

# Information Service in the Big Data Era and Development Strategies for University Libraries

Dong-Xia WANG

Data Center for Economics in Shandong University of Finance and Economics

Jinan, Shandong, China 250014

1173874666@qq.com

**Key Words:** the Big Data Era; Information Service; University Libraries

**Abstract:** Under the background of big data era, the comprehensive linked information service, consistent dynamic information service, self-adaptive personalized information service system and knowledge information service for auxiliary decision-making are the dominating directions for the information service development in the future. In order to provide qualified information service and meet the need of users, university libraries should collect and analyze more native data and provide more tools for knowledge discovery. What's more, they should also keep continuous creativity, implement the data values and promote knowledge innovation.

## Introduction

The emergence of Cyberspace as well as the popularity of Internet of Things and computer technology results in rapid data generation speed, automotive data collection and low-cost data storage and transportation, which has increased sharply the volume of data and means the Big Data Era is coming[1]. The Digital Universe Study by IDC claims that the volume of data created and replicated in 2011 is 1.8ZB, and the study also predicts a volume of 35ZB by 2020[2]. People have realized that the discovery and adoption of data with vast volume, high variety, fast velocity and low values will be the main pattern for social wealth growth. In this way, the coming of Big Data Era also indicates the coming of tide with productivity development and customer surplus [3]. In the Big Data Era, the questions that how the information service based on data should develop and how the libraries with information service deal with such an era require further consideration and study.

## Information Service in the Big Data Era

### Consistent Dynamic Information Service

In the Big Data Era, chances emerging from the flow of vast data will be lost fleetly or be taken away by others, and no one wants delayed information for shares, air ticket price and road conditions, which will be harmful. Many data vary from time, which requires a consistently changing information service to keep its validity. Then how to analyze such a vast stream of data becomes the biggest problem. However, with the fusion and completion between cloud, Hadoop and excavation technology of data, the dynamic information service in many fields has come true. Cloud can provide vast storage space and super computing power for big data, and Hadoop can realize the retrieval of distributed data, and excavation technology of data provides accurate model for data analysis. The combination of them provides guarantee for the consistent dynamic information service. In the Big Data Era, consistent information service not only provides immediate information, but also makes real-time prediction such as shares price prediction and tickets price prediction.

### Self-adaptive Personalized Information Service

Human behavior has heterogeneity in time and space, so the users' needs for information are complex, multi-dimensional, heterogeneous, changing or even conflicting, which makes the realize of these needs become a big challenge[4]. However, in the Big Data Era, human's behavior will

leave “data footprint” in the information system, which is stored in different systems and is blended and integrated through automatic technology. With such integration, a life trace and panorama is represented and data are thus explained with each other, which makes human’s behavior predictable [5]. In the Big Data Era, the personalized information service provided for users are combined and finished dynamically in three methods. The first one is to analyze the data and get personal behavior rules with the function of independent filter; the second one is to integrate and analyze the data and get the common behavior model of certain kind of people with the function of cooperative filter; the third one is to analyze the data and get the possible information action according to the context, which has a function of context filter. Three methods are connected and corrected constantly so that self-adaptive personalized information service can be provided for users. In the Big Data Era, the key of self-adaptive personalized information service lie in the acquisition of feedback information. So in order to develop the reliability and accuracy of personalized information, exploratory recommendation service should be offered forwardly and consistently to users.

### **Comprehensive Linked Information Service**

In the Big Data Era, people tend to grasp related information comprehensively in fear that false decisions are made due to lacked information or missed opportunities and the information service in this era must meet this need. With the improvement of linked data technology, to meet this need through big data gradually comes true. Linked data are lightweight semantic web technology and the value lies in the transfer of unstructured data and structured data with different standards into structured data with the same standard with the help of RDF model so that the machine can understand these data. By describing semantics of linked data, the computer can recognize and deal with the data automatically, which makes the internet involved into semantic web realize data resources share. By integrating linked data in different areas and using visible instrument, computer can thus provide comprehensive linked information service. For example, if information of real estate is required, the related information such as education, shopping, dinning, traffic, finance, entertainment, environmental protection and security are also provided. What’s more, the concrete data as well as the present data and previous data and even future data are also required. These data are integrated with linked data technology and provided to users in a visible way so that users can make decisions according the comprehensive information provided[6].

### **Knowledge Information Service for Auxiliary Decision-Making**

In the Big Data Era, what users need is not listed data, but useful data for decision-making and the more intelligent the better, which reveals the decisive value of big data. The value of big data lies in the automation and intelligence in searching for models, rules and characteristics from the vast data, which can thus help the users to make more accurate decisions and gain higher values. That’s to say, it is born from the perspective of decision-making [7]. In the Big Data Era, both for enterprise and personnel, the decision should be made according to fact and accurate judgment instead of mere experiences, otherwise, the opportunities or advantages will be lost. The provision of such a information service needs abilities from three aspects. The first one is the ability of data fusion, that is, the ability to fuse rapidly structured, half-structured and unstructured data, to deal with the vast data effectively. The second is the strong processing ability, that is, the ability to find knowledge with high value-added by combining analytical technology such as network analysis, data stream mining, time series analysis and visualization and Hadoop technology. The third one is the ability of innovative thinking, that is, big data thinking, which refers to an awareness of finding new model and information from big data and requires the support of creative scientists. In a word, data+ technology+ thinking are what knowledge information service for auxiliary decision-making in the Big Data Era needs.

## **Development Strategies for University Libraries**

### **The Importance of Data Collection Lies in Localization**

Data itself is a kind of document, which includes abundant knowledge. So data collection is the important duty of libraries. However, libraries do not need collect all the data and there are specialized institutions for data collection in every field, which are more professional and accurate. The data collected finally comes to several people through share and exchange and then are sold out to libraries. So libraries need only choose the data they need. However, libraries should collect two kinds of localized data, resource data and the user's data. For resource data, scientific research data is of this kind and is collected through scientific methods such as observation, detection, experiment and investigation by senior researchers in their research activities [8]. Resource data has high values and libraries have duty to collect them. University libraries should urge the universities to establish deposit payment and employment management mechanism that can balance the benefits of research funding departments, research departments and researchers, and accomplish the storage and application of scientific data through the establishment of corpus. Of course, some university libraries have practices in this aspect, such as Datastar by library of Cornell University, which permits the publication and storage of data, which further realizes the cooperation and share of data [9]. Besides scientific data, local economic data should also be collected. Of course, these data cannot be collected by libraries and local government department has collected and stored them, which can be received through share and cooperation. Since localized data cannot be collected comprehensively and timely, they become a big fortune of university libraries in the future. Another kind of data is user's data, including the interaction data between the user and resource, such as mobile internet data and sensor data, social network interactive data and indirect behavior data such as learning data of students and teaching data of teachers. User's data mainly provides reference for the accurate personalized service of libraries. So various users' data should be integrated and provides self-adaptive personalized information service. The localized data collected by libraries can be used by others through share, but the control power should not be lost.

### **The Importance of Data Storage Lies in Curation**

In the big Data Era, data is a kind of resource. Data must be organized and managed for exploration and application, just like the books in the library, which need go through a process of interview, content, storage, circulation and evaluation, and this is so-called data curation. Data curation is a kind of management with both function of storage and supervision, working through choosing, storing and maintaining the digital asset of the system [10]. To deal with the fast-emerging data, university libraries in some countries such as America, England, Canada and Australia start the data management as well as challenging practice and research. Meanwhile, they also explore some mature models of data curation. For example, DCC in English put forward a kind of model which can be divided into three layers: the first layer is the traditional research process and publication process; the third layer is a high-level process [11]. Based on the life cycle of academic communication, ANDS put forward the concept of data curation continuum and divided academic publication process as private research domain, shared research domain and public domain, further describing the border of data management [12]. University libraries must adopt mature data curation mechanism to make an effective choice, storage and maintenance of data, and reach an effective sharing of data. The choice of data, just like the interview, must be organized through a comprehensive evaluation according to the development of the libraries, short-term and long-term need as well as the valuation of data. As a result, the people who carry out such a process and a structured system must be required to integrate the data and form a data network. What's more, there is a cradle-to-grave process in the data, which proves the data as a dynamic process and requires a consistent choice of the data to save the storage space and costing. In a word, data curation is the basis of the comprehensive linked information service, consistent dynamic information service, self-adaptive personalized information service system and knowledge information service for auxiliary decision-making, and reveals the destiny and role of libraries in the Big Data Era. The libraries should change its role, provide support for data storage and explore

data resources, which can absorb the attention of the researchers and increase the values and competitive power of libraries.

### **The Importance of Data Analysis Lies in the Instrument Application**

In the Big Data Era, libraries should provide not only service of one-site search, but also the cluster analysis, hotspot predication, network analysis, visualized analysis, citation analysis and knowledge correlation analysis so that some valuable knowledge can be found. Although library needs complex analysis instrument, these instruments are explored professional system and data companies with more professionals, better mechanism and more abundant funds than libraries, which can ensure the improvement of these instruments. Interestingly, some libraries have explored or are trying to explore some system but the sustainability is worrying. The internationally influencing knowledge discovery systems are Summon by serials Solution Company, Primo by ExLibris Company and EBSCO Discovery Service by EBSCO Company. These systems are popularly applied in university libraries of the world and China Superstar Company puts forward the Superstar Discovery system by overcoming the weak points for Chinese knowledge. These knowledge discovery systems are based on a vast number of documents and provide a high-value document discovery and deep knowledge discovery as well as visible knowledge linkage by analyzing the attribute data and relation data of these documents. With the help of cloud service by Saas, these systems are distributed in the libraries and these libraries choose the suitable analysis instruments and provide service for users. Meanwhile, the analysis instruments can provide intelligentized service such as dynamic information service and knowledge information service for auxiliary decision-making. As a result, libraries must make a comprehensive evaluation and choice and staff in the libraries should also grasp these techniques.

### **The Importance of Data Service Lies in Knowledge Innovation**

The information resources used by libraries are not for information service provision, but for the finding of valuable knowledge, further providing knowledge-value-added service. In face of overcritical users, libraries must keep their abilities of innovation otherwise they may be disbanded by their users. In other words, the staff in the library must have a sense of Big Data, which can help them find the opportunities and new knowledge as well as the correlation between cross-academic knowledge. With such a sense, knowledge in different subjects can intersect with each other and knowledge innovation can thus emerge. Although the self-service ability of university users is strong, their ability of data analysis is weaker than those professional. As a result, they want to gain more help from the library staff, especially the meaning of the knowledge through data analysis, which requires an innovation of data analysis instruments. Even users able to serve self totally need the help of the library staff because they need the most valuable and dynamic knowledge with time going on, which rely on the efforts of data scientists. By data cirution including constructing model, searching for model and analysis as well as prediction [13], these high-quality scientists use their advantages and strong points to keep innovation and meet users' need. As a newly-developed technology, the Big Data requires related talents which are few at the present time. So university libraries should have crisis awareness and forward looking, and make training and introduction plan for those professionals who are the manifestation of the core value and also provide knowledge information service for auxiliary decision-making and implement the data value.

### **Conclusion**

In the Big Data Era, the distribution of data industrial chain forms and the exchange and sharing become the mainstream. As a result, professional institutions for related service emerge constantly and the information service in the Big Data Era presents the developing trend of the comprehensive linked information service, consistent dynamic information service, self-adaptive personalized information service system and knowledge information service for auxiliary decision-making. The coping strategies of university library are: the importance of data collection lies in localization, the importance of data storage lies in curation, the importance of data analysis lies in the instrument

application, and the importance of data service lies in knowledge innovation.

## References

- [1] Bigdata. The Next Frontier for Innovation, Competition and Productivity, McKinsey Global Institute (2011).
- [2] Yuhong Cui. Research on the New Role of Library in E-Science Environment: Scientific Data Management, *Library Magazine*. (2012).10:20-24.
- [3] F. Gantz & D. Reinsel. The 2011 Digital Universe Study: Extracting Value from Chaos. Available on <http://www.emc.com/collateral/demos/microsited/emc-digital>. (2011).
- [4] Xiaobin Huang & Huixin Zhong. The Construction of Enterprise Competitive Intelligence System Based on Big Data, *Intelligence Magazine*. (2013) 3:37-43.
- [5] Cheng Hong. A Study on the Current Situation of Scientific Data Service Abroad, *Library Magazine*. (2012) 10:31-34.
- [6] Ling Jin. From Support to Decision: the Big Data Realize the Business Value of Enterprise. Available on <http://www.ciotimes.com/bigdata/72565.html>
- [7] Xuwei Pan. Self-adaptive Personalized Information Service: a Method Based on Contextual Cognition and Ontology, *Journal of Library Science in China*. (2009)11:41-47.
- [8] Zhihong Shen. Open CSDB: An Applied Study of Linked Data in the Science Corpus, *Journal of Library Science in China*. (2012)5:17-26.
- [9] Wanlu Shi & Ren Shuhuai. Data Curation: a New Movement of Library Service, *Library Magazine*. (2012)10:24-27,34.
- [10] Zipei Tu . The Big Data, Guangxi Normal University Press. (2012) 161.
- [11] Cuijuan Xia. The Publication and Realization of Linked Data---Drupal as a Case, *Journal of Library Science in China*. (2012)1:49-57.
- [12] Xiaolin Zhang. Right Management of the Storage and Propagation of the Contents in Institutional Repository, *Journal of Library Science in China*. (2012).
- [13] Huixin Zhong. The Development Direction of Information Service in the Big Data Era and Coping Strategies of University Libraries, A Academic Memoir of Annual Meeting by Library Club of Guangdong. (2013)11:8-17.