digital curation education in terms of course offerings, especially considering only seven institutions offered digital preservation courses in 2003 (Gracy & Croft, 2011). Although our analysis concerned only 36 courses from 22 institutions, we found that more than half of ALA-accredited institutions offered one or more courses on digital preservation/curation. This reflects higher education's recognition of the growing need to train students with relevant skills and knowledge.

Despite this progress, many courses were still offered at an introductory level, with the exception of some institutions—such as the University of Maryland, San Jose State University, University of North Carolina at Chapel Hill, University of Illinois Urbana Champaign—where more specialized educational options related to digital preservation or curation were available. While introductory courses may be a good starting point, they may not be sufficient to educate students with the practical skills and hands-on experience needed to apply theories and concepts to a real work context. Thus, educating students with the practical skills for the day-to-day work in curation may still be challenging at these introductory level courses, as Botticelli et al. (2011) have argued. Additionally, as introductory-level classes, about one-third of the courses did not require any prerequisites, so it may be necessary to spend more time covering foundational or related concepts.

Perhaps because of the introductory nature of most courses, only about one-third of the classes integrated digital preservation tools and technologies into the course content through class activities. An even smaller percentage (7%) of assignments incorporated technology for student coursework. While some courses covered recent technologies, classes fitting this category were relatively low in number considering the role of preservation technology in practical work. Indeed, the courses that we examined rarely utilized laboratories. Previous studies have already pointed out challenges in integrating technology, especially through laboratory or field experience, into the current curriculum (Bastian et al., 2010). Overcoming these challenges often requires more human and financial resources, and sometimes more physical and network resources. While recent tools supporting digital preservation work are important, integrating these technologies can provide more practical perspectives and skills to students, so they need to be more thoroughly pursued.

Many courses attempted to help students develop critical and analytic skills, understand and analyze problems, and suggest solutions by letting them work with real problems. For instance, group projects emerged as the most common type of assignment, which often

**Table 5: Top 5 resources in each resource category**

| | Journal title | Frequency (*N* = 616) | Percentage |
|---|---|---|---|
| 1 | *International Journal of Digital Curation* | 81 | 13.1% |
| 2 | *D-Lib Magazine* | 80 | 13.0% |
| 3 | *American Archivist* | 69 | 11.2% |
| 4 | *Archivaria* | 46 | 7.5% |
| 5 | *Archival Science* | 34 | 5.5% |
| **Total** | | **310** | **50.3%** |
| | Conference title | Frequency (*N* = 83) | Percentage |
| 1 | iPres | 10 | 12.0% |
| 2 | International Digital Curation Conference | 7 | 8.4% |
| 3 | Archiving Conference, Society for Imaging Science & Technology | 6 | 7.2% |
| | Joint Conference on Digital Libraries | 6 | 7.2% |
| 5 | IFLA Conference | 4 | 4.8% |
| | Museums and the Web | 4 | 4.8% |
| | The State of Digital Preservation | 4 | 4.8% |
| **Total** | | **41** | **49.2%** |
| | Report publisher | Frequency (*N* = 374) | Percentage |
| 1 | Council on Library and Information Resources (CLIR) | 45 | 12% |
| 2 | Digital Preservation Coalition | 29 | 7.8% |
| 3 | Northeast Document Conservation Center | 23 | 6.1% |
| 4 | Washington DC: CCSDS | 20 | 5.3% |
| 5 | Digital Curation Centre (DCC) | 19 | 5.0% |
| **Total** | | **136** | **36%** |
| | Web resource publisher | Frequency (*N* = 523) | Percentage |
| 1 | Library of Congress | 53 | 10.1% |
| 2 | Library of Congress. The Signal Blog | 22 | 4.2% |
| | Digital Curation Centre (DCC) | 22 | 4.2% |
| 4 | Inter-University Consortium for Political and Social Research (ICPSR) | 12 | 2.3% |
| 5 | Council on Library and Information Resources (CLIR) | 11 | 2.1% |
| **Total** | | **120** | **22.9%** |

involved working hands-on with local partners. When collaborating with a partner was not possible, a good number of assignments asked students to work with given scenarios, whether from real cases or realistically fabricated ones.

Our analysis of the course topics indicates that the courses were commonly built around the preservation actions throughout the digital information lifecycle (i.e., *selection*, *access*, *appraisal*), accompanied by the preservation methods and techniques (i.e., *metadata and emulation*). While our analysis presents several topics common within digital preservation across different classes, we also found less emphasis placed on topics such as policy and ethics, despite their importance, which aligned with the general tendency in LIS education, lacking in courses in ethics or courses with ethics content (Garnar, 2016). Both terms, policy and ethics, appeared only once in relation to two topics: policy appeared with Topic 1, *preservation actions throughout digital information lifecycle*, and ethics appeared with Topic 4, *archival approaches to preservation*. Many previous studies have underscored the ethical consideration for digital preservation work, when selecting digital resources for preservation, retaining their integrity, and determining preservation responsibilities (Berger, 2009). Also, preservation policy plays a vital role in developing a strategic roadmap and encouraging proactive and responsible preservation practices (Mannheimer, Yoon, Greenburg, Feinstein, & Scherle, 2014). It is critical to teach professional ethics in digital preservation, especially because ethics become even more significant when dealing with content that is related to personally identifiable information. While it is not surprising that the term "ethics" appeared only in relation to *archival approaches to preservation*, considering that many ethical decisions are involved in archival work, more emphasis on digital preservation work in general should be made to benefit students fully.

It is important to note that digital preservation education is still deeply rooted in archival science. This is not surprising, as the archival approach was dominant when we examined course approaches from the syllabi. This was also strongly supported by the course content, as three of the top five journals used in classes were major archival journals, and two of the top five conference proceedings used in classes were archival conferences. Additionally, Topic 4 focused specifically on the *archival approaches to preservation*, and two other topics (Topic 3 and Topic 5) included archives-related items in the top twenty terms. As Conway (2014) has argued, because archival science contributed much to digital preservation education, some articles from archival studies were appropriated to outline the core archival concepts related to digital preservation, such as authenticity, provenance, and integrity. While other articles were concerned with preservation models and standards and their implications (e.g., OAIS), many others addressed different preservation issues that had broader implications outside of an archival context.

Our analysis of reading resources indicates that most courses updated the course readings and followed the fast-changing trends of digital preservation. Many articles were from the 2010s, and even more articles were from 2018, although several articles that address the core concepts were from the late 1990s. We also noticed that there is still less consensus on the core set of literatures across different syllabi, as the overall distribution of all reading resources was pretty equal (unlike the major topics taught in courses, which were easily identifiable). As Botticelli et al. (2011) have argued, the field itself is rapidly changing and

developing an abundance of literature, which could be one reason for this lack of consensus. The most commonly utilized course readings were textbooks or tutorial-based articles from a few leading institutions in the field, which may also suggest the need to develop more teaching-focused resources.

Our analysis also notes the influence that many organizations have on course content due to their significant contribution to the field. For instance, the Digital Curation Centre (DCC) was a major contributor of course content. The different types of resources published from DCC were heavily used as class readings, including journal articles, conference articles, reports, and web materials—all of which were listed as one of the top five resources. Furthermore, the DCC's Digital Curation Lifecycle Model was the third most frequently used reading resource, which suggests a significant impact. The content produced from the Library of Congress was also utilized a lot, mostly through their website and blogs.

Finally, we also found that general digital preservation education lacks topics on diversity. Library professionals are committed to diversity and inclusion as a foundation of the profession. The lack of a key focus in developing diversity as part of the curriculum has been an issue in overall LIS education (Adkins, Virden, & Yier, 2015; Jaeger, Subramaniam, Jones, & Bertot, 2011) and thus appears to be an issue in digital preservation courses across different institutions. Digital preservation requires not only a technical approach but also social approaches, and it should provide perspectives that entail diversity issues, such as implications for underrepresented populations and communities. In addition, current courses were heavily utilizing materials from North America and the United Kingdom. While it is true that these regions have been leading the development of the field, students would benefit from being exposed to international work and perspectives in the field.

## Conclusion

Although this study contributes to the understanding of one facet of digital preservation education through course syllabi analysis, it also has some limitations. Due to the limited availability and access of the course syllabi, this study analyzed only 36 syllabi from 22 institutions among a pool of 59 courses from 35 institutions, which limited the study results. Also, syllabi provided different levels of detail for each course. Some factors, such as assignment descriptions and in-class activities, were accordingly excluded from our analysis. Finally, since we addressed only courses with a sole focus on digital preservation, we did not count other courses that may teach related skills and knowledge, such as metadata, archives, policy, digital collections, etc. A study by La Beaud (2017) adopted this broad approach and examined other related courses, which may serve as a complementary analysis to our study.

In addition, while topic modeling is an excellent tool for analyzing large amounts of unstructured data in order to find patterns, there are several limitations to topic modeling, and there are specific limitations to LDA. First, the topics themselves can be difficult to interpret and require the incorporation of human judgment to determine the meaning of each topic. Additionally, because LDA uses soft clustering, terms can appear in more than one topic (Song & Ding, 2014, p. 253), making it difficult to determine themes for each topic. Lastly, for this particular corpus, the sample was quite small. Topic modeling is more often conducted on larger samples of text which likely would provide more diverse results.

Lastly, our study examined only a graduate-level curriculum, whereas post-graduate or continuing education programs are important to have a full understanding of digital preservation education. Continuing education especially played a key role when demands for individuals with digital preservation skills greatly exceeded the supply and when existing practitioners sought an opportunity to train themselves in this area. Investigating post-master's certificate in digital preservation as well as curriculum on institutes that have a focus on digital preservation, such as DigCCur (https://ils.unc.edu/digccurr/), would provide deeper insights on digital preservation education.

Our study investigates digital preservation courses in LIS by using the content analysis method on course syllabi. Our analysis demonstrates the significant growth in digital preservation education, but it also presents several areas that need to be improved, such as more integration of technology into the courses, providing advanced courses, developing core sets of literature from the vast amount of existing literature, and developing more content for teaching issues related to ethics and diversity.

Ayoung Yoon, *Indiana University Purdue University Indianapolis (IUPUI), School of Informatics and Computing, Department of Library and Information Science*, is an assistant professor in the School of Informatics and Computing (SOIC) at Indiana University Purdue University Indianapolis (IUPUI). Her research areas include data curation, digital preservation, and data sharing/reuse. Email: ayyoon@iupui.edu

Angela P. Murillo, *Indiana University Purdue University Indianapolis (IUPUI), School of Informatics and Computing, Department of Library and Information Science*, is an assistant professor in the SOIC at IUPUI. Her research focus is scientific data management, scientific data cyberinfrastructure, and data science education. Email: apmurill@iu.edu

Paula Anders McNally, *Indiana University Purdue University Indianapolis (IUPUI), School of Informatics and Computing, Department of Library and Information Science*, is associate faculty in the SOIC at IUPUI. She holds a PhD in history and teaches courses in the field of archival science. Email: pjmcnall@iu.edu

## References

Adkins, D., Virden, C. & Yier, C. (2015). Learning about diversity: The roles of LIS education, LIS associations, and lived experience. *The Library Quarterly*, *85*(2), 139–149. https://doi.org/10.1086/680153

Bastian, J., Cloonan, M. V., & Harvey, R. (2011). From teacher to learner to user: Developing a digital stewardship pedagogy. *Library Trends*, *59*(4), 607–622. https://doi.org/10.1353/lib.2011.0012

Bastian, J., Harvey, R., Mahard, M., & Plum, T. (2010). Building a virtual archives and preservation curriculum laboratory at Simmons College: A case study in collaborative construction. *Journal of Education for Library and Information Science*, *51*(4), 241–250. Retrieved from http://www.jstor.org/stable/25764641

Berger, S. (2009). The evolving ethics of preservation: Redefining practices and responsibilities in the 21st century. *Serials Librarian*, *57*(1–2), 57–68. https://doi.org/10.1080/03615260802669086

Blei, D., Ng, A., & Jordan, M. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, *3*, 993–1022.

Botticelli, P., Fulton, B., Pearce-Moses, R., Szuter, C., & Watters, P. (2011). Educating digital curators: Challenges and opportunities. *International Journal of Digital Curation*, *6*(2), 146–164. https://doi.org/10.2218/ijdc.v6i2.193

Constantinescu, N. (2010). Adapting the information professionals to the digital collections universe. *Eprints in Library and Information Science (E-LIS)*. Retrieved from http://eprints.rclis.org/14755/1/Adapting_the_information_professionals_to_the_digital_collections_universe.pdf

Conway, P. (2014). At the nexus of analog and digital: Introduction to papers from a symposium of preservation educators. *Preservation, Digital Technology & Culture (PDT&C)*, *43*(1–2), 2–8. https://doi.org/10.1515/pdtc-2014-1001

Costello, K. L. (2010). Digital preservation education in iSchools. Paper presented at iConference 2010, University of Illinois, Champaign-Urbana. Retrieved from https://www.ideals.illinois.edu/bitstream/handle/2142/15054/Costello_iSchools_Abstract.pdf?sequence=2

Duff, W., Marshall, A., Limkilde, C., & van Ballegooie, M. (2006). Digital preservation education: Educating or networking? *The American Archivist*, *69*(1), 188–212. https://doi.org/10.17723/aarc.69.1.0437j178u17665v5

Fulton, B., Botticelli, P., & Bradley, J. (2011). DigIn: A hands-on approach to a digital curation curriculum for professional development. *Journal of Education for Library and Information Science*, *52*(2), 95–109. Retrieved from http://www.jstor.org/stable/41308885

Galloway, P. K. (2014). Commentary on integrating research and teaching. *Preservation, Digital Technology & Culture (PDT&C)*, *43*(1–2), 51–53. https://doi.org/10.1515/pdtc-2014-1011

Garnar, M. (2016). Professional principles and ethics in LIS graduate curricula. *Journal of Intellectual Freedom and Privacy*, *1*(2–3), 6–12. https://doi.org/10.5860/jifp.v1i2-3.6166

Gracy, K. F., & Croft, J. A. (2011). Quo vadis, preservation education? *Library Resources & Technical Services*, *50*(4), 274–94. https://doi.org/10.5860/lrts.50n4.274

Harvey, R. (2010). Digital curation: A how-to-do-it manual. New York, NY: Neal-Schuman.

Hedges, M. (2014). International Master's program in digital curation at King's College London and Humboldt-Universität zu Berlin. *Preservation, Digital Technology & Culture (PDT&C)*, *43*(3), 137–138. https://doi.org/10.1515/pdtc-2014-1019

Jaeger, P., Subramaniam, M., Jones, C., & Bertot, J. (2011). Diversity and LIS education: Inclusion and the age of information. *Journal of Education for Library and Information Science*, *52*(2), 166–183.

Kim, J. (2015). Competency-based curriculum: An effective approach to digital curation education. *Journal of Education for Library and Information Science*, *56*(4), 283–297.

Kolbe, R. H., & Burnett, M. S. (1991). Content-analysis research: An examination of applications with directives for improving research reliability and objectivity. *Journal of Consumer Research*, *18*(2), 243–250. https://doi.org/10.1086/209256

La Beaud, E. (2017). Analysis of digital preservation course offerings in ALA accredited graduate programs. *SLIS Connecting*, *6*(2), Article 9. https://doi.org/10.18785/slis.0602.09

Lee, C., & Tibbo, H. (2011). Where's the archivist in digital curation? Exploring the possibilities through a matrix of knowledge and skills. *Archivaria*, *72*(12), 123–168.

Lee, C. A., Tibbo, H. R., & Schaefer, J. C. (2007). Defining what digital curators do and what they need to know: The Digccurr project. In Proceedings of the 7th ACM/IEEE-CS Joint Conference on Digital Libraries (pp. 49–50). New York, NY: ACM. https://doi.org/10.1145/1255175.1255183

Mahard, M., & Harvey, R. (2013). Digital stewardship education at the graduate school of library and information science, Simmons College, 1–4. Paper presented at the DigCurV 2013, Rheinisch-Westfaelische Technische Hochschule Aachen * Lehrstuhl Informatik V. Retrieved from http://researchbank.rmit.edu.au/view/rmit:23347

Mannheimer, S., Yoon, A., Greenberg, J., Feinstein, E., & Scherle, R. (2014). A balancing act: The ideal and the realistic in developing Dryad's preservation policy. *First Monday*, *19*(8). https://doi.org/10.5210/fm.v19i8.5415

Moen, W., Kim, J., Warga, E., Wakefield, J., & Halbert, M. (2012). iCAMP: Building digital information curation curriculum. In *Proceedings of 2012 iConference* (pp. 648–650). Retrieved from digital.library.unt.edu/ark:/67531/metadc86944/

National Research Council. (2015). *Preparing the workforce for digital curation*. Washington, DC: The National Academies Press.

Recker, A., Engelhardt, C., Oßwald, A., & Strathmann, S. (2014). Training and education in digital curation in Germany, Austria, and Switzerland. In *Archiving Conference, Archiving 2014 Final Program and Proceedings* (pp. 160–163). The Society for Imaging Science and Technology.

Ross, S. (2012). Digital preservation, archival science and methodological foundations for digital libraries. *New Review of Information Networking*, *17*(1), 43–68. https://doi.org/10.1080/13614576.2012.679446

Shankar, K., & Cushing, A. (2016). Creating an evidence-based digital curation curriculum for Ireland: Case study at University College Dublin. *An Leabharlann: The Irish Library*, *25*(2). Retrieved from http://researchrepository.ucd.ie/handle/10197/8124

Song, M., & Ding, Y. (2014). Topic modeling: Measuring scholarly impact using a topical lens. In Y. Ding, R. Rousseau, & D. Wolfram (Eds.), *Measuring scholarly impact*, (pp. 235–257). Cham, Switzerland: Springer International. https://doi.org/10.1007/978-3-319-10377-8_11

The Stanford Natural Language Processing Group. (n.d.). Stanford topic modeling toolbox. Retrieved from https://nlp.stanford.edu/software/tmt/tmt-0.4/

Tibbo, H. (2015). Digital curation education and training: From digitization to graduate curricula to MOOCs. *International Journal of Digital Curation*, *10*(1), 144–153. https://doi.org/10.2218/ijdc.v10i1.352

Yakel, E., Conway, P., Hedstrom, M., & Wallace, D. (2011). Digital curation for digital natives. *Journal of Education for Library and Information Science*, *52*(1), 23–31.

## Appendix A: List of textbooks and their frequency counts in the course syllabi (*N* = 11)

- Harvey, R. (2010). *Digital Curation: A How-To-Do-It Manual*. New York: Neal-Schuman. (4)
- Harvey, R. (2012). *Preserving Digital Materials* (2nd ed.). Berlin: De Gruyter Saur. (2)
- Corrado, E. M., & Moulaison Sandy, H. (2017). *Digital Preservation for Libraries, Archives, and Museums* (2nd Ed.). Lanham, MD: Rowman & Littlefield Publishers. (1)
- Gladney, H. (2007). *Preserving Digital Information*. Berlin: Springer. (1)
- Ball, A. (2010). *Preservation and Curation in Institutional Repositories*. Bath: Digital Curation Centre. (1)
- Oliver, G and Harvey, R. (2016). *Digital Curation* (2nd Ed.). Chicago: American Library Association. (1)
- Sitts, M. K. (Ed.), (2000). *Handbook for Digital Projects: A Management Tool for Preservation and Access*. Andover, MA: Northeast Document Conservation Center. (1)

## Appendix B: List of topical terms

| Collection | Models | Institutional |
|---|---|---|
| Preservation | Legal | Sustainability |
| Selection | Issues | Users |
| Archival | Data | Research |
| Materials | Management | Workflows |
| Disaster | Strategies | Objects |
| Planning | Policy | Content |
| Archives | Representation | Forensics |
| Digital | Use | Properties |
| Curation | Reuse | Standards |
| Access | Ethics | OAIS |
| Records | Digitization | Information |
| Appraisal | Plans | Tools |
| Authenticity | Project | Systems |
| Metadata | Media | File |
| Web | Challenges | Collections |
| Migration | Concepts | Significant |
| Emulation | Storage | Risk |
| Formats | Repository | Lifecycle |
| Repositories | Trusted | |

## Appendix C: Term probability per topic

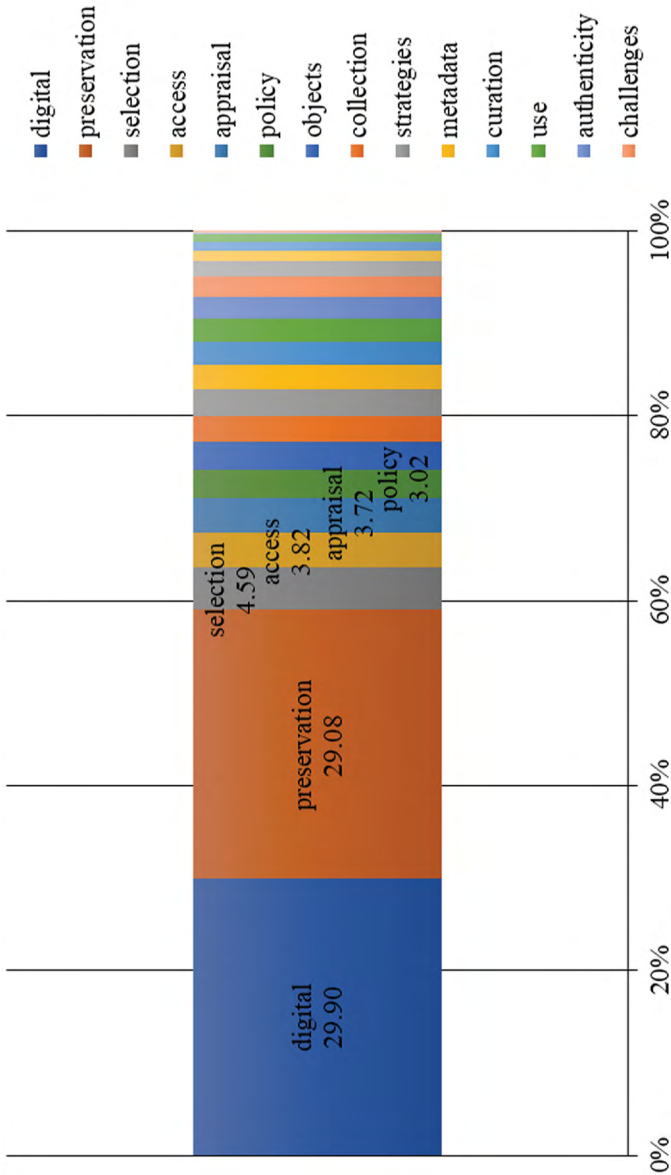


**Figure A1:** Topic 1: Preservation actions throughout digital information lifecycle

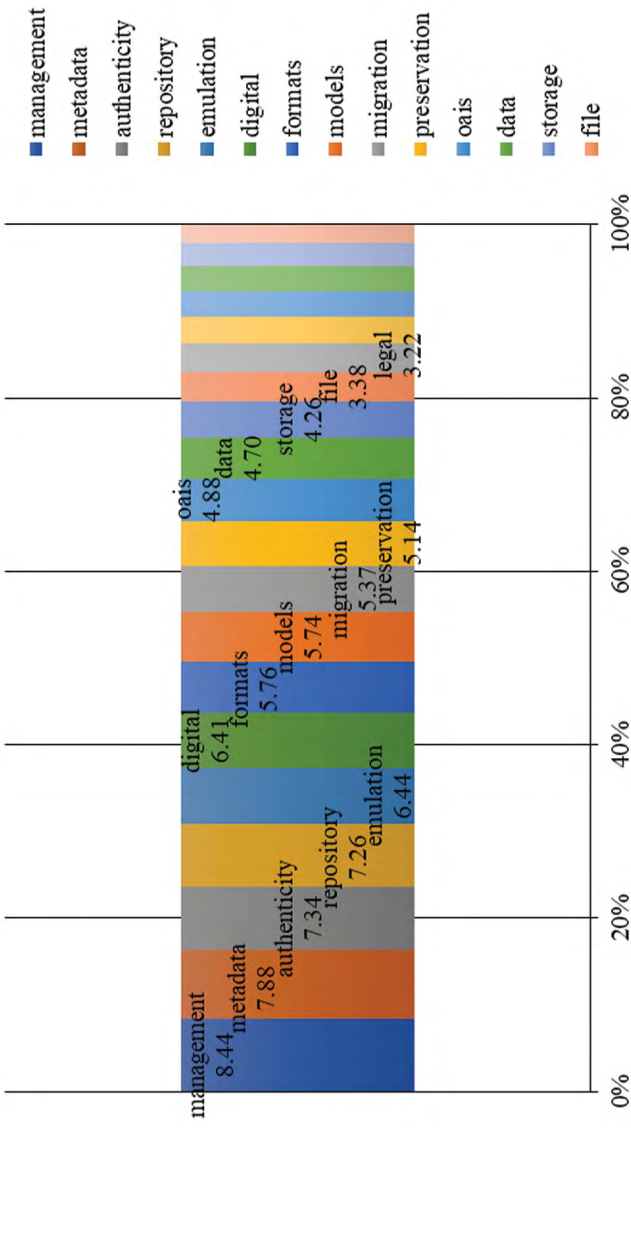


**Figure A2:** Topic 2: Preservation methods

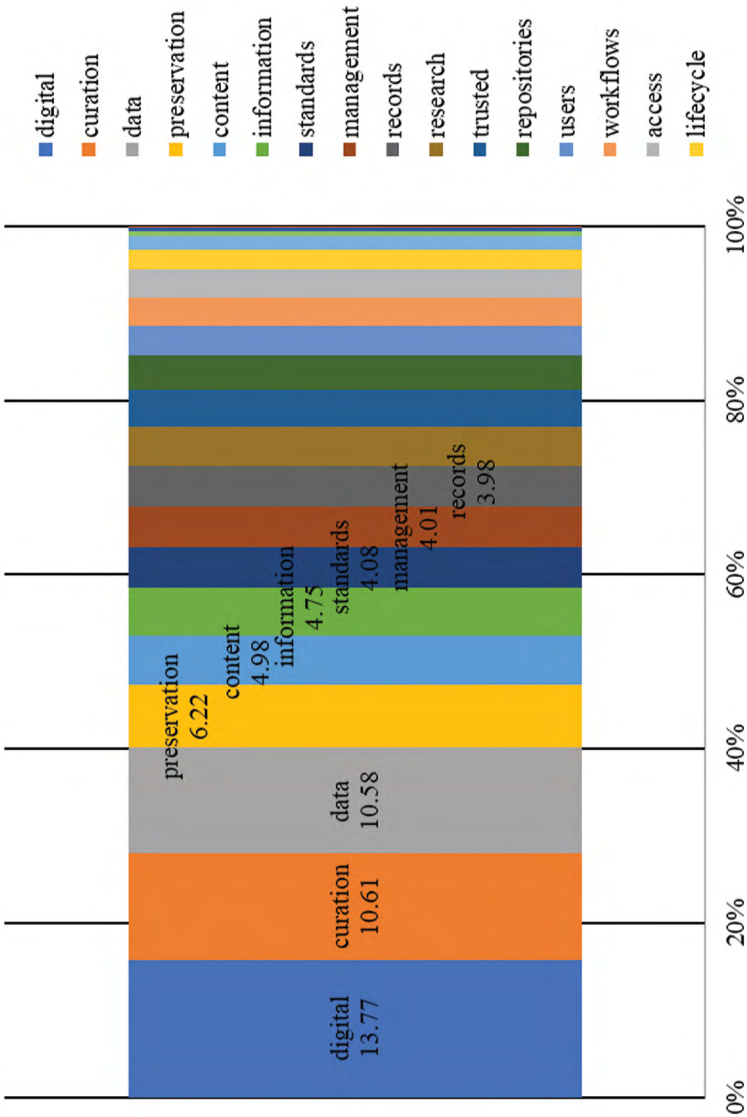


**Figure A3:** Topic 3: Digital curation

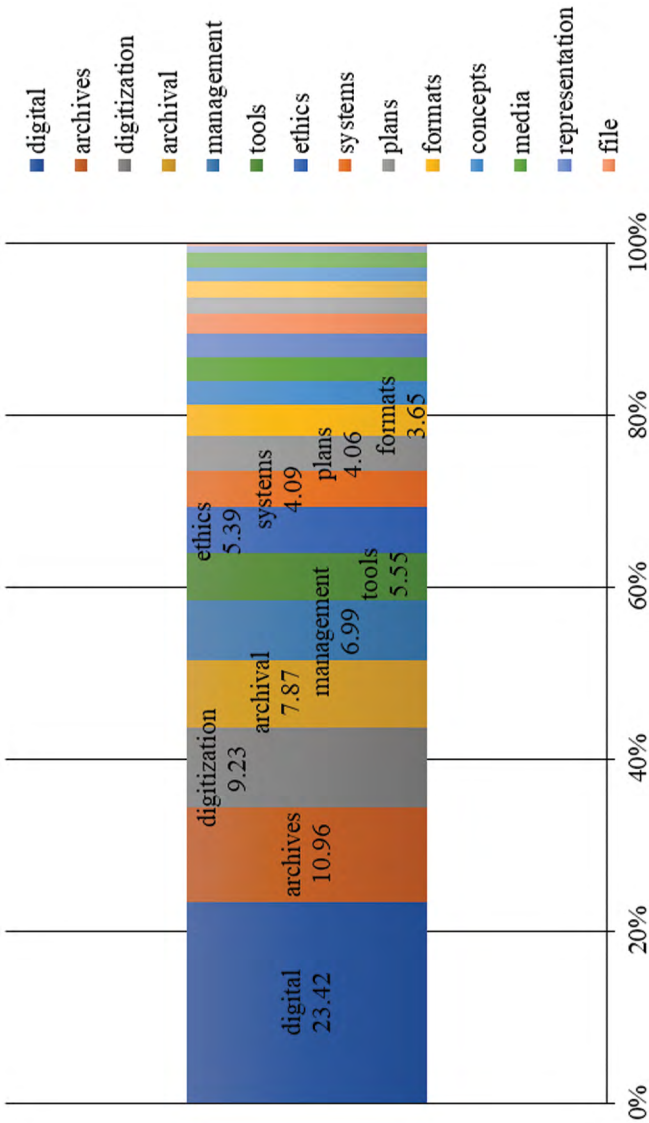


**Figure A4:** Topic 4: Archival approaches to preservation

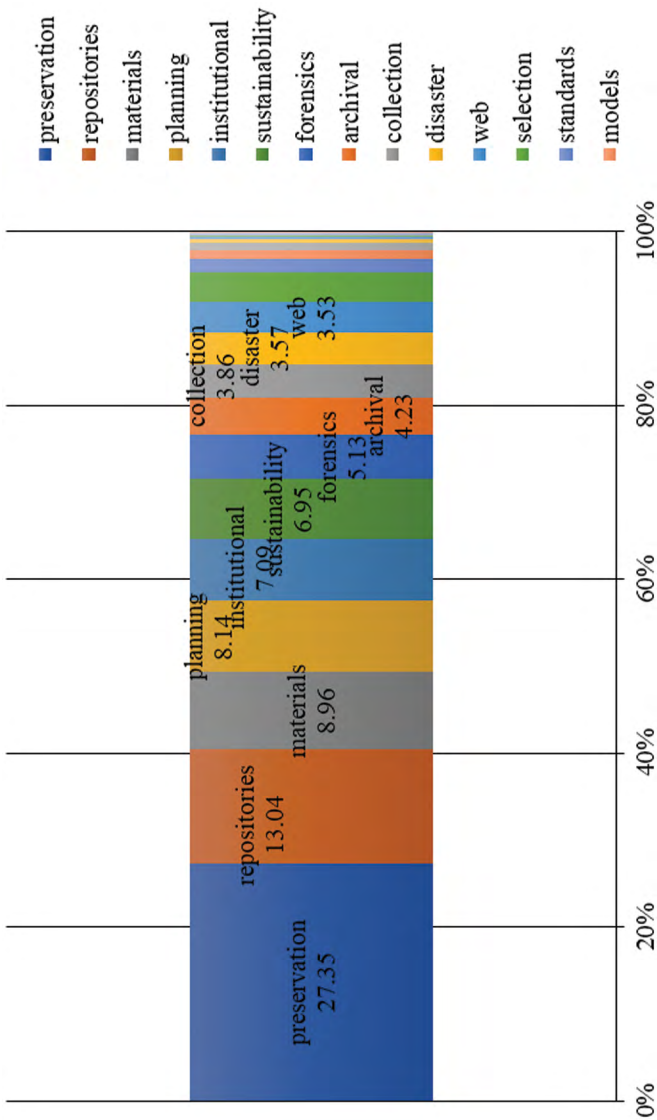


**Figure A5:** Topic 5: Repositories