The Journal of Academic Librarianship 45 (2019) 22–30

Contents lists available at ScienceDirect

The Journal of Academic Librarianship

journal homepage: www.elsevier.com/locate/jacalib

Problems and Changes in Digital Libraries in the Age of Big Data From the Perspective of User Services

Shuqing Li[a,*], Fusen Jiao[a], Yong Zhang[a], Xia Xu[b]

[a] Department of Information Management and Information System, College of Information Engineering, Nanjing University of Finance & Economics, Nanjing 210023, China
[b] Department of Business Administration, School of Management, Nanjing University of Posts and Telecommunications, Nanjing 210023, China

ARTICLE INFO

Keywords:
Big Data
Digital library
User service
Service innovation

ABSTRACT

Based on the investigation of the position of user service for constructing digital libraries in the big data era, this paper points out that not only data resources of modern digital library have the characteristics of big data, but also the existing library services need to use big data methods to achieve reform and innovation, including resource transferring, resource utilization, social identity, thinking innovation. We focus on the importance of user services and types of big data resources that digital libraries can utilize, which include big data within libraries such as user behavior data and digital literature resource, and other big data outside libraries such as scholarly big data. We also examine the problems and potential of digital libraries in the age of big data relative to data, technology, services, and users. Using existing big data resources and considering the characteristics of current users' needs from the perspective of users, more effective ideas and methods to improve existing services in digital library can be put forward. At the same time, it is the personalized need of users in the age of big data that constitute the driving factor for the development of digital library from resource-sharing service to user-oriented service.

Introduction

Libraries are important social institutions that help people access various information resources. With the continuous development of information technologies, libraries have been evolving constantly, and this has greatly expanded library services and improved their efficiency and effectiveness. American scholars proposed the concept of a digital library in the early 1990s. After years of practical application and development, digital libraries have gradually become an important aspect in the development of modern libraries.

The development of digital libraries is directly related to the development of information technologies, particularly Internet technologies. A digital library is an innovative library service that uses information technology, and nearly every major development stage of a digital library is accompanied by major technological changes. From the perspective of information-based civilization, information technology has enabled libraries to go through the traditional stages of automation and develop digital libraries. It is expected that innovations will continue as libraries respond to technological developments, such as mobile Internet services and artificial intelligence (Wang, 2017). From this perspective, big data will inevitably have a profound impact on the services provided by modern digital libraries.

With the promulgation of the Chinese State Council's Guiding Opinion on Actively Promoting the Internet + Action in July 2015 and the Chinese National 13th Five-year Plan in March 2016, emerging information technologies, such as big data and artificial intelligence, are attracting increasing attention. For example, the 2017 Chinese Government Work Report refers to artificial intelligence (Li, 2017). Correspondingly, in the digital library field, the application of new technologies is gradually receiving increasing attention. The IFLA (The International Federation of Library Associations and Institutions) Trend Report ranks artificial intelligence as one of four major technology trends and argues that artificial intelligence has three major implications relative to the future of libraries, i.e., next-generation browsers that extend beyond keyword searches and semantic analysis of web content, integrated speech recognition and machine translation to support real-time multilingual translation, and multiple and complex translation and identification of cloud services for complex multivariate web content (IFLA, 2017). The 2017 New Horizon Report Library Edition included artificial intelligence as one of six important technological developments for the library community (NMC, 2017).

Therefore, we must reexamine the impact that digital information technologies, such as the Internet, big data, and artificial intelligence, have had on the development of digital libraries. Based on this analysis,

* Corresponding author.
E-mail address: leeshuqing@gmail.com (S. Li).

https://doi.org/10.1016/j.acalib.2018.11.012
Received 11 July 2018; Received in revised form 20 November 2018; Accepted 26 November 2018
Available online 05 December 2018
0099-1333/

we propose a direction and route planning for the next phase of digital library development. This article focuses on the development of digital libraries in big data environments and the important role that personalized information services play.

Thinking about the relationship between digital library and big data

Further understanding of big data

The literature indicates that the big data concept was first conceptualized in 2001 by Laney in his research notes. Laney believes that an important feature of big data is that it cannot be processed effectively by traditional data management tools (Laney, 2001). Currently, ways to combine and integrate big data to improve business designs and service management are being considered. In the following, we consider "big" and "data" separately.

Big is easy to understand. Data resources and types have increased significantly. Most big data-processing methods emphasize high-speed and efficient use of large amounts of data. To extract value from big data, effective and efficient data-processing technologies are required. Since 2010, with the maturity and widespread application of cloud computing and artificial intelligence, big data has rapidly shifted from theoretical research to technical application. Typically, human knowledge acquisition from data involves a standard process shown in Fig. 1.

The conceptual level of data is relatively low because it refers to original data resources extracted from the objective world without processing or analysis. Information, which forms the basis for people's thinking and decision-making, is the product of effective data processing. Data analysis results can be conceptualized as a higher-level knowledge system. Obviously, the formulation of big data seems like a big step back in terms of data. However, the concept of big data reflects a conscious emphasis on the data. The huge amount of data currently available is unprecedented and has led to new research conditions and opportunities.

As mentioned previously, traditional data are often limited to a single stage of interaction within a data-processing life cycle. Once a single life cycle ends, the data lose value. With the emergence of big data, new data associations are being identified, and new data-related business requirements are emerging. With big data, data association can lead to the discovery of new knowledge (Qian, Cheng, Liang, & Wang, 2015). Cross-sectoral, cross-business, customized, and personalized needs continue to emerge. The value that can be obtained continues to attract attention. The potential and effectiveness of data-based decision-making is strengthened continuously, and the data dividend is gradually realized (Li, Lv, & Li, 2016).

Do digital libraries have big data?

Demand for big data is growing in the digital library field (De Mauro, Greco, & Grimaldi, 2016). However, relatively few studies have investigated digital libraries relative to big data. One reason for the lack of such studies is that many people think traditional database management systems can handle the daily data storage and business processing requirements of digital libraries (Xu, Du, Wang, & Liu, 2017). For example, the National Geological Library of the People's Republic of China only stores approximately 710,000 data items, which is considerably less than common commercial big data systems. Some scholars believe that distributed systems are unnecessary (Hessman, 2013); therefore, we consider an important question: Do digital libraries possess and need big data?

Some studies have investigated basic research questions relative to the application of big data in digital libraries, such as "How," "When," and "If" questions (CNI, 2015). The basic nature of these questions reflects the confusion that many people have about whether big data exist in digital libraries.

Big data do not have a fixed threshold size, and different disciplines and applications may have different definitions. Moreover, various data reduction techniques have been developed. For example, to reduce storage requirements and retain important information, OCR can identify text in an image.

Most current information systems are incapable of storing, processing, or analyzing big data. Even if such systems are updated routinely, they may not be able to cope with increasing amounts of data. For example, due to the growing number of Internet users and web sites, the number of visitors and scope of their use are also growing, and this growth trend appears to be accelerating (Chen et al., 2015). Some data may still be relatively small compared to big data. For example, many digital libraries initially store data collections submitted by many individual researchers. Although the total amount of these data is not large, it demonstrates the same overall rapid growth trend as big data and can exceed the capabilities of existing library systems (Salo, 2010). In addition, with the continuous growth of multimedia resources, such as images and videos, the data types in a digital library have become increasingly diverse.

The widespread use of linked data on the web has also significantly increased the amount of data that digital libraries must handle (Brandon, 2013). In addition, the data connections between academic literature in digital libraries are receiving increasing attention. In addition, the number of connections between different datasets is increasing. Such connections provide more opportunities for coauthor analysis and co-citation analysis, etc. These new data connections enrich the existing data possessed by digital libraries (Teets & Goldner, 2013). Of course, big data provide an effective decision-making function, and effective use of such resources results in a smaller input-output ratio in a library. In fact, some libraries have begun to use this big data decision support function to perform input-output analyses (Fister, 2015).

Big data analysis also provides libraries with more effective ways to fully use such data. The value of big data primarily relates to resource purchase decisions, personalized reader services, hot spot analysis, and the creation of shared academic environments. Some studies have suggested that big data can have a huge impact on data mining, data management, data visualization analysis, auxiliary decision-making mechanisms, and reader behavior analysis (Wu, Su, & Deng, 2013). Given focused scientific analysis, it has become easier for libraries to utilize big data to implement knowledge services (Zhang, 2016). Therefore, as the main gathering place of human knowledge, libraries are gradually storing increasing amounts of big data. Not only has the size and type of data reached the big data standard, but integration and comprehensive utilization of external big data have become key to improving service levels in existing libraries. Therefore, digital libraries are expected to play an increasingly important role in big data analysis and information utilization.

However, compared to other domains, research into big data relative to libraries is limited. For example, the overall research competitiveness is generally weak, research efforts are scattered, and there is a lack of empirical research. In particular, there is insufficient emphasis

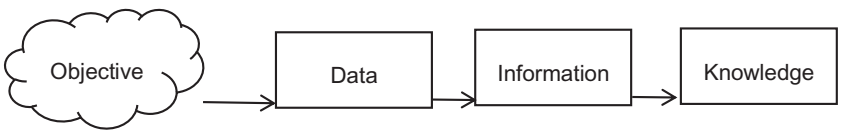

Fig. 1. Standard knowledge acquisition process.

on the practical application of big data in libraries (Lu, 2014).

There are many reasons for this. For example, some scholars believe that digital libraries tend to be self-sufficient relative to organizational management and often do not consider emerging technologies (ProQuest, 2013). Some scholars argue that the reasons may be associated with budget constraints (Horstmann & Witt, 2017). Even though hardware costs are declining rapidly and software usability is increasing, it remains difficult to determine whether such changes will help alleviate the difficulties associated with insufficient funds for digital library construction (Wang, Xu, Chen, & Chen, 2016). Some scholars think that information security and privacy protection issues are unavoidable when considering big data applications. For example, private information may leak if user information is used to identify user interests. Such risks will also have an impact on the extensive application of big data technology in digital libraries (Lee, 2013). Even the advantages of traditional data resources have produced some disadvantages relative to large-scale research into and the application of big data in the field of libraries. For easy use, such data must be transformed, e.g., digitization of paper documents, and integrated with other data. However, these data have experienced various development stages and various information processing technologies have been adopted. Therefore, unlike emerging fields, transforming and integrating traditional data resources often incurs significant processing and conversion costs, which represents another difficulty for digital libraries relative to the application of new information technologies.

Impact of big data on digital libraries

The emergence of big data exacerbates the problem of information overload that has been associated with the development of the web. The costs and difficulties associated with efficient use of information resources are increasing. For example, from a user perspective, big data can produce a growing requirement for weak information among users (Carole, 2008), which is often characterized by fuzzy structures, unclear knowledge scope, a lack of clear and systematic retrieval and discovery steps, and the need to dynamically explore a great number of documents to achieve only partial satisfaction. However, in the face of complex and dynamic research issues, the requirement for weak information is becoming increasingly important and common.

In terms of technological talent in the big data era, there is growing need for data specialists, such as data engineers with big data-processing knowledge and skills, data analysts who can model big data and perform application analyses, and data stewards who can manage and discover valuable data and ensure data availability. Data stewards include data archivists, subject librarians, and other similar professionals (Boulton, 2014). Librarians have helped researchers collect and analyze scientific data for a long time. In the age of big data, subject librarians are becoming more specialized data librarians. Skill-related knowledge for data librarians includes open data licensing agreements, intellectual property rights, rights management, data management plans, resource utilization, data analysis and application, big data deployment and related architectures, management of institutional repositories, data reference, and data publication (Journal Center of the School of Information Management, Wuhan University, 2016).

In addition to changing user information requirements, the impact of big data on user thinking may be also apparent. In fact, thinking determines action. Big data not only provide the means and the possibility of action, big data also provide a completely new way of looking at the world based on the acquisition of as many facts and data as possible to make judgments. The more data we acquire, the more likely we are to eliminate the uncertainty of information, and knowledge with greater value can be created. Therefore, we must consider data content, data found in business, and how such data are used. In many cases, we have some data but do not know how to use the data. Therefore, big data thinking can be considered a type of data literacy. In other words, it is a conscious willingness to understand data significance in all industries, talk to data experts to ensure that data produce value, and enhance competitiveness and improve the effectiveness of management decisions. Data literacy can be considered a new manifestation of information literacy in the age of big data. It is obvious that advances in technology lead to changes in thinking that ultimately lead to shifts in business-related change management (Huang & Li, 2016).

Big data thinking also promotes data management, analysis, use, and services in digital libraries. These significant changes will help librarians identify data processing and data service requirements and can also help librarians understand emerging responsibilities. Big data environments require a change in library management thinking, i.e., from resource management to data management (Gao, 2015). In other words, based on the acquisition and analysis of big data, big data thinking involves the potential relationships among data and library services. This type of thinking has four primary characteristics: regularity, unbiasedness, relevance, and openness (Tian, 2015). Clearly, in relation to digital libraries, big data thinking will facilitate management and service innovation (W.W. Zeng, 2014).

Big data in digital libraries

Importance of user services

The big data era has resulted in new challenges for digital library services and opportunities for transformation. Scientific research is shifting to data-intensive research, and data science has become increasingly important in scientific research. As the service center for documentary information and data resources, libraries face internal and external pressure to change. Libraries must adapt to user requirements and improve content delivery, ease of use, and service responsiveness (Coelho, 2011).

From a service resource perspective, traditional digital libraries often only function as a provider of data resources and play a leading role in knowledge dissemination. However, they do not pay sufficient attention to individual user interests and data content requirements. With the rapid development of big data technologies, the library community has experienced a profound shift from the traditional mass-service model to a personalized service model. The service mode and content of digital libraries have gradually shifted from literature-based to user-based, and from general to personalized services. This shift enables library users to acquire knowledge more effectively and facilitates further optimization of information resources (Li, 2012; L.X. Wang, 2015). The ability to address personalized user knowledge requirements has become more urgent. For example, users are now more interested in autonomously tagging information content, reorganizing knowledge, sharing it widely, and enhancing interactions with other users through online reference tools.

There are three reasons why the big data era has resulted in a need for personalized services in digital libraries.

1) The continuous production of huge amounts of data makes obtaining effective information more difficult. The information overload problem is becoming increasingly prominent relative to limited user information acceptability and time costs. Therefore, finding content that users are genuinely interested in from large-scale data resources and filtering irrelevant information to reduce unnecessary information screening costs has become key to improving user satisfaction in digital libraries.
2) The ever-increasing amount of data leads to ever-increasing data connections. Such connections can not only improve our understanding of data and facilitate ways to find target data more effectively and efficiently, but also provide the necessary and basic conditions for further exploration and analysis of hidden values which traditional single-data resources cannot provide. In large amounts of data, there are a great number of associations among the data, such as the associations among user social data, associations

between users and users, associations among users and resources, and associations among different resources. Such connections allow users to obtain the required service content more easily and quickly. In addition, such connections can generate new user information requirements and can be used to create new types of information services by combining existing user interest patterns.

3) Users obtain and analyze data to obtain knowledge related to a particular application. The understanding and application of the knowledge content is determined by the data and also depends on the specific application environment and current information requirements. Links, interactions, and integration of semantic and application relationships will have a significant impact on user understanding of the obtained data (Zhang, 2005). We often refer to such external information as a context. Personalized information services consider combinations of both contextual information and the user service content.

By providing personalized services, digital libraries can greatly enhance the diversity of user service and provide users with relevant information resources, which reduces user time costs and the costs associated with organizing library information (Gu, 2010). The design of a fully functional and user customizable digital library service system involves cognitive and behavioral design processes. In addition, employing an effective, user-centric human-computer interface can generate a sense of user independence, which is a key factor of user satisfaction (Ferran, Mor, & Minguillón, 2005).

Libraries are an important component of public social service systems. The advent of big data has driven changes to the traditional service modes of digital libraries; however, it has also resulted in unprecedented challenges. To satisfy increasing user demand for personalized service, digital libraries must keep up with technological developments and enhance the application of big data technologies to optimize system construction. Although digital libraries were relatively slow to research both big data and individualized applications of big data, related research and applications have attracted significant attention. Although libraries are faced with transformation requirements, we believe that many of the difficulties faced by existing traditional data resource services in libraries can be resolved using big data technologies and personalized services.

Big data resource in digital libraries

Big data in digital libraries

Not only do digital libraries need to explore how to use large external data resources, their internal data are increasingly showing the characteristics of big data (Li & Zhang, 2013). Understanding a digital library's internal big data resources is a prerequisite of effective use of the resources.

In a digital library, big data primarily comprise electronic and document resources, user information data, such borrowing information and browsing history, and various formalized data that are gradually increasing in library information services (X.Y. Wang, 2015).

For example, the National Library of China has the largest digital document repository and service base in China with more than 1000 terabytes of digital resources and a growth rate of 100 terabytes per year (NDLC, 2018). At the end of 2011, the National Library held 561.3 TB of digital resources. In 2007, it held only 200 TB (DLPP, 2018), which represents a five-fold increase in in a decade. This amount of data is significant. It represents a great deal of knowledge that can reveal relationships among various knowledge themes, entity objects, and carrier forms, as well as research elements, scientific literature, science and technology projects, event activities, experts and scholars, product technologies, organizational structures, and presentations (J.X. Zeng, 2014).

Digital libraries also have a wealth of user data resources, such as user registration information, user behavior information (such as browsing, retrieval, and downloading histories), user interaction information, and other logs. By analyzing these user data and combining them with user social information on the web, we can perform more comprehensive and accurate analysis of library user reading habits, resource utilization behavior, and network usage.

These data represent valuable resource wealth; however, unlike other resources, the value of data resources is often reflected in the process of sharing and promoting data. For example, in 2012, the Harvard University Library publicly published big data about more than 12 million books provided by 73 library branches. Each collection offers up to 100 different values for each attribute, including data, manuscripts, maps, videos, and audio. Users can access the data through the U.S. Digital Public Library (Audrey, 2012). Through long-term relationships with many different types of libraries, in April 2012, OverDrive (Ohio, USA) claimed in its first Big Data Report that they can provide an amount of user usage data collected from these libraries and collaborative e-book sites, such as Buy It Now. These data include the circulation of e-books, reader book browsing and download history, library sites' daily traffic logs, and other information. They freely provide these data to publishers and other libraries that have a working relationship with their library. In addition, using data analysis and mining technologies, they found that the circulation of e-books has significant influence on publishing houses and book dealers, i.e., publishers can determine which books to publish according to readers' electronic book browsing and download histories, and distributors can use this information to automatically determine acquisition bids, build reader-recommended bibliographies from a user perspective, and organize marketing activities from a publisher perspective (Spolanka, 2016).

With the continuous development of library technology and data resources, modern digital libraries can more easily collect various usage data than before, including the usage of various databases and user feedback, such as information on social media. Libraries have developed many useful evaluation tools and have integrated quantitative and qualitative statistical data from library surveys and user usage into existing data resources (Ichiko, 2010). With the advent of big data, an increasing number of libraries prefer to use sophisticated tools, such as learning analytics for data analysis and research performance analysis, compared to traditional data analysis methods, such as statistical methods (Cox & Jantti, 2012).

We should also pay attention to different voices. For digital library services, are there limitations and problems with the combination of big data? This type of big data analysis method can be understood as a typical data-driven research methodology. The characteristics of the data will play a very direct role in related research. Some scholars argue that the introduction of external data sources for digital library analysis carries with it some potential risks. For example, there are some shortcomings in big data information relative to social media, such as privatization, amateurization, and balkanization. In addition, the use of library user logs is an effective and typical big data analysis method; however, due to a lack of user motivation, information requirements, and specific information about their true meaning, simply using such logs often leads to oversimplification and misjudgments for related research, which can be also be considered a common manifestation of amateurization (Niu, Zhang, & Chen, 2014).

In addition, limited by the present technical capabilities of the digital library, simply pursuing the acquisition and combination of big data does not often have a direct impact on process transformation and service improvement. Therefore, the digital library must adapt to the actual requirements of big data processing as quickly as possible and even introduce related evaluation tools to measure the effect of using big data resources. For example, the University of Washington Library launched the “Making the Numbers Speak” project to visualize key data resources using visualization tools and later introduced the Balanced Scorecard Strategic Management Framework to identify useful big data for their own library development between 2010 and 2011 (Qin, 2014).

Big data outside digital libraries

In the face of big data resources outside digital libraries, the more realistic and feasible choice for a digital library is using a third-party service for effective data processing and analysis, such as digital resource database providers and Internet platforms. Such service providers can offer very professional and comprehensive big data resource services based on advanced technological development capabilities with the accumulation of existing data resources. This digital library resource is often referred to as scholarly big data (Huang, Lu, Cheng, & Gui, 2016).

For example, in November 2017, Baidu Scholar (launched June 2014) announced on its official website that it had collected 1.29 million worldwide academic sites, indexed 1.2 billion worldwide academic sites, built 400 million academic documents, created four million homepages of Chinese scholars, and collected 10,000 Chinese academic periodicals with three million keywords. In November 2017, Baidu Library claimed to provide 190 million web documents, nearly 700,000 video course resources, and over one million academic journal articles. Microsoft Academic reported 50 million records in 2010 with annual average increases of 2.7% to 13.6% for different databases (Larsen & Ins, 2010). At the same time, many research literature resources can be obtained free of charge on the web. Some scholars have stated that the proportion of free scientific literature accounted for 43% between 2008 and 2011 (Eric, Didier, & Philippe, 2013). Many well-known IT companies involved in the information retrieval of academic web resources provide users with convenient and efficient services, such as Google Scholar, PubMed, ArXiv, and CiteSeer.

Academic big data have many different characteristics from general big data. For example, academic big data are often highly correlated such that there are many intrinsic links of great analytical value, such as citations, co-occurrence of authors, and co-occurrence of author institutions and publishers, which have always been an advantage in library and information sciences. In addition, ambiguity is obvious in the data due to the diversity of subject terms and professional expressions. For example, the widespread use of acronyms often brings out challenges related to synonyms and polysemy. Same author's names or same affiliations and other external literature information (especially in English) are often written in a variety of different format. In addition, only recently has the diversity of citation formats been resolved. What is more complicated is that copyright protection and intellectual property rights also limit the dissemination and usage of relevant literature resources (Williams, Wu, Choudhury, Khabsa, & Giles, 2014).

Innovation in digital library service modes in the big data age

Problems

Throughout history, libraries have been constantly innovating and revolutionizing themselves to adapt to changes in human society and technology (Zhang, 2001), and, in the big data era, the existing functions of digital libraries are facing new changes. Some problems have reached a point that demands the “theory of extinction of the library” proposed by American librarian Lancaster must be considered (Lancaster, 1982). With the widespread use of Internet technology and mobile devices, the library is no longer the only information service organization. The library's relatively conservative management system gradually weakens its own core competitiveness; therefore, rapid development of information technology once again leads to wide-scale reintroduction and discussion of the “theory of extinction of the library” (Luo & Yao, 2014).

From the perspective of historical development, libraries have always been archival organizations with books as the primary resource, and the degree of association with society has been relatively small, which has made the development of library science and related disciplines face similar situations. For example, library, intelligence, and archival sciences have higher rates of self-citation and citation by similar subjects in the humanities (92.53% and 84.64%, respectively) (Hao & Lu, 2017). This indicates that the knowledge of these disciplines is relatively enclosed and has less impact on related disciplines. Such related academic information is often restricted to dissemination within limited channels. As a result, progress in the field of libraries has been relatively slow in terms of acceptance and assimilation of external professional knowledge and technology. For the digital library, a large number of rapidly changing information technologies exacerbates this problem. Therefore, even in the digital library stage, libraries are accustomed to adopting substance-oriented and relatively isolated digital strategies, which is another reason library development tends to be conservative (Wu, 2017).

It is obvious that the present library service model can be considered an extension of the traditional library model. Libraries primarily focus on traditional literature rather than information content; thus, the operating model of libraries remains highly dependent on the traditional academic exchange system based on commercial publishing. Most digital libraries tend to focus on digitization of documents, the organization and storage of digital documents, document retrieval, and document delivery. It is undeniable that this traditional model has value; however, overreliance on or restriction of such resources and services puts the future of the library at risk (Zhang, 2011).

Several problems that should be explained separately are outlined in the following.

1) Resource delivery: The digital library has changed. In the past, the transfer of resources was performed by librarians. In a modern context, the bulk of this work is performed by the users themselves. Borrowing paper documents can be fully automated. Accessing digital resources has become a basic learning skill, and most users have mastered the skills and technology required to discover and access various information resources. In fact, the status of libraries as a former information service center is being marginalized. OCLC's Library Awareness 2010 showed that very few people used library portals to find information (the 2005 survey also showed only 1%) (Gauder, 2011). This dilemma faced by the library precisely reflects the need for library service reform in the age of big data.
2) Resource utilization: Digital libraries can provide the functions and forms of digital resource services that a traditional library cannot. However, with the continuous development of academic big data on the web, current user intent and methods to use the digital library's resources are also being weakened constantly, and users prefer to give priority access to other digital resources on the web. The importance of social discovery is increasing; however, the importance of libraries relative to providing credible resources is diminishing. Many scholars have found that an increasing number of users employ web search engines to obtain information. In contrast, users of library academic resources have shown a declining trend (Chua & Goh, 2010). For example, CNKI, Google Scholar, and other tools are often used to acquire literature, and web encyclopedias are used to interpret the concept of knowledge. Relative to library lending services, i.e., the traditional core service, e-book service platforms, such as Google Books, have become more convenient for users. In addition, mobile reading and other mobile applications provide users with more choice.
3) Social recognition: In a social context, the value of libraries and librarians is decreasing in the age of big data. For example, the subject librarian, as a former assistant researcher that engages literature discovery and innovation analysis, is no longer the mainstream of library services due to the continuous development and changes in user needs. The Ithaka Institute of the United States performed faculty surveys over three years and found that user identification of the library as an information portal gradually declined over the survey period (Long & Schonfeld, 2010). However, the library as a storage or preservation institution remains essentially unchanged, and the library as a “purchaser” is gradually increasing. Moreover,

the function of teaching and research support, which is considered important by library directors, has not been universally accepted by faculties. Therefore, many scholars have pessimistic views about the development of libraries in the context of emerging technologies (Li, 2002).

4) Change in thinking: Relative to libraries, the traditional concept of identifying knowledge as a collection of resources, identifying user need as access to literature, and identifying the provided service as the retrieval and acquisition of resources has been completely divorced from practical modern requirements. However, unlike technological changes with very short iterative cycles, changes in thought are both more difficult and longer lasting. The resulting misunderstanding and habitual inertia can often only be identified and changed after problems occur. After investigating a number of research librarians, OCLC believes that the value of libraries, library-related human resources, and library technology will face a crisis that could have a huge impact (Michalko, Malpas, & Arcolio, 2010).

Possibility of change

Asking questions and facing difficulties does lead to improvement. We should further change our thinking and clearly identify the direction of the next move. For example, in the 1993–2000 strategic plan released by the British Library, the “Library Collaboration” chapter was replaced by a chapter that discusses “Leadership, Partnerships and Collaboration” and highlights cooperation (Donlon, 1993). In August 2013, IFLA's Declaration on Library and Social Development emphasized that libraries should realize their own value in the participation of social development and further emphasize cooperation with social development rather than focusing on only libraries and reading (Wu, 2017). All of these will better demonstrate the social functions and occupational values of libraries.

In fact, through observing the overall history of library development, many scholars believe that the library has always been able to make new acquaintances and survive through constant “evolution” by adapting to change. Changes in libraries will never stop because the need for information and knowledge will never stop (Fang, 2013). For example, Changping concluded that public libraries and the national economy are in a synergistic development relationship based on the anatomy of the related elements of public libraries and association analysis of the national economy (Hu & Luo, 2005).

This view has a long history. When the 62nd IFLA General Assembly was held in Beijing in 1996, relevant Chinese scholars put forward a similar view. For example, Jianzhong proposed that information service is the core value of a library. Libraries should actively meet the challenge of informatization and constantly consolidate their status as information centers (Wu & Koenig, 1996). Jingzheng also thought that libraries should adapt to the trend of informatization to realize their value and serve society with multifunctional, networked, multi-carrier, and intelligent modes relative to the rise of the knowledge economy and the development of network information (Li & Ma, 1999). In 2004, the annual meeting of the Library Society of China reshaped the spirit of the Centenary Library. In 2007, the core value of the library was an important issue that aroused heated discussion, where the transformation of the library's spirit relative new technological conditions was emphasized. The library gradually realized that service and value are its most precious features rather than technology. Some scholars argued that there is an imbalance between tools and values in the study of library science. It is believed that the integration of tools and values should be the focus of library science development in the 21st century, and the study of library value should be the focus of future research (Xiao, 2004). For example, for public libraries, the “public” essence is the maximum value, and the function of library services includes archive preservation, promotion of the economy, and the improvement of cognition (Liu & Wen, 2012).

Some scholars suggested that the work of librarians should be considered an extension that complements the work of data scientists in the age of big data. For example, the knowledge and experience of librarians relative to understanding user needs can provide data scientists with a better basis for information services. In addition, librarians' archive management and data curation skills can also provide data scientists with the possibility of exploring the long-term value of data resources. Even in the midst of an overlap with data scientists, librarians can provide advantages relative to their ability to control data, data structures, and data conversion (Stanton, 2012).

The American Library Association considers that a library's basic mission is to provide unrestricted access to information. In other words, with the help of professional librarians, libraries use existing and emerging science and technology to collect, organize, store, archive, and save information to provide a variety of information access routes and services (Huang & Hu, 2012). Obviously, the big data era expounds new characteristics of this mission. Modern libraries, especially digital libraries, should learn to use various methods to collect, organize, store, file, and save big data to realize more effective and convenient information services. We have sufficient reason to believe that the big data era provides new opportunities relative to library development.

Changes to digital libraries relative to user services

Libraries have begun to use big data thinking to examine their resources and services, and research into the effective integration of big data by digital libraries is emerging. For example, in 2009, Yongcheng and Huilan of Donghua University Library first proposed the concept of data security and the big data in library circulation (Luo & Chen, 2009). Some scholars have proposed that the construction of a digital library requires a big data environment relative to changes in data environments, changes in scientific research methods, changes in user information literacy, and information technology development (Chen, Qian, & Dai, 2014). From a technical perspective, others scholars have proposed the use of advanced technologies, such as Service Oriented Architectures and cloud computing, to realize a new library service system with all-media resource management capabilities, business management capabilities, and resource-discovery capabilities within the entire domain (Yin & Liu, 2013).

However, scholars have different opinions about what specific aspects of change to include. Such scholars generally believe that there will be great diversity in the utilization of big data and in the reform of existing library service modes. Some scholars propose that library reform under a big data environment should proceed relative to three aspects, i.e., resource construction, technology application, and library service. For resource construction, it is necessary to expand the scope of resources, increase the breadth of resource integration, and increase the depth of resource processing. In terms of technology application, semantic technology should be emphasized, the application of clustering technology should be strengthened, data analysis technology should be widely used, and retrieval technology should be improved. In terms of services, digital library services should be enriched by shifting services from a common passive model to a more proactive, automatic, and personalized model (Su, 2015). Some scholars think that changes in five specific aspects are particularly obvious: (1) changes in the international data environment require digital libraries to manage big data; (2) changes to scientific research methods require digital libraries to support data-driven research environments; (3) the transfer of the innovation model requires digital libraries to meet the needs of business development; (4) changes in user information literacy require digital libraries to meet the needs of knowledge search; and (5) digital libraries must adapt to development and changes in information technology to upgrade service platforms (Chen et al., 2014).

According to OCLC's Information Context framework, service revolution in digital libraries in the big data era can be summarized from three main aspects: the basic information environment, the behavior of

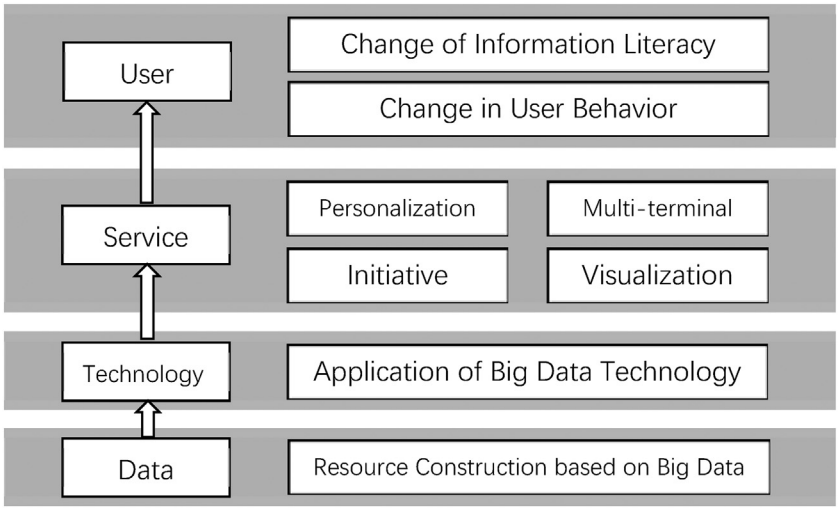

Fig. 2. Methods and trends of digital library transformation in the big data era.

information users, and the basic operation mechanism of the information service (OCLC, 2007). According to this idea, we show an overall framework in Fig. 2.

We can describe the overall function of a digital library as a process of "data–technology–service–user". This organization is also close to the four core components of the internal construction of digital libraries, i.e., resource construction, platform construction, new media services, and standards construction (Han, 2016). Each step of this process in a big data environment has its own improvement direction and change method.

1) Data in traditional digital libraries primarily include literature data, digital resource collections, database resources, and other forms. Therefore, the construction of digital library resources based on big data should emphasize two goals. The first is to use big data to improve the storage and utilization of existing data resources, integrate big data resource into existing digital library resource systems, and enrich the existing data size and type. The second is to integrate newly generated data in new data formats and associated data on the web with the existing data resources of digital libraries. Such data resources provide the possibility of improving traditional services, and they can also provide new service forms and methods.
2) Technology is an indispensable part of digital libraries. The development of the digital library involves the continuous application of information technology. Traditional technology platforms can be improved by technology required for big data processing, such as data acquisition, storage, analysis, and mining technologies. New technology solutions, such as distributed frameworks, parallel computing, big data, and artificial intelligence, will be a foundation of ongoing digital library innovation.
3) Service can be understood as a process in which a digital library can provide data resources directly or indirectly to users. It can also reflect the values of the application of technology in a library. In the big data era, it is possible to identify individual interest patterns of users such that services can be adapted to the changing information needs of users. Therefore, a traditional one-to-many service mode will gradually evolve into a more personalized one-to-one service mode. As a result, each user will have their own digital library, and the digital library can provide proactive services, such as personalized recommendations according to the user's interests. At the same time, we consider the possibility of user access to multi-device terminals to improve and enhance service levels in all aspects. Visualization allows users to access digital library services in a more intuitive and convenient manner. In future, various technologies are expected to become available, such as virtual reality and wearable devices.
4) The user is the object of digital library services. However, the goal of a digital library service is to satisfy the user's information needs; thus, it is more important to consider current user requirements from the user's perspective to more effectively propose ideas and methods to improve existing services. In addition, individual user needs drive the development of digital library services from resource-sharing to user-oriented services (Wu, 2009). For example, for general library users, existing studies have shown that the information literacy of library users has undergone great changes with the ongoing popularization of information technology. Scientific researchers served by subject librarians have data resource and data-processing capabilities that librarians do not. Therefore, the role of "helping users" in a traditional library service should be shifted to "prompting users" and "suggesting to users." However, McKinsey predicted that nearly one-half of data scientist jobs in the United States will be vacant in 2018 (Manyika et al., 2011) because training data scientists incurs great costs. In fact, this situation is the same in the library field because, for librarians to adapt to big data-processing requirements, they must acquire complex expertise in related fields, such as statistics, computer science, and information science. However, short-term rapid training cannot satisfy such requirements (De Mauro et al., 2016).

The user is the most important target of library services. In the past, we put forward the "customer first;" however, true action is not enough. Therefore, to enhance user satisfaction and improve existing service processes and methods, the perspective must be a primary consideration.

Conclusion

It is worth stressing that, relative to traditional and digital libraries, users are not only the target of the service, i.e., users are also a valuable resource if standing on the side of resource. The big data era allows the possibility to fully understand and connect users. Interaction between users and libraries is not only to meet the information needs of users but also continuously provide more user resources to libraries. By exploiting user resources, digital libraries can have a broader perspective of the construction of data resources. Obviously, this user-centric digital library transformation model can provide good opportunity for personalized service development. This change in the characteristics of library services can put forward higher requirements of utilization of big data, and guide the direction of changes in library services.

Acknowledgments

This work was supported by the Chinese National Social Science Foundation 16BTQ030 (2016).

References

Audrey, W. (2012). Strata Week: Harvard Library releases big data for its books. Retrieved from http://radar.oreilly.com/2012/04/harvard-book-data-cloudera-hadoop-splunk-ipo.html.

Boulton, G. (2014, December). Open data and the future of science. *Septentrio conference series (no. 1)*.

Brandon, B. (2013). Talk about big data: How the library of congress can index all 170 billion tweets ever posted. Retrieved from https://www.networkworld.com/article/2162603/uc-voip.

Carole, L. P. (2008). Research practices and research libraries: Working toward high impact information services. Retrieved from https://www.ideals.illinois.edu/bitstream/handle/2142/9742/researchpracticieslibraries.ppt.pdf?sequence=3.

Chen, C. F., Qian, O., & Dai, Y. Z. (2014). Study on the construction of digital library in the age of big data. *Library and Information Service, 58*(7), 40–45.

Chen, H. L., Doty, P., Mollman, C., Niu, X., Yu, J. C., & Zhang, T. (2015). Library assessment and data analytics in the big data era: Practice and policies. *Proceedings of the Association for Information Science and Technology, 52*(1), 1–4.

Chua, A. Y. K., & Goh, D. H. (2010). A study of web 2.0 applications in library websites. *Library & Information Science Research, 32*(3), 203–211.

CNI (2015). Video: Analytics and privacy. Retrieved from https://www.cni.org/news/video-analytics-and-privacy.

Coelho, H. S. (2011). Web 2.0 in academic libraries in Portuguese public universities: A longitudinal study. *Libri, 61*(4), 249–257.

Cox, B., & Jantti, M. (2012). *Discovering the impact of library use and student performance. Deputy vice-chancellor (academic) - Papers*.

De Mauro, A., Greco, M., & Grimaldi, M. (2016). A formal definition of Big Data based on its essential features. *Library Review, 65*(3), 122–135.

DLPP (2018). About digital library promotion project. Retrieved from http://www.ndlib.cn/szzyjs2012/201201/t20120113_57990.htm.

Donlon, P. (1993). British library. For scholarship, research and innovation: strategic objectives for the year 2000: London: the British library. *Alexandria, 6*(1), 85–86.

Eric, A., Didier, A., Philippe, D., et al. (2013). Proportion of open access peer-reviewed papers at the European and world levels-2004-201. Retrieved from http://www.science-metrix.com/pdf/SM_EC_OA_Availability_2004-2011.pdf.

Fang, S. H. (2013). Study on future development mode of academic libraries in digital information environment. *Journal of Academic Libraries, 31*(4), 31–35.

Ferran, N., Mor, E., & Minguillón, J. (2005). Towards personalization in digital libraries through ontologies. *Library Management, 26*(4/5), 206–217 (12).

Fister, B. (2015). Big data or big brother? Data, ethics, and academic libraries. *Library Issues: Briefings for Faculty and Administrators, 35*(4).

Gao, L. (2015). Big data thinking and the discovery of knowledge resources in library. *Library and Information, 159*(01), 122–124.

Gauder, B. E. (2011). *Perceptions of libraries, 2010: context and community. A report to the OCLC membership.* 6565 Kilgour Place, Dublin, OH: OCLC Online Computer Library Center, Inc.

Gu, L. P. (2010). Research on models of user behavior driven personalized services. *New Technology of Library and Information Service, 26*(10), 1–9.

Han, Y. J. (2016). *China library development report, rural library volume.* Chinese National Library Press.

Hao, R. Y., & Lu, W. H. (2017). A quantitative research on the crossing of disciplines in china's humanities and social science. *Journal of Jiangsu University (Social Sciences Edition), 19*(1), 85–92.

Hessman, T. (2013). *Putting big data to work. 262(1418)*, Ind. Week9.

Horstmann, W., & Witt, M. (2017). Libraries tackle the challenge of research data management. *IFLA Journal, 43*(1), 3–4.

Hu, C. P., & Luo, X. C. (2005). The thinking of strategy based on coordination between public library and economy. *Library Development,* (5), 12–15.

Huang, D. Y., & Hu, X. P. (2012). Investigation and analysis on cloud computing applications in overseas libraries. *Library and Information Service, 56*(13), 109–113.

Huang, R. H., & Li, B. Y. (2016). Data literacy education: Expansion of information literacy instruction in the big data era. *Document, Information & Knowledge, 1*, 21–29.

Huang, Y., Lu, W., Cheng, Q. K., & Gui, S. S. (2016). The structure function recognition of academic text—Application in academic search. *Journal of the China Society for Scientific and Technical Information, 35*(4), 425–431.

Ichiko, M. (2010). The value of academic libraries: A comprehensive research review and report. *Journal of Information Science & Technology Association, 61*(1), 384–385.

IFLA (2017). Advances in artificial intelligence. Retrieved from https://trends.ifla.org/literature-review/advances-in-artificial-intelligence.

Journal Center of the School of Information Management, Wuhan University (2016). The first Chinese data librarian training class was successfully held in Beijing. Retrieved from http://www.dik.whu.edu.cn/CN/column/item167.shtml.

Lancaster, F. W. (1982). Libraries and librarians in an age of electronics. *Information Resources Press,* 229.

Laney, D. (2001). *3D data management: Controlling data volume, velocity and variety. 6(70)*. META Group Research Note.

Larsen, P. O., & Ins, M. V. (2010). The rate of growth in scientific publication and the decline in coverage provided by science citation index. *Scientometrics, 84*(3), 575–603.

Lee, J. M. (2013). Understanding big data and utilizing its analysis into library and information services. *Journal of the Korean BIBLIA Society for Library and Information Science, 24*(4), 53–73.

Li, B. Y., & Zhang, X. Y. (2013). The problem of large data in the construction of digital library. *Information Sciences, 11*, 26–29.

Li, J. Z., & Ma, H. Q. (1999). Improving library services of information in social sciences. *Journal of the Library Science in China, 25*(6), 25–28.

Li, K. Q. (2017). Report on the work of the government-delivered at the fifth session of the 12th National People's Congress of the People's Republic of China on March 5, 2017. Retrieved from http://language.chinadaily.com.cn/2017-03/17/content_28591593.htm.

Li, L. (2012). Discussion on personalized service mode of digital library in colleges and universities. *Lantai World, 23*, 91–92.

Li, Y., Lv, P., & Li, L. (2016). Research on personalized services of university library based on the big data mining and decision analysis system. *Document, Information & Knowledge, 2*, 60–68.

Li, Z. X. (2002). Discussion on the disappearance of library in the digital age. *Information and Documentation Services,* (1), 41–44.

Liu, Y., & Wen, T. X. (2012). A survey of the research on the new library movement in China in twenty-first century. *Library, 4*, 46–51.

Long, M. P., & Schonfeld, R. C. (2018). The Ithaka S + R library survey 2010: insights from U.S. academic library directors. Retrieved from http://wcm-dev-web1.cul.columbia.edu/content/dam/librarywebsecure/behind_the_scenes/assessment/forums/ithakapresentation_assessmentforum_04122011.pdf.

Lu, J. (2014). The evaluation and prospects of big data research in library in China. *Library Journal, 33*(1), 20–25.

Luo, X. C., & Yao, M. (2014). Evolution of public library value: According to research perspective changes of value system. *Journal of the Library Science in China, 40*(3), 27–36.

Luo, Y. C., & Chen, H. L. (2009). Research on anonymous publishing of book circulation information based on demand. *Fujian Computer, 25*(8), 100–101.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., et al. (2011). *Big data: The next frontier for innovation, competition, and productivity.* McKinsey Global Institute.

Michalko, J., Malpas, C., & Arcolio, A. (2010). *Research libraries, risk and systemic change.* OCLC Research.

NDLC (2018). About the national digital library of China. Retrieved from http://www.nlc.cn/dsb_footer/gygt/lsyg/index_2.htm.

Niu, X., Zhang, T., & Chen, H. L. (2014). Study of user search activities with two discovery tools at an academic library. *International Journal of Human Computer Interaction, 30*(5), 422–433.

NMC (2017). NMC horizon report 2017 library edition. Retrieved from http://cdn.nmc.org/media/2017-nmc-horizon-report-library-EN.pdf.

OCLC (2007). The view from OCLC programs and research. Retrieved from http://www.oclc.org/content/dam/research/partnership/highlights/infocontext_v1-0.pdf.

ProQuest (2013). Data mining "big data": A strategy for improving library discovery. Retrieved from http://www.proquest.com/blog/2013/241902501.html.

Qian, Y. H., Cheng, H. H., Liang, X. Y., & Wang, J. X. (2015). Review for variable association measures in big data. *Journal of Data Acquisition & Processing, 6*, 1147–1159.

Qin, S. J. (2014). Process data analytics in the era of big data. *AIChE Journal, 60*(9), 3092–3100.

Salo, D. (2010). Retooling libraries for the data challenge. *Ariadne, 64*.

Spolanka (2016). Overdrive announces a series of "big data" reports. Retrieved from http://www.noshelfrequired.com/overdrive-announces-a-series-of-big-data-reports.

Stanton, J. (2012). Big data and the library professional. *Journal of the Library Administration and Management Section, 8*(2), 22–24.

Su, X. N. (2015). Opportunities and challenges faced by digital libraries in the era of big data. *Journal of the Library Science in China, 41*(6), 4–12.

Teets, M., & Goldner, M. (2013). Libraries' role in curating and exposing big data. *Future Internet, 5*(3), 429–438.

Tian, J. P. (2015). The service innovation and development of library in the era of big data. *The Library Journal of Henan, 2*, 22–23.

Wang, C., Xu, S., Chen, L., & Chen, X. (2016, June). Exposing library data with big data technology: A review. *Computer and Information Science (ICIS), 2016 IEEE/ACIS 15th international conference on* (pp. 1–6). IEEE.

Wang, L. X. (2015). Personalized books recommender system for university library. *Modern Information, 35*(12), 41–46.

Wang, S. W. (2017). Information civilization and library development trend. *Journal of Library Science in China, 43*(5), 4–20.

Wang, X. Y. (2015). Analysis of and reflection on the application of big data in university libraries. *Journal of Academic Library and Information Science, 33*(1), 55–60.

Williams, K., Wu, J., Choudhury, S. R., Khabsa, M., & Giles, C. L. (2014). Scholarly big data information extraction and integration in the CiteSeerχ digital library. *IEEE, International Conference on Data Engineering Workshops* (pp. 68–73). IEEE.

Wu, J. Z. (2009). On information need and service for digital library client. *Hebei Sci-Tech Library Journal, 22*(5), 59–61.

Wu, J. Z. (2017). Re-discussion of ten hot topics in the development of librarianship. *Journal of the Library Science in China, 43*(4), 4–17.

Wu, J. Z., & Koenig (1996). The value and benefit of information services – The sixth dialogue on the future of the library. *Library Magazine,* (2), 34–36.

Wu, K., Su, X. N., & Deng, S. H. (2013). Big data, cloud computing and user behavior analysis. *Digital Library Forum. 6. Digital Library Forum* (pp. 19–23).

Xiao, X. M. (2004). Seeking the integration of instrumental rationality and value rationality—The development direction of library science in the twenty-first century. *Library, 1*, 1–4.

Xu, S., Du, W., Wang, C., & Liu, D. (2017). *The library big data research: Status and*

directions. 5(3), 77–88.
Yin, H., & Liu, W. (2013). On the new generation of library service system: Functions, visions and prospects. *Journal of the Library Science in China, 39*(5), 26–33.
Zeng, J. X. (2014). The "big data" of the digital library. *Digital Library Forum, 4*, 1.
Zeng, W. W. (2014). How university library deals with the challenge of big data. *Library Work in Colleges and Universities, 4*, 41–43.
Zhang, H. (2016). Analysis on the present situation of big data research in university libraries in China. *Library Work and Study, 7*, 46–50.
Zhang, X. L. (2001). Mechanisms of digital library: Evolution of paradigms and its challenges. *Journal of the Library Science in China, 27*(6), 3–8.
Zhang, X. L. (2005). From digital library to e-knowledge mechanism. *Journal of Library Science in China, 31*(4), 5–10.
Zhang, X. L. (2011). The trends that will disrupt digital libraries. *Journal of the Library Science in China, 37*(5), 4–12.