A content analysis on the use of methods in online user research

Elke Greifeneder

Berlin School of Library and Information Science,

Humboldt-Universität zu Berlin

Unter den Linden 6

10099 Berlin, Germany

Greifeneder(at)ibi.hu-berlin.de

1. Introduction

Purposeful data results from an expressed purpose in combination with an adequate method. Data gathering is an essential part of online user studies, and every method has its areas of application and its limitations: quantitative surveys are limited in their ability to detect causal relations; with qualitative interviews broad generalizations are risky. In library and information science, user research is a domain in which we gather large amounts of data. But is our data really "purposeful"? Already in 1972, Frank Heidtmann (p. 36-37) made the criticism that we use inadequate research techniques and that these research techniques are – independent of their appropriateness – used in an inaccurate and invalid way.

In 2002, Denise Troll Covey interviewed participants from the Digital Library Federation (DLF) about their use of and experience with methods in user research. She stated that "Libraries are struggling to find the right measures on which to base their decisions. DLF respondents expressed concern that data are being gathered for historical reasons or because they are easy to gather, rather than because they serve useful, articulated purposes" (p. 2-3). All studies are "assessing use and usability of online collections and services" (p.7-8, 17, 24 and 35) – they all deal with online user research. Online user research means our studies focus on an (online) digital library environment. Hence, our users are online. What does this mean for our research? Digital library users are no longer tied to a local place. Online users of digital libraries are multi-local, multi-lingual and live in multiple time-zones. Getting "purposeful data" in online user research requires that the research be done online because the users are there.

2. Related Work

Most of the research on users and digital libraries analyzes a single digital library and its users. The content analysis illustrates this with a sample list of 70 publications (Greifeneder, 2009). The

number of publications on specific methods is limited. Edgar (2006), Homewood (2003), Xia (2003), and Nicholson (forthcoming, 2010) discuss general research design issues. Studies like "The virtual scholar: the hard and evidential truth" try to draw a general picture of online users. The described content analysis grows out of Troll Covey's study (2003) that addresses the relation of purposes and methods in online user research.

3. Research problem

Not all methods are currently usable online; focus groups are difficult in an online environment, as are interviews and ethnographic observations. Surveys or log file analyses are on the other hand more easily used for online studies (Greifeneder, 2010). All the studies Troll Covey analyzed in her research were about an online environment, but she did not ask whether the study itself happens online or offline. Another limitation of her study is her population: 24 relatively rich libraries that can afford staff time and money for intensive user research. This paper addresses both limitations. With the analysis of international publications on digital user research during the last ten years it has a broader scope and it considers whether the method was used on- or offline. The content analysis also examined the stated purposes in online user research and the relationship between that purpose and the result. Due to space considerations, the results of the latter analysis of purposes will be separately published.

4. Research Design

The method used to answer these questions is a content analysis using thematic coding. Within the analysis, six categories haven been used: 1) method, 2) purpose, 3) result, 4) offline/online 5) object of study and 6) type of researcher. In an additional step, each category was divided into several subcategories. For example the main category "methods" splits into subcategories including surveys, focus groups and log file analyses. In this research, "in-vivo" codes have been used that derive directly from the data. For example an author writes "the purpose of this study is to measure the use of our electronic services with a survey". Thus the purpose is "use", the object is "electronic services" and the method is a "survey". A second step encodes these original statements into abstract groups for further analysis. Most qualitative data analysis programs require the full documents to be imported, but this was not possible with these publications. Therefore the analysis was done by hand in Microsoft Excel. Attached is an illustrative example of four of the publications that were analyzed. The examples show the content analysis process, including the original in-vivo codes, a first coding (that summarized several different descriptions to the code "use"), and finally an enhances numeric encoding for computer-based measurement.

If a study used two methods, for example an interview and a survey, it counted both as a survey and an interview. As a result, the total of all methods is higher than the number of publications. This approach considers the fact that some methods are applied in combination in order to collect separate kinds of data, for example the study by John Crawford (2003) used focus groups, a survey and a citation analysis in order to "monitor off-campus usage of EIS [electronic information services], the use of passworded databases, and the freely available internet" (p. 1). 11 % of all publications (8 of the 70 studies) make use of a combination of methods: mainly surveys, interviews and usability-tests are combined.

The aim of this content analysis is to identify methods we use in online user research. The analysis is based on the following sampling criteria:

The five article databases are DABI, E-LIS, DissOnline, ProQuest and LISA. DissOnline, the German thesis repository, and ProQuest Dissertation Abstract, the international database for US/Canadian thesis, cover the current research on a high scholarly level. LISA, Library and Information Science Abstracts, is the international database for the LIS field and serves as a supplier for peer-reviewed articles. E-Lis, E-Prints in Library and Information Science (LIS), is an Open-Access Repository and covers for this analysis the part of current user research mainly within conference proceedings and serves as a European counter balance to the mainly US research in LISA and ProQuest. Finally, DABI is a German journal database that contains only abstracts of articles that have been published in the German community. These articles are mainly non peer-reviewed. They are used in this analysis to include the current user research in smaller (and poorer) institutions.

The publications that have been taken into consideration contain applied user research in a digital library environment and examine only online services. The sampling excluded the opening hours of the local institution, satisfaction with local staff, and general observations about human behavior. All publications post-date 1998[1]. The sample contains no unpublished conference contributions and no slides from conference presentations. For the retrieval process, the databases classifications were used. In DABI these were the keywords "Benutzerforschung", " Studie", "Umfrage", "Befragung" and "Fragebogen"[2], in E-LIS the subject terms "use studies" and "user studies", within LISA the descriptor combinations " [use AND Digital Libraries]", as well as "[studies AND Digital

---

[1] In 1999, the DSL technology is introduced in Germany and in 2000 the typical Swiss man spent 12.5 minutes a day online. The broad use of digital collections started only after 1998
[2] translation: user research, study, poll, questionnaire, survey

Libraries"]", "[survey AND Digital Libraries"] and "[focus groups]. DissOnline offers no classification within LIS, nor does ProQuest, so the search query "[all thesis within LIS AND 1998-2009] served as basis for sorting in subject to the sampling criteria.

The content analysis included a total of 13 thesis, 13 conference papers and 44 articles, of which 52 % are peer-reviewed articles, 34 % are non-peer-reviewed articles from DABI (that means in German and research from smaller institutions) and 14 % are non-peer-reviewed articles from E-LIS.

5. Research limitations

Of course, these 70 publications are not the whole research that is done in online user research in LIS. Much of user research is done without publishing the results, because user studies are often conducted in order to make concrete improvements in a system or in a service, rather than with any intention to publish the results as research. There are no exact numbers how big this grey zone is. A member of the German National Library reported in a discussion that they have averaged one user research study per year since 2000 – but only two were published and can be found using the criteria above. In this example the ration of published to unpublished studies is about 1 to 4 or 25%. The 70 publications in this study offer only a picture, showing tendencies in methods and purposes. It may be reasonable to guess that only the best and most interesting projects result in a publication. If that is true, then the analysis paints a picture of user research in LIS that shows only its best parts.

6. Results of the analysis

6.1 Methods used for online user research studies

Most of the online user research is done by students (36 % of all studies). That is not a surprising number, since user research is a popular subject for bachelor, master and doctoral theses. The other groups are librarians working in libraries, who do 31 % of all studies, and LIS school faculty and staff, who do 26% of all studies. In 7 % of the cases, an external non-LIS-professional conducted or helped with the study. As mentioned earlier, one study might have several people from different groups working on it. Each group counted one. It is striking that in those studies where the result matched the purpose, an external expert person was generally involved.

The most frequently studied of online user research has been electronic resources with 33 %. Such electronic resources are often licensed and not accessible without appropriate authentication. In critical financial times, libraries need to prove the value of these electronic resources. Free digital

libraries are the second mostly studied object with 22 %. Only 9% of the studies cover the local OPAC.

Based on the data, libraries use fewer surveys[3] and a greater variety of methods than may be obvious. Nonetheless 43 % of all user-research methods are still surveys. (Remember that a study using both a survey and a focus group, counts both for the survey and for the focus group categories.) This means that in nearly half of all the studies, researchers used a survey. Log file analysis follows in popularity with 18 %, followed by interview with 12 % (see graphic 1). An additional interesting fact is that surveys are used by all kind of researcher groups, but that students favor log file analysis.
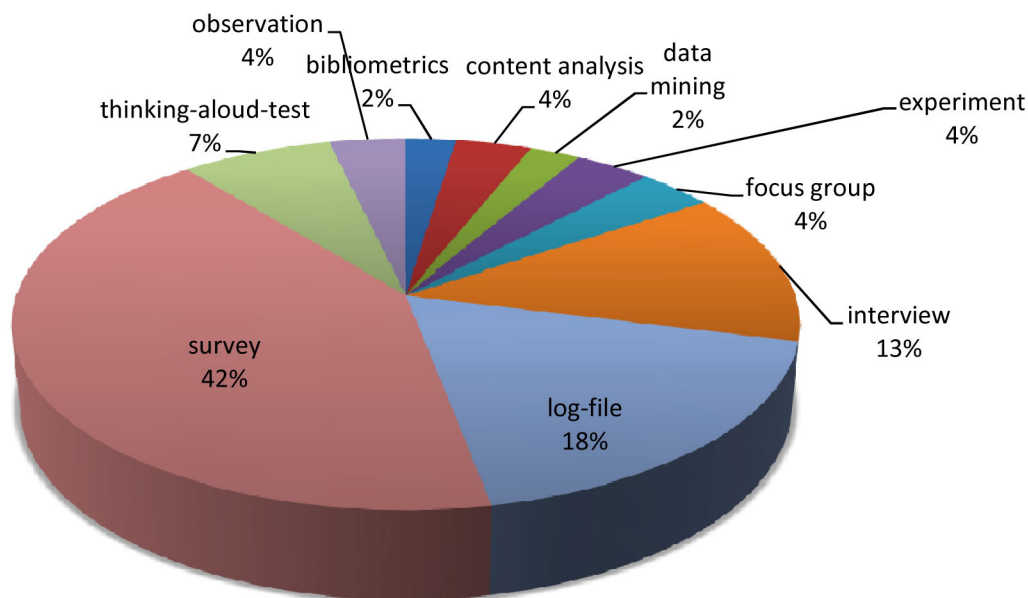
Graphic 1: Methods used for online user research studies

Troll-Covey offers different results in her study, but there too surveys are used by most of the researchers. While "most of the DLF respondents" (p. 35) used log files in her study, not even one-fifths does it in the current analysis. Even more obvious are the differences in qualitative methods such as focus groups: "more than half of the DLF respondents" (p. 17) use them contrast to a use of 4 % in the current analysis. One possible reason for these differences may be that DLF members have more money to spend for qualitative research and for a multiple-methods approach than others have. It could also be that people said in the interviews that they did a focus group that was really only a meeting with student assistants. The reason for the differences could also be found in the

---

[3] Surveys in this analysis mean a questionnaire with a (mostly highly) structured form. Surveys do not include open-ended interviews

published data. Perhaps people actually did interviews at the beginning of their research, but did not mention them in the final publication. In any case the numbers in this analysis are not what people have said they did in an interview, but what they took seriously enough to publish under their own name.

6.2 Offline or Online: Online User research today

The initial question for this content analysis was where the research has taken place. The analysis shows that still nearly 50 % of the online user research is held offline (for example, in focus groups in local libraries). The highest online rates come from the university faculty and staff in Library and Information Science (see graphic 2).

	offline	online
student	48 %	52 %
librarian	45 %	55 %
university faculty and staff members in LIS	39 %	61 %

Graphic 2: Online User Research - offline or online?

A second important question is: which methods are specifically used online? The content analysis reveals that the only method that is used predominantly online is log file analysis – a method where no human interaction is necessary during the data collection process. Similar reasons hold for online-surveys, which also rely only on machine-based interactions. The high number of offline surveys in the graphic 3 may be an outgrowth of the state of user research at the beginning of the analysis period, due to the domination of the survey method at this time. The first email survey dates back to 1999, the earliest online survey in the publication set took place in 2001. The first think-aloud test (in the publication set) happened only in 2002, and the first interview was done in 2003. Methods like interviews or focus groups currently take place nearly entirely offline. Only one of the studies used an online interview. Only in 21% of the cases did researchers use a qualitative method (focus groups, interviews or observations) – more than double that number used surveys.

	log file	survey	interview	focus group
offline	20 %	66 %	91 %	100 %
online	80 %	34 %	9 %	

Graphic 3: Methods used offline and online

Qualitative methods traditionally require human-to-human communication between the interviewer and the interviewee to be able to reformulate a question or to respond to a specific answer in order to get deeper insights into behavior. An example of qualitative research in usability engineering comes from the construction of persona and scenarios for a digital library – for both, deep insight into sample users is needed, rather than data about the whole population. Quantitative methods may be used afterwards to check if the persona or the scenario matches the population. Despite artificial intelligence experiments, machines currently cannot effectively conduct unstructured interviews on their own.

The problem is that quantitative research designs require knowledge about the user's context to be able to ask the right questions and to interpret the data in the right way. Do closed answer-sets offer the options that users would provide or do they only the questioner's perspective? Can log files be analyzed without knowing the full social context of the users' actions? If most people select new offerings, does this mean that they want that particular information or are they merely browsing? The analysis above shows that surveys are used for many more purposes than all the other methods and that they are often used as an all-purpose research tool for need assessments, user typologies, perception studies, satisfaction testing – even for testing usability.

7. Conclusion

Although researchers may use quantitative methods, they tend to articulate purposes like user typologies or need assessments that implicitly demand qualitative methods with an interactive human presence. If the purpose is to know users and the context in which they use a digital library, human-mediated inquiries need to substitute for surveys and log file analysis. As Notess (2004) says: "Part of the problem is that the log files do not tell us anything about user motivation or rationale. For instance, we noted that only 11% of user sessions used bookmarking. But we do not know why the other 89% did not make use of this feature."(online document).

In several of the studies that were considered, the authors themselves commented that they better need qualitative data for their own purposes or that they gathered within a survey qualitative data but did not know how to interpret the results. For example Lehnard-Bruch (2005) used some open-ended questions in her survey and had to admit that "the analysis of the open-ended questions like 'what can the library do to improve its service?' and 'what are additional offers you would like to see in your library?' proved to be difficult, because the answers are very heterogeneous and because

many answers are only mentioned once"[4] (p. 147). Sarah Diepolder (2003) obviously had problems in analyzing her data and had to refer to interpretations that were pure speculation: "The high number of identified changes in the search strategy may be a result of [...] it also might be a result of [...]"[5] (p. 29). Jina Choi Wakimoto (2006) used a mostly closed-ended survey and adds in his publication: "Perhaps most telling were the comments received from 38 percent of survey respondents" (p. 129).

It is necessary to add that researchers using qualitative methods also had problems: Don MacMillan (2007) used a focus group and said that "We found the analysis and reporting stage more time-consuming and difficult than expected. This stage requires advance planning and use of qualitative analysis methodology."(p. 430) .The problem is not only the need to use qualitative methods but to know how to use them. More research about online qualitative methods and more training is necessary.

References

Alcaín, Mª Dolores, Piedad Baranda, Luis Rodríguez Yunta, Adelaida Román, and Ángel Villagrá. 2001. Evaluación de la base de datos ISOC a través de un estudio de usuarios: Homenaje a José María Sánchez Nistal. *Revista Española de Documentación Científica* 24, no. 3. p. 275–88, DOI= http://eprints.rclis.org/9741

Crawford, John. 2003. The use of electronic information services by students at Glasgow Caledonian University: background to the project and introductory focus groups. *Library and Information Research* 27, no. 86, p. 30-36, DOI= http://eprints.rclis.org/34137

Edgar, Bill 2006. Questioning LibQUAL+: Critiquing its Assessment of Academic Library Effectiveness. In*: Proceedings of the ASIS&T Annual Meeting*. DOI= http://www.asis.org/Conferences/AM06/papers/112.html

Homewood, Janet. Huntington, Paul and Nicholas, David. 2003. Assessing used content across five digital health information services using transaction log files. *Journal of Information Science* 29 (6).

Greifeneder, Elke. 2009. Sample Online User Research in LIS, DOI= http://www.elke.greifeneder.de/UserResearch

Greifeneder, Elke. 2010. The need for qualitative methods in online user research in a digital library environment. In: *Proceedings of the iConference 2010 University of Illinois, Urbana Champaign*, p.

---

[4] Original citation: „Die Auswertung der offenen Fragen „Was könnte die Landesbibliothek an ihrem Serviceangebot verbessern?" und „Welche weiteren Angebote würden Sie sich wünschen?" gestaltete sich als schwierig, da die Antworten sehr heterogen sind und es viele Einzelnennungen gibt."

[5] Original citation: "Die hohe Zahl der dabei ermittelten Strategieänderungen (70 Prozent) kann möglicherweise daraus resultieren, dass [...] es könnte aber auch […]".

444-445. DOI: http://nora.lis.uiuc.edu/images/iConferences/2010papers_Allen-Ortiz.pdf

Harmeyer, Dave. 2007. Online virtual chat library reference service: A quantitative and qualitative analysis. Pepperdine University, California

Heidtmann, Frank. 1972. Zur Theorie und Praxis der Benutzerforschung: Unter besonderer Berücksichtigung der Informationsbenutzer von Universitätsbibliotheken. München-Pullach: Verl. Dokumentation.

Lehnard-Bruch, Susanne, Barbara Koelges and Ute Bahrs. 2005. Benutzerumfrage der Landesbibliotheken im Landesbibliothekszentrum." *Bibliotheken heute* 1 (3), p.147-48.

Diepolder, Sarah. 2003. Was ist eine Körperschaft?: Umfrage zur Opac-Nutzung an der Universitätsbibliothek Tübingen. *BUB* 55, no. 1, p. 28-30

MacMillan, Don, Susan McKee, and Shawna Sadler. 2007. Getting everyone on the same page: a staff focus group study for library web site redesign." *Reference Services Review* 35, no. 3, p. 425-433.

Mayr, Philipp. 2004. Website entries from a web log file perspective: a new log file measure. In *AoIR-ASIST 2004 Workshop on Web Science Research Methods*, Brighton, DOI= http://eprints.rclis.org/2831/

Nicholas, David. 2010. The virtual scholar: the hard and evidential truth. In Digital Library Future: User perspectives and institutional strategies, ed. Ingeborg Verheul, Anna Maria Tammaro, and Steve Witt, vol. 146. Munich: K.G.Saur, in print.

Nicholson, Scott. 2005. A framework for Internet archeology: Discovering use patterns in digital library and Web–based information resources. *First Monday* 10 (2). DOI= http://firstmonday.org

Notess, Mark. 2004. Three looks at users: a comparison of methods for studying digital library use. *Information Research* 9, no. 3

Riolo, Massimo. 2001. Le risorse elettroniche di un sistema bibliotecario: analisi e monitoraggio del loro utilizzo. Diploma Universitario thesis, Università degli Studi di Milano Bicocca (Italy), DOI= http://eprints.rclis.org/2793/

Troll Covey, Denise. 2002. Usage and Usability Assessment: Library Practices and Concerns. Washington, DC: Council on Library and Information Resources, DOI= http://www.diglib.org/pubs/dlf096/dlf096.htm

Wakimoto, Jina Choi, David S. Walker, and Katherine S. Dabbour. 2006. The myths and realities of SFX in academic libraries. *Journal of Academic Librarianship* 32, no. 2

Xia, Wei. 2003. Digital library services: perceptions and expectations of user communities and librarians in a New Zealand academic library. *Australian Academic and Research Libraries* 34 (1), 56-70. DOI= http://www.alia.org.au/publishing/aarl/34.1/full.text/xia.html

Attachement

Document type	year	Title	Method	Details about method	Researcher type	Purpose of study	Relation purpose and result	Object of study	Offline/ online
thesis	2007	Online virtual chat library reference service: A quantitative and qualitative analysis	12 content analysis	analysis of reference transcripts; builds on other studies and guidelines	1 student	7 to provide a theoretical conceptual model of best practices for the reference interview	5 purpose does not match result; it is a method of measurement	7 reference service	0 offline
thesis	2001	Le risorse elettroniche di un sistema bibliotecario: analisi e monitoraggio del loro utilizzo	2 log file analysis	log file analysis and online questionnaire and costs analysis	1 student	2 + 3 find out how users use the system AND analyze use-cost-relation AND satisfaction measurement	3 purpose match result, - but very static analysis of use	2 electronic resources	1 online
Conference Paper	2004	Website entries from a web log file perspective: a new log file measure	2 log file analysis	Web Entry Factor (WEF)	1 student	2 use + user patterns	3 purpose matches result, but log file analysis of webpage, no user behavior or patterns but more general usage frequencies	1 institutional web page	1 online
Journal paper	2001	Evaluación de la base de datos ISOC a través de un estudio de usuarios : Homenaje a José María Sánchez Nistal	1 survey	paper and email questionnaire sent to users and to reference librarians, open and closed questions	2 LIS school faculty	2+ 3 + 9 information about the use + satisfacion measurement + measurement of quality	2+ 3 + 6 purpose matches result, but use is only defined as how often someone clicks on something and satisfaction is measured in asking directly how satisfied someone is with something	23 internal electronic resources	0 offline