Evaluating User Needs for a Research Workbench

Hugh Bedford
bdfhug001@myuct.ac.za
University of Cape Town
Rondebosch, Cape Town, South Africa

ABSTRACT

There is an abundance of information and research available online, which is ever increasing. Digital libraries provide a central portal to access different types of information but often lack the services to digest and analyse the information effectively. There is a growing demand for enhanced services that allow researchers to perform their tasks more efficiently. After considering academic researchers, specific user needs and tools were identified that would form part of a researcher workbench.

KEYWORDS

user requirements, digital libraries, NDLTD, research workbench

1 INTRODUCTION

The invention and innovation of the internet over the past 25 years has lead to the digitisation of publishing and the proliferation of content online [10]. There is a growing amount of research and evidence that researchers are bogged down by the research that is available [4]. It was and still is becoming increasingly more challenging for researchers to grasp the conclusions of literature [4]. There is a need for a tool to digest and manage the vast majority of online information that is now present [7]. Digital libraries, basically store materials in electronic format and manipulate extensive collections of those materials effectively, provide a method to address this need [14]. There are now a large number of digital libraries, such as the Internet Archive [1] or Open Library [2], which are freely accessible online. This paper focuses on digital libraries focusing on academic literature, notably the Networked Digital Library of Theses and Dissertations (NDTLD). The NDLTD has a mission of promoting adoption, creation, use, dissemination and preservation of electronic theses and dissertations (ETDs) [11].

This review aims to investigate user needs and requirements of digital libraries and, in particular, the needs of researchers using academic repositories and ETDs services. The methods of reviewing results of the research are old and are not adequate [4]. The paper will first consider user needs and then consider the architecture of digital libraries enhanced services. Finally, the paper will look more closely at NDLTD and academic repositories.

2 USER NEEDS AND REQUIREMENTS OF DIGITAL LIBRARIES

Several studies have investigated and researched user requirements and preferences of digital libraries to understand how they are used and how they should be designed to meet the needs of the target user best. Lombardi explained that helping users find resources effectively and efficiently online with several other disorganised

---

[1] www.archive.org
[2] www.openlibrary.org

resources is the main priority of an academic library [21]. When designing and developing a digital library, it is vital to take a user-centred approach since the product needs to directly serve the needs of its users [9]. It was generally found to be essential to understand user needs and preferences to ensure a digital library web portal was of a sufficient level of quality and there was interoperability of the services provided [1]. Both Agosti and Orio and Agosti et al. mentioned adapting to evolving user needs and their needs changed after acquiring new knowledge [1, 2].

The rest of this section aims to categorise similar user requirements mentioned in the literature. The categories are each critical features that could enhance the Networked Digital Library of Theses ad Dissertations [6].

2.1 User Support

The literature presented several examples of users not making use of all existing features on digital libraries unless they were experienced users [2, 3, 9]. Agosti and Orio studied user requirements for effective access to digital archives of manuscripts [2]. They found that to improve interaction with the service; contextual aid should be provided to the user [2]. If the user is less experienced, they will need support to select the appropriate criteria for a more effective search [3]. When users have a strongly defined search, the search is generally effective; however, when they move to more uncertain searches, the search becomes less effective [3]. Digital libraries should add support structures to the service to aid users, which will allow them to gain the maximum benefit [2, 3, 19].

Although there should be user support, a digital library should be easy to learn [9, 19]. An easy to use and learn digital library will encourage user interaction and continued use.

2.2 Resource Recommendations

The literature presents evidence that there is a user need for digital libraries and Electronic Theses and Dissertations (ETDs) to recommend other literature that may be of interest to the user based on what the user to looking for. Agosti and Orio found that professional researchers wanted to consider resources from other online collections or similar resources within the same collection [2]. Researchers often need to see most of or all the resources from the same subject area and are not able to [9]. Users also wanted to be able to list important resources [9]. This, however, presents challenges because the digital library will have to rank resources on levels of importance, which brings in the issue of what criteria to use. Instead, users should be able to see frequently accessed resources and be informed of new material that has been added so they can see the most up-to-date material available [9]. The vast number of ETDs available on NDTLD make it very hard for researchers to determine which to read [13]. Recommendation features like the

ones mentioned above could make it easier for researchers to find potentially relevant papers.

Overall, the user interface should make clear to users the different choices available for resources that meet their criteria [1].

2.3 User Collaboration

There is an increasing demand for users and researchers to be able to collaborate in some form. Kani-Zabihi, Ghinea and Chen said that digital libraries currently have a lack of support for users to work and collaborate [9]. There is a need for different collaboration tools that make it easier for researchers to share ideas and work together.

Users found that being able to add comments to resources and see feedback from other users would make digital libraries a more effective research tool [2, 9]. The ability to add comments and in-line annotations on resources may increase the involvement of non-domain users and would help new researchers to engage with the content more effectively [2, 19]. Researchers wanted to have the ability to collaborate with other researchers on a specific topic [19].

2.4 Visualisation of Results

Visualisation of results and how the results are presented are important for users' engagement, especially new users of the digital library. If query results are presented in a clear and more visual format, it might appeal to and stimulate novice, and non-domain users of the digital library [1]. Not only would it be appealing to novice users, but a more straightforward presentation of results would also make digital libraries more efficient for all users overall [20]. Sweetnam et al. also found that there was a preference amongst the studied users for a visual-based interface for finding information [19].

Digital libraries can also use visualisations for summary techniques of ETDs, not just a visualisation of query results. Concept maps are a highly effective summary technique that can be used to promote a visual summary of ETDs, which would help promote learning and understanding [13]. The concept map could be used as a tool to help researchers determine whether a resource is relevant to the research being conducted or not [13].

2.5 Search Tools

Finally, users need tools that aid the search of resources and understand and find text easily. First and foremost, a search must return reliable results [9, 19]. This is the most important part of a digital library, but there are many ways a digital library can be improved. More powerful search features would allow researchers to be efficient and complete tasks faster [20]. There is a need for a faceted search that allows users to search content across all metadata categories [19]. This would provide users with the ability to search more accurately for the resources they are looking for.

ETDs are written in many languages [13]. On a global library such as NDTLD, there are ETDs written in several different languages, which means there may be a relevant paper. Still, the language barrier prevents researchers from finding suitable material. A research tool should translate ETDs, which would make more information available to researchers using the service [13]. Agosti

et al. found it was important for users to be able to change the user interface of the digital library to their home language when using The European Library (TEL) [1].

Users wanted the ability to create a profile on digital libraries [9]. This would enable users to select their areas of interest, which would mean users would be able to be recommended content based on their interests and commonly searched topics. Having a profile would allow the user to bookmark the results of a search, log their interactions with a particular resource and have to ability to organise their bookmarks for future reference [19].

3 DESIGN AND ARCHITECTURE OF DIGITAL LIBRARIES

The design and architecture of digital libraries are essential to consider when developing high-level services that will interact and rely on the underlying architecture of the digital library system. There is a growing demand for simpler and more manageable tools and services for the management of content, and the underlying architecture makes a lot of it possible [12].

3.1 Overall Design

The design of different digital library systems differs from system to system. Still, they essentially contain a user interface layer, a service layer with necessary services for user interaction and a repository layer that stores and manages the digital resources [12]. There is a need for these separate components since those who provide high-level services were seldom the parties that provide the data or resources [7, 16]. The parties that provide these services to the user are referred to as a service-provider [18]. The service providers operate in the service layer and harvest metadata from data providers, which operate in the repository layer, and use the metadata to provide these services [18].

Interoperability became a means of concern since the format of metadata and protocol to harvest the data varied among data providers, meaning interoperability is now a significant aspect to consider when building a digital library [16]. The Open Access movement arose to address this issue. The Open Access movement made it mandatory for digital libraries adhering to their philosophy to make metadata for their content available via the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) [16, 18]. OAI-PMH is a protocol that operates as a client-server model and makes requests as URL-encoded parameters that are sent over an HTTP transport layer [16]. The response (metadata) is received as XML documents that use the ETDMS standard, which is a derivative of the Dublin Core Metadata Format [16]. Although the Dublin Core Format is not popular to be used within repositories, it makes use of 15 elements to describe almost any resource, which means it is best for interoperability between repositories [16, 18].

The process whereby a digital library collects metadata from remote repositories in XML documents via OAI-PMH is referred to as harvesting[18]. Digital libraries harvest the metadata from remote repositories, store it locally and then perform searches on the local data [18]. Figure 1 is a diagram illustrating how harvesting takes place. NDLTD, which is this paper's focus, uses harvesting to collect its metadata instead of other methods such as federation. NDLTD harvests its data from remote repositories, which are
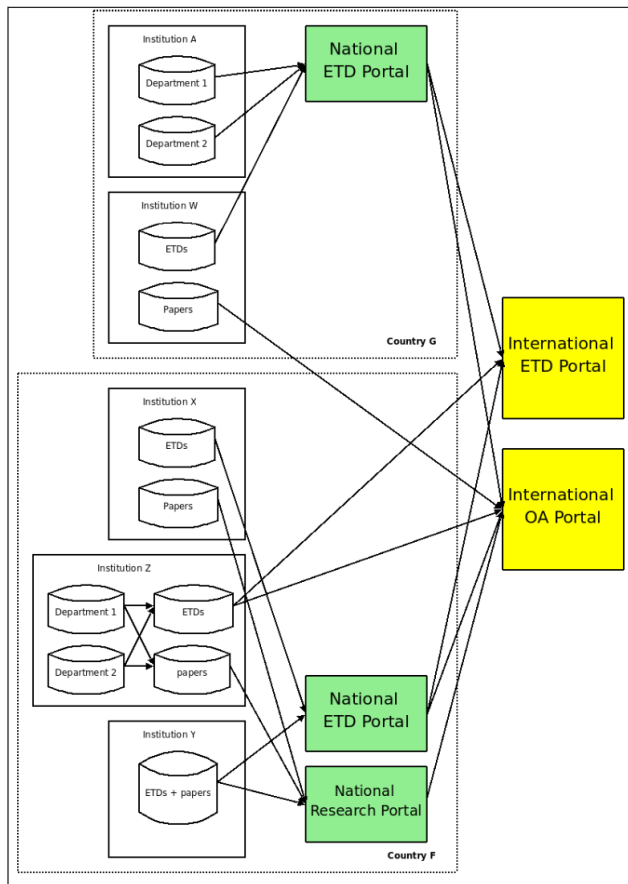
Figure 1: Network of ETDs and Data Providers [16]

Figure 2: Harvesting Metadata from Repositories [18]

generally national ETD portals [16]. This can be seen in Figure 2. Harvesting, seen in Figure 1, is the process that occurs between the national repository and the international repository, as seen in Figure 2 [18]. NDLTD harvests the metadata, which features can then use to provide higher-level services to the user.

3.2 OAI-PMH Support for Possible Digital Library Services

The OAI Protocol is flexible enough, and with the richness of the metadata, developers can implement several services into digital libraries such as NDLTD. Firstly, the protocol would be able to support reference linking [18]. Users would have the functionality of navigating from one electronic resource to another that it references. This would make it easier for the user to navigate to and view electronic resources in similar subject areas. Secondly, it is possible to add annotations [18]. Users could have the ability to add annotations to existing documents, which is mentioned in section 2.3 and is desired functionality amongst digital library users. Another desired feature discussed in section 2.5 that is deemed possible by OAI-PMH is profile-based filtering [18]. Users could have the ability to indicate interests, and then all resources matching
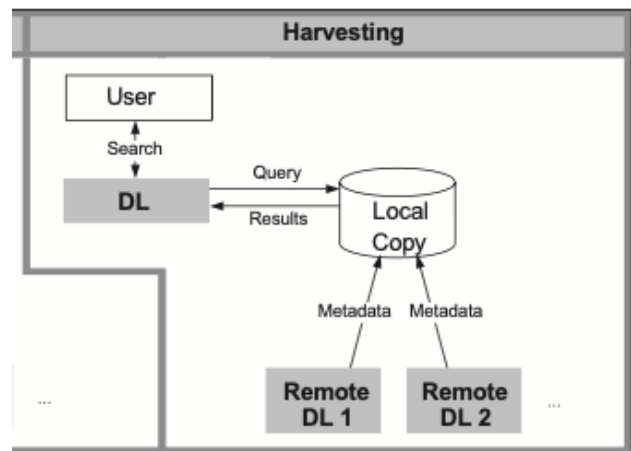
those interests would be presented to the user. Finally, user browsing could be enhanced [18]. Similar resources could be grouped and presented to the user using categorical data available due to the protocol. This, again, is a feature that users of digital libraries would like available, as mentioned in section 2.

4 NDLTD AND INSTITUTIONAL REPOSITORIES

4.1 Use of NDLTD

The Networked Digital Library of Theses and Dissertations (NDLTD) is a collaborative effort of universities around the world to promote creating, archiving, distributing and accessing ETDs, as well as to encourage local advancement and adoption of digital library technologies [7]. At the heart of NDLTD is electronic theses and dissertations and being able to accurately search and browse them from combined collections of NDLTD members is the main service [6, 7]. Users make use of discovery tools to find ETDs across the various collections but, beyond discovery, NDLTD does not provide users with many other interaction services [17]. The service could offer higher quality results of a specific nature, and this could give the users specialised services such as those mentioned in section 2, to navigate and interact with the resources [17].

4.2 User features for NDLTD

There are several requested features for NDLTD and other national NDLTDs worldwide, many of which are similar to the user requirements for digital libraries mentioned in section 2 but focused on NDLTD. After considering literature focused on NDLTDs, the features deemed necessary for NDLTDs include extraction of citations to cross-link papers, user recommendations, personalisation of information access, multi-language support, and finally, full-text search [6, 8]. The internet allows for ETDs to be accessed anywhere in the world, so features that let for queries in several languages and queries in one language and returned documents in another language would be of better use to any user of NDLTD [8].

4.3 Institutional Repositories

Similarly to NDLTDs, these repositories can be complex to search and navigate [19]. Existing platforms and tools for digital archives tend only to provide basic tools for interacting with digital objects. If more complex tools are available, they are generally only available for a more specific type of collection or repository [19]. Software packages that can be used for institutional repositories are often inadequate to manage all the materials in the repository and do not provide services that allow users to interact with the repository effectively [15]. Software packages available include EPrints and DSpace [22].

5 DISCUSSION

It is essential to consider what users need are and how features would fit into the current architecture of Networked Digital Library of Theses and Dissertations to create an effective research tool. It is evident from section 2 that there are many features that users of digital libraries need and require but are not available to them.

There is a lack of services available to researchers, making it hard to research and analyse information effectively. There have been a few attempts at research workbenches, such as iOpener Workbench, which has the aim of allowing researchers to understand and comprehend scientific research considerably faster [5].

In his 1945 essay "As We May Think", Bush explained how our methods of reviewing results of the research are generations old and are inadequate for their purpose [4]. This has largely remained unchanged. There are central portals to access research, such as NDLTD mentioned in Section 4, but a lack of services to do so efficiently and quickly.

After considering the user needs, the architecture of digital libraries, including the protocols used and how the system is composed was analysed to determine how and if such features could be implemented. The component approach to the design of digital libraries, namely NDLTD, allows for a services to be implemented on top of the data with no changes to how the data is gathered. Furthermore, OAI-PMH is powerful enough that it makes most of the features above possible.

After diving deeper into NDLTD itself, it was found that many desired enhanced services overlap with the broader digital library features identified earlier, along with some additional specific features such as citation extraction to cross-link papers.

6 CONCLUSION

To conclude, there are many user requirements for users of digital libraries. Few digital libraries have both extensive resources and powerful services that allow users to interact with those resources effectively [18]. Users don't easily find resources related to their informational needs, and there is a lack of services that could make the process easier [18]. Several users don't even make use of the existing features adequately due to a lack of contextual aid [2].

The architecture of digital libraries allows for enhanced services to be built on top of the repository layer and not affect any of the data, making it easier to develop services without having to re-engineer the whole system. The OAI-PMH makes it possible to implement several features into digital libraries due to the protocol's flexibility.

Upon analysis of the user requirements and then the architecture of digital libraries, which included the NDLTD, it is possible to develop a comprehensive tool that enhances the services that NDLTD offers to meet the majority of, if not all, the needs of researchers that make use of NDLTD.

REFERENCES

[1] 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 (Dec. 2010), 225–238. https://doi.org/10.1007/s00799-011-0075-7

[2] Maristella Agosti and Nicola Orio. 2012. User Requirements for Effective Access to Digital Archives of Manuscripts. *Journal of Multimedia* 7 (04 2012). https://doi.org/10.4304/jmm.7.2.217-222

[3] George Buchanan, Sally Cunningham, Ann Blandford, Jon Rimmer, and Claire Warwick. 2005. Information Seeking by Humanities Scholars. *Research and Advanced Technology for Digital Libraries: 9th European Conference* 3652, 218–229. https://doi.org/10.1007/11551362_20

[4] Vannevar Bush. 1945. As We May Think. *The Atlantic Monthly* 176, 1 (July 1945), 101–108. https://www.theatlantic.com/magazine/archive/1945/07/as-we-may-think/303881/

[5] Cody Dunne, Ben Shneiderman, Bonnie Dorr, and Judith Klavans. 2010. iOpener Workbench: Tools for rapid understanding of scientific literature. In *Human-Computer Interaction Lab 27th Annual Symposium, University of Maryland, College Park, MD*.

[6] Edward Fox, Gail McMillan, Hussein Suleman, Marcos Gonçalves, and Ming Luo. 2004. *Digital Libraries: Policy Planning and Practice.* Ashgate Publishing, Chapter Networked Digital Library of Theses and Dissertations(NDLTD).

[7] Edward Fox, Hussein Suleman, Remesh Gaur, and D Madalli. 2003. Design Architecture: An Introduction and Overview, Design and Usability of Digital Libraries: Case Studies in the Asia Pacific. *Information Science Publishing* (2003), 22–37.

[8] Yi Jin. 2004. The development of the China Networked Digital Library of Theses and Dissertations. *Online Information Review* 28 (10 2004), 367–370. https://doi.org/10.1108/14684520410564299

[9] E Kani-Zabihi, G Ghinea, and S Chen. 2006. Digital libraries: what do users want? Digital libraries: what do users want? Digital libraries: what do users want? Digital libraries: what do users want? Digital Libraries: what do users want? *Online Information Review* 30, 4 (2006), 395–412.

[10] Norbert Lossau. 2004. Search engine technology and digital libraries-libraries need to discover the academic internet. *D-Lib magazine* 6 (2004).

[11] NDLTD. [n.d.]. Networked Digital Library of Theses and Dissertations: Misson, Goals, and History. http://www.ndltd.org/about

[12] Lighton Phiri and Hussein Suleman. 2013. Flexible Design for Simple Digital Library Tools and Services. In *Proceedings of the South African Institute for Computer Scientists and Information Technologists Conference* (East London, South Africa) *(SAICSIT '13)*. Association for Computing Machinery, New York, NY, USA, 160–169. https://doi.org/10.1145/2513456.2513485

[13] Ryan Richardson, Venkat Srinivasan, and Edward Fox. 2008. Knowledge discovery in digital libraries of electronic theses and dissertations: An NDLTD case study. *Int. J. on Digital Libraries* 9 (11 2008), 163–171. https://doi.org/10.1007/s00799-008-0046-9

[14] Michael Seadle and Elke Greifeneder. 2007. Defining a digital library. *Library Hi Tech* 25 (06 2007), 169–173. https://doi.org/10.1108/07378830710754938

[15] Sarah Shreeves and Melissa Cragin. 2008. Introduction: Institutional Repositories: Current State and Future. *Library Trends* 57, 2 (2008), 89–97.

[16] Hussein Suleman. 2011. Interoperability in digital libraries. In *E-Publishing and digital libraries: Legal and organizational issues.* IGI Global, 31–47.

[17] Hussein Suleman. 2012. The NDLTD Union Catalog: Issues at a Global Scale. Universidad Peruana de Ciencias Aplicadas (UPC).

[18] Hussein Suleman and Edward Fox. 2001. The Open Archives Initiative. *Journal of Library Administration* 35, 1-2 (2001), 125–145. https://doi.org/10.1300/J111v35n01_08 arXiv:https://doi.org/10.1300/J111v35n01_08

[19] Mark S. Sweetnam, Maristella Agosti, Nicola Orio, Chiara Ponchia, Christina M. Steiner, Eva-Catherine Hillemann, Micheál Ó Siochrú, and Séamus Lawless. 2012. User Needs for Enhanced Engagement with Cultural Heritage Collections. In *Theory and Practice of Digital Libraries*, Panayiotis Zaphiris, George Buchanan, Edie Rasmussen, and Fernando Loizides (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 64–75.

[20] 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* (San Antonio, Texas, USA) *(DL '00)*. Association for Computing Machinery, 238–239. https://doi.org/10.1145/336597.336674

[21] John V.Lombardi. 2000. Academic Libraries in a Digital Age. *D-Lib Magazine* 6, 10 (2000), 1–7.

[22] Gerard Westrienen and Clifford Lynch. 2005. Academic Insitutional Repositories. *D-Lib Magazine* 11, 9 (2005), 1–6.