Differentiated digital library evaluation in a hierarchical model stemming from its operational scope and complexity

WU Jianhua* & WANG Zhaohui

Library of Huazhong Normal University, Wuhan 430079, China

Abstracts   Digital libraries are complex systems and this brings difficulties for their evaluation. This paper proposes a hierarchical model to solve this problem, and puts the entangled matters into a clear-layered structure. Firstly, digital libraries (DLs thereafter) are classified into 5 groups in ascending gradations, i.e. mini DLs, small DLs, medium DLs, large DLs, and huge DLs by their scope of operation. Then, according to the characteristics of DLs at different operational scope and level of sophistication, they are further grouped into unitary DLs, union DLs and hybrid DLs accordingly. Based on this simulated structure, a hierarchical model for digital library evaluation is introduced, which evaluates DLs differentiatingly within a hierarchical scheme by using varying criteria based on their specific level of operational complexity such as at the micro-level, medium-level, and/or at the macro-level. Based on our careful examination and analysis of the current literature about DL evaluation system, an experiment is conducted by using the DL evaluation model along with its criteria for unitary DLs at micro-level. The main contents resulting from this evaluation experimentation and also those evaluation indicators and relevant issues of major concerns for DLs at medium-level and macro-level are also to be presented at some length.

Keywords   Digital libraries, Evaluation, Hierarchical model, Unitary digital library, Union digital library, Hybrid digital library

1   Introduction

The evaluation of DLs is part of the activities of DLs research and construction. It came forth in the form of a project evaluation with the start-up of Digital Library Initiative (DLI) of America, followed by the research and practice in-depth with the development of DLs.

The evaluation activities of DLs in United States (US), United Kingdom (UK) and Europe (EU) concentrate on three aspects. The first is on organization activities

* Correspondence should be addressed to Wu Jianhua (Email: wujh@mail.ccnu.edu.cn).

and the standards, such as DigiQUAL and E-Metrics of Association of Research Libraries (ARL), eVALUEd and SCONUL (Society of College, National and University Libraries) of UK, EQUINOX and DELOS of EU, and COUNTER has become the standard for the whole world to participate in. The second is on the project evaluation. It originated in US and practiced with the establishment of DLI. Professor Marchionini of the University of North Carolina has taken a long period of time for the evaluation study on the Perseus Digital Library.[1] Some pioneer participants of DLI had done valuable studies on the evaluation problems of DLs at the final phase of DLI development, from the point of view that "digital libraries are sociotechnical systems".[2] The program of evaluating digital library in the higher education such as that of eLib of UK was open to the public. People were encouraged to take part of it extensively. The useful data stemming from investigations, focus groups, observations and field studies, were collected. Such evaluation undertakings helped the DL system designers to understand better users' interests and issues of concerns. The third aspect is on the general study of evaluation-related issues and on DL evaluation practice. These studies include but not limited to the DL evaluation theory, criteria for DL evaluation, etc., for which Saracevic and Fuhr have done much work.[3–7]

The DLs evaluation in China emphasizes particularly on the quantitative measurement, usage evaluation of electronic resources, evaluation theories, and the evaluation of some particular systems and services including the usability and the developing systems.[8–9]

DLs are a host of complex systems. Their complexity is represented in the following five ways: 1) Multi-disciplinary coverage. Fox listed 65 subject fields related to DL studies[10]. 2) Co-existence of different levels of operational sophistication and stages of development. Researches on and practice of DLs are being conducted by library and information professionals concurrently independent of each other's stage of progress that having been made in a particular DL development project. For instance, some people are studying the basic theory and technology, whereas some others maybe working in the field of practical applications or there are still yet more others who maybe putting DLs in a testing situation to study their social consequences. It is quite common that previous research results are to be tested again in a changed or simulated new operational situation and have them revalidated in a more meticulous manner. 3) Distributed heterogeneous information systems. DLs can be seen as the different information systems with different subject matter and subject to different standards, distributed and embedded into the Internet. Furthermore, new DLs are appearing on the scene and some of them, whether they be old ones and/or new ones, are merging vertically or horizontally. Such phenomena give rise to new and complicated problems in doing research on and/or in field practice for DL-related theories, technologies, intellectual

National Science Library,
Chinese Academy of
Sciences

property rights and economics. 4) Tremendous differences in operational scope and complexity among DLs. 5) Diversified forms of DLs.

It is impossible to evaluate such complicated systems with a uniform set of criteria and measurement. Therefore, it is necessary to evaluate DLs differentiatingly such as making an analytical study at different scope categories and at different sophistication levels. Reviewing the current literature of DL evaluation theories and practices, one can find that professional people have not yet developed a feasible DL evaluation system that is equally applicable to all different categories of DL operational scope and levels of sophistication. This paper is an attempt to remedy such a deficiency by proposing an integrated DL evaluation system for any one single DL and/or for them all as a class. This paper is organized as follows. In section 2, the operational scopes and types of DLs are analyzed hierarchically. In section 3, a hierarchical model for DLs evaluation is proposed. In section 4, a 3-layered evaluation criteria system is presented in detail.

2　A differentiated analysis in gradations of operational scopes and organizational types of DLs

The distinct differences in operational scope and diversity of types present difficulties for DL evaluation. Obviously, one cannot use a uniform set of criteria and measurement indicators to evaluate DLs that differ in scale and type. In order to develop a DLs evaluation system that suits varied scopes and types of DL operation, it is necessary for one to clarify and define the scope and type of a given DL.

2.1　DL operational scope and its impact on evaluation

DLs can be classified into five groups, i.e. mini DLs, small DLs, medium DLs, large DLs, and huge DLs, as showed in Fig 1.

Take academic digital libraries as an example.

The smallest academic DL can consist of only one journal. D-Lib Magazine may be considered as its representative. That can be categorized into mini DLs.

Some DLs are established based on a set of publications held by some academic societies, for example, ACM (Association for Computing Machinery) Digital Library. Which is surely larger in scale than those built with only one journal. It can be rated as small DLs.

Some academic DLs collect mass information resources, however, with limited types, for example, China Academic Journals Full-text Database (CAJ) held by China National Knowledge Infrastructure (CNKI) and ChongQing VIP in China and Elsevier ScienceDirect in the Netherlands, focusing on the journal collections. CNKI and VIP collect almost all journals published in mainland China, totaling more than 6,000 journal titles. Elsevier ScienceDirect collects over 1,500 important
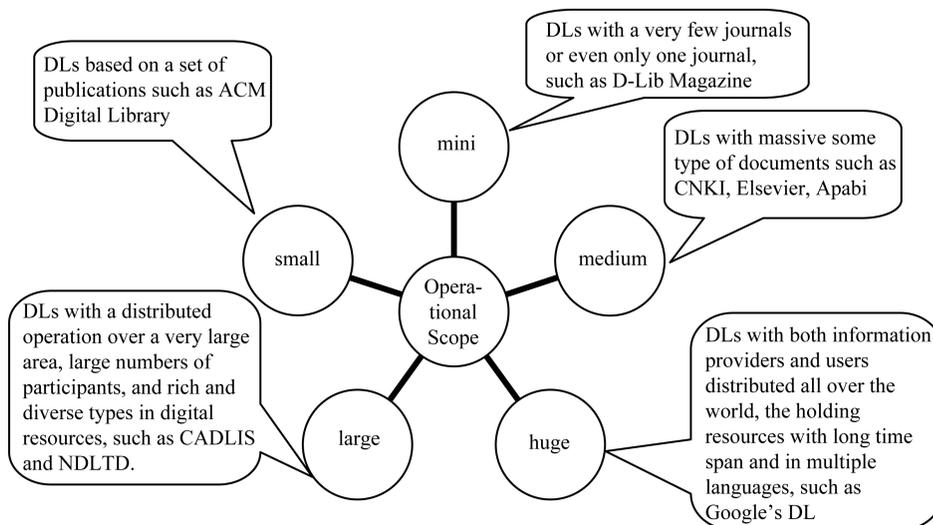
Fig. 1   The Operational scope of digital libraries

journals in scientific fields. For another instance, Apabi eBook, SuperStar Digital Library, and Sursen Digital Library focus on the book collection. They are much larger in scale of quantity than small DLs, and could be treated as medium DLs.

Some DLs are a distributed operation over a very large area, with large numbers of participants, and with rich and diverse types in digital resources. For example, China Academic Digital Library & Information System (CADLIS), the National Digital Library of China, the National Culture Information Resource Sharing Project sponsored by the Ministry of Culture, all can be seen as the large DLs. Furthermore, there is another kind of DLs that are built together by several institutions. Although the type of resources is relatively unitary, the amount is very large and a wide range of subjects are covered. Networked Digital Library of Theses and Dissertations (NDLTD) is an example of this kind, which can also be defined as a large DL.

For another kind of DLs, both the information providers and the information users are distributed all over the world. The time-span of collection building has gone through a very long period of time. Many languages are involved. The resources are rich and diverse in subjects and formats. They can be categorized as huge DLs. The ambitious digital library project of Google that is currently being constructed belongs to this huge kind of DL. People all over the world are looking forward to its successful completion.

One can anticipate that, in order to meet people's information demand and also for their easy access to information on a world-wide scale, all types of DLs now under construction will probably take steps to reassemble themselves into becoming a part and parcel of some huge DLs.

From the above analysis, one can find that the scales of DLs are determined by a few factors, i.e. subject types and amounts of resources, time span of collection coverage, space for housing resources, language variety of resources and/or whether the DLs are built by one entity player or by multiple players in a consortium. The richer the types of resources, the larger the amounts of resources, the longer the time span and the bigger the onsite storage space, the more language varieties, the DLs will be more complicated and their operational scope will be undoubtedly larger. If there are fewer types of resources, fewer amounts of resources, limited time span and space for collection building and collection housing, fewer types of languages, the DLs will be less complicated and their operational scope will be smaller in comparative terms.

Take a DL specializing in thesis and dissertation collections as an example. For the DL in a specific university, although there maybe a long time span in its collection coverage, if the resources are from a unitary source and are centralized, and running independently, it is small in scale of operation and can only be classified as a mini or a small DL. For instance, the thesis and dissertation database of CNKI or Wanfang collects resources from almost all universities and academic institutions of China, but the resources of its collection are centralized in management and the database runs independently. It can be classified as a small DL. In the case of China Academic Library & Information System (CALIS), the thesis and dissertation collection of which is different from the front. The two-layer architecture for CALIS is established. All colleges and universities in China are taking part in this project and responsible for building their local resources. The national administration center located in Tsinghua University takes charge of constructing the central metadata database and the portal. This is a DL, which is built by more than 1,000 participating libraries in this library consortium. The amount of its resources in academic theses and dissertations is vast. It can be classified as a medium DL. NDLTD is being built in a way similar to CALIS' DL of thesis and dissertation collection, However, the participating libraries are from all the world. Its collection consists of multiple languages. It is more complicated in administering this huge collection of such diversity as in subjects, in languages and in publication formats, etc. than that of CALIS, Therefore, it is grouped with large DLs.

The main problem of evaluating a DL is that there are differences in DLs' scope of operation, which automatically presents a thorny issue of what kind evaluation criteria should be applied to these different categories of DLs. As we know, people pay attention to different elements of different DLs in terms of their operational scopes. They actually use different criteria in their evaluation of various types of DLs.

Mini DLs, such as D-Lib Magazine that builds its digital resources from a single source, collects all resources available from a single source; it naturally does not

have a problem of completeness in its collection specialty. People are usually more concerned about their service performance, such as the speed of response time, service reliability, retrieval functions and effects, the quality and value of publication contents, and the position and influence in the dominant subject field that these particular mini DLs operate.

For small DLs such as the ACM portal and the collection of some academic theses and dissertations of a particular university, it is very critical for users that the strength of their resources in certain targeted subject areas is of enormous significance and unique value. Furthermore, people will pay attention to the completeness of their collections, the availability of their resources, their accessibility, and the cost for getting these resources.

For the same reason, people's concerns of DL usage vary vis-à-vis the operational scope (i.e.: medium, large or huge) of DLs, which will also result in variation for their evaluation criteria. For example, CAJ of CNKI and VIP collect large numbers of journals in China, but what about the completeness and significance of these resources to a particular researcher? Are all core journals and significant journals in full text available? How much time and cost will it take to get a single copy of a paper? Is the quality of content high or low? Is every journal collected from its first issue of publication to the latest issue available? Is the journal holding information up-dated quickly? What is the situation about the reliability of contents? How much will it cost to build a mirror site? Is the view and download information available for statistical purpose? Take CALIS' DL of thesis and dissertation for example again. This is a union catalog set-up of participating DLs. What about the interoperability and availability for the participating institutions in this consortium? How much time and cost will it take from the time one searches for the information of a thesis or dissertation to the time when he gets the full text? Are the links reliable? Is it easy to use? Are thesis and dissertations collected comprehensively and completely? Are they of high quality? For NDLTD that collects thesis and dissertations from all over the world, in addition to the issues of concerns listed above, the multi-language service function is another dimension to be evaluated.

2.2 An analysis of the organizational types of DLs within a five-tiered ranking hierarchy

Nowadays, there are thousands of DLs of all types that are in existence. According to the scope of operation and level of sophistication, within a five-tiered ranking hierarchy (mini, small, medium, large and huge DLs) as discussed above, DLs can be further classified into unitary DLs, union DLs and hybrid DLs.

There is a common feature shared by mini DLs, small DLs, and medium DLs. They are all unitary DLs. There is only one entity to build and run it. It is usually

built upon the single digital library system and conforms to the same standards and specifications.

Large and huge DLs are usually union DLs. There are many institutional players to build and run them. They are built upon distributed heterogeneous digital systems, and may follow the same standards and specifications. However, owing to the fact that there are multiple players to operate these union DLs, local operational differences may exist in some of these DLs. It can only become an integrated DL by means of installing an interoperable mechanism.

Hybrid DLs are neither the same as unitary DLs nor similar to union DLs. Most of them are affiliated with some large institutions and have diverse types of resources. Their resources are distributed in different digital library systems, which may follow different standards and specifications. This brings much more difficulties for users to use their rich resources, because what they are facing is not an integral DL. A great amount of work must be done to reorganize and to integrate the resources in a seamlessly way so as to provide users with a one-stop shopping effect.

3 Hierarchical model for DLs evaluation

In addition to the operational scopes and types of resources, the granularity of DLs is another important factor that influences evaluation. Take the evaluation of digital resources as an example. If the granularity is different in size in those documents being evaluated, the attributes of these documents that people care about will be different, and the elements for which they will measure and evaluate will be different also. What is to be measured really depends on the purpose of the evaluation. For pictures, the emphasis is on such elements as the format, the resolution, the size, and the storage space required. But for pictorial and/or graphic databases, people care more about the pertinence of the contents, the richness, the completeness, the uniqueness, the speed of transmission, and the immediate availability, etc.

Evaluation is an undertaking by those people who attempt to recognize the value of certain objects and events to them. Things usually have many different kinds of values to different people. In value evaluations, one has to have a clear understanding about his or her own objectives and then decide on certain specific elements or aspects for evaluation. When one is going to evaluate a DL, is he to evaluate the operational scope, the level of complexity in resources management, its socio-cultural influence, or its impact on learning, research and knowledge creation? Or the value emphasis is more narrowly focused on some of the technical aspects such as the quality of its information resources, its user interface, its information accessibility and its services rendered? Obviously, it will be more complicated to evaluate a DL comprehensively than to evaluate it only in some aspects. It is our contention that a 3-tier evaluation model may be the right answer to deal with the complicated issue of DL evaluation in all their variant forms and operational scopes.

National Science Library, Chinese Academy of Sciences

Specifically, this model deals with DL evaluation problems by assigning them into three separate categories in accordance with their operational scope; namely, micro-level, medium-level and macro-level. DLs in each category contain a range of different and unique elements to be factored into their evaluation.

3.1　Micro-level evaluation of DLs

Evaluating DLs at micro-level means that to evaluate either a relatively small-scale DL in all its operational aspects or a relatively large-scale DL focusing narrowly on a particular area of operation. It is simpler to evaluate some targeted DL operations in relatively narrowly focused areas as the scope of evaluation strictly limited to a narrow range. There are only a few evaluative indicators and variables to be taken into consideration. The complexity of the evaluation task is relatively low.

The evaluation of mini DLs and small DLs can be always dealt with as a micro-level situation, regardless whether the targeted areas for evaluation lie in the overall DL operation, or in the quality of its resources, the function and performance of its information retrieval service, its information accessibility, the frequency of library usage, and/or a given library's impact on users. The reason is that the targeted objects for evaluation are basically unitary DL. The types and quantity of these unitary DLs' resources, the time and space span for their collection building and housing, the number of users and the range of their distribution are all limited in scope. It is easy to set up or to modify an evaluation indicator system for such DLs, and it is also easy to obtain adequate evaluation data.

Medium DLs are also unitary DLs. They have larger number of collections and users. Their users are distributed in a relatively wider area than those of mini DLs and small DLs. The evaluation of this kind of DLs still belongs to a micro-level evaluation situation regardless whether it is to evaluate a single attribute or multiple attributes of a medium DL. Take Elsevier ScienceDirect as an example, it can be treated as a reputable commercial DL, which collects more than 1,500 peer-reviewed journals in the scientific field. It can be classified into the medium DL. If one is going to evaluate only some specific attributes such as the disciplinary coverage of the database resources or the system's usability, it is a micro-level evaluation. But if one is to evaluate it in a comprehensive and exhaustive way from all possible angles, which may include but not limited to such perspectives as its academic impact and value, etc., it is thus indeed beyond the scope of a micro-level evaluation.

For large DLs and huge DLs, if one is just to evaluate any one single attribute or a range of quantifiable operational aspects such as the system's usability, the performance, the quality of digital resources, the satisfaction level of users, etc., it is still considered as a micro-level evaluation situation.

3.2 Medium-level evaluation of DLs

A medium-level evaluation aims at evaluating large DLs in all their operational aspects, or at one particular operational aspect of a huge DL. The targeted objects under evaluation are of considerable magnitude. The activities of evaluation are of a wide range. The information appertaining to the evaluation is abundant and there are numerous variables. Thus, the complexity of a medium-level evaluation is on a high order.

For a comprehensive evaluation of a medium DL, or for an evaluation of some medium DLs in a comparative perspective, such undertakings present an issue of high complexity invariably. Thus, they belong to a medium-level evaluation situation.

The evaluation of hybrid DLs and union DLs belong to medium-level evaluation. For example, academic libraries, public libraries, research libraries all poss large quantity of digital resources. Most of their digital resources are commercial databases, along with a few self-built digital resources on the side altogether. It is of high complexity to evaluate the digital resource service system in these medium DLs for their digital resources are diverse and enormous in quantity.

The multi-attribute evaluation of one aspect of large DLs and huge DLs, and the comprehensive evaluation of hybrid DLs and union DLs, can all be grouped together into the category of medium-level evaluation if the evaluation process is not very complicated nor it is very difficult to get the data required for the evaluation.

3.3 Macro-level evaluation of DLs

The macro-level evaluation of DLs is an extremely complicated undertaking. The targeted objects that are to be evaluated are huge in operational scope. The evaluation undertakings involve a very wide range of activities. The information and informational variables appertaining to the task of a macro-level evaluation is just overwhelming. Besides, it is a very complicated process to set up an evaluation indicator system. It is also difficult to obtain and process those information necessary for the evaluation task on hand. These are some of the major problems for conducting a macro-level evaluation.

The evaluation of large and huge DLs from an all-embracing perspective including but not limited to the architectural service, technological aspects, sociocultural impact, or the evaluation in a national or regional comparative context, can be considered as a macro-level evaluation. The evaluation of this macro-level kind takes place at very large DLs.The complexity of the macro-level evaluation task is indeed very high.

Furthermore, the evaluation of DLs at all developmental levels, and the impact of DLs on human society, can be considered as macro-level evaluation, too.

The evaluation model and the level discussed above are tentative and preliminary. It's better to define first and foremost a proper level category for the evaluation

task according to its purpose, the scale of the targeted objects for evaluation, the complexity of contents for evaluation, and then to device appropriate measures to do the job and to achieve the goal. The more complicated the problems, the more efforts are required for getting the evaluation job done well and also in conformity with the established standards.

4   A hierarchical structure of the DL evaluation criteria

4.1   A summary of the existing DL evaluation criteria

4.1.1   Fuhr's holistic view

Fuhr proposed a description scheme and a set of DL evaluation criteria based on DELOS conceptual model of digital library. These criteria take a holistic view and they try to evaluate all types of DLs in a comprehensive way. Fuhr initially focused his evaluations of DLs in three aspects; namely, data or digital resources collection, technology, users and usage.[11] Later he developed a new scheme covering the system in operation, subject contents of the digital resources and user.[7]

4.1.2   Lynch's perspective of developmentalism

Lynch believes that a successful DL ought to be a DL capable of sustained development. There are three critical factors associated with it. The first is the sustainability of adequate financial support. The second is the sustained development of management and controlling techniques. The third is to have a clearer idea about the profile (i.e.: who they are and where they are) of the DL's potential users so as to customize its services in a more satisfactory manner to the information needs of its clientele.[12]

4.1.3   The 5S model-based qualitative evaluation of DLs

Goncalves designed a model for qualitative evaluation of DLs based on the 5S theory.[13] This model set up an evaluation system of 21 quantifiable indicators to evaluate a smallest DL in six aspects; namely, the catalog, the resources collection, digitized items, metadata specifications, repository and services.

This model takes the view that DLs are similar to information systems, which organize and manage digital objects. It pays special attention to the quality of information and service. It is suitable for the evaluation of mini or small DLs in a unitary-level situation.

4.1.4   5S model-based users' perception on a successful DL

Shen critically and selectively adapted Goncalves's quality model, and developed a more satisfactory model with a set of new criteria to measure the degree/extent of success of DLs.[14]

These new criteria are focused on the DL's information quality (IQ), quality of the information system (QS), performance expectation (PE) and social influence (SI). Shen's model examines information quality and system quality of a given DL from users' perspective. It is believed that users' satisfaction will lead to frequent use of the given DL, which is a positive indication to the successfulness of this particular DL.

4.1.5   DL evaluation standards as practiced in China

The criteria developed by DL researchers in China pay more attention to the DL resources collection, applicable information technologies, service quality and administration. In this set of criteria for DL evaluation, those indicators that associated with administration are listed either separately or combined with those indicators for the evaluation of a given DL's resources collection, information technologies and/or professional services.[8,15–16]

4.1.6   DL evaluation from the perspective of its electronic resources performance

Electronic resources performance has been discussed thoroughly in literature. Xu constructed a system of evaluation indicators that are based on specialists' opinions. It consists of five elements such as contents of resources, retrieval system, usage, cost, and services.[9] This kind of criteria usually takes commercial databases such as CNKI, Elsevier ScienceDirect as targeted areas for DL evaluation purpose,

4.1.7   Liu Wei's criteria for hybrid DLs evaluation

Based on the objectives and capabilities of urban libraries and also the factors of the evaluation standards for traditional libraries into an overall consideration, Liu Wei et al. developed a set of evaluation criteria as the matrix for the evaluation of hybrid DLs.[17]

Compared with other evaluation indicator systems, the most significant feature of Liu's criteria is that he uses individual database aggregated rather than counting digital documents as units for quantification and measurement. This is because Liu's criteria are specifically custom-tailored to evaluate hybrid DLs, which invariably fall into the category of medium-level evaluation situation.

Summing up the above criteria, one can find that, other than the criteria proposed by Liu Wei, which aim at medium-level valuation of DLs, all other criteria are constructed for micro-level evaluation, thus are more suitable for the evaluation of unitary DLs. None of the criteria clearly and unambiguously manifest itself with regard to its suitability for a particular level of evaluation (i.e.: micro-level, medium-level or macro-level). Each criterion is actually better suited for evaluating

certain particular DL functional areas such as on resources or on technology and so on. All criteria established so far are needed to be further abstracted, refined in order to form an integral DL evaluation system that has universal appeals and practical applicability.

4.2 Differentiated evaluation criteria for DLs at different hierarchical gradations

From the above analysis about the complexity of DL evaluation, we have come to realize the fact that DL users pay attention to different issues and apply different criteria for DLs at different levels. Most of existing criteria are actually for DLs at micro-level. Some criteria fail to differentiate the level that the targeted objects being evaluated really belong to. As a result, many existing or proposed indicators systems are not practical. So it is necessary to create a new integrated DL evaluation system by taking into full consideration of the hierarchical nature of DLs in terms of their operational scope and complexity. This new integrated DL evaluation system should be able to measure more accurately DLs operation, large and small in all their shapes and forms.

The DELOS conceptual model proposed by Fuhr would be of directive significance in terms of developing a set of DL evaluation criteria. He holds that people should take a holistic view to do DL evaluation. Information system, collection resources and human resources should be integrated. There are five aspects which need to be considered particularly. The first is the functional capabilities and efficiency of the information system. The second is the interaction between users and the information system. The third is the support for different sites/points of access and usage strategies such as searching, browsing, and navigating. The fourth is logistic support. The last is situational and contextual factors such as organizational and group forming issues. The first three aspects belong to problems of system design, and the last two are of problems associated with circumstances of users.

Summarizing and comparing the existing DL evaluation criteria will help us identify the main issues that DL users are paying great attention to:

- Users-resources-information system-DL usage frequency (Fuhr)
- Users-administration-sustainable development (Lynch)
- Resources-organization of resources-services (5S quality model)
- Information quality-system quality-performance expectancy-social influence (5S success model)
- Resources collection-technology-services-administration (Evaluation standards as practiced in China)

- Collection contents-cost-information retrieval capability of the information system-system usability-usage frequency of DLs-data vendors' services provided (Xu Ge)
- Resources-services-administration-users perception (Liu Wei)

From the listing, one can find that users, resources, services, system or technology are DL functional areas that most people are concerned about. Information technologies and their usability are the focus of attention at an earlier operational stage of DLs. With the development of more applicable information technologies, DLs systems are getting more mature; resources and services become the new focus. Administration is embodied in aspects of resources organization and system functions and services, so it is not necessary to be listed separately in the indicators though it may be integrated with other DL functional areas. What users demand includes resources, system functions and services, the degree of understanding and meeting users' demand are all embodied in resources, system and services. So, resources, system and services are three major aspects of DLs. Resources is the foundation of services. System is a platform to organize and to provide resources. In meeting DL users' information demands, related services and interface functions are created from the information system. Accordingly, resources, system and services can be taken as the three most essential indicators of our proposed DL evaluation system.

Based on the three basic indicators as groundwork, combined with the objectives and requirements of evaluation, it is easy to deduce different criteria for a 3-tiered evaluation system.

4.2.1   The evaluation criteria for unitary DLs at micro-level

Library activities associated with its resources include the collection development and the organization and utilization of its resources. Resources collection is to collect all kinds of relevant materials according to the DL's stated missions, objectives and the targeted users' information demands. It has the library's collection all cataloged in accordance with the established professional standards so as to make it ready for public use. The quality of the resources collecting and cataloging work will determine whether the DL can satisfy users' resources and information demands. The resources collected should be in compliance with the DL's service policy, and has a high relevance to the DL's established subject specialty and priority. Resource formats should be rich, and the subject matter of resources should be coherent, complete and authoritative. Resources organizing should follow established professional standards and specifications so as to develop an accurate bibliographic record and to make the collection easily accessible. Specifying metadata, performing cataloging, enlisting digital objects into collection, and constructing collections

into a repository are all resources organization activities. The use frequency of a DL resources can be measured by the number of registered users, amount of time spent in browsing and downloading, etc.

DLs can be seen as a kind of information system, so the evaluation criteria for information system are applicable. Efficiency, effectiveness, reliability are basic indicators.

Services are functions and interface developed in the system aiming at users demand. Service functions include basic functions and personalized functions. Basic functions are those that every unitary DL should possess, such as browsing, navigating, retrieving, help and consulting, altering, downloading, printing, emailing, document managing, etc. Just as Lynch stated, DLs provide environments for doing active work and helping people to cooperate, make decision and analyze information. Personalized functions are designed for special groups of users with special preference, tastes, and demands under particular condition. Users make use of system functions via interface. The basic principle for interface design is its being made pleasing to the eye, and also clear, and easy to use.

Systems' performance is reflected in service quality. It is not necessary to have service quality listed separately; it can be integrated with service indicators.

According to the above analysis for unitary DLs at micro-level, the evaluation indicators system is formed as shown in Table 1. Comparing these criteria with Fuhr's five important aspects, one can find that all key elements are embodied in service indicators. According to DLs' unique characteristic that "Content is the king" proposed by Fuhr, resources are the most important aspect to be evaluated. The indicators for content are embodied in resources indicators.

4.2.2　Main issues associated with the evaluation criteria at medium-level

The evaluation criteria for unitary DLs at micro-level discussed above also suits comprehensive evaluation of mini DLs, small DLs and medium DLs. For the evaluation of large and huge DLs, in addition to the fundamental factors in the evaluation indicators, some special factors should be taken into account. Large and huge DLs usually can be classified into hybrid DLs, union DLs, regional DLs, global DLs instead of unitary DLs. The primary feature of which is the heterogeneous and distributed systems, in which, issues of interoperation and standardization need to be addressed. Compared with the evaluation at micro-level, the resources are in different granularity, too. Those problems should all be considered.

Now, take university DLs for illustration.

University DLs can be categorized as a kind of hybrid DLs, the evaluation content and indicators are significantly different from those of unitary DLs'.

Table 1   Evaluation indicators for unitary DLs at micro-level

1st tier indicators	2nd tier indicators	3rd tier indicators	Measuring methods
Resources	Collection	Pertinence Completeness Authoritativeness Satisfaction rate of users	Measuring and numerating
	Organization	Metadata conformity Metadata completeness Catalog completeness	Testing
	Utilization	Number of registered users Amount of browsing Amount of downloading	Measuring and numerating
System	Efficiency	Responding time	Testing and calculating
	Effectiveness	Recall, precision	
	Reliability	Rate of successful services	
Services	Basic function	Browsing, navigating, searching, help and consulting, alerting, document managing printing, emailing	Testing
	Personalized functions	Team supporting, self-organization, filtering, refining and extracting, text mining	Testing
	Interface	User-friendly, clear, well-suited	Experiencing
	Service quality	See system indicators	

The catalog reveals the available resources in a given DL. The rate of resources that have already been organized into the WebPAC reflects the degree of digitization at cataloging level. It is a basic evaluation indicator.

The allocation ratio of digital resources for disciplines of the university reflects the whole digital resources satisfaction level of a university library. The amount of document delivery shows the level for utilizing the resources outside the university library, and the level devoted to its' alliances.

The accessibility to digital resources is a radical problem which puzzles university libraries and their patrons. Commercial databases are usually restricted to be visited on campus by means of IP range, and this brings inconvenience for users to use. This problem can be solved by setting a proxy server or by means of Virtual Private Network (VPN) technology. However, many university libraries in China failed to solve the problem.

And, it is essential to build the resources navigation system and to reorganize the resources, because resources of university DLs are diverse, numerous, and jumbled. One-stop search and virtual reference service are also basic functions that university libraries should provide. The usability of university DLs' websites and use of digital resources are important contents to be evaluated.

In summary, the situation that whether a university DL meets users demand should be reflected in the evaluation criteria, which includes resources, functions, and efficiency.

4.2.3 Main indicators of the evaluation criteria at macro-level

DLs have significantly changed mankind's style of production, learning, working and living, and have improved the efficiency of activities. Evaluation serves the function of guidance. Building a scientific evaluation indicators system will not only help improve DLs' quality and level of sophistication, but can also help improve the level of development for DLs in a given region or in a country.

The evaluation indicators at macro-level should include the impact of DLs on learning and research, on the development of people's potentials, on productions, social life, social benefits and economic benefits, etc.

People's information demands are changing. When current information needs are met, more new needs may arise. The DL evaluation criteria reflect not only the objectives of evaluation, but also the demand of mainstream DL users. Therefore, the evaluation criteria should change in keeping pace with the changing objectives of the DL evaluation undertakings and the changing needs of the mainstream DL users. The criteria, if remaining stagnant, will not work well. The criterion for different targeted DL evaluation objects can be deduced from the basic criterion according to the level of the objects to be evaluated. In other words, based on the basic essential criteria, one can make his or her choice to form a set of criteria that meets the requirements of a particular DL and is easy to use.

5 Conclusion

By analyzing the operational scopes and types of DL hierarchically, this paper builds a 3-tiered DL evaluation system of micro-level, medium-level, and macro-level. Aiming at problems at each different level, a differentiated evaluation criteria system is developed. It puts complicated problems of DL evaluation into a clear gradational structure.

Guided by the hierarchical model proposed in this paper, the authors have completely analyzed 211 academic DLs in China and have constructed a set of evaluation indicators for their evaluation. After our further consultations with a few specialists in the field, we are able to have some of those DL evaluation indicators further refined. Analytic Hierarchy Process (AHP) method has been used to calculate the weighted coefficients. The finalized indicator system has been applied in our empirical and comprehensive evaluation of 4 representative DLs among the 211 major academic DLs. The tally of the evaluation results is in conformity with the reality and it also reflects the level of DL development as a whole among those 211 major academic DLs in China. At the same time, the evaluation results show the right approach to improve the DL[18]. The empirical study indicates that these DLs' evaluation based on the hierarchical model being proposed here by these two authors is scientific and effective in guiding the DL evaluation in all of their different levels and categories.

National Science Library,
Chinese Academy of
Sciences

References

1   Marchionini, G. Evaluating digital libraries: A longitudinal & multifaceted view. Library Trends, 2000, 49(2):304–333.
2   Bishop, A. P., Van House, N. A., & Buttenfield, B. P. (Eds). Digital library use: Social practice design and evaluation. Cambridge, MA:The MIT Press, 2003:1–24.
3   Saracevic, T. Digital library evaluation: Toward an evolution of concepts. Library Trends, 2000, 49(3):350–369.
4   Saracevic, T. Evaluation of digital Libraries: An overview. In Presentation at DELOS Workshop on the evaluation of Digital Libraries. Padova, Italy. 2004, October:4–5.
5   Bertot, J. C. & Davis, D. M. Planning and evaluating library networked services and resources. Westport, Conn.: Libraries Unlimited, 2004:1–22.
6   Andrews J. & Law D.(Eds). Digital libraries: Policy, planning and practice. Aldershot, Hants, England: Ashgate Publishing Limited, 2004:113–126.
7   Fuhr N., Taskonas, G., & Aalberg, T., et al. Evaluation of digital libraries. International Journal on Digital Libraries, 2007, 8(1):21–38.
8   Qiao, H. Study on digital libraries evaluation (in Chinese) (thesis). Beijing: Peking University (2005).
9   Xu, G. Study on evaluation methods and its application of electronic resources performances of university libraries in China (in Chinese) (thesis).Chengdu: Xinan Jiaotong University (2006).
10 Fox, E. A., Akscyn, R. M., & Furuta, R. K. et al. Digital Libraries. Communications of the ACM, 1995, 38(4):23–28.
11 Fuhr, N., Hansen, P., & Mabe, M., et al. Digital libraries: A generic classification and evaluation scheme. In Constantopoulos, P. & Ingeborg, S. (Eds). Proceedings of the 5th European Conference on Research and Advanced Technology for Digital Libraries. Darmstadt, Germany: Springer, 2001:187–199.
12 Lynch, C. Colliding with the real world: Heresies and unexplored questions about audience, economics, and control of digital libraries. In Bishop, A. P., Van House, N. A., & Buttenfield, B. P. (Eds). Digital library use: Social practice in design and evaluation. Cambridge, MA:MIT, 2003:191–218
13 Goncalves, M. A., Moreira, B. L., & Fox, E. A., et al. "What is a good digital library?"—A quality model for digital libraries. Information Processing and Management, 2007(43):1416–1437.
14 Shen, R., Vemuri, N. S., & Fan, W., et al. What is a successful digital library? In Gonzalo, J., Costantino, T., & Verdejo, M. F. et al. (Eds). Proceedings of the 10th European Conference on Research and Advanced Technology for Digital Libraries. Alicante, Spain: Springer, 2006:208–219.
15 Huo, L. P. & Zeng, Y. The construction of digital libraries evaluation indicators system. Library and Information Service (in Chinese), 2002(4):19–22.
16 Wang, H. J. Exploring the digital libraries evaluation indicators system. Library Work and Research (in Chinese), 2004(1):13–17.
17 Liu, W., Lou, X. Y., & Zhang, C. J. Study on digital libraries evaluation. Library and Information Service (in Chinese), 2007(5):21–24.
18 Wu, J. H. Digital libraries evaluation approach (in Chinese). Beijing: Science Press, 2009:204–220.

(*Copy editor*: *Ning LI; Language revison: Charles C. YEN*)