

Local Histories in Global Digital Libraries: Identifying Demand and Evaluating Coverage

Katrina Fenlon

Graduate School of Library and Information Science
University of Illinois at
Urbana-Champaign
501 E. Daniel St.
Champaign, IL 61820
Kfenlon2@illinois.edu

Virgil E. Varvel Jr.

Graduate School of Library and Information Science
University of Illinois at
Urbana-Champaign
501 E. Daniel St.
Champaign, IL 61820
Vvarvel@illinois.edu

ABSTRACT

Digital collections of primary source materials have potential to change how citizen historians and scholars research and engage with local history. The problem at the heart of this study is how to evaluate local history coverage, particularly among large-scale, distributed collections and aggregations. As part of an effort to holistically evaluate one such national aggregation, the Institute of Museum and Library Services (IMLS) Digital Collections and Content (DCC), we conducted a national survey of reference service providers at academic and public libraries throughout the United States. In this paper, we report the results of this survey that appear relevant to local history and collection evaluation, and consider the implications for scalable evaluation of local history coverage in massive, aggregative digital libraries.

Categories and Subject Descriptors

H.3.7 [Information Storage and Retrieval]: Digital Libraries – collection, user issues, dissemination.

Keywords

Local history, collections, metadata, aggregations, evaluation

1. INTRODUCTION

Digital collections of primary source materials have potential to change how citizen historians and scholars research and engage with local history [14]. State libraries and national aggregations have rushed to realize this potential, by bringing together and exposing previously isolated, local collections on the Web.

Intensive digitization programs have targeted local, special collections at public libraries. State libraries have cooperated to build state- or region-centric databases. National and international aggregations, including the Digital Public Library of America (DPLA, <http://dp.la/>) and Europeana (<http://europeanadigital.eu/>), seek to capitalize on these efforts by integrating local historical materials from different places with one another. How successful these efforts are, from varying perspectives, is unclear. How can

we evaluate the success of such collections in order to guide future digital library development?

The problem at the heart of this study is how to evaluate local history coverage, particularly among large-scale, distributed collections and aggregations. As part of an effort to holistically evaluate one such national aggregation, the IMLS DCC (<http://imlsdcc.grainger.illinois.edu/>), we conducted a national survey of reference service providers at academic and public libraries throughout the United States. In this paper, we report select results of this survey that are relevant to local history and collection evaluation. The full survey results are in a forthcoming report [2]. In brief, regardless of whether their library was small or large, public or academic, respondents to our survey expressed great zeal for local history above all other topics, for improved local history coverage, and for improved geographic browsing facilities.

In our discussion of survey results, we consider how to begin meeting local history demand through curatorial approaches to collection and aggregation development. We suggest that curatorial development to meet a specific demand, like that evident for local history, must be guided by collection evaluations. We describe challenges to evaluating and leveraging local history, which present ripe opportunities for future research.

2. BACKGROUND

2.1 Local History and Users

Local history is known to be popular among library users in both public and academic settings, as it occupies “a borderland between knowledge that is personal...and knowledge that is generalizable and therefore worthy of scientific attention” [14]. It has therefore engendered much attention in the collection development literature, especially for public libraries [12,13]. Recent work explores infrastructures for local history and digital collection development [7,14,17]. However, the literature is weaker on specific development and evaluation methods for digital libraries or aggregations of collections. While the general public, from students to citizen historians to genealogists, are presumed to be the primary audience for local history, scholarly historians have an increasingly recognized stake in the discoverability and accessibility of digital heritage materials. Historians are known to use websites of repositories and other electronic means to locate primary materials for research [15], and scholars have been shown to gather together digitized primary sources into thematic research collections for intensive study [9]. It has been demonstrated that contextual information about objects and how they are organized is imperative to scholarly

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

JCDL '13, July 22–26, 2013, Indianapolis, Indiana, USA.

Copyright is held by the owner/author(s). Publication rights licensed to ACM.

ACM 978-1-4503-2077-1/13/07...\$15.00.

navigation and interpretation of digital primary sources [3]. As a result, there is an increasing emphasis on the curation of digital primary sources for scholarly use [4].

How scholarly practice engages local history collections is not well understood. A collection may function both as a local history resource and, depending on its content, as a resource for broader, un-localized interest. The local function of any given collection, as opposed to its more general research utility, is not always well defined or leveraged in practice (in description, search, browse, or interactive facilities). This problem is compounded when resources are pulled out of their local contexts, which happens when distributed collections are aggregated into Europeana, DPLA, or DCC. Libraries tend to focus on “global approaches to digital collection building by producing expansive gateways for all their user communities” [9]. How do we continue the work of making local collections globally discoverable and accessible, without sacrificing the local context of collections, or compromising the strength of collections as local history resources? Methods and tools for the evaluation and exploitation of digital local history collections, especially in the context of massive aggregations, are yet rare (for one counterexample, see [10].)

In addition to results of our survey, which offer new evidence on the value of local history to public and academic library users, this paper gives an exploratory analysis of the challenges of assessing not only geographic, but local history coverage in cultural heritage metadata. We aim, with this paper, to lay the groundwork for future research on evaluating and exploiting local history in large-scale aggregations.

3. DCC SURVEY

The Institute of Museum and Library Services (IMLS) Digital Collections and Content (DCC) is among the largest cultural heritage digital libraries in the U.S., aggregating more than 1,500 collections and 1 million items from a diverse range of institutions in 44 states. DCC maintains a co-searchable registry of heritage collections and repository of item-level metadata, which together have served as both a unique public resource and a test bed for research into metadata interoperability, topic modeling and other advanced techniques for improving subject access to digital libraries, interface design, and collection evaluation.

3.1 Survey Method

In 2011/12, the DCC project conducted a survey to evaluate the DCC aggregation from the perspective of reference service providers at public and academic libraries. This study asked fifteen questions designed to help determine (1) the primary user groups for DCC and topics of interest; (2) how DCC compares to similar digital resources, in terms of both scope of coverage of history and perceived quality; (3) the effectiveness of collection metadata; and (4) suggested improvements to the resource.

The research used two reference frames: public and academic libraries. To reduce internal bias, this data collection purposefully excluded institutions from the sample that have contributed content to the IMLS DCC aggregation. We randomly sampled within each of these two frames. One reference service provider was targeted to represent each institution.

In total, 1,274 academic libraries and 1,192 public libraries were sampled from a total population of 3,689 academic (according to most recent NCES Academic Libraries Survey [11]) and 8,870

public libraries (according to combined IMLS and CIRSS listing [16]). The survey data was collected using Limesurvey open-source web survey software. As usual in this domain, the response rate was low despite multiple rounds of follow-up and sample expansion. 420 responses were used in the final analysis: 168 academic and 252 public library respondents. Cutoffs for the categories of small, medium, and large for public libraries were standardized by the American Library Association’s Definition of Levels for Organizational Membership. Categories for academic libraries were derived from the National Center for Education Statistics’ Academic Library Statistics reference. Medium libraries comprised approximately 30% and 23% of academic and public library respondents, respectively. Large libraries comprised approximately 20% and 26% of academic and public library respondents, respectively.

Method and survey results are fully elaborated in a technical report for the Institute of Museum and Library Services [2]. This paper focuses on the implications of a few questions that asked about users’ topical interests and needs in local history from the resource and general collection evaluation.

3.2 Survey Results Relevant to Local History

Responses to the question, “Please list 5 – 10 examples of historical topics of interest to your users”, were particularly relevant to understanding user requirements for local history representation in digital libraries. This question garnered 408 responses comprising more than 2,100 distinct topical terms or phrases. Academic respondents were fewer but offered a greater diversity of topics (see Table 1). Approximately 50% of each of public and academic respondents represented small libraries.

The diverse spread of responses – from ancient history to surprising assertions like, “businesses looking for historical photos as decoration (seriously)” – made manual coding into anything more than general, umbrella topics infeasible. Coding was iterative, and grounded by patterns in responses; therefore codes exhibited some semantic overlap. For example, the subject “Civil War” is ontologically within “Military history”. However, each counted independently as the coding aimed to align as closely as possible with expressed topics. The distributions of codes for both public and academic libraries turned out to be long tailed, as expected.

Table 1. Comparative quantitative summary

Library Type	States Represented	Total Responses	Distinct topics	Codes assigned
Academic	48	160	710	166
Public	47	248	1,008	155

The top historical subject cited by both public and academic libraries was local history. See the distribution of top 5 topics in Figures 1 and 2. Local history accounted for 31% of public library topics and 13.5% of academic library topics. For academic libraries, responses coded as local history were in the vein of, “History of specific industries of the region”, “History of the peoples settling in surrounding areas”, and “Regional special collections”. For public libraries, responses included, “Citizens who have gone on to make a difference in our country”, “What was life like here when my ancestor was alive?”, and “When was this county formed?”.

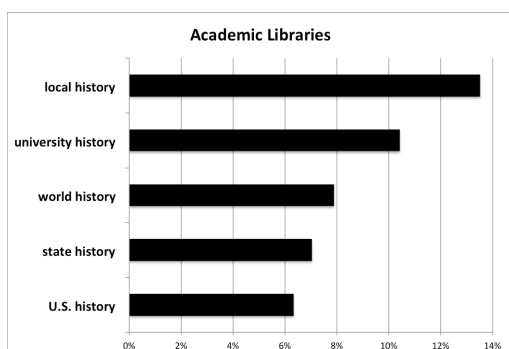


Figure 1. Code occurrences, Academic snapshot, n = 710

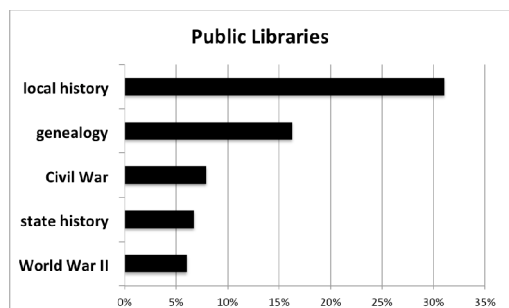


Figure 2. Code occurrences, Public snapshot, n = 1574

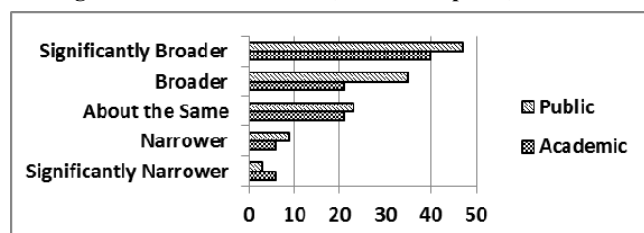


Figure 3. DCC Geographic coverage compared to other digital history resources

These already dramatic figures grow when clusters of topics are considered. Topics closely related to local history – namely state history, local architecture, and cemeteries – all fall into the top ten public library responses. When combined with local history terms, this family of codes accounts for approximately 45% of all public library responses. Responses were well distributed among states; 89% of represented states expressed interest in local history. A comparable cluster – with the addition of University history as a localized topic – accounts for 35% of academic library responses with 67% of responding states included. For both academic and public libraries, regardless of size and apparently location, local history is by far the most cited topic of historical interest to library users. Local history and its offshoots together account for between one third and one half of both academic and public library topics of interest, as reported by reference librarians.

Corroborating our results from the question of topics of interest were results of, “What kind of materials could be added to make [the resource] more valuable to you and your service community?” While the number of responses was too slim to claim generality, it is striking that nearly 50% of responses to this question (or 12% of total respondents) suggested the addition of local history materials. When asked how well the browsing categories match the interest of users, respondents, both academic and public, suggested topical and regional browsing would be

useful. In addition, in response to another question about how to improve interface and functionality, respondents called for improved geographic browsing functionality, for clickable maps, narrower geographic categories for browsing (especially cities), and tools that reveal the relationships between items, e.g., “For instance, a church building postcard for Moline, Illinois is linked to church buildings, postcards, Moline, Illinois, and the church's denomination”. Significantly, both academic and public library respondents considered overall geographic coverage of DCC to be broader than other digital history resources, as shown in Figure 3.

4. EVALUATING LOCAL HISTORY

The survey produced unprecedented qualitative evidence of library users’ overriding interest in local history. Our results suggest that, when feasible, the addition of content and facilities for local history exploration have the greatest potential to add value to a cultural heritage digital library, whether intended for public or academic audiences.

Evaluating the success of a library in serving local history research, in order to set priorities for strategic development, will depend on evaluations of local history coverage. Such evaluations could also help enrich metadata and build refined browsing mechanisms (e.g., with geocoding and more precise mapping).

To curate collections or aggregations as local history resources for public or scholarly use, administrators must understand the topical, geographic, and temporal coverage of collections and the aggregation as a whole. These analyses can be laborious, but emerging, automatic evaluation techniques offer some help. Hillmann (2008) describes metadata quality problems endemic to aggregation and reviews attempts to systematize scalable evaluation techniques. However, Hillmann’s techniques are quality-centric, while we seek to add descriptive evaluation for higher-order, curatorial work. How do we help administrators perform the kinds of evaluation necessary to curating cultural heritage collections at scale? The concept of curation for digital collections needs further elaboration ([10] offers some guidance). Besides activities focused on metadata quality, curation in an aggregation context will rely on strategic, policy-driven collection development, based in turn on analyses of the topical landscape of aggregations, topical density of constituent collections, collection interconnectedness, diversity, and small/large collection complementarity.

Evaluation techniques specific to geographic information largely arise from Geographic Knowledge Discovery (GKD) and data mining. Data quality evaluation and visualization techniques – often geared at improving search or browse [1,7] – are both prominent in the GKD literature, but are usually applied to structured geospatial information designed to support scientific inquiry [4]. These techniques have limited application to unstructured or semi-structured text in cultural heritage metadata records.

In response to the survey results obtained, which offer new evidence on the value of local history to public and academic library users, we are conducting an exploratory analysis of the challenges of assessing local history coverage, which subsume the challenges of GKD and metadata quality evaluation. Future work aims to help aggregators serve the local history needs of users. Specifically, we will confront the following challenges to evaluating local history in large aggregations of cultural metadata:

A minority of metadata records contains readily identifiable spatial information. For example, at DPLA's launch in April 2013, only one fifth of its 2.5 million items appeared to have codifiable geographic information in dedicated metadata fields, so that they could be made accessible through a browsing map and menu. Location information is often embedded in free-text descriptions or subject terms, which impede extraction. In addition, records may contain multiple instances of spatial information. The difficulties of location disambiguation are notorious, besides which libraries employ multiple geographic naming conventions. How can administrators get a read on the scope of coverage, if only a small fraction of records contain readily accessible geographic information?

Descriptions of spatial coverage vary wildly in granularity, and browsing facilities and maps struggle to comprehend the variety. Assuming spatial information can be automatically gleaned from records, the resulting mass of data is difficult to interpret. Common visualization techniques, while critically important, often fail to represent the full range of geographic information in cultural heritage records. For example, DPLA employs a cutting-edge interactive map on which items are grouped by location; the larger the group, the more visible the items become on the map. Groups with fewer items only appear at higher levels of zoom. Local results, therefore, can effectively disappear, becoming invisible without cumbersome zooming and dragging. In addition, while records variously name regions and counties and landmarks, only city and state information tends to be pinpointed on maps. The metadata most useful for local historical inquiry may obscure items in the results of search and data visualizations. Critical mass, in principle, determines the discoverability of local items by browsing and maps. But most places rarely obtain a critical mass of items within an aggregation.

Breadth of geographic coverage alone does not determine a collection's usefulness for local history research. To describe a collection as being *about* local history is to imply that it achieves contextual mass [8] centered on the theme of locality. This is a much stronger statement than saying a collection has many items from one place. It implies that a collection has been curated to facilitate research on the subject of local history. We need improved understanding of other factors that contribute to an aggregation's local history utility, such as the range of types of materials available, how items and collections complement or contextualize one another, and temporal coverage.

5. CONCLUSION

Local history holds proven interest for academic and public library users, yet remains problematic in digital cultural heritage aggregations. Evaluations of local history are impeded not only by the difficulties of wrangling geographic information in heterogeneous descriptions, but by persistent gaps in our understanding of collection curation for digital libraries. In future work, we will address conceptual and technical challenges of exploiting local history in large-scale aggregations of geographically distributed collections.

6. ACKNOWLEDGEMENTS

IMLS supported this research. The authors wish to acknowledge PI Carole Palmer, along with Jacob Jett and Megan Senseney for their contributions to experimental design and implementation.

7. REFERENCES

- [1] Baziz, M., Etcheverry, P., Marquesuzaa, C., & Sallaberry, C. (2008). Accessing heritage documents according to space criteria within digital libraries. *Journal of Digital Information Management*, 6(1).
- [2] DCC. (2013). *Report on the Institute of Museum and Library Services Opening History Evaluation Survey*. K. Fenlon, V. Varvel, C. Palmer (Eds.). CIRSS Report DCC1301. (Forthcoming)
- [3] Duff, W. M., & Johnson, C. A. (2002). Accidentally found on purpose: Information-seeking behaviors of historians in archives. *Library Quarterly*, 72(4), 472–496.
- [4] Flanders, J., & Muñoz, T. (n.d.). An Introduction to Humanities Data Curation. In *DH Curation: A Community Resource Guide to Data Curation for the Digital Humanities*. Champaign, IL. Retrieved from <http://guide.dhcurator.org/intro/>
- [5] Gervais, M., Bedard, Y., Levesque, M., Bernier, E., & Devillers, R. (n.d.). Data quality issues and geographic knowledge discovery. In H. J. Miller & J. Han (Eds.), *Geographic Data Mining and Knowledge Discovery*. 2009: CRC Press.
- [6] Hillmann, D. (2008). Metadata Quality: From Evaluation to Augmentation. *Cataloging & Classification Quarterly*, 46(1).
- [7] Lenstra, N. (2010). The history and future of local history infrastructure: Open access and commodification of local history in the United States and the Anglosphere. Presented at the 5th International Conference on Cooperation and Promotion of Information Resources in Science and Technology: Information Sharing in U-Era, Beijing, China. Retrieved from <http://hdl.handle.net/2142/17434>
- [8] Palacio, D., Cabanac, G., Sallaberry, C., & Hubert, G. (2010). On the evaluation of Geographic Information Retrieval systems. *International Journal on Digital Libraries*, 11(2), 91–109. doi:10.1007/s00799-011-0070-z
- [9] Palmer, C. (2004). Thematic Research Collections. In *A Companion to Digital Humanities*. Blackwell Publishing.
- [10] Palmer, C. L., Zavalina, O., & Fenlon, K. (2010). Beyond size and search: Building contextual mass in digital aggregations for scholarly use. In *Proceedings of the ASIS&T Annual Meeting*. Pittsburgh, PA.
- [11] Phan, T. Hardesty, L., & Sheckells, C. (2011, December). *Academic Libraries: 2010*. Report NCES 2012-365. US Department of Education, Washington, DC: National Center for Education Statistics.
- [12] Phillips, F. (1995). *Local History Collections in Libraries*. Libraries Unlimited.
- [13] Reference and User Services Association. (2006). RUSA guidelines for establishing local history collections. ALA. Retrieved from <http://www.ala.org/rusa/resources/guidelines/>
- [14] Reid, P. H., & Macafee, C. (2007). The philosophy of local studies in the interactive age. *Journal of Librarianship & Information Science*, 39(3).
- [15] Tibbo, H. (2003). Primarily history in America: How U.S. historians search for primary materials at the dawn of the digital age. *The American Archivist*, 66(1), 9–50.
- [16] Varvel, V., & Hughes, K. (Eds.). (2011). *Public Library Data Service Statistical Report 2011*. Chicago, IL: American Library Association.
- [17] Weessies, K. W. (2012). Local history maps in full text resources. *Journal of Map and Geography Libraries*, 8, 230–241. doi:10.1080/15420353.2012.700300