

# Linking Special Collections to Classrooms: A Curriculum-to-Collection Crosswalk

Sonia Yaco, Caroline Brown, and Lee Konrad

## ABSTRACT

Specialized primary source holdings, not only manuscripts and books but also audio and moving images, are difficult to discover, often requiring users to navigate multiple search tools. These discovery challenges arguably lead to underutilization of specialized primary source holdings in the higher education curriculum. Faculty often include collections in their syllabi only if they have a direct relationship with an archivist or know of specific relevant collections. Similarly, archivists have the most success matching collections to courses when they have built individual relationships with professors, becoming familiar with course content. Particularly at a time when academic libraries are under increasing pressure to link their holdings to student outcomes, a new discovery paradigm to augment personal relationships is needed. This article suggests a conceptual model that would provide a mix of traditional methods and new data mining tools to increase access points to curricular content. The article consists of two parts: a review of existing methods, both human and computer, for connecting curriculum to library resources and a pilot of a software curriculum-to-collection crosswalk that matches course content to specialized primary source holdings via subject. The crosswalk creates recommendations of specialized primary source holdings relevant to specific courses for use by special collections librarians and archivists in working with faculty and students.

© Sonia Yaco, Caroline Brown, and Lee Konrad. 

## KEY WORDS

Academic libraries, Faculty, Students, Bibliographic instruction, Data mining, Information-seeking behavior

Faculty underutilize primary source library and archives holdings, including audio and moving image collections, in curricula. Anecdotal evidence suggests this is due to instructors' lack of familiarity with holdings and discovery challenges. Although libraries and archives are making strides in increasing the discoverability of holdings, many users must still navigate multiple search tools to find relevant collections.<sup>1</sup> Library marketing efforts are often as fragmented as holdings' discovery tools, using a different outreach technique for each silo—manuscript collections, university archives, audiovisual holdings, electronic databases, digital collections, and so on. A small number of university courses, such as history and research methods, routinely include specialized primary source holdings in their syllabi. A few courses such as digital humanities focus almost exclusively on primary sources. Each repository has a core of faculty who have discovered relevant materials for their courses, often because they have relationships with librarians or archivists, or because librarians or archivists have approached them directly.

Discovery of courses that might utilize specialized holdings is also challenging, even at small colleges. Few university course catalogs have robust search capabilities, and many lack the ability to do simple keyword searches. Librarians and archivists have the most success matching collections to courses when they have built individual relationships with professors, becoming familiar with course content. To create these relationships, librarians and archivists reach out to as many faculty and departments as possible.

How can we build more connections between faculty and curators? How can we better link our collections to curriculum? We suggest that the solution lies in a mix of traditional techniques to improve our integration into the classroom and the development of new automated tools to improve discovery of curricula relevant to primary source collections.<sup>2</sup>

### Current Practice to Build Connections with Instructional Faculty

The literature assumes that an involvement in undergraduate, graduate, or postgraduate courses is a core part of the role of university archives or special collections and is one of the key ways they contribute to university life. William J. Maher argued, "Through archival support of teaching and research, the institution is aided in accomplishing its basic mission of communicating and expanding knowledge" and, in addition, "It illustrates the key role of archives in supporting research, which is so highly valued in academic settings."<sup>3</sup>

As authors such as Eleanor Mitchell, Helen Tibbo, and Nicolas C. Burckel discussed, providing digital resources can be an effective way of engaging with students.<sup>4</sup> However, in a 2005 survey of major research institutions, Anna Elise Allison found that only 4.88% of archives and special collections had used online

tutorials in teaching, and the majority of literature continues to focus on face-to-face contact with students.<sup>5</sup> This contact can take a variety of forms. For Maher, this might be answering inquiries and helping students with research projects, and he suggested being proactive: "If the archives finds that it does not have enough undergraduate students use, it should contact instructors . . . to suggest a class assignment."<sup>6</sup> Maher and many others also focused on the provision of an orientation session during which archivists or special collections librarians explain their collections, outline any sources that are particularly relevant to the course, and explain finding aids and policies and procedures.<sup>7</sup> During research for her master's thesis, Allison found that an impressive 96.47% of the archives, manuscripts, and special collections departments at major research institutions that responded to her survey provided preliminary classroom instruction or orientation sessions for undergraduate students. Allison described the types of activities undertaken, the most popular being a lecture with handouts and time for student questions.<sup>8</sup>

Exploring primary sources can play a key role in developing the critical thinking skills of students, interdisciplinary skills that students can apply throughout their university careers. Marcus C. Robyns discussed using primary sources to teach research skills and to create independent learners, and argued that this should be embedded at an early point in students' careers.<sup>9</sup> Louise Kennedy made a similar point about special collections librarians and archivists being in a unique position to contribute to the problem-based learning and embedding of information literacy skills prevalent in many courses today.<sup>10</sup>

Elizabeth Yakel and Deborah Torres have tried to be more specific in analyzing the knowledge required to work effectively with primary sources.<sup>11</sup> Barbara Rockenbach agreed that special collections librarians are in a position to do more by devising "inquiry-based learning exercises" which are more specifically targeted to the course objectives.<sup>12</sup> Using examples from learning theory and case studies at Yale, she discussed how to increase student engagement and critical thinking and argued, "The characterization of archives as laboratory creates an experimental space where hands-on experience in analyzing, asking questions of, and telling stories with primary source documents are possible."<sup>13</sup>

Robyns pointed out that some educators and others might have issues with archivists or librarians being involved in developing classes or courses that go beyond a simple introduction to collections. He said, "Many archivists have argued that being a teacher goes beyond the mandate of archival management and that the responsibility for teaching thinking and research skills should be left to properly trained faculty," and some think that "this approach might jeopardize the archivist's role as a neutral arbiter in the research process."<sup>14</sup> Ken Osborne would disagree. While he mainly discussed working with primary and secondary schools, he argued strongly that archivists have a role as educators

and are missing an opportunity if they do not accept this.<sup>15</sup> Kennedy found that while initially some confusion existed over the role of the archivist in terms of dissemination and teaching, most academics responded positively to their potential involvement in developing courses.<sup>16</sup>

This type of engagement, however, may be solely with one or two classes, with the impetus coming from the archivist or the librarian. It can be difficult to convince academics in some disciplines of the value of archives or primary sources. Kennedy focused on the history department and found that any collaboration often happened because of personal connections between academics and curators, something that may be more likely in the humanities and social sciences.<sup>17</sup> Allison found that the vast majority of those she surveyed worked with English and history classes, but her research did indicate that archivists and librarians are working with an encouragingly broad range of subject areas.<sup>18</sup> She suggested that classes can “raise awareness among students of the power of archives as repositories of information, knowledge, history and identity which reflect the societies in which they were constructed”<sup>19</sup> and that “learning how documents can mediate between the past and the present will be one of the most important lessons students, especially those not specializing in history, can learn in their college years.”<sup>20</sup> If we accept Allison’s suppositions, then archives and special collections potentially can contribute directly to a wide range of courses, particularly given the variety of subject areas represented in them. Robyns’s example of Northern Michigan University demonstrated this clearly. By showing faculty how they can use archives and primary sources to develop critical thinking and research skills, he has been able to contribute to classes across a range of disciplines as diverse as chemistry and nutrition. Robyns provided some useful practical tips for archivists, such as emailing faculty directly and meeting with academics to discuss the relevance of collections to specific classes.<sup>21</sup> Maher also looked beyond the history student and advocated creating specialized finding aids such as subject indexes and “guides to specific subjects (e.g., cultural anthropology, urban politics, civil rights) documented in several series and collections.”<sup>22</sup>

Like Robyns, Rockenbach believes in being active. She described how “aggressive” outreach techniques, such as identifying relevant courses and contacting professors with examples of collections and exercises, achieved a 25% success rate.<sup>23</sup> Examples of successful activities include having students research and produce online exhibitions, encouraging students to ask questions of primary sources, and comparing primary and secondary sources.

Whatever techniques are in place, a minority of students use special collections. Jessica L. Wagner and Debbi A. Smith’s survey found that most were not aware of university archives, had not used digital collections, and did not know what an archives contains. They suggested other ways of reaching students,

such as through connections with extracurricular activities and student societies and encouraging donations from them.<sup>24</sup> Kennedy also mentioned acquisition as a way of collaborating with faculty and enabling the development of new research strands.<sup>25</sup>

A clear potential exists for faculty and students to use special collections and archives in a number of subject areas. Such use could add demonstrable value to the learning experience of students across campus. An active approach and personal contacts have helped librarians and archivists contribute to the curriculum, but a more targeted and informed approach could be possible by using the tools described later in this article.

While a significant amount of literature explores how librarians and archivists promote the use of primary sources to faculty, research on how faculty find primary sources for curricular use is limited, including only passing mentions of how faculty find manuscript collections for teaching materials. Christine Borgman and her colleagues have begun to address this gap in their work on digital libraries; otherwise, the only applicable literature covers information-seeking behavior of faculty researchers.<sup>26</sup> Borgman et al., working with geography faculty at the University of California–Santa Barbara on the multi-year Alexandria Digital Earth Prototype (ADEPT) project, offered this insight about the difficulties of learning how faculty search for teaching materials:

Most are able to articulate their information seeking in support of research better than they can explain how they seek information for their teaching. In most cases, the teaching and research activities appear to reinforce each other. While pursuing research materials, they encounter items of value for teaching. Conversely, some try out ideas for research in their teaching, so the information they gather for a course may serve both purposes.<sup>27</sup>

Generalizing from a small sample of geography faculty using digital libraries, who teach on the same topic as they research, to all faculty using special collections carries some risk. However, the study suggests that understanding how faculty look for information in their research applies to how they find information to use in teaching.

Helen Tibbo and Ian Anderson examined the information-seeking behavior of historians in parallel studies in the United States and the United Kingdom, respectively.<sup>28</sup> Both studies suggested that archives need a new type of discovery system. Tibbo found archival methods of discovery inadequate for history researchers and suggested a broader approach including outreach: “Repositories must move beyond provision of access and bibliographic instruction.”<sup>29</sup> Anderson suggested creating a retrieval system in line with historians’ information-seeking behavior:

We now need to develop on-line archival systems that are part finding aid, part expert system, and part intelligent agent able to conceptualize, mediate, and tailor the information provided. It may take something of a leap in imagination, but it is not impossible to visualize a system that has such features. Perhaps the closest example we have to such a system is Amazon. . . . When viewing a particular title, one is able to see what other titles people who purchased that title bought, readers are able to post their own reviews alongside those from Amazon and rate the title with a simple star system. . . . Here the on-line retail system is seeking to replicate the purchasing behavior in a bookshop.<sup>30</sup>

In a similar vein, Borgman et al. supported digital library search capability that mirrors teaching methods. Geography faculty organize their teaching around concepts, not around a specific geospatial object, so the ability to search by concept in addition to date and region is necessary.<sup>31</sup> Maria Cristina Pattuelli, in studying the use of digital libraries by high school history teachers in North Carolina, found that an understanding of the context of educational use is crucial to designing such a library, including creating metadata to reflect how teachers think of learning objects.<sup>32</sup>

One way to discover the characteristics of faculty research and teaching is by analyzing data. Scott Nicholson used bibliomining, the “application of statistical and pattern recognition tools to the data associated with library systems,”<sup>33</sup> to understand linkages between scholars and citations in scholarly articles.<sup>34</sup> In Sweden, Irene Wormell, using bibliometric methods, compared electronic serial holdings to faculty research areas “to help information resource managers map the emerging new subject developments and new research areas implied in each department’s dynamic scientific and social life. Portal managers and librarians, in spite of their ambition, may not have easy access to this kind of information through their traditional communication channels.”<sup>35</sup>

Traditional communication methods such as formal outreach, informal networking, attending campus events, and following campus news can lay the groundwork for understanding faculty research and curricular needs but may provide uneven information. The underlying assumption of much of this literature is that archivists and curators need to have a thorough knowledge of their institutions’ curricula and faculty research interests to devise appropriate finding aids or to link collections effectively to subject areas.

Some librarians have used systematic approaches to analyze curriculum. In the late 1960s, William E. McGrath and Norma Durand assigned Library of Congress call numbers to all courses in the University of Southeastern Louisiana catalog.<sup>36</sup> McGrath and Durand used their analysis for collection evaluation and development. Later authors explored techniques, applications, and sources of course subject assignment often referred to as “course analysis.”<sup>37</sup> More recently, authors have studied syllabi without assigning subject headings for collection

evaluation and collection development.<sup>38</sup> Syllabi studies provide data on the inclusion of library resources and current library use to anticipate future library use.<sup>39</sup> Librarians and archivists study syllabi to identify opportunities for outreach, collaboration, and instruction, particularly in the area of information literacy.<sup>40</sup>

The techniques described above, while providing useful information, are time consuming; automation of the process has the potential to provide a faster and more effective approach. Literature from fields such as computer science, educational sciences, and decision science provides intriguing but limited discussions of the use of data mining<sup>41</sup> of course content. Computer scientists from the University of New Brunswick used data mining to analyze syllabi and other learning objects (for example transcripts, course catalogs, and so on) to automate the process of transferring credits between institutions and recommending courses for further study. In "Information Extraction from Syllabi for Academic e-Advising," the authors described the techniques and tools used for the data extraction and suggested that these methods are generalizable across a variety of domains.<sup>42</sup> While the end results were not fully successful (the authors could not completely automate the application, as a need remains for some manual intervention and interpretation), the process the authors used to create the application serves as an excellent framework and provides guidelines for future exploration.

Although the developing field of educational data mining primarily deals with student outcomes, some research involving course content and library resources may be instructive in improving discovery of course content.<sup>43</sup> Changeii Tang et al. used student data (background, academic record, and interests) and course data (structure and content) to create a personalized distance education course. They provided five algorithms to match personally identifiable structured student data to a small body of secondary sources.<sup>44</sup>

Marwah Alaofi and Grace Rumantir used student data to build personalized library search results in response to the large volume of resources available in digital libraries.<sup>45</sup> Andreas Geyer-Schulz et al. utilized individual patron data to create a recommender system at Universitat Karlsruhe and reported cost reduction, improved service, and better collection management.<sup>46</sup>

Big data and data mining are not without challenges. Jennifer Fu, working with campus-level Geographic Information Systems (GIS) data, saw promise and challenge in pulling together campus data sources:

From my perspective, a smart campus is one that is efficient, intelligent, and environmentally sustainable. If geo-spatial solutions can be integrated with various campus information systems including student information, course catalog, course materials and syllabi, faculty research and publications, and library catalogs and research databases (e-books, and journals), it would



greatly enhance its usability. The biggest challenge lies in the history of various existing information systems, which are not necessarily interoperable with each other. Integrating them in a single or a series of smart campus databases could be a very difficult task if not impossible. The solutions might lay in APIs and web services that can be published, or central indexed for retrieval.<sup>47</sup>

## Our Vision

We propose a new discovery paradigm to increase utilization of specialized primary source holdings in a broad range of academic disciplines. We envision a curriculum-to-collection crosswalk software tool or app that would match course content to specialized primary source holdings via subject.

The crosswalk would recommend collections relevant to courses or vice versa. Faculty and students could search using a course title and find relevant audiovisual and other materials. Students interested in a particular collection could search for related courses. Repository administrators wanting to demonstrate the utility of their holdings to a specific college could utilize the crosswalk. Archivists would use the app to augment existing outreach efforts. Librarians and archivists could use new faculty orientation, library liaison programs, and other outreach to encourage faculty and students to use the crosswalk.

A variety of data mining techniques potentially applies to our work. It appears likely that we will need to employ a mix of them to accomplish our goal of automatically creating relationships between collections and curricular data. The crosswalk will need a variety of sources at each university for each type of data. Collection data may be in multiple on-campus and consortia catalogs. Curricular sources may include university catalogs, class schedules, websites, and course management systems. Catalogs and class schedules could provide a list of classes. Department names and course titles could provide initial subject keyword data. We could use reading lists and textbooks assigned to each course to understand the subject of the course itself.

It is easy to imagine how the application of these techniques could enhance our efforts to promote the use of materials to communities beyond those who visit the archives in person. In short, the ability to use data techniques to explore archival collections against other data sets could provide insights into the existence of related collections, the identification of potential collaborators, and an ability to analyze emerging scholarly trends against our respective archival collections. Such analysis could in turn allow us to be highly active in our collection and outreach efforts—particularly as we would have established strong data-centric connections indicating that a given set of archival documents would/could/should be of interest to a particular scholar, researcher, or student.



Moreover, the use of data mining in the archival context offers the potential to explore relationships between our collections and data sets that are important to the institutions and people on whose behalf we work. In addition to identifying relationships between, say, archival collections and a given set of courses in a given institution, this crosswalk could also facilitate the discovery, awareness, and use of materials by user communities within and beyond our respective institutions. This emerging world of linked data and reuse of information opens an opportunity to place archives and special collections at the heart of our institutions' information hubs.

We are in the process of developing a full conceptual model that is not specific to one brand of catalog data or course data, or to one country. For this article, we will describe a pilot that uses a small group of courses and collections at the University of Illinois at Chicago and one data mining method. We will employ the experience of the pilot to discuss inherent opportunities and challenges that we can foresee at this point in our research.

## Methodology

To begin to test the feasibility of a curriculum-to-collection crosswalk, Sonia Yaco, special collections librarian at the University of Illinois at Chicago (UIC) conducted a pilot study of a data mining system. The pilot used 101 undergraduate and graduate social justice-related courses, which are taught in a cross-section of UIC colleges and departments (30 departments in total). Yaco chose social justice courses because UIC's Special Collections and University Archives Department has strong holdings in social reform, ranging from the Jane Addams Hull House papers to the Chicago Urban League. The pilot had 4 goals: collect course metadata, assign subjects for courses, find relevant manuscript collections, and use pilot recommendations in bibliographic instruction and outreach.

### GOAL 1: COLLECT COURSE METADATA

Tracking down possible sources for course data at UIC and at many universities is not unlike following the plot of a poorly written yet complicated murder mystery. Obtaining data required interviewing university administrators about what kinds of course data might exist; finding the office(s) in charge of each source; gaining access to front-end and back-end data; navigating user interfaces and search engines—or lack thereof; learning the format, content, and update interval of the data; and, finally, extracting the data. The online UIC catalog, accessible to the public, lists and describes all courses that may be taught depending on faculty availability. The catalog supplied narrative

descriptions for pilot courses. The Course Request System (CRS), accessible only to authorized administrators and faculty, contains a decade of records for new, changed, and deleted courses. CRS includes course objectives, descriptions, “Weekly Topics,” and “Sample Sources” (e.g., books, videos, articles), as well as course prerequisites and corequisites. The schedule of classes, accessible only to those with university credentials, lists instructors, class locations and times, and required books.<sup>48</sup> “Books” can include information in a variety of formats, including e-documents, printed books, and AV materials. The schedule shows all courses offered in upcoming semesters, depending on student enrollment. The University Library catalog lists course reserves—assigned reading materials such as print and e-books, electronic documents, websites, and movies for courses in the current semester. Instructors decide whether to post their reading lists in course reserves and whether the lists will be restricted to the instructor and enrolled students. The UIC bookstore constructs a list of course materials for courses each year, which it provides to the library upon request. Although assigned course materials include information in a variety of physical and digital formats, for the sake of simplicity, we will refer to these generically as *textbooks*.

Yaco exported selected data from these sources into a Microsoft Access database. Of the 101 pilot courses, 100 included narrative data and 52 included textbook data. The 52 courses had 281 total and 264 unique textbooks.

## GOAL 2: ASSIGN AUTHORITY-CONTROLLED SUBJECTS TO COURSES

Yaco used 2 methods to assign authority-controlled subject terms to courses: harvesting textbook subjects and asking a UIC cataloging librarian to evaluate course descriptions. Yaco hypothesized that the subjects of assigned textbooks would be a good indicator of course content. A combination of factors stymied using automated queries against the University Library catalog to obtain subjects for these titles including multiple data sources; out-of-date ISBNs; and irregular titles, author names, and data formats. Additionally, the University Library does not own many of the assigned textbooks, so they are not listed in the catalog. To compile subjects for these titles, Yaco searched internal and external library catalogs and commercial databases to find the correct titles, ISBNs, and subjects; cut, pasted, parsed, and cleaned the data with Excel and OpenRefine; and, finally, connected subjects to titles and courses in Microsoft Access. Due to the labor-intensive nature of this process, Yaco did not obtain subjects for all textbooks assigned to pilot courses. She looked up subjects for 140 of the 264 textbooks, netting 618 unique subjects for 52 courses. The librarian assigned subject terms for 4 courses, adding another 8 subjects. In total, 55 courses covered 626 unique subjects (one course had subjects from

both librarian assignment and textbooks) (see Table 1). Of the 30 departments in the pilot, 19 departments with 86 courses had course-subject data identified. The 11 departments without course-subject data account for only 15 courses.

**Table 1. Subject Assignment for Courses by Method**

Method	Courses	Subjects
Derived from textbook subjects	52	618
Assigned by librarian	4	8
Unique total	55	626

**GOAL 3: CREATE RECOMMENDATIONS OF RELEVANT COLLECTIONS**

Two separate overlapping systems describe manuscript collections at UIC. The Special Collections and University Archives Department uses Archivists’ Toolkit to catalog archival collections and create Encoded Archival Description (EAD) finding aids. Staff—archivists, clerical employees, graduate assistants, and undergraduate students—assign subjects to collections as part of processing. Subject terms come primarily from Library of Congress Subject Headings (LCSH) but also from the Thesaurus for Graphic Materials, Medical Subject Terms, and local sources. Collections with multiple or large finding aids may have multiple resource records. Archivists’ Toolkit at UIC currently has 614 resource records and 1,627 subjects.

The University Library catalog, a Voyager system, provides intellectual access to general holdings and manuscript collections. Librarians and clerical employees assign LCSH to manuscript collections and other holdings. The catalog includes MARC records for 522 manuscript collections with 1,521 subject terms. The MARC records include collection-level narrative descriptions from the scope and contents and biography/history portions of finding aids. Because Archivists’ Toolkit and the UIC library catalog had different metadata structure and content, Yaco extracted and combined metadata from both for the pilot.

Yaco used Microsoft Access to compare the identified course subjects with collection subjects and descriptions. To facilitate matching, she normalized subject terms by removing punctuation. After she generated the matches, she removed duplicate recommendations caused by a collection and a course having more than one subject in common. For instance, queries recommended the Gary Urban League records collection twice for the course Social Work in a Multicultural Society—once for the subject “Race” and once for “Race relations.” The first set of queries, matching course and collection authority-controlled subjects, found 56 unique subjects from 24 courses. These matches yielded 746 recommendations of collections for courses. The second set of queries, matching

course subjects against collection narrative metadata, found 44 unique subjects from 27 courses and yielded 870 recommendations of collections. In total, 885 collection recommendations were generated for 29 courses.

Yaco used several techniques to test the appropriateness of the computer-generated suggestions. Inclusion in syllabi is one gauge of whether instructional faculty view the collections as relevant. Of the 29 courses with recommendations, only 7 had accessible syllabi. No manuscript collection is included in those syllabi. However, rather than indicating that faculty did not consider any collections relevant to their courses, this may instead demonstrate faculty's lack of awareness of the library's holdings and how manuscripts can be used in courses.

Inclusion in librarian-created course research guides and training materials is another measure of relevancy. Only one of the pilot courses with recommendations, History and Theories of Feminism, has a research guide. The guide lists primary and secondary sources but does not include any of the computer-generated recommendations.<sup>49</sup> A list of suggested collections that UIC archivist Gretchen Neidhardt created aligns more with the computer-generated recommendations. Neidhardt developed her list for training sessions with library liaisons to academic departments using online collection descriptions in early 2014. However, of the 691 computer-generated recommendations for humanities and social science courses, only 77 were included in her list. Again, this does not necessarily mean that the computer-generated suggestions were not appropriate. Instead, it shows that traditional methods match courses and collections with limited success. For more discussion of the relevance of the recommendations, see the findings below.

#### GOAL 4: USE PILOT RECOMMENDATIONS IN BIBLIOGRAPHIC INSTRUCTION AND OUTREACH

Liaison librarians and archivists provide bibliographic instruction at UIC. Liaisons work with a set of departments and colleges, serving as subject bibliographers and providing bibliographic instruction. Liaisons also create LibGuides (online research guides) for and do outreach to their assigned departments. In coordination with the head of Reference Services and Resources who supervises liaisons, Yaco sent emails, broken down by college and department, to liaisons with collection recommendations to encourage them to incorporate the computer-generated recommendations into bibliographic instruction and outreach (see Figure 1). She also sent emails to archivists who provide bibliographic instruction for courses across all departments.

Collection recommendations for liaisons by dept., course, subject				
Liaison	Dept	Course	Subject	Collection
Paula	SOC	105	Social Problems	
		215	Sociology of Childhood and Youth	Forest Preserve District of Cook County records
				Lawndale Community Development Conference collection
				Society of Separationists collection

**FIGURE 1.** Yaco sent collection recommendation reports via email to liaison librarians.

Findings and Results

GOAL 1: COLLECT COURSE METADATA

This pilot study revealed just how hard it is to discover course content. Search capabilities of the multitude of sources at UIC that describe course content are rudimentary. The sources contain overlapping, conflicting, and duplicate information, with different degrees of currency. Data gathering was labor intensive. Yaco was able to collect some metadata for all 101 courses, but not all data types for all courses. Some courses lacked textbooks or narrative data, which limited the ability to match their content to collections.

GOAL 2: ASSIGN AUTHORITY-CONTROLLED SUBJECTS

Using textbook subjects and direct cataloging of courses, Yaco assigned authority-controlled subjects to 55 of the 101 pilot courses. To illustrate our findings, we will use Urban Revitalization and Gentrification, an urban planning and policy course. The catalog describes the course as follows:

UPP 544. Urban Revitalization and Gentrification. 4 hours.

Urban change in U.S. cities since World War II that is associated with socio-economic restructure under globalization. The course examines restructure under the new global order and its impact on cities and urban planning and

different social groups. Course Information: Graduate standing in Urban Planning and Policy or consent of the instructor.<sup>50</sup>

The cataloging librarian used this description to assign the Library of Congress subjects “City planning,” “Urban policy,” and “Urban renewal.” Table 2 shows the subjects derived from the course’s textbooks. With the possible exception of “Postmodernism,” all of the terms appear to be relevant to the course. Three terms assigned by the librarian overlap those derived from textbooks: “Urban policy,” “Urban renewal,” and “City planning,” which suggests that the textbook subjects appropriately describe the course.

**Table 2. Subjects for Urban Revitalization and Gentrification**

Cities and towns—United States	Human geography
Cities and towns	Polarization (Social sciences)
City and town life—United States	Postmodernism
City and town life	Sociology, Urban—United States
City promotion	Spatial behavior
Civilization, Modern—1950	Tourism and city planning
Gentrification—Government policy	Tourism
Gentrification	Urban policy
	Urban renewal

**GOAL 3: CREATE RECOMMENDATIONS OF RELEVANT MANUSCRIPT COLLECTIONS**

By querying course subjects against library catalog subject terms and collection abstract and scope-and-content narrative fields, Yaco created 885 recommendations for almost a third of the pilot courses (29/101). Many of the recommendations seemed appropriate (see Table 3).

**Table 3. Collection Recommendations for Urban Revitalization and Gentrification—Selected**

Chicago Council on Urban Affairs (TRUST, INC.) records
Fort Dearborn Project records
Friends of the Parks collection
Greater Lawndale Conservation Commission records
Metropolitan Planning Council records
Southport Neighbors Association records
University Village Association records

Using recommendations from faculty, librarians, and archivists to evaluate the relevancy of these computer-generated recommendations had mixed results. Seven syllabi available for pilot courses included no Special Collections holdings, even though all require a research paper or research project. Only one syllabus mentioned the library and then only in a general note that textbooks may be found there. In the one pilot course with recommendations for which there is a LibGuide, History and Theories of Feminism, no correlation exists between the collections listed in the guide and computer-generated recommendations for the course. This gap is due to different selectors and different selection criteria. The computer chose collections using subjects from the course's textbooks, and an archivist picked collections to include in the LibGuide using her knowledge of the collection. The textbook subjects are modern terms like "feminism," whereas the archival collections, which primarily document gender inequality around the 1890s, use terms such as "social reformers" and "women."

The list of suggested collections that UIC reference archivist Neidhardt created for library liaisons overlaps somewhat with the computer-generated recommendations. Neidhardt evaluated each of the 467 computer-created recommendations she had not included and determined that 308 were relevant. For example, her original list included two-thirds (12/19) of the computer-generated collections for the class Urban Revitalization and Gentrification. After reviewing the computer-generated list, she found 5 others to be relevant. She rated one, "Theater program collection," as inappropriate for the class and one she could not evaluate because it was not described online.

Neidhardt and Yaco analyzed the 159 recommendations Neidhardt rated as irrelevant. Some were false hits—the computer matched a course subject *race* to a collection description containing the word "Grace." Other recommendations were inappropriate because the subjects that matched were too broad, such as geographic headings, "women," and "Americans." Neidhardt suggested providing ways for users to evaluate the relevancy of a recommendation. This could include a relevancy ranking/rating covering the type of match (subject term, description, or both), the matching subject term(s), and the number of matched terms.

#### GOAL 4: USE PILOT RECOMMENDATIONS IN BIBLIOGRAPHIC INSTRUCTION AND OUTREACH

The University Library is beginning to incorporate the crosswalk and pilot recommendations in practice and planning. Liaisons are discussing these recommendations with faculty for inclusion in their syllabi as part of efforts to increase library use. Archivists plan to use the recommendations in outreach



and bibliographic instruction. In discussions with manuscript donors, Yaco found them to be receptive to the idea of the crosswalk as a tool to increase access and use of their collections. The library's Strategic Plan, finalized in February of 2016, includes actions to implement the crosswalk in assessment, bibliographic instruction, and outreach:

Goal 1, Maintain and expand core collections that respond to the changing research, educational, and health care information needs of the university community.

Strategic initiative 3: Add new and leverage existing special collections and university archives that respond to the instructional and research needs of the campus.

... Explore applying methodology of collections/curriculum crosswalk pilot to assess collections further.

Utilize recommendations from collections/curriculum crosswalk pilot in bibliographic instruction.

Utilize recommendations from collections/curriculum crosswalk pilot to do targeted outreach to faculty, department.<sup>51</sup>

## Discussion

When Yaco did the data analysis at UIC (January–March 2015), none of the sources she used had global keyword search capability. By June 2015, the university had added search functions to the undergraduate and graduate catalogs. However, other information about course content that could be quite useful to students and faculty as well as to librarians and archivists, such as CRS's "Course Objectives" and "Weekly Topics," are still inaccessible.

The pilot has the biggest potential impact on the relationship between the library and the university curriculum. Course content managers in the Office of Academic Programming and the library have begun conversations to include the library in the course proposal and creation workflow. In addition, we are discussing creating a short list of subject terms to apply to courses and manuscript collections to facilitate matching. These subjects will be a subset of those that the library currently uses in its catalog. Faculty or program directors would add subjects to courses. The university community and the public will be able to see these subjects in the course catalog and/or schedule of classes.

The methods used for data gathering in this pilot curriculum-to-collection crosswalk are not scalable at UIC. However, plans are underway for a new course metadata system at UIC. Course Information Management (CIM) will remove the redundancies and conflicts in current systems. CIM will also track

course dependencies across all colleges within UIC. Having a single point of access would make collecting course metadata for the full crosswalk more feasible. Harvesting textbook subjects was particularly labor intensive in this pilot. After the fact, Yaco realized that Library Systems could produce a report of textbooks and subjects from course reserves. She also found that OCLC's cataloging software includes a function that allows batch lookups of subjects for ISBN. Using this function, Yaco was able to create recommendations for one additional department and two additional courses. While significantly quicker than the manual process Yaco used in the pilot, it still required data manipulation.

Data currency is another challenge. The course data sources have overlapping, conflicting, and duplicate information, with different degrees of currency. Given the small sample size of the pilot, we could easily have checked which courses are currently being taught, but doing so for all courses would be difficult. Recommending collections for courses no longer taught is of no use. Plans by the Office of Academic Programming to systematically monitor the currency of courses in the UIC catalog will minimize this problem.

Recommending collections based on out-of-date textbook lists is another problem. No one audits the textbook disclosure mandated by the Higher Education Opportunity Act, leaving no way to know if textbooks lists are up to date. Learning objects such as syllabi are more robust indicators of actual course content, but we would need to use sophisticated data mining to extract subjects from narrative data.

This pilot created collection recommendations that librarians and archivists are using as a tool in outreach to faculty. Investigating whether these recommendations are useful to faculty for curriculum development and to students in their research is the next step. A broader study of information-seeking behavior by faculty in curriculum development is also needed.

## Conclusion

By linking university data in new ways, the creation of a curriculum-to-collection crosswalk provides new opportunities to link audiences to our collections. Such a tool has broad implications for course and collection transparency for the academic community. It could allow instructional faculty to add value to their classes and better exploit the unique resources held by their institutions. The crosswalk could also facilitate research activity, assist archivists in providing appropriate collections to be used in teaching, and inform students and their advisors seeking to select courses.

Would the crosswalk lead to faculty using archival holdings in classes more often? The simple answer is that we are hopeful. A more nuanced answer is not possible because of the lack of research on how faculty currently find

curricular primary source materials. We do know that librarians and archivists have a history of trying many methods to increase their connection to faculty. More knowledge of which courses could be using our collections seems likely to help information professionals in their outreach.

A multitude of data silos for courses, collections, and subjects make the pilot methods of data gathering unscalable at UIC. If other universities have complex or straightforward data structures for tracking course content and collection holdings, it is likely that each has its own unique mix of data sources. Each also has its own practices for data stewardship, data governance, and so forth. The specialized nature of collections and curricular data will make it difficult to develop a generalized software tool for use across disparate institutions. However, Yaco, Arkalgud Ramaprasad, and Saleha Rizvi are using the crosswalk to create an ontological framework to guide the design of institution-specific software.

The crosswalk has the potential—on a grand scale—to put archival collections in context by finding relationships with faculty and student needs, with other holdings, and with other educational resources. Providing that context could help alumni and donors to understand archives and help institutions to form more meaningful partnerships with their outside communities. Rather than sitting in a corner with a sole, but much beloved history methods course, our collections can be at many tables, serving many roles, to the mutual advantage of all.

## NOTES

Acknowledgements. Preliminary discussions with Natasha Samreny led to the idea of using textbook subjects to assign subjects for courses. Thanks to Saleha Rizvi for her insight into strategies for data mining and to Stephen E. Wiberley Jr. for providing sage advice in reviewing multiple drafts of this paper. Thanks also to Gwen Gregory, Joelen Pastva, Viola Fox, Gretchen Neidhardt, and other colleagues at University Library at University of Illinois at Chicago for technical assistance for the pilot. University librarian Mary Case's and Paula Dempsey's support for implementation of the crosswalk is particularly appreciated.

<sup>1</sup> As evidenced by the increasing amount of literature on finding aids and discoverability, see for example Celeste Chapman, "Observing Users: An Empirical Analysis of User Interaction with Online Finding Aids," *Journal of Archival Organization* 8, no. 1 (2010): 4–30; and Christopher J. Prom, "Using Web Analytics to Improve Online Access to Archival Resources," *The American Archivist* 74 (Spring/Summer 2011): 158–84.

<sup>2</sup> We focus in this article on collaboration with faculty in the areas of learning and teaching, not in research or in areas of collecting, although these functions often overlap.

<sup>3</sup> William J. Maher, *The Management of College and University Archives* (Chicago: Society of American Archivists, 1992), 10 and 127.

<sup>4</sup> Eleanor Mitchell, Peggy Seiden, and Suzy Taraba, eds., *Past or Portal? Enhancing Undergraduate Learning through Special Collections and Archives* (Chicago: Association of College and Research Libraries, 2012), ix contains useful case studies exploring aspects of using collections in courses; Helen Tibbo, "The Impact of Information Technology on Academic Archives in the Twenty-First Century," in *College and University Archives*, ed. Christopher J. Prom and Ellen D. Swain (Chicago:

- Society of American Archivists, 2008), 39–43; Nicholas C. Burckel, “Academic Archives: Retrospect and Prospect,” in *College and University Archives*, 12.
- <sup>5</sup> Anna Elise Allison, “Connecting Undergraduates with Primary Sources: A Study of Undergraduate Instruction at Archives, Manuscripts and Special Collections” (master’s thesis, University of North Carolina, 2005), <https://cdr.lib.unc.edu/indexablecontent/uuid:cc3864a9-10e4-467e-8b56-8171e400ab7b>.
  - <sup>6</sup> Maher, *The Management of College and University Archives*, 261.
  - <sup>7</sup> See for example Maher, *The Management of College and University Archives*; and Xiaomu Zhou, “Student Archival Research Activity: An Exploratory Study,” *The American Archivist* 71 (Fall/Winter 2008): 476–98.
  - <sup>8</sup> Allison, “Connecting Undergraduates with Primary Sources,” 32.
  - <sup>9</sup> Marcus C. Robyns, “The Archivist as Educator: Integrating Critical Thinking Skills into Historical Research Methods Instruction,” *The American Archivist* 64 (Fall/Winter 2001): 363–84.
  - <sup>10</sup> Louise Kennedy, “Partners or Gatekeepers? Archival Interactions with Higher Education at University College Dublin,” in *Archives and Archivists 2*, ed. Alisa C Holland and Elizabeth Mullins (Dublin, Ireland: Four Courts Press, 2013), 223.
  - <sup>11</sup> Elizabeth Yakel and Deborah Torres, “AI: Archival Intelligence and User Expertise,” *The American Archivist* 66 (Spring/Summer 2003): 51–78.
  - <sup>12</sup> Barbara Rockenbach, “Archives, Undergraduates, and Inquiry-Based Learning: Case Studies from Yale University Library,” *The American Archivist* 74 (Spring/Summer 2011): 298.
  - <sup>13</sup> Rockenbach, “Archives, Undergraduates, and Inquiry-Based Learning,” 301.
  - <sup>14</sup> Robyns, “The Archivist as Educator,” 364.
  - <sup>15</sup> Ken Osborne, “Archives in the Classroom,” *Archivaria* 23 (Winter 1986–87): 16–40.
  - <sup>16</sup> Kennedy, “Partners or Gatekeepers?,” 229.
  - <sup>17</sup> Kennedy, “Partners or Gatekeepers?,” 228–29.
  - <sup>18</sup> Allison, “Connecting Undergraduates with Primary Sources,” 30–31.
  - <sup>19</sup> Kennedy, “Partners or Gatekeepers?,” 225.
  - <sup>20</sup> Maher, *The Management of College and University Archives*, 257.
  - <sup>21</sup> Robyns, “The Archivist as Educator.”
  - <sup>22</sup> Maher, *The Management of College and University Archives*, 105.
  - <sup>23</sup> Rockenbach, “Archives, Undergraduates, and Inquiry-Based Learning,” 302.
  - <sup>24</sup> Jessica L. Wagner and Debbi A. Smith, “Students as Donors to University Archives: A Study of Student Perceptions with Recommendations,” *The American Archivist* 75 (Fall/Winter 2012): 538–66.
  - <sup>25</sup> Kennedy, “Partners or Gatekeepers?,” 226.
  - <sup>26</sup> Christine L. Borgman, Gregory H. Leazer, Anne Gilliland-Swetland, Kelli Millwood, Leslie Champeny, Jason Finley, and Laura J. Smart, “How Geography Professors Select Materials for Classroom Lectures: Implications for the Design of Digital Libraries,” *ACM/IEEE Joint Conference on Digital Libraries Proceedings* (Tucson, Ariz.: JCDL, 2004), 179–85; Christine L. Borgman, Laura J. Smart, Kelli A. Millwood, Jason R. Finley, Leslie Champeny, Anne J. Gilliland, and Gregory H. Leazer, “Comparing Faculty Information Seeking in Teaching and Research: Implications for the Design of Digital Libraries,” *Journal of the American Society for Information Science and Technology* 56, no. 6 (2005): 636–57.
  - <sup>27</sup> Borgman et al. “Comparing Faculty Information Seeking in Teaching and Research: Implications for the Design of Digital Libraries,” 642.
  - <sup>28</sup> Ian G. Anderson, “Are You Being Served? Historians and the Search for Primary Sources,” *Archivaria* 58 (2004): 81–129; H. R. Tibbo, “Primarily History in America: How U.S. Historians Search for Primary Materials at the Dawn of the Digital Age,” *The American Archivist*, 66, no. 1 (2003): 9–50.
  - <sup>29</sup> Tibbo, “Primarily History in America,” 29.
  - <sup>30</sup> Anderson, “Are You Being Served?,” 113.
  - <sup>31</sup> Borgman et al., “How Geography Professors Select,” 182.

- <sup>32</sup> Maria Cristina Pattuelli, "Teachers' Perspectives and Contextual Dimensions to Guide the Design of N.C. History Learning Objects and Ontology," *Information Processing and Management* 44, no. 2 (2008): 635–46.
- <sup>33</sup> Scott Nicholson, "The Bibliomining Process: Data Warehousing and Data Mining for Library Decision Making," *Information Technology and Libraries* 22, no. 4 (2003): 146.
- <sup>34</sup> Scott Nicholson, "The Basis for Bibliomining: Frameworks for Bringing Together Usage-Based Data Mining and Bibliometrics through Data Warehousing in Digital Library Services," *Information Processing and Management* 42, no. 3 (2006): 785–804.
- <sup>35</sup> Irene Wormell, "Matching Subject Portals with the Research Environment," *Information Technology and Libraries* 22, no. 4 (2003): 162–63.
- <sup>36</sup> William E. McGrath and Norma Durand, "Classifying Courses in the University Catalog," *College and Research Libraries* 30 (November 1969): 533–39.
- <sup>37</sup> H. Vernon Leighton, "Course Analysis: Techniques and Guidelines," *The Journal of Academic Librarianship* 21, no. 3 (1995): 175–79; Gwenn S. Lochstet, "Course and Research Analysis Using a Coded Classification System," *The Journal of Academic Librarianship* 23, no. 5 (1997): 380–89; Jeremy Sayles, "Course Information Analysis: Foundation for Creative Library Support," *Journal of Academic Librarianship* 10, no. 6 (1984): 343–45; Yelena Pancheshnikov, "Course-Centered Approach to Evaluating University Library Collections for Instructional Program Reviews," *Collection Building* 22, no. 4 (2003): 177–85.
- <sup>38</sup> Cindy Shirkey, "Taking the Guess Work out of Collection Development: Using Syllabi for a User-Centered Collection Development Method," *Collection Management* 36, no. 3 (2011): 154–64; Rick Jon Bean and Lynn Marie Klekowski, "Course Syllabi: Extracting Their Hidden Potential (at DePaul University's Suburban Campus Libraries)," *Sixth Off-Campus Library Services Conference Proceedings* (Mount Pleasant, Mich.: Central Michigan University, 1993), 1–11; Renee Nesbitt Anderson, "Using the Syllabus in Collection Development," *Technicalities* 8, no. 1 (1988): 1–3; Joseph McDonald and Lynda Basney Micikas, "Collection Evaluation and Development by Syllabus Analysis: The Must-Ought-Could (MOC) Method (at Holy Family College)," in *Acquisitions '90*, ed. David C. Genaway (Canfield, Ohio: Genaway and Associates, 1990), 289–316.
- <sup>39</sup> Nancy H. Dewald, "Anticipating Library Use by Business Students: The Uses of a Syllabus Study," *Research Strategies* 19, no. 1 (2003): 33–45.
- <sup>40</sup> Lisa M. Williams, Sue Ann Cody, and Jerry Parnell, "Prospecting for New Collaborations: Mining Syllabi for Library Service Opportunities," *Journal of Academic Librarianship* 30, no. 4 (2002): 270–75; Cheri Smith, Linda Doversberger, Sherri Jones, Parker Ladwig, Jennifer Parker, and Barbara Pietraszewski, "Using Course Syllabi to Uncover Opportunities for Curriculum-Integrated Instruction," *Reference and User Services Quarterly* 51 (Spring 2012): 263–71; Katherine Boss and Emily Drabinski, "Evidence-Based Instruction Integration: A Syllabus Analysis Project," *Reference Services Review* 42, no. 2 (2014): 263–76; Sammie Morris, Lawrence J. Mykytiuk, and Sharon A. Weiner, "Archival Literacy for History Students: Identifying Faculty Expectations of Archival Research Skills," *The American Archivist* 77 (Fall/Winter 2014): 394–424.
- <sup>41</sup> The SAA publication *A Glossary of Archival and Records Terminology* defines data mining as "The process of identifying previously unknown patterns by analyzing relationships in large amounts of data assembled from different applications." Richard Pearce-Moses, *A Glossary of Archival and Records Terminology*, <http://archivists.org/glossary/terms/d/data-mining#.V5jZMKI3hgg>.
- <sup>42</sup> Yegven Biletsky, J. Anthony Brown, and Girish Ranganathan, "Information Extraction from Syllabi for Academic e-Advising," *Expert Systems with Applications: An International Journal* 36, no. 3 (2009): 4508–16.
- <sup>43</sup> For an overview of the educational data mining, see C. Romero and S. Ventura, "Educational Data Mining: A Survey from 1995 to 2005," *Expert Systems with Applications* 33 (2007): 135–46.
- <sup>44</sup> Changjie Tang, R. W. H. Lau, Qing Li, Huabei Yin, Tong Li, and Danny Kilis, "Personalized Courseware Construction Based on Web Data Mining," *Web Information Systems Engineering, 2000. Proceedings of the First International Conference*, vol. 2 (Washington, D.C.: IEEE, 2000): 204–11.
- <sup>45</sup> Marwah Alaofi and Grace Rumanitir, "Personalisation of Generic Library Search Results Using Student Enrolment," *Information Journal of Educational Data Mining* 7 (2015): 68–88.
- <sup>46</sup> Andreas Geyer-Schulz, Andreas Neumann, and Anke Thede, "An Architecture for Behavior-Based Library Recommender Systems," *Information Technology and Libraries* 22, no. 4 (2003): 165.

- <sup>47</sup> Jennifer Fu, "Smart Campus—My Experiences and Perspective," *Advancing the Spatially Enabled Smart Campus, Position Papers* (Santa Barbara: University of California, Santa Barbara, Center for Spatial Studies, 2003), <http://escholarship.org/uc/item/79d3127j>.
- <sup>48</sup> An amendment to the *Higher Education Act of 1965*, the *Higher Education Opportunity Act*, enacted in 2010, requires universities to list the titles and prices of required textbooks for every course "when feasible" as a way of helping students anticipate the true cost of courses when enrolling (part C of title I [20 U.S.C. 1015] section 133, D).
- <sup>49</sup> Valerie Harris, "GWS 292: History and Theories of Feminism: Tips for Finding Primary Resources," 2015, University Library, [http://researchguides.uic.edu/Moruzzi\\_GWS292](http://researchguides.uic.edu/Moruzzi_GWS292).
- <sup>50</sup> University of Illinois at Chicago, "Graduate Catalog/Graduate Course Descriptions/Urban Planning and Policy (UPP) Courses," *Graduate Catalog*, 2016, <http://catalog.uic.edu/gcat/course-descriptions/uppl/>.
- <sup>51</sup> University of Illinois at Chicago, *University Library Strategic Plan January–December 2016*.

## ABOUT THE AUTHORS



**Sonia Yaco** is an assistant professor at University Library at the University of Illinois at Chicago and the head of Special Collections and University Archives. She has a courtesy appointment in the Department of History. From 2007 to 2013, she was the Special Collections librarian and university archivist at Old Dominion University in Norfolk, Virginia. She is the cofounder and senior advisor to the Desegregation of Virginia Education (DOVE) project. Yaco was previously the president of Anlex Computer Consulting. She is a regular contributor to scholarly archival journals on topics that include building cultural heritage collections that reflect the diversity of society and innovative methods to improve the discoverability of archives using emerging technology. She holds a master of arts degree from the School of Library and Information Studies at the University of Wisconsin–Madison.



**Caroline Brown** is the program leader at the Centre for Archive and Information Studies, University of Dundee. She is also university archivist and joint head of the department of culture and information at the university. She has served on national and international academic and professional bodies including committees for the Archives and Records Association, UK and Ireland. She is currently a trustee for the Scottish Council on Archives, a director of the Scottish Archive Network, a member of the International Council on Archives Section on Archival Education, and a panel member for the Scottish Graduate School for Arts and Humanities. She publishes and speaks on a wide range of recordkeeping issues.



**Lee Konrad** began his professional career at University of Wisconsin–Madison in 1986 working in the outreach division of the Office of International Studies and Programs. He attended UW–Madison as a graduate student in the early 1990s and was subsequently hired as a librarian by the General Library System in 1993. He worked for several years at College Library as a computer and media services librarian. In 2000, he was asked to serve as head of the General Library System's newly formed Digital Collections Group. Konrad subsequently served as director of Arts, Humanities, and Social Science Libraries from 2005 to 2010 and, since 2010, as UW–Madison Libraries' associate university librarian for technology strategy and data services. He holds a BA in history and an MA in library and information studies, both from UW–Madison.