Usability Testing of the Virtual Data Center

Jessica Hovater, M. Krot, D. L. Kiskis
University of Michigan
1301 Beal Ave.
Ann Arbor, MI 48109-2122
{jhovater, krot, dlk}@umich.edu

Ilona Holland and Micah Altman
Harvard University
G-4 Littauer Center, North Yard
Cambridge, MA 02138
Micah_Altman@Harvard.edu

ABSTRACT
In this paper we present an overview of the usability issues related to the Virtual Data Center (VDC) digital library. Usability testing and heuristics were an integral part of the development process from the inception of the project and throughout its evolution. The developers of the VDC have used the findings from usability evaluations to identify several prominent usability issues present within the digital library. Test results continue to exert a positive and valuable influence on both system redesign and the implementation of new features.

Keywords
Usability, digital library, open source, distributed systems, World Wide Web

INTRODUCTION
The Virtual Data Center (VDC) is an open-source, web-based digital library for the management and dissemination of social science research data. It is being developed as a collaborative effort between researchers at the Harvard-MIT Data Center (HMDC) at Harvard University and the University of Michigan's School of Information and College of Engineering (<http://TheData.ORG>). The VDC is designed to facilitate the retrieval and dissemination of quantitative social science research data. It provides support for distributed and federated collections of data through features such as distributed security and federated metadata harvesting [1]. It also provides tools for numerical and graphical data analysis. The VDC uses XML metadata specificiations created by the Data Documentation Initiative (DDI), which has its roots at Inter-university Consortium for Political and Social Research (<http://www.icpsr.umich.edu>). By using this widely accepted specification, the VDC will allow libraries and data centers around the world to share collections of data with one another.

The usability issues present with the VDC are three-fold. First, the VDC uses all open-source software (OSS) and is itself open source. The maintainers and installers of OSS have a notorious reputation for presupposing that users have high levels of technical expertise [2]. There are few examples of usability techniques being applied to open source projects, with the Gnome Usability Project being one notable exception [3]. Second, the VDC is web-based, carrying with it a number of usability constraints not faced by non-web applications. Finally, the VDC is a digital library. With more and more students and researchers turning to online resources rather than physical libraries to conduct their work, the importance of usable digital libraries has risen significantly. Digital libraries, if presented in a usable way, can greatly enhance the productivity of researchers by making a much larger amount of data accessible and by providing extended functionality such as data analysis and download. Yet this greater power adds complexity to user interfaces, and best practices in this area are still in early development.

The goal of the VDC is increased data access and research efficiency, and the development team has put a renewed focus on the usability testing and enhancement of the system. We conducted user-needs analyses using focus groups and surveys before commencing the project. After developing the software to a sufficient level to meet these needs, the development team once again returned to usability studies to assess the progress of the VDC in a deliberate attempt to improve the usability of the system for the next release. Areas of particular focus include improved documentation, navigability, metadata presentation, and searching functionality.

INITIAL USABILITY WORK ON THE VDC
In the year 2000, the VDC development team at Harvard University conducted several tests to gain preliminary feedback on the usability of the HMDC website, a precursor of the VDC. While the workflow, technology, and processes were quite different than the (then) proposed VDC, the studies were important for gathering information about how users conducted research and how the VDC might solve existing problems.

Focus Groups
In January 2000, the team conducted focus groups with students from a digital library design class at Harvard. The overarching theme that could be derived from results of these focus groups was that the cognitive time necessary to

complete a task was orders of magnitude greater than the computation time necessary. This was illustrated by giving the same task to an experienced research assistant and to students from the focus group, considered to be less-experienced, "average" users. The task that was assigned was to find the percentage of women who had voted for Clinton for President in 1996. The research assistant found the answer from the system in less than five minutes by finding the data in the related study and running a cross-tabulation analysis in the system. However, only one student from the focus group was able to find the answer, and it took this student several hours to do so.

Findings

Students from the focus group cited several usability problems as being the reason why this simple task was so difficult to complete. First, the language in the system was specialized to people who had experience with this type of research data, making it difficult for the average user to recognize what they were looking for. Second, the typology of resources available through the system was unfamiliar to the users. Users were not familiar with fundamental differences between studies collected using different research methods. For example, the data that provided the answer to the question was most likely to be found in survey data but not in census data. Without sufficient knowledge of the collection and what existed there, the users were not able to make educated and efficient choices about where to search for the necessary data. Third, incomplete documentation kept the users from learning all that was available on the system. Since information about cross tabulations was available only in a long "Frequently Asked Questions" document, many users did not realize that they could run such analyses on the system. Due to these usability burdens, the task was very difficult for the users to complete even though the actual process for finding the answer was not time-consuming.

User Surveys

In November 2000, the HMDC, jointly with the Harvard Graduate School of Education, conducted extensive user studies evaluating the existing version of the HMDC website. The goals of these studies were to identify:

- Current patterns of use among those seeking data
- Qualities of other sites that users find helpful
- Traits of users seeking data
- Usability issues within the (then) existing website

For these studies, team members conducted hour-long interviews and questionnaires of 53 advanced researchers, professors, and graduate students with online research experience.

Findings

The team members conducting the surveys reported three main findings that resulted from these studies:

- Users are increasingly conducting data research via the web instead of from the physical library. In fact, most of those surveyed conducted research for data primarily via the web.
- In digital libraries, users want less navigation in favor of a more direct route to the data.
- Users want raw data instead of an author's interpretation of that data.

This study showed the increasing importance of digital libraries as more researchers turn to online resources to conduct their research. These online libraries provide ubiquitous access to data and can also provide advanced support for data analysis. Interestingly, the subjects in these studies who listed the library as their primary resource rather than online resources stated that the availability of reference help was the main reason that they continued to do research at a library. This result underscores the importance of developing usable digital libraries if they are to truly enhance research and data accessibility. The digital library must, in essence, do the job of not only the library, but of the reference librarian, as well.

A minimalist approach to usable interfaces seemed to be preferred by the users, who decried the use of frames and extensive navigation and insisted on obvious and direct routes to the data. This means making the collections and the searching mechanisms visible and easily accessible to the user.

An advantage of digital libraries is the ability to display data to the user in a number of useful formats. Users of these studies appreciated graphs, dynamic charts, descriptive tables, and written summaries. However, the format that was preferred the most was raw data, accompanied by bibliographic information documenting how that data had been used in published research. Most users preferred to conduct detailed analysis with their own tools, and wanted the library to provide sufficient online analysis functionality to explore the data and convert it to a form that was usable by standard statistical packages.

CURRENT USABILTY WORK WITH THE VDC

In January of 2002, we began a second iteration of usability testing to evaluate the current VDC. First, we conducted cognitive walkthroughs so as to thoroughly assess the usability of the system ourselves, before presenting the system to test users. Several prominent usability issues in relation to the functionality and navigability of the system were identified by these evaluations, and the VDC development team made a round of changes and enhancements. Several months later, with a stable and enhanced VDC in place, we conducted usability tests with test users.

Cognitive Walkthroughs

Our goal in conducting cognitive walkthroughs was to evaluate the system by methodically walking through the system as an actual user would. At each step in a process, we considered a series of questions to analyze the usability

at that specific point in the process. We questioned whether an actual user would know what to do at each step, and whether they would be presented with appropriate feedback [4]. We evaluated basic user tasks such as logging into the system, searching for a study, and performing extractions on data.

Findings

After completing the walkthroughs, several major and minor usability issues were evident: lack of documentation, unfamiliar language, and inefficient search functionality. Some of the language used in the VDC is specific to either the VDC or to digital libraries in general (e.g. "collection" and "permissions") and could be foreign to a user. Likewise, some features supported by the VDC were confusing and lacked explanatory documentation. For these reasons, we noted a definite need for help documentation to define some of the language in the system and to instruct the user on how to complete certain tasks. As a result of this recommendation, the VDC team developed context-specific help documentation as recommended by Nielsen [5] and others. Hyperlinks embedded in each page are linked to the help topics specific to that page. Through this mechanism the developers attempted to first provide the most relevant help documentation to the user, while also providing access to comprehensive documentation. In addition to the help system, we re-evaluated the documentation and directions, re-structured interfaces, and added concise in-line help to several interfaces.

The search interface reflected several usability issues, and in one case required the removal of a feature. Initially, the VDC supported a way for users to browse studies by author name. Although this was useful in theory, walkthroughs determined that inadequate adherence to metadata standards by the metadata content providers seriously handicapped the ability of the VDC to provide this feature in a useful manner. The feature was subsequently disabled, although it remains available for small collections with high-quality metadata.

We felt that the search form did not match users expectations based on other search engines. The search functionality within a digital library is crucial to the usefulness of the library and to the productive use of the library by researchers. Therefore, we stressed the importance of powerful and useful search features. As a result of these recommendations, the VDC development team implemented a "quick search" feature that works similar to Internet search engines such as Google™, which allow users to type keywords into a single text box to perform a search across several fields. This feature facilitates quick searching for users who may have more general search criteria in mind. The existing search form then became an "advanced search". Users who know more information about studies for which they are searching may specify search terms that exist in the title or the author list, for example.

User Testing

Several months after the cognitive walkthroughs and subsequent modifications to the VDC, we conducted formal usability tests with five graduate students from the School of Information at the University of Michigan. The purpose of these tests was to expose the VDC to novices who typified potential users of the system, and to evaluate their success with using the system. The tests were conducted at the University of Michigan Media Union Usability Laboratory, which facilitated the recording of the session by capturing the screen onto videotape along with the test user's voice. Test users were given a list of tasks to attempt and were instructed to work through the tasks as thoroughly as possible. While they worked, they verbalized their thoughts and explained their reasons for acting upon the system as they did. We evaluated the test results by noting where test users became confused, what common mistakes they made, and what routes through the system they took while attempting to complete each task. These tests were immensely helpful because they produced empirical evidence of how typical users will interact with the system and what usability issues may cause problems for the user.

Findings

Our first finding was that the documentation still needed improvement. Most of the test users had used other digital libraries regularly but were still unfamiliar with some of the terms used in the VDC. Almost all of the test users were confused by the concept of a "collection". They asked what the difference was between the collections and were not sure in which collection they should search for a study. Although there was some documentation about collections, the users noted that this documentation was vague and did not help them to understand the concept of collections. We believe this is, in part, because the collection functionality has not yet been fully utilized in the production system. These findings are consistent with the findings from initial focus groups and demonstrate that the VDC has to do even more to help users and curators conceptualize collections.

Although most of the users had experience in data analysis, they still had many questions about the process of extractions and performing cross tabulations in the VDC. They felt that the minimal explanation on the pages was not sufficient for them to feel confident performing data analysis. Some users suggested providing different levels of documentation for different levels of user proficiency. For example, this would allow experienced researchers to conduct powerful data analyses, while still supporting beginners through the process of conducting analyses.

Several test users observed that the presentation of metadata about a study could be improved so as to facilitate better searching and browsing of studies. After searching

for a study, the users wanted to know more about the study than the basic metadata shown on the results page, before committing to downloading the data. Users said that they would be interested in having the ability to browse additional metadata, and particularly to see information such as dates when the study was submitted to the VDC, the location where the study took place, or more information about the subjects in the study. One solution is to allow users to control the level of detail presented and the presentation format.

The users noted that the search functionality could be improved further. They tried to perform more powerful keyword searches than the system supported and were expecting the search to be more like other search engines with which they are familiar. We have since added a query-builder that supports powerful search functionality such as Boolean operators, phrase searching, and stop-words.

The majority of the users found that the search view that allowed them to browse subject headings was the most useful. Unless they were trying to search for a study by a particular name, they found it to be more helpful to be able to browse a list of subjects, and then to search for studies under that subject.

FUTURE WORK

The VDC development team continues to enhance the documentation within the system, including both documentation within the pages and the separate help pages. The team is creating a glossary of terms to supplement the help documentation. We expect this to be particularly valuable for the terms used in data analysis and manipulation. For longer processes, such as performing data analysis, users may benefit from a more systematic, step-by-step description of the process within the page. Furthermore, it might be useful to provide optional links within the pages that would direct beginners to more basic and detailed documentation. The team has also implemented a correspondence form within the system to allow the user to communicate questions or problems to a site administrator.

The development team continues to create more clear and direct navigation. The system now includes a history bar that leaves a trail of all of the pages that the user has visited to get to their current location, allowing the user to backtrack as necessary. There is also a "Recently Viewed Studies" bar, which many of the test users cited as being very helpful. In this bar all of the studies that the user has looked at in the system remain visible and accessible. These and other navigation tools have greatly improved the navigational ability of the user within the system.

The VDC development team improved the searching and browsing capabilities in the system. We have implemented more powerful searches, and are adding integration with topic thesauri. More browsing capabilities within collections are being explored, including keyword browsing, since this is a feature that the test users overwhelmingly appreciated and found useful. And, in response to user demand, we are implementing cross-collection and cross-federation searches.

In response to users' requests to see more about the studies, we have developed alternative (e.g. 'printer-friendly') layout options, and are developing a mechanism whereby users can also control the amount of detail presented in various descriptions of studies and results sets. Furthermore, we are developing heuristics for enhancing the metadata, which would both enhance presentation and make searching and browsing more effective.

Through all of these tests on the usability of the VDC, we have drawn a distinct list of crucial usability features. Overall, test users have expressed interest and enthusiasm for the VDC. The VDC enables valuable searching and data analysis activities. However, in order to really harness the full potential of a digital library such as the VDC, we must continue to work with real users, and real content, and to continue to adapt and enhance the interface based upon real usage.

ACKNOWLEDGEMENTS

Supported by NSF Award #IIS-9874747.

REFERENCES

1. Altman, M., et. al. Overview of The Virtual Data Center Project and Software. *JCDL '01: First Joint Conference on Digital Libraries*

2. Nichols, David M., et al. "Usability and Open-Source Software Development" in *Proceedings of the Symposium on Computer Human Interaction*, (eds.) Kemp, E., et. al., 6 July 2001, ACM SIGCHI New Zealand. 49-54

3. The Gnome Usability Project. *Web site*. <URL: http://developer.gnome.org/projects/gup/> (06/26/02)

4. Nielsen, Jakob. *Usability Engineering*. Academic Press, San Diego, CA, 1993.

5. Wharton, Cathleen, et. al., "The Cognitive Walkthrough Method: A Practitioner's Guide." in Nielsen, Jakob, and Mack, R. eds., *Usability Inspection Methods*, 1994, John Wiley and Sons, New York, NY