Two hundred twenty-nine health sciences libraries (HSLs) worldwide were surveyed regarding the availability of digital collections, evidence of the type of digital collections, level of access, software used, and HSL type. Of the surveyed libraries, 69% (n=157) had digital collections, with an average of 1,531 items in each collection; 49% (n=112) also had institutional repositories. In most cases (n=147), these collections were publicly available. The predominant platforms for disseminating these digital collections were CONTENTdm and library web pages. Only 50% (n=77) of these collections were managed by the health sciences library itself.
INTRODUCTION

Worldwide, more and more libraries of every size, shape, and specialty are either currently offering or in the process of publishing digital collections consisting of digitized historical and archival items, institutional repository content, and research data [1]. Many health sciences libraries (HSLs) are a part of this trend [2]. As holders of unique items of historical and research interest to a wide range of potential users, HSLs are poised to become major players in the realm of digital initiatives and collections. This survey attempts to quantify and describe digital collections related to and provided by HSLs.

BACKGROUND

While numerous HSLs offer sizable digital collections and repositories—which the authors define as managed collections of information, with associated services, stored in digital formats and accessible over a network [3]—there is little published literature surveying the scope and content of digital collections [1]. In-depth literature searches in Library and Information Science Abstracts (LISA), Library Information Science & Technology Abstracts (LISTA), PubMed, PubMed Central, and EMBASE revealed mostly case studies of individual initiatives or suggested best practices for digitization, dissemination, preservation, and promotion of digital collections in HSLs. For example, Mix’s 2011 article on the sixth edition of Samuel Hahnenmann’s Organon der Heilkunst discussed both the significance and provenance of the notable homeopathy text, as well as the process of creating a digital homeopathy collection using a combination of technical solutions [4]. Welch discussed collaboration between the Waring Historical Library and the School of Nursing at the Medical University of South Carolina, which resulted in a digital library becoming an advocacy tool that ultimately enhanced the Waring Library’s value to its parent institution [5], but neither went further in establishing the state of digital collections in the health sciences.

A 2007 article by Ismond and Shiri compared 6 medical digital libraries but focused on collections intended for use by researchers or health care professionals and excluded digital libraries with a historical focus or those whose collections primarily consisted of images and videos [2]. In 2006, Kahl and Williams surveyed digital library projects at 111 Association of Research Libraries (ARL) member libraries. They found that over 80% of English-language ARL libraries had published digital projects with an average of 12.6 projects per library, and most offered unrestricted access. However, about half of the digital projects (collections of items) were published by only about 15% of the institutions surveyed, and over 25% of the total projects did not contain any descriptive metadata. It was also noted that these projects were primarily image and text based. Finally, the authors found that while collections are growing in many subject areas, they also found a lack of broad overviews in the published literature regarding digital library project access, which suggested that further research and focus on exploring digital library projects of a particular type was needed [1]. The authors’ literature review revealed that no similar study focusing on digital collections that HSLs offer has been published to date.

In an attempt to provide a similar census of digital projects in HSLs, the authors undertook a survey of more than 200 HSL web pages. The survey collected data on the number of HSLs involved in digital projects, the type of projects being digitized, the software being utilized, and the average number of items per collection in order to create an international snapshot of digitization by examining the number, type, size, and impact of digital collections in HSLs on a global level.

METHODS

Data collection took place between August and December 2012. To generate an authoritative and comprehensive listing of HSLs, the authors generated an international list from two sources: members of the Association of Academic Health Sciences Libraries (AAHSL) listed in the thirty-fourth edition of the Annual Statistics of Medical School Libraries in the United States and Canada [6], published by AAHSL, and libraries with collections listed in the US National Library of Medicine’s (NLM’s) Directory of History of Medicine Collections [7].

The authors divided the list of HSLs and reviewed each library’s website for evidence of answers to the following:
- Are any digital collections available?
- Is there an institutional repository? Is it separate from historical digital collections?
- What kind of access is offered (open, closed, limited to institution, or by request)?
- What database or management software (defined by PC Magazine as “software that is used to manage text, images, audio, and video content for a Web site” [8]) is used to provide the digital collections?
- What type of HSL (academic, hospital, other) publishes the collections?
- Who is responsible for managing the collections (i.e., HSL or outside entity)?

Each HSL was surveyed by a scan of its website. The authors looked for evidence of digital collections in the form of a link from the library home page referring to another page containing information on digital collections, digital archives, history, or special collections. If no link could be discovered, the authors searched the library website using the library’s website search box and/or site map. If searching yielded no results, the authors attempted to contact the library directly via online chat or email to the individuals listed in the contact information for each website. The number of institutions contacted by online chat or email was not recorded, but less than
half of the institutions that were surveyed required such contact.

After determining the availability of digital collections on an HSL website, the authors then described the collections using the aforementioned metrics. Although the presence of an institutional repository was noted, items residing in institutional repositories (including electronic theses and dissertations) were not counted for this survey, and in-depth data collection was limited to mainly English-language websites. The number of total “items,” defined as unique digital materials (either “born digital” or digitized analog materials) with a single item record (for example, books such as anthologies counted as one item), was counted and recorded. The content management software used for each collection was determined from the available online documentation or via chat and email. During the survey period, data were collected on a shared, private Google Doc spreadsheet and exported to Microsoft Excel for analysis at the conclusion of data collection.

In addition to data describing digital collections, the parent institutions and uniform resource locators (URLs) of digital collections were also collected. When it could be determined, the authors also recorded data on management responsibilities for each digital collection, noting if the collection was built and maintained by the staff of the HSL itself or if it was managed by an outside entity, such as a parent institution or local consortium.

RESULTS

Two hundred twenty-nine HSLs worldwide were included in this analysis, 70% (n=159) of which were located in the United States, 7% (n=17) in Canada, and 23% (n=53) in other countries listed in the NLM Directory of History of Medicine Collections [7]. Of the surveyed libraries, 69% (n=157) had digital collections, with an average of 1,531 items per collection; and 49% of the surveyed libraries (n=112) also had institutional repositories. Of those with digital collections, 83% (n=190) were academic HSLs.

In most cases (n=147), these collections were publicly available. The predominant platforms for disseminating these digital collections were CONTENTdm (n=42 of 157, 27%) and library web pages (n=35, 22%). However, a number of other software platforms were used by individual libraries. Only 50% (n=77 of 157) of these collections were managed by the HSL itself.

Size of collections varied greatly across libraries. Collections of great magnitude—such as Georgetown University’s Bioethics Research Library [9], London’s Wellcome Library Images Collection [10], and the Digital Library Medic @ from Bibliothèque Interuniversitaire de Santé, Paris [11]—included well over 40,000 items per collection. The University of California–San Francisco’s (UCSF’s) Legacy Tobacco Documents Library [12] alone included more than 10 million documents concerning the historical development of and scientific research on tobacco products. These large collections resulted in an average collection size of 95,510 items per collection from raw data.

To have a better idea of average collection size, the authors applied the Excel TRIMMEAN function after data collection. TRIMMEAN calculates the mean taken by excluding 20% of data points from the top and bottom tails of a data set [13]. Once outlying data were excluded from the analysis, collection size averaged 1,530 items per collection. The TRIMMEAN included all libraries included in survey (n=229), regardless of whether they had a digital collection or not.

To look for more patterns from the raw data, the authors extracted a list of 157 libraries that responded “Yes” to the question “has a digital collection.” The top and bottom 20% (n=31) were excluded from the list to remove outlier data. The remaining libraries (n=95, 61%) were then sorted by average size of digital collection, and the top and bottom 10% of libraries (n=9.5, rounded up to 10) were compared.

Digital collections from the top 10 libraries had an average collection size of 338 items. They were more likely to use CONTENTdm as a digital project platform (n=7), originate from an academic institution (n=9) in the United States (n=8), and have an institutional repository (n=6) as part of their digital collections. Conversely, digital collections from the bottom 10 libraries had an average collection size of 6.7 items and were more likely to use a web page on the library website as a digital project platform (n=7), more likely to originate from an academic institution (n=4) or museum (n=3) outside of the United States (n=7), and less likely to have an institutional repository (n=6). Collection size did not have an effect on who managed the digital collection. Libraries from the highest and lowest percentiles were both equally likely to be managed by an HSL (n=5) or a department or library outside of the HSL (n=5).†

DISCUSSION

The data suggest a wide proliferation of digital collections published by both HSLs and their various parent institutions. These collections include historical items optimized for online viewing, born-digital items, theses and dissertations, reports, research data collections, and multimedia. The large percentage of HSLs with institutional repositories also correlates with increasingly favorable attitudes toward open access and data sharing among academic institutions. The most important implication from the analysis of the data is the evidence of wide variation in HSLs’ participation in digitization efforts. The differences in content platforms, collection size, and management responsibility for digital content suggest that further support for these digitization efforts would be welcome by those with limited resources to create collections.

† The full results table is available from the authors on request.
LIMITATIONS

During the course of this survey, the authors identified several challenges that may lead to limitations in the data. Notably, a lack of common terminology to define and identify digital collections and the tendency in library web design to hide special collections under several layers of links and sub-pages may have led the authors to inaccurately surmise an HSL did not offer any digital collections. On several occasions during data collection, the authors contacted an HSL (mostly through online chat) to determine if digital collections were available. Another limitation of the survey is the lack of representation of HSLs from beyond Europe and the continental United States. The NLM Directory of History of Medicine Collections does not include libraries located in China, Japan, or Russia (among other countries) and in addition requires institutions listed in the directory to be able to answer reference questions and, if necessary, fulfill interlibrary loans [7], which limits the comprehensiveness of the international list of HSLs. In addition, the authors lack reading comprehension skills for languages other than English and were unable to effectively read web pages offered by overseas HSLs in detail. While translation services now offered by web browsers did allow for some access, it is possible that collections were overlooked.

CONCLUSION

Digital projects and initiatives are becoming more and more prominent in HSLs, delivering unique, valuable content to communities and user groups worldwide. HSLs both big and small have embraced digital historical and institutional collections, from hospital libraries to NLM, which publishes its own History of Medicine collections [7]. Even if they do not wish to house digital items locally through a commercial content manager such as CONTENTdm or an open-source solution such as DSpace, HSLs can utilize free, web-based tools such as the Internet Archive <http://www.archive.org> and the Medical Heritage Library <http://www.medicalheritage.org> to share unique collections. Items can also be easily stored and disseminated through library websites.

Whatever the method of presentation, HSLs should always keep in mind that the ability to locate a digital collection is key. The greatest digital collection is of no use if users cannot easily find it either through library web pages, catalogs, or Internet search engines. HSLs should also promote digital collections adequately so that they are visible to the widest possible range of user groups. One never knows when historical items may provide the all-too-valuable missing link for someone’s research.

IMPLICATIONS

Future research should include similar surveys of HSLs in South America, Europe, Africa, Asia, and Australia in order to provide a complete picture of the state and prevalence of digital historical and institutional collections in the health sciences. Such research could lead to a central registry of all digital collections offered by HSLs worldwide as a service to the research community. The future of digital collections in the health sciences is limited only by the creativity and motivation of health sciences librarians, which is in no short supply.

REFERENCES

1. Kahl C, Williams S. Accessing digital libraries: a study of ARL members’ digital projects. J Acad Lib. 2006 Jul;32(4):364–9. DOI: http://dx.doi.org/10.1016/j.acalib.2006.03.003.
2. Ismond KP, Shiri A. The medical digital library landscape. Online Inf Rev. 2007;31(6):744–58. DOI: http://dx.doi.org/10.1108/14684520710841748.
3. Arms WY. Digital libraries [Internet]. Cambridge, MA: MIT Press; c2000. Chapter 1, Background [cited 21 Oct 2013]. <http://www.cs.cornell.edu/wya/diglib/MSI999/Chapter1.html>.
4. Mix LA, Cameron K. From Hahnemann’s hand to your computer screen: building a digital homeopathy collection. J Med Lib Assoc. 2011 Jan;99(1):51–6. DOI: http://dx.doi.org/10.3163/1536-5050.99.1.009.
5. Welch JM, Hoffius SD, Fox EB. Archives, accessibility, and advocacy: a case study of strategies for creating and maintaining relevance. J Med Lib Assoc. 2011 Jan;99(1):57–60. DOI: http://dx.doi.org/10.3163/1536-5050.99.1.010.
6. Association of Academic Health Sciences Libraries. Annual statistics of medical school libraries in the United States and Canada, 2010–2011. 34th ed. Seattle, WA: The Association; 2012.
7. US National Library of Medicine. Directory of history of medicine collections [Internet]. Bethesda, MD: The Library; 8 Mar 2010 [updated 11 Sep 2012; cited 21 Oct 2013]. <http://wwwcf.nlm.nih.gov/hmdirectory/directory/locations.cfm>.
8. The Computer Language Company. PCMag.com encyclopedia: definition of content management system [Internet]. New York, NY: Ziff Davis; 2013 [cited 21 Oct 2013]. <http://www.pcmag.com/encyclopedia/term/40273/content-management-system>.
9. Bioethics Research Library. Bioethics Research Library [Internet]. Washington, DC: Georgetown University; 2013 [cited 21 Oct 2013]. <http://repository.library.georgetown.edu/handle/10822/503786/>.
10. Wellcome Images [Internet]. London, UK: Wellcome Library [cited 21 Oct 2013]. <http://wellcomeimages.org>.
11. Vincent JF. Digital Library Medic @ [Internet]. Paris, France: BIU Health [cited 21 Oct 2013]. <http://www.biusante.parisdescartes.fr/histmed/medica.htm>.
12. UCSF Library, University of California, San Francisco. Digital collections [Internet]. San Francisco, CA: Regents of the University of California; 2013 [cited 21 Oct 2013]. <http://www.library.ucsf.edu/collections/digital/>.
13. Office: TRIMMEAN function [Internet]. Seattle, WA: Microsoft; 2013 [cited 21 Oct 2013]. <http://wwwoffice.microsoft.com/en-us/excel-help/trimmean-function-HPh010342968.aspx>.

AUTHORS’ AFFILIATIONS

Keith M. Pickett, MLIS, kpickett1@tulane.edu, Research Support and Education Librarian; Maureen M.
Knapp, MA, AHIP, mknapp2@tulane.edu, Research Support and Education Librarian; Rudolph Matas Library of the Health Sciences, Tulane University, 1430 Tulane Avenue, SL-86, New Orleans, LA 70112-2699

Received July 2013; accepted November 2013