

# Application thinking of Associated Data in Library

Jiliang Wang

Weifang University Library, Weifang 261061, China

wjl\_wfu@163.com

**Abstract.** Associated data has broad application prospects in library service work, including the release, discovery, integration, knowledge retrieval, data access, and academic exchange of library resources. The article analyzes in detail the practical significance, specific methods and main problems that should be paid attention to in the library application related data, and provides effective practical guidance for library application related data.

**Keywords:** library; associated data; practical analysis.

## 1. Introduction

The concept of associated data was first proposed by Tim Berners-Lee, the father of the Internet, in "Associated Data Notes" in 2006, in which he analyzed the development and evolution of the Web and proposed the idea of developing a data network, and the core of the data network. And the key is to associate data. At the 2009 TED conference, he once again proposed that the associated data is a box of data, and when linked through open standards, many new things and new applications can be emanating from it [1]. The core technology for associating data is the use of the Resource Description Framework (RDF) for the organization and exchange of resources. The Semantic Web technologies such as URI, HTTP, and RDF publish, share, and correlate related data resources on the network, which is an optimization strategy for transforming from a file network to a data network [2]. The basic principle is to implement a dynamically associated knowledge object network with a standard-based knowledge representation and retrieval protocol and a gradually expandable mechanism, and support knowledge discovery and knowledge organization based on this. The library is the main organization of knowledge organization and service provision. It has a large amount of structured data, and the emergence of related data provides a broad application prospect for library knowledge organization, discovery and service. The application of related data by foreign library circles has attracted the attention and promotion of the world library community.

## 2. Practical Significance of Library Application Associated Data

### 2.1 Conducive to the Fusion of Library Data Resources

With the rapid development of network information technology, libraries must continuously expand their service methods and means to help readers discover the intrinsic relevance of resources and obtain useful resources to improve the quality of library services. Foreign libraries have widely used resource discovery services to expand their directory search interfaces, browse the results of dynamic updates, and display more collection information [3]. Associated data can provide a structured data foundation for expanding library bibliographic information, providing readers with new types of knowledge access services; linking library links to a wider range of information sources, not only helping readers access the information resources of the library, but also It can help readers to obtain more information resources outside the library; the library can also publish the library resources as associated data, so that readers can link back to the library. This requires libraries to continuously integrate various existing data resources, provide high-quality interrelated data and information resources for readers to access, and strengthen information institutions such as archives, museums, cultural and art museums, and information companies. Exchange and collaboration, strengthen the exchange and reuse of data resources between the two parties, and actively integrate, release and develop data.

## **2.2 Conducive to Strengthening Scientific Research and Academic Exchanges**

In the network environment, libraries are increasingly involved in scientific research and academic exchange activities. Subject librarians help researchers conduct citation analysis of research results. The library develops and utilizes institutional warehousing systems to help knowledge institutions preserve, manage, and reuse research materials that have been published and not published. Some library institutional warehousing software began to support semantic services and associated data, and to achieve semantic data storage and semantic services by providing a unified URI identification and relationship between resources for knowledge resources in the warehouse [4]. In particular, the digital library resource system begins to consider the interoperability of the system, not only the interoperability with other institutions' warehousing, but more importantly, the interoperability with other types of services, which requires the support of associated data. In academic research, researchers hope to see the results of research being cited and reused, and hope to find more relevant original literature, raw data and discover more authors in the same field through research results. Library-associated data can help researchers find this information, find more information about the author, and even get information on how to fund the project, and allow users to download the original research data.

## **2.3 Conducive to Expanding the Library Service Platform**

Associated data is of great significance for libraries to publish bibliographic data and normative data on WEB. It is the basic technology for the Internet to develop and standardize the control of online resources and digital objects in the Semantic Web era. It is one of the core technologies for the library to publish and construct information resources [5]. Associated data not only reduces the complexity of integrating distributed heterogeneous data sources in WEB networks, but also promotes new applications of library-associated data and extends the impact of library service platforms. The library has the advantages of data resources accumulated in the long-term literature, can enhance the connection with the external data world, expand the service scope of the library's knowledge discovery platform, and continuously improve the service level of the library. At the same time, it can also fully explore the relevant professional knowledge of the existing library field, such as metadata models, models, standards and protocols, redefine requirements, compile guidelines, and develop new standards, publish all kinds of data and norm files of the library to the Internet in the form of linked data, improve the impact of library data on the Internet, and make the library data occupy an important seat on the Internet.

## **2.4 Conducive to Improving Library Service Capabilities**

Strengthening social service functions and supporting local socio-economic development are one of the important functions of libraries. Associated data exists in the form of public WEB, which can meet the needs of Internet data sharing. In the associated data cloud diagram, there are many associated resources that can be used as the library's resource library and knowledge base. The associated open data in the cloud map can be used to expand and correlate the library resources, providing users with richer and more accurate search results. Because of the associated data, libraries can provide users with high-quality access to knowledge resources, provide links to other institutional resources, and expand the use of knowledge sources, thereby expanding the scope of library data and enhancing the social service capabilities of libraries. Associated data can connect the resources of the library with the external world, realize one-time writing and repeated use, and can open the extended service in a wider scope, providing rapid utilization and reuse within the system or nationwide. Continuously improve the social value of library resources.

### **3. The Practice of Library Application Associated Data**

#### **3.1 Providing Associated Data Release of Literature Resources**

Publishing library resources as associated data is the primary application area for associated data in libraries. As a specialized organization for information collection, organization and arrangement, the library's MARC data, specification documents, subject headings, etc. can be published as any connected data. At present, many libraries have begun to adopt RDF and related data, such as the Swedish National Library, the Library of Congress, OCLC, and the German National Economic Library. The Swedish Union Catalogue (LIBRIS) is the world's first joint catalogue of bibliographic data published as associated data, mainly managed by the Royal Swedish Library, opening up approximately 6.5 million bibliographic records and 200,000 documentary records (personal names, place names, subject headings) of more than 200 member libraries, create links from the federated directory to DBpedia, and provide valuable experience for the library community to develop linked data applications. The Library of Congress has linked all of the Congressional title table LCSH in the SKOS format and provided the download of the LCSH vocabulary, becoming another successful example of associated data applications. In addition to publishing the authoritative document LCSH as associated data, the Library of Congress also creates associations between different resources (such as LCSH and LIBRIS). [6] OCLC uses the SRU service to provide associated data for the VIAF (Virtual International Specification Document) project. Through the release of associated data, the library provides a broader space for the network communication of information resources.

#### **3.2 Extending Information Resource Discovery with Associated Data**

Currently, many libraries attempt to extend their directory retrieval interface and functionality by implementing resource discovery services. The Resource Discovery Service can view dynamic results and display more collection information. Since resource discovery is primarily achieved through subject headings and other data in MARC records, there are limitations in revealing depth. Although, the library bibliography has added information beyond the bibliography, such as catalogues, book covers, and links to comments. However, this method is still limited to the bibliographic level. When data is imported into a directory or linked to a specific record, the record does not provide subject-based information, or details about the location and author. Or, although more additional information about topics and authors can be provided, there is no link to a particular collection. Linked data allows libraries to be associated with a wider range of information resources, not limited to the resources themselves, and can expand information such as locations and people. By increasing the scope of information provided, it not only helps users to determine whether they are the required books or journals, but also helps users to obtain information outside the library. The library can also publish its own resources as open, linked data, allowing users to link back to the library. Thereby expanding the scope of library resources and discovering more useful related information resources.

#### **3.3 Using Linked Data to Realize Heterogeneous Resource Fusion**

In the network environment, the exchange, integration and reuse of data are increasing. In the future, the associated data of the library will come from distributed and heterogeneous library data sources on the Internet, and data integration must be performed before users can browse and utilize them. With fused linked data, it is becoming easier for libraries to identify and exchange data with other organizations. Libraries are also paying more and more attention to the role of data fusion. More and more libraries will play a role in data fusion. Cooperation between libraries and information institutions and data storage organizations will become more and closer. The data fusion is becoming a reality. For example, Europeana is the digital resource portal of the European National Library, which incorporates metadata from a number of libraries and cultural institutions in Europe about library objects, uses SKOS and other metadata models, and is developing a semantic retrieval service for digital resource portals [7]. However, from the current application of associated data, most of the associated data from different data sources are rarely further integrated with each other. Therefore,

the integration and amalgamation of entity data from different data sources into different vocabulary data mapping will become an important challenge for future library-associated data applications.

### **3.4 Using Associated Data for Knowledge Retrieval**

The existing library search service system has some shortcomings in author search and topic search, such as incomplete search for different forms of authors, no extended search service, and no relevant knowledge retrieval and discovery. Associated data provides a good way to improve the service capabilities of the library's existing search system. The use of canonical documents and vocabularies can provide extended search services (such as semantics, episodes, and subordinate word expansion), and publish topic specification documents, person name specification documents, etc. as associated data to enable knowledge retrieval, which will be distributed online and in different languages. The collation and integration of the specification documents can improve the effectiveness of the retrieval service system [8]. Due to the associated nature of the associated data, the associated data is inseparable from knowledge organization, knowledge processing, and knowledge services since its inception. Many researchers have conducted relevant explorations. The use of associated data to generate semantic annotations for common patterns extracted from text documents, the method can be applied to the field of biological network analysis. The role of connected data in promoting knowledge organization and knowledge processing still has many places to explore. The library can think and research in these aspects and explore the potential value of connected data.

### **3.5 Provide Cross-institutional Data Access and Reuse**

Associated data has obvious advantages in the organization and unified management of massive data resources, providing a strong inter-agency cooperation and data sharing capability between libraries and academic communities, and also provides a convenient underlying data access method for third parties. Associated data helps to achieve write-once, multiple-use, and resource sharing between different organizations. The library reuses bibliographic data for a variety of purposes, such as generating new resource listings and recommending resources by topic or course. Associated data is structured and can be quickly re-used, and associated data helps provide open extension services on a larger scale, such as across the country and across all libraries of the same type. The University of Huddersfield in the UK shares 13 years of reader circulation data and readers' recommended data under the Open Data Commons License, providing downloads of anonymous information sets about readers' borrowings to other libraries. Comparing this kind of data across the museum, you can find readers in different regions to borrow trends, evaluate the development strategy of collections, etc. [9]. Some multi-institutional collaborative projects (such as Europeana, LOACH, NDIIPP, OREChem, etc.) use flexible data to collaborate through the use of linked data principles to maximize data sharing and data management convenience, enable internal and institutional alliances to maximize data reuse.

## **4. Problems in Library Application Associated Data Should be Noted**

### **4.1 Interface Interactivity**

The interaction between the data application interface and the user will directly determine the efficiency of the service, and the library user has high requirements for the friendly performance of the user interaction interface. From the perspective of the library user, the associated data provides it with a data that integrates the distributed heterogeneous data sources. However, associative data is not a specific technology, but a model, a framework for data organization and sharing. The associated data browser allows different users to browse between different data sources. Currently, the function and friendliness of the associated data browser and its search engine are still very limited, and cannot meet the requirements of library users. How to improve the friendly interactivity of the associated data browser will be a major challenge for future connected data. In the improvement, the navigation mechanism provided by the hypertext browser can be used to design a similar navigation control mechanism on the associated data browser, so that the user can move forward and backward; add and remove buttons, etc. can also be integrated on the associated data browser.

## **4.2 System Reliability Issues**

The application of linked data is open and social. A data application platform based on associated data will integrate external data and functions extensively, so library-associated data services will become more integrated and dependent on the external social information environment. The associated application of the associated data is not a common link, but a deep data embedding, or even functional embedding; if the external data link breaks, it will directly affect the stability and functional realization of the local system. Because it builds the stability, reliability, and integrity of its own systems on uncontrollable external systems, it challenges the reliability of library services [10]. Of course, this problem is not inherent in the associated data, but the associated data has a strong dependence on external data, and the reliability of the connection is particularly prominent. At the same time, the stability of third-party services will directly affect the reliability of associated data applications.

## **4.3 Service Resource Integration Issues**

Resource integration issues with associated data and other Web services will be another major challenge for library-associated data applications in the future. Since there is very little associated data for currently available applications, especially reliable associated data is less. When integrating data, it is inevitable to integrate data published in other ways. The data processing methods of these different sources are different, resulting in the complexity of system design. The related data of the library in the future comes from the distribution of multiple regions and even global networks, and heterogeneous data sources. Data integration must be carried out before users can browse. However, according to the current application process of associated data, there is little further integration with each other.

## **4.4 Data Link Maintenance Issues**

The increasing participation of data in the associated data network and the increasing size of the application determine that the links of the associated data are not static, and the relationship between different associated data sets will become more and more complex. After the establishment of the link relationship between the library's associated data sets, timely and effective maintenance of these dynamically changing links is an important means to play the role of linked data and maintain the link relationship between data sets. Currently, the network Web architecture of associated data allows the existence of dead links, but too many dead links will undoubtedly cause the client to issue a large number of invalid HTTP requests, thereby reducing the efficiency of use. Therefore, the library should periodically scan and scan the URI links of the associated data. When the data and links change, the related links should be updated or deleted in time.

## **4.5 Personal Privacy Protection Issues**

At present, the protection awareness of users' personal privacy is getting stronger and stronger. Associated data provides users with a data set that integrates heterogeneous data and distributed data, which facilitates user access to information, but also brings related privacy infringement risks, which will greatly limit the breadth of application of associated data. The solution of this problem requires many efforts, such as the public to enhance the protection of their privacy awareness, the identification of private information after the acquisition of related data, and the relaxation of relevant provisions on privacy infringement at the legal level. But to really solve it, it also requires an effective combination of information technology and legal means.

## **5. Conclusion**

Associated data, as an implementation of the Semantic Web, is leading a worldwide network revolution that promotes the correlation of data from different data sources by publishing and linking distributed structured data on the Web. For libraries, the advantages of applying associated data can provide strong technical support and assistance in library resource integration, information



dissemination, retrieval services, knowledge access, and academic exchanges. Although at present, the associated data still has many challenges in data integration, link maintenance, privacy protection, etc. However, in the near future, with the solid foundation of the associated data technology foundation, the completeness of the system structure and the release method, the association is simple. The application of data in libraries will have a broader perspective.

## References

- [1]. Berners-Lee T. Linked data [EB/OL]. [2013-05-18]. <http://www.w3.org/DesignIssues/LinkedData.htm>.
- [2]. Bizer C. the Web of linked data [EB/OL]. [2013-05-18]. <http://www.wiwiiss.fu-berlin.de/en/institute/pwo/bizer/research/publications/Bizer-WebDB-WebOfLinkedData-Talk.pdf>.
- [3]. Heath T, Hausouase M, Bizer C. How to publish linked data on the web [EB/OL]. [2013-05-08]. [http://videolectures.net/iswc08\\_heath\\_hp1dw/](http://videolectures.net/iswc08_heath_hp1dw/).
- [4]. Huang Yongwen. A Review of the Application Research of Associated Data in the Library [J]. Modern Library and Information Technology, 2010(5):1-7.
- [5]. Liu Wei. Associated Data: Concept, Technology, Application and Prospects [J]. Journal of Academic Libraries, 2011 (2): 5-12.
- [6]. Haslhofer B. Linked Data is an Attempt to continue the Well-established Information Organization Tools Known in Libraries [EB/OL]. [2013-05-28]. <http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf>.
- [7]. Concordia C, Gradmann S, Siebinga S. Not (just) a Repository, nor (just) a Digital Library, nor (just) a Portal: A Portrait of Europe as an API [EB/OL]. [2013-05-08]. <http://www.ifla.org/files/hq/papers/ifla75/193-concordia-en.pdf>.
- [8]. Huang Yongwen. The Framework of Linked Data Application and Suggestions for Constructing Linked Data Applications [J]. Digital Library, 2011, (9): 7-13.
- [9]. Pattern D. Free Book Usage Data from the University of Huddersfield [EB/OL]. [2013-05-28]. <http://www.daveyp.com/blog/archives/528>.
- [10]. Lin Haiqing, Lou Xiangying, Xia Cuijuan. Library Linked Data: Opportunities and Challenges [J]. Journal of Library Science in China, 2012(1): 59-67, 112.