

Literature review on NDLTD and the necessity for an advanced research workbench

Dean Williams
WLLDEA003

June 13, 2021

Abstract

In this review we discuss the NDLTD and how its global search can be improved upon by an advanced research workbench of tools. We first outline what NDLTD is as well as its purposes, goals, benefits, usage internationally and its architecture. We also address the problems with NDLTD and digital libraries in general. We address common user requirements of DLs and how they are applicable to the NDLTD. We then discuss potential tools to help address user requirements as well as make NDLTD's global search a much easier and more effective experience. We discuss the benefits of these tools and how they could be implemented.

Keywords: NDLTD, Workbench, Digital Library, ETD, Research

1 Introduction

Digital libraries (DLs) are a direct response to the ever increasing need to manage the vast quantities of electronic data we produce, collect, and consume [10]. The Networked Digital Library of Theses and Dissertations (NDLTD) is a collaborative effort of universities all around the world to promote creating, archiving, distributing and accessing Electronic Theses and Dissertations (ETDs) [10]. It provides users with access to a grand collection of ETDs, spanning over a hundred universities [7]. While the benefits of access to such a large collection are numerous, there is lots of room for improvement. This literature review hopes to shed light on the current state of NDLTD (as well as other digital libraries) and how it can be improved upon by the implementation of an advanced research workbench as a higher-order service to extend the NDLTD global search.

2 Networked Digital Library of Theses and Dissertations

Theses and dissertations are particularly useful to researchers, however most of them languish in obscurity in university libraries. The best way to make use of

theses and dissertations is through electronic publication and online access [1]. The Networked Digital Library of Theses and Dissertations (NDLTD) provides online access to collections of numerous electronic papers all over the world. These collections are, however, managed as independent projects and as such are very loosely linked [2].

2.1 Purposes and goals

The NDLTD was created with the purpose of aiding graduate education.

Its main goals are to broaden access to student research worldwide, make it easier for universities to host a digital library of their own, and allow students worldwide much freer access to information [7].

2.2 National Projects

Several countries are actively taking (or have taken) part in projects involving the NDLTD. Some of the largest contributors include: the USA, Germany, Australia, India and Brazil [7]. Others include China and Spain.

- **USA:** In 1996, the US Department of Education awarded a grant to Virginia Tech which resulted in the creation of NDLTD [7].
- **Australia:** The Australian Digital Theses Project has standardized on SGML and PDF as document formats [7].
- **India:** The Vidyanidhi project, sponsored by India's National Information System for Science and Technology (NISSAT), is a national effort to create and maintain a digital library of Indian theses [7].
- **Brazil:** Biblioteca Digital de Teses e Dissertações (BDTD) is collaborative effort to provide national access to ETDs in Brazil [13].
- **China:** To improve local accessibility, the China Networked Digital Library of Theses and Dissertations (CNDLTD) project has been carried out by CALIS (China Academic Library and Information System), a federation of academic libraries with over 152 members [1].

2.3 Metadata

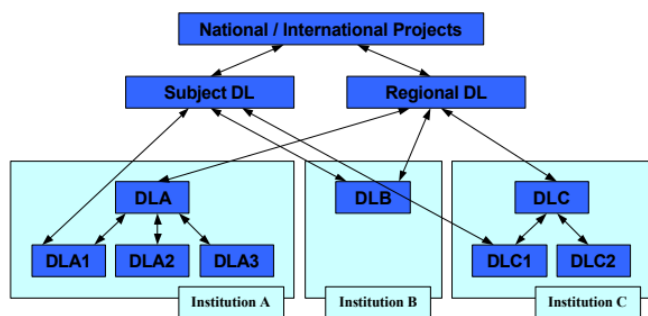
Electronic theses or dissertations (ETDs) are typically described using a metadata standard such as ETD-MS, which is based on the Dublin Core standard. This standard was developed by the NDLTD Standards Committee after years of international discussion [7]. This metadata standard is widely used across the world (such as in the CNDLTD [1]) and as such allows for interoperability between DLs.

While the Dublin Core is seldom the best choice for metadata for any given repository, its generality makes it suitable for interoperability[2].

2.4 Architecture

DLs are based on the client-server model where the data and metadata are stored on one or more centrally-located servers, which are typically accessed by Web clients [9]. A DL also requires an OS alongside software to manage data and user requests. Digital libraries often make use of RAID technology to safeguard data and are regularly backed up, preferably at offsite locations [9].

NDLTD follows the typical networked digital library (NDL) standard for system architecture (shown below).



Generalised architecture of an NDL [9].

As the scale of a DL network increases, remote searching becomes less viable as a basic interoperability mechanism [10]. A solution is to use metadata harvesting to create central collections of metadata, then provide remote search interfaces on these collections.

2.5 Open Archives Initiative

The Open Archives Initiative (OAI) is dedicated to solving problems of digital library interoperability [2].

In most cases ETDs can only be located by an end-user if the archive containing them is consulted directly. Ideally, one should be able to access all ETDs through a unified interface. The Open Archives Initiative's Protocol for Metadata Harvesting (OAI-PMH) is a HTTP

based client-server protocol that provides an application-independent interoperability framework based on metadata harvesting [12]. NDLTD has adopted the OAI-PMH to allow for accumulation of metadata from all members into a single unified collection [7]. All members are encouraged to support the protocol to enable access to their ETD metadata.

2.5.1 Harvesting and federation

NDLTD has made use of two methods to achieve basic interoperability among repositories: Harvesting and federation[2].

- **Harvesting:** Harvesting involves collection of metadata from remote repositories to be stored locally and then performing searches on the local copy of the metadata [2].
- **Federation:** Federation is a more expensive mode of operation in terms of network and system constraints. The DL sends the search criteria to multiple remote repositories and the results are gathered, combined, and presented to the user [2].

Initially when numbers were small, NDLTD used the federation technique, however as membership expanded NDLTD shifted to using the harvesting method, making use of the OAI-PMH [10].

2.6 Problems with NDLTD and DLs

There is still a considerable difficulty when trying to access information from DLs and the NDLTD. Sweetnam et. al. states that DLs are typically monolithic and can be difficult to navigate [3]. Existing platforms for DLs tend either to provide basic tools for interacting with a range of collections, or more complex tools that are tied to a very specific type of collection [3].

Richardson et. al. states that the vast number of ETDs available make it difficult for users to determine which dissertations (or sections thereof) to read [4].

DLs (such as NDLTD) only provide core functionalities, such as search, retrieval, and access to information objects, which are often not able to meet application specific requirements [6]. In a study by Theng et. al., less than 50% of users found that DLs helped them achieve their goals quicker than searching traditional libraries [5]. They also found that 46% of users thought that DLs should provide better and clearer displays of results.

Often, problems with finding important research material arise from the nature of the DL's search engine. Most DLs require users to have specific authors or titles in mind when they search. Buchanan et. al found that when information seeking moved from a strongly-defined goal into more uncertain areas, problems rapidly emerged [8]. Precise searches required careful selection of search

criteria. This careful selection was only observed in a few users [8].

3 User requirements of Digital Libraries

Requirements vary based on the type of user. In particular, differences in the level of knowledge/experience with search functionality, and information/communication technologies in general, have proven to be factors that influence users' needs, and the acceptance of a system [3].

According to a study done on user requirements for DLs by Sweenam et. al. [3], common user requirements among all users include:

- 1) The need for a reliable full text search.
- 2) The ability to bookmark search and filter configurations so that they may be easily used again later.
- 3) Visualisations of collections to allow for easier and useful interaction. Map based visualisations were identified as especially useful.

3.1 Log of interactions

Users often found themselves in a position where they were not able to recover a previously noticed fact or idea [3], as such the importance of a log of interactions within the DL environment was stressed. This log of interactions can be used as means of retrieving important search criteria entered by the user and even allow for automatic bookmarking.

3.2 Annotations and comments

Professional researchers and apprentice investigators both highlighted the importance of being able to add in-line annotation to collection items [3]. Additionally users expressed the importance of highlighting and annotating chunks of text as well as sections of images. Being able to link annotations together is another requested feature [3].

3.3 Sharing and Collaboration

Professional researchers stressed the importance of sharing specific subsets of their annotations with students [3]. Additionally, apprentice investigators expressed the benefit of a feature that would allow them to "follow" more senior academics [3].

3.4 Skeuomorphism

A skeuomorph is a derivative object that retains some necessary ornamental designs from the original object. In many cases, users suggest that a DL should have features

resembling a traditional library [3, 5, 11], with features such as:

- a **librarian** who helps students find necessary information [5].
- a **virtualised physical collection** that incorporates traditional library methods such as storing relevant information in close proximity. Being able to see all papers related to the same subject [11].
- papers should be **categorised** based on their subject [11] and stored/represented based on their category.

4 Discussion of potential Tools

There are very few digital libraries that have both extensive collections and effective services. Often users do not easily find the resources related to their particular information need [2]. According to "What do users want?" by Kani-Zabihi et. al., a user-centred development approach should be adopted when designing tools for DLs.

4.1 Recommendations Tool

Such a tool could provide recommendations for individual users to suggest relevant papers and information based on other users activity [7]. Another alternative is profile based filtering, which stores research interests of individual users and then checks for relevance among newly submitted ETDs [7].

The usage of a log of interactions alongside this recommendations tool could prove especially useful in rendering DLs (like ND LTD) more transparent [3] and provide useful aid with finding important resources.

4.2 Profile Based Filtering

Users indicate a set of interests and then all objects corresponding to those interests would be presented to them on a continuous basis. This mode of operation is perfectly suited to the OAI protocol because of the inherently incremental nature of harvesting [2].

4.3 Cross-linking Tool

Citations can be extracted from ETDs to generate a mapping between papers that are linked [7]. A visual representation of linked documents could provide a much easier to understand and navigate interface than the current methods available.

4.4 Collection Visualisation Tool

In the study by Sweetnam et. al., the importance of a visualisation for collections was expressed by many users [3]. According to Richardson et. al., between 52% and 96% of learners are visual learners [4]. Thus to allow

for better comprehension for the vast majority of users, a visualisation tool is necessary. Such a tool could take the form of a concept map [4] or a “Family-Tree” like structure [3]. Other methods for visualisation should be explored.

Users showed a preference for concept maps generated for each part of a document as opposed to one generated using the entire document [4]. Richardson et. al. [4] showed that visualisation (in the form of concept maps) can be particularly useful at overcoming the language barrier. They conducted a cross language experiment with a university from the USA and a Spanish speaking university from Mexico and found that concept maps greatly aided with helping users to determine if a document is relevant to the information they need. Such concept maps were generated automatically and proved highly successful [4]. However, they took significant computation time (10-20 minutes).

Another visualisation that can be implemented is the visualisation of query results. Agosti et. al. [14] suggests that presenting search results in a more visual format should be more appealing and stimulating for novice users who access the system.

Since the majority of users accessing NDLTD are novice users, this visualisation might prove extremely useful.

4.5 Search and Filter Configuration Bookmarking Tool

As discussed in the user requirements, many users found that they were often unable to retrieve complex and specific search/filter configurations that they were previously using to find information [3]. A simple solution is to allow users to save their search/filter configurations via a bookmarking tool. Users should be able to store multiple configurations and label them in different categories. Search and filter configurations should also be accessible via a log of interactions, so that information may be easily retrieved upon accidental closure or exiting of the NDLTD page as well as accidental overwriting.

4.6 Annotations Tool

A tool to allow users to annotate objects in a DL and share these annotations is a much requested feature [3, 15], one which would prove quite useful for the NDLTD. Adding such a service would likely require the construction of a separate annotation database [2]. In leveraging the OAI protocol, such a separate database could itself be an OA— then any entry in the OA of annotations would refer back to records in other existing OAs [2]. Metadata would then be generated using the source OA and the annotation OA.

In the study by Sweetnam et. al. [3], users stressed the importance of being able to export annotations.

The effective use of a DL requires the possibility of adding researcher’s comments [15]. The ability to access different researchers’ comments and annotations can be done seamlessly through the usage of a built in social media tool. This would allow users to access comments and annotations of researchers they follow or have the ability to share comments with other users.

4.7 Built in Social Tool

As stated in the user requirements, many users feel that a way to socialise with other scholars is an essential feature that should be added to most DLs [3]. Such a feature would be relatively simple to implement and could draw inspiration from existing social medias like Twitter or Instagram. Allowing users to create profiles, follow other profiles and share ETDs with other users would be among the many features such a tool would provide. Currently, the academic search engine “Google Scholar” allows users to create public profiles [16]. However, this feature does not allow for social interaction between users.

5 Conclusions

Although the NDLTD has many benefits and has provided users with a massive access to information, it has its pitfalls. These pitfalls include difficulty accessing specific information due to the overwhelming number of ETDs available, a lack of sufficient collection visualisation, a lack of social interaction between scholars/users, the lack of a log of interactions as well as a means of saving configured filter information. These problems can be solved by implementing an advanced research workbench to extend the NDLTD global search. This workbench would provide useful tools such as a recommendations tool, profile based filtering, a cross-linking tool, a collection visualisation tool, a log of interactions alongside a bookmarking tool and a built in social media tool to help with sharing information. Providing any of these tools would greatly aid in the usability of the NDLTD global search and hence aid the spread and access of information across the world.

References

- [1] Jin, Y., 2004. The development of the China Networked Digital Library of Theses and Dissertations. *Online Information Review*, 28(5), pp.367-370.
- [2] Suleman, Hussein Fox, Edward. (2008). The Open Archives Initiative. *Journal of Library Administration*. 35. 125-145. 10.1300/J111v35n01_08.
- [3] Sweetnam M.S. et al. (2012) User Needs for Enhanced Engagement with Cultural Heritage Collections. In: Zaphiris P., Buchanan G., Rasmussen

- E., Loizides F. (eds) *Theory and Practice of Digital Libraries*. TPD L 2012. Lecture Notes in Computer Science, vol 7489. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-33290-6_8
- [4] Richardson, W.R., Srinivasan, V. Fox, E.A. Knowledge discovery in digital libraries of electronic theses and dissertations: an NDLTD case study. *Int J Digit Libr* 9, 163–171 (2008). <https://doi.org/10.1007/s00799-008-0046-9>
- [5] Yin Leng Theng, Norliza Mohd-Nasir, and Harold Thimbleby. 2000. Purpose and usability of digital libraries. In *Proceedings of the fifth ACM conference on Digital libraries (DL '00)*. Association for Computing Machinery, New York, NY, USA, 238–239. DOI:<https://doi.org/10.1145/336597.336674>
- [6] Castelli D., Pagano P. (2002) OpenDLib: A Digital Library Service System. In: Agosti M., Thanos C. (eds) *Research and Advanced Technology for Digital Libraries*. ECDL 2002. Lecture Notes in Computer Science, vol 2458. Springer, Berlin, Heidelberg. https://doi.org/10.1007/3-540-45747-X_22
- [7] Fox, E.A., McMillan, G., Suleman, H., Gonçalves, M., Luo, M.,(2004). *Networked Digital Library of Theses and Dissertations(NDLTD)*. In J. Andrews D. Law (Eds.), *Digital Libraries: Policy, Planning and Practice* (pp.M167T187). Hants, England: Ashgate.
- [8] Buchanan G., Cunningham S.J., Blandford A., Rimmer J., Warwick C. (2005) *Information Seeking by Humanities Scholars*. In: Rauber A., Christodoulakis S., Tjoa A.M. (eds) *Research and Advanced Technology for Digital Libraries*. ECDL 2005. Lecture Notes in Computer Science, vol 3652. Springer, Berlin, Heidelberg. https://doi.org/10.1007/11551362_20
- [9] Suleman, Hussein and Fox, Edward A. and Madalli, Devika (2003) *Design and Implementation of Networked Digital Libraries: Best Practices*, Proceedings of DRTC Workshop on Digital Libraries: Theory and Practice, March 2003, Bangalore, India.
- [10] Fox, Edward A. and Suleman, Hussein and Gaur, Ramesh C. and Madalli, Devika P. (2005) *Design Architecture: An Introduction and Overview, Design and Usability of Digital Libraries: Case Studies in the Asia Pacific*, 22-37, Information Science Publishing.
- [11] Kani-Zabihi, E., Ghinea, G. and Chen, S.Y. (2006), "Digital libraries: what do users want?", *Online Information Review*, Vol. 30 No. 4, pp. 395-412. <https://doi.org/10.1108/14684520610686292>
- [12] Webley, Lawrence and Chipeperekwa, Tatenda and Suleman, Hussein (2011) *Creating a National Electronic Thesis and Dissertation Portal in South Africa*, Proceedings of 14th International Symposium on Electronic Theses and Dissertations, 13-17 September 2011, Cape Town, South Africa.
- [13] Southwick, Sílvia. (2006). The Brazilian electronic theses and dissertations digital library: Providing open access for scholarly information. *Ciência da Informação*. 35. 10.1590/S0100-19652006000200011.
- [14] Maristella Agosti, Franco Crivellari, Giorgio Maria Di Nunzio, and Silvia Gabrielli. 2010. Understanding user requirements and preferences for a digital library Web portal. *Int. J. Digit. Libr.* 11, 4 (December 2010), 225–238. DOI:<https://doi.org/10.1007/s00799-011-0075-7>
- [15] Agosti, Maristella Orio, Nicola. (2012). User Requirements for Effective Access to Digital Archives of Manuscripts. *Journal of Multimedia*. 7. 10.4304/jmm.7.2.217-222.
- [16] Google Scholar About Page: <https://scholar.google.com/intl/en/scholar/citations.html>