

Are There Any Differences in Data Set Retrieval Compared to Well-Known Literature Retrieval?

Kern, Dagmar; Mathiak, Brigitte

Postprint / Postprint

Konferenzbeitrag / conference paper

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Kern, D., & Mathiak, B. (2015). Are There Any Differences in Data Set Retrieval Compared to Well-Known Literature Retrieval? In S. Kapidakis, C. Mazurek, & M. Werla (Eds.), *Research and Advanced Technology for Digital Libraries* (pp. 197-208). Springer. https://doi.org/10.1007/978-3-319-24592-8_15

Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

gesis
Leibniz-Institut
für Sozialwissenschaften

Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Mitglied der

Leibniz-Gemeinschaft

Are there any Differences in Data Set Retrieval compared to well-known Literature Retrieval?

Dagmar Kern¹, Brigitte Mathiak^{1 2}

¹GESIS – Leibniz Institute for the Social Sciences, Cologne, Germany

²University of Cologne, Cologne, Germany

dagmar.kern@gesis.org, b.mathiak@uni-koeln.de

Abstract. Digital libraries are nowadays expected to contain more than books and articles. All relevant sources of information for a scholar should be available, including research data. However, does literature retrieval work for data sets as well? In the context of a requirement analysis of a data catalogue for quantitative Social Science research data, we tried to find answers to this question. We conducted two user studies with a total of 53 participants and found similarities and important differences in the users' needs when searching for data sets in comparison to those already known in literature search. In particular, quantity and quality of metadata are far more important in data set search than in literature search, where convenience is most important. In this paper, we present the methodology of these two user studies, their results and challenges for data set retrieval system that can be derived thereof. One of our key findings is that for empirical social scientists, the choice of research data is more relevant than the choice of literature; therefore they are willing to put more effort into the retrieval process. Due to our choice of use case, our initial findings are limited to the field of Social Sciences. However, because of the similar characteristics for data sets also in other research areas, such as Economics, we assume that our results are applicable for them as well.

1 Introduction

Inspired by the successes of Bioinformatics, much research infrastructure has been built in the last few years to allow all scientific disciplines to archive and share scientific data sets. This infrastructure has been extraordinarily successful, in the sense that there are now much more data sets available than ever before. The trend of journals to force authors to share their data sets before publishing their articles based on these data sets also contributes to that. DataCite¹, probably the biggest catalogue, lists over two million data sets. With the increase of available data sets, the finding of information becomes more difficult for the researcher and thus, an important issue for further research in the field of data set retrieval. Even with a large time budget simple reading through everything is no longer a viable option. Most of the current data catalogues utilise systems which were originally designed for literature or web site

¹ <https://www.datacite.org/>

searches, e.g. SOLR². But do these data set retrieval systems really satisfy users' needs? We conducted two user studies and found answers to that question. In a lab study with 7 participants and in telephone interviews with 46 participants we focussed on data set retrieval for a specific domain, the Social Sciences.

In this paper, we describe the methodology of the user studies, discuss the results and derive challenges that developer of data set retrieval systems are facing. Although we assessed user behaviour in a particular system for social scientists, we believe that many of our observations are transferable to other similar systems in different domains. In particular, our key insight is that users are willing to put higher effort in the searching and selecting process of data sets than they do when searching for literature. They also expect high-quality metadata to support this process. We assume that this is true for other data retrieval scenarios as well.

2 Motivation

For the Social Sciences there are different relevant archives that provide research data sets. The most relevant international archives for Social Science research data are the ICPSR³ in the USA, the UKDA⁴ in Great Britain and the NESSTAR⁵ software. One of the most known data catalogue in Germany is the DBK⁶. Other rather highly specialised data catalogues only provide one data set or a collection of data sets (SOEP⁷). Many archives are now linked to each other, exchange metadata and share infrastructure and sometimes even software. Almost all current archives now have some sort of online representation of their data. Typically, those systems are similar to nowadays digital literature retrieval systems. There are some systems that attempt to transcend this general trend and try to develop tailored services for data sets. E.g. OECD and NESSTAR focus on browsing and visualisation of the data. Others focus on rich metadata for the data sets, like the ICPSR and most specialised archives. The retrieval component is an integral part of all systems.

When treating metadata of data sets like metadata of publications the integration of such information in a library system can be made easily [15]. When taking a closer look at both types of metadata (cf. Fig. 1), it can be seen that there are some basic similarities. However, in order to use data for scientific analyses, researchers need more detailed information about the data set and its structure. It has to be understood which values belong to which question and in which context the data has been captured. Furthermore, for the reusability of data sets and the comparability of results gained through analysis based on the data set it is essential to know which other analyses have been already performed with the data set.

² <http://lucene.apache.org/solr/>

³ <https://www.icpsr.umich.edu>

⁴ <http://www.data-archive.ac.uk/>

⁵ <http://www.nesstar.com/> NESSTAR is the meta-search portal of many European archives.

⁶ <https://dbk.gesis.org/dbksearch/index.asp?db=e>

⁷ http://www.diw.de/en/diw_02.c.221180.en/research_data_center_soep.html

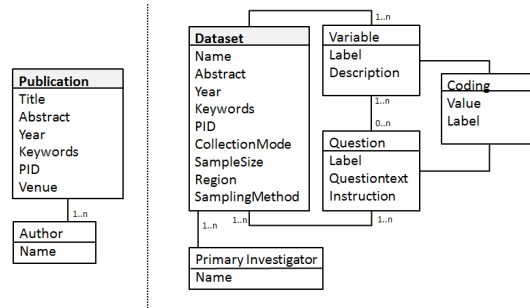


Fig. 1. Comparison of metadata between publications and data sets. In both cases, we restricted ourselves to what could arguably be the most relevant fields.

3 Related Work

That there is a need for sharing data for reuse is an almost common agreement in the scientific community (e.g. see [9]). However, to the best of our knowledge no studies on the behaviour of users of data set repositories in the field of the Social Sciences have been published yet. However, studies on systems in related fields, such as History [7] or in the Humanities [6] have been made. The results of these studies show that the users asked first for more accessibility, then for more and richer metadata and then for more convenience, including domain-specific applications. This stands in strong contrast to literature retrieval, where convenience is by far the most commonly named requirement [4]. Established theory on information seeking [17] stresses the time constraints. Optimising task speed, usability and simplicity has ever since been a major driver in the development of information system user interfaces. Accordingly, [5] points out that the importance of convenience is highly dependent on the context.

In the field of the Social Sciences, there are a few studies performed to understand the seeking behaviour of this target group in general (e.g. [14],[18]). However, data set retrieval is often not an explicit subject of discussion. Most of the presented results focus on literature retrieval. In our research, we try to fill this gap by applying standardised methods for collecting information about user requirements and user behaviour. While Diane Kelly [12] provides a good overview of different evaluation techniques in the field of interactive information retrieval in general, we present in the following the methods that are relevant for our conducted studies. One method is the “simulated work task situation” introduced by Borlund and Ingwersen [2]. A short cover story is used to describe a realistic information seeking situation that motivates the test person to use the system under investigation. Results gained through this method reflect real information needs [3]. In case of our data set retrieval system, this is a task such as “Find one or more relevant data sets that fit your actual research topic”. In the context of the simulated work task situation the “Think-aloud” method [8] can be deployed to learn more about the users’ current needs and requirements. Subjects are asked to articulate what is on their minds when using the system. This data can be captured by recording software and inexpensive computer microphones. While

applying this technique it has to be kept in mind that participants might have difficulties to put their thoughts in words [12].

Interviews belong to the most common form to gather data that indicate user needs [16, S. 221]. The interviews are often made face-to-face but also telephone interviews serve this purpose. Interviews can be divided into structured, semi-structured or unstructured interviews depending on how strict the interviewer is following the interview guide. After gathering the data the analysis of the qualitative data follows. Usually, the responses are assigned to set of categories to be able to summarise them. The quality of the analysis depends on how the data are collected. A transcription of interviews followed by an analysis based on these notes is better than an analysis based on notes that are taken during a non-recorded interview [12].

4 Research context

In an on-going project for integrating access to different sources of information (publications, data sets, web sites, etc.) we follow a user-centred design approach [10]. According to the first step “specify context of use” [10], we performed an analysis of the data catalogue of GESIS - Leibniz-Institute for the Social Sciences⁸, called DBK. The DBK contains study descriptions and the associated data sets. Its main function is to serve as a central point for downloading and purchasing data sets, including the legal and commercial infrastructure for handling sensitive and confidential data. It also serves as a reference guide for additional documentation and metadata. The information about a data set in the database contains bibliographic citation, content and methodology descriptions, lists of errata and versions, the list of primary publications about the data set and the link to the associated questionnaires and codebooks as well as the link to the data set download. So far, a comprehensive list about publications that are based on the data set is not available. This metadata for each data set is displayed in a detailed view (see Figure 2). The search function of the DBK is a full text search based on a set of predefined fields of the relational database it is built upon. The underlying software is Open Source⁹.

To gain some preliminary insights into the users’ behaviours, we started our investigations with a query log and session analysis based on the web servers logging files. We were able to identify some differences to the typical query pattern that we know from our literature system sowiport¹⁰. Known-item queries constitute roughly two thirds of all queries, which is much more than in the case of our literature system. Additionally, session analysis shows that many users come in through direct links, but then spend a long time on the page with detailed information about the data set (see Figure 2), by-passing the search engine completely.

⁸ <http://www.gesis.org/>

⁹ <https://dbk.gesis.org/DBKfree2.0/Defaulten.htm>

¹⁰ <http://sowiport.gesis.org/>

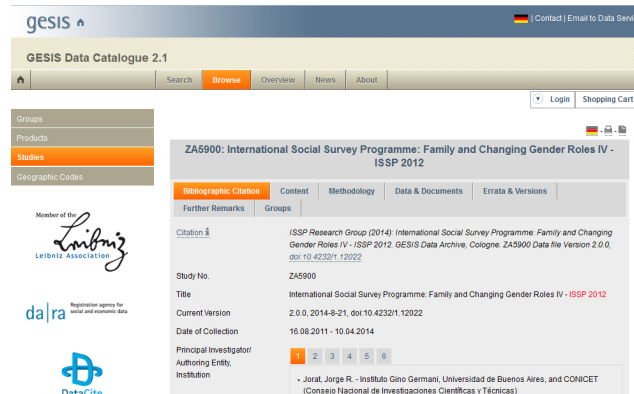


Fig. 2. Detailed view of a data set.

When searching, author names, a kind of information users of our literature system often search for, are not used. Although these insights already exposed some information about the differences of literature and data set retrieval, we decided to perform more detailed user studies based on semi-structured interviews to get a clearer picture on the information seeking process and the users' intentions themselves.

5 Methodology of the user studies

We conducted two studies, a lab study with seven participants and telephone interviews with 46 participants. Our goals were, to find out how Social Science researchers actually search for data sets, what their requirements are, how they achieve their goals and whether they are satisfied with the search path they used. The studies were conducted with German-speaking users of the DBK.

5.1 Lab study setup

The lab studies were performed in single sessions where each participant sat together with an experimenter in front of a computer. After a short introduction the participants were asked to use the DBK to search for one or more data sets fitting their current research questions. They were observed while using the DBK for ten minutes and were asked to think aloud, while working on their individual tasks. The audio as well as the screen were recorded. Afterwards, we conducted a semi-structured interview during which the participants were encouraged to use the data catalogue to recall some actions they had performed in the past or to examine functions they had not used before. These single sessions took about 45 minutes each.

5.2 Telephone Interview setup

In addition to the lab study, telephone interviews were conducted. It was the same interview guide as used in the lab study. The questioning was recorded and transcribed afterwards. The telephone calls took between 7 and 25 minutes. In contrast to the lab interviews, the participants were not asked to use the DBK, but they were allowed to if they wanted. However, we expected that they answered the question based on their memory of recent use of the portal.

5.3 Subjects

For the telephone interviews, DBK users were asked by email (sent out to about 600 registered users), if they were willing to take part in a questioning. The interviews were conducted with 46 participants (11 female, 35 male). The youngest interview partner was 26 and the oldest 80 years old (mean age 41). Ten participants worked or used to work as a professor at a University, 18 were postdocs, 17 held a master or an equivalent degree and one held a bachelor degree. The main research interests lay in Social Sciences (20) or Political Sciences (12).

For the lab study, seven participants (3 female, 4 male) were recruited from the sample above, based on geographical closeness. They were between 26-43 years old. The group consisted of six PhD candidates and one postdoc. Their research interests were in Social Sciences (4) and Political Sciences (3). It must be mentioned that six of them had affiliations with GESIS, e.g. as research assistants, which may have introduced an institutional bias. However, since the lab participants were much more critical about the catalogue than the telephone interviewees, we believe that this bias at least did not result in underreporting of problems.

6 Results of the User Study

Based on the transcriptions of the interviews we summarised and analysed users' behaviour as well as their requirements on a data set retrieval system. In this paper, we focus mainly on the responses that help us to answer the question "Are there any Differences in Data Set Retrieval compared to well-known Literature Retrieval?" Detailed information about further results will be presented in another context. As the interview structure was the same in both studies the results are presented together. The interview guidelines were not followed strictly in all cases, e.g. when a preceding answer clearly made a follow-up question irrelevant. This occurred rarely, though. The numbers of responses to each question never dropped below 38.

6.1 Lab study observations

Participants of the lab study were asked to search for data sets relevant to their research by using the catalogue. The most often used strategy was to use the simple search as an entry point and then to browse through the often very long result list. In

most cases the initial query was not refined. The subjects inspected each entry dutifully. They sometimes consulted additional documentation, such as questionnaires, to evaluate the relevance of entries. Two participants entered previously known survey programmes and then proceeded with browsing through the result list in a similar manner as the other participants did. All, except one participant, found a relevant data set within the allotted ten minutes, however, in many cases, it was implied that further investigations would be needed to properly assess its applicability. Most participants felt that there is a lack of additional filtering options. However, the filtering possibilities offered, were not used at all. It should be noted here that there were no complaints on the rather lengthy process of evaluating relevance. In digital library systems, the users are much quicker to judge positive document relevance, e.g. 26.49s on average in the system described in [11]. From our experience of the lab study, we claim that for data sets this time is much higher. As our participants felt that it was part of the normal research process.

6.2 Frequency of use

In our interview, the first questions aimed at assessing the participants' frequency of usage of the DBK. With the exception of those subjects, who use the DBK for teaching most participants used the system rarely (1-3 times per year, also cf. Fig. 5). Since all participants were active researchers, we assume that they use library systems much more often.

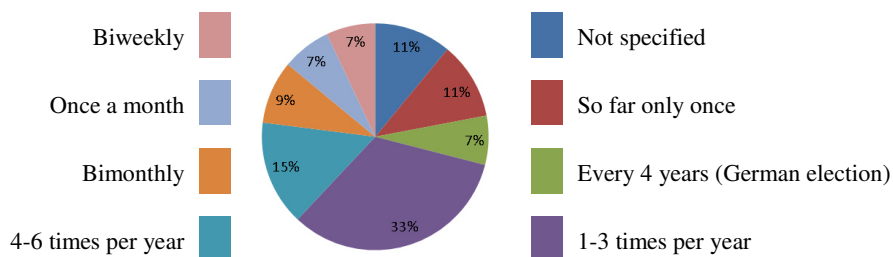


Fig. 3. Frequency of usage

6.3 Metadata of data sets vs. metadata of literature

The results of our user study provide insight in the relevance of the different metadata for data sets. In the following we summarise the key findings in pointing out the differences of metadata for data sets and literature which seem to be quite similar at the first glance. For data sets, it can be said that the title is in most cases more a brand name than an accurate description of the contents. Furthermore, the publication year is irrelevant for data sets. Highly relevant, on the other hand, is the time frame to which the data refers to. What the abstract is for the publication the list of keywords and the table of content are for the data sets. Both metadata provide keywords but with a different granularity. For a publication, it is important in which journal or pro-

ceedings it is published and for the data set, this information can be seen as the context in which the data are collected for example under the umbrella of the comprehensive survey programme. While for publications the authors are very important, for data sets the primary investigators seem to play a rather minor role for the data set search process. Not one participant mentioned them in the interviews.

Furthermore, data set retrievers rely on the additional documentation like code-books, method reports or questionnaires to be able to analyse the data. The quality of these documentations is essential. Other aspects that are very important for the users of data sets are the mode of questioning, the sampling size and the involved countries.

Similar to references for publications data set often based on specific item and scales. Documentation about these not only helps to understand the data collecting process, but also helps to estimate the quality of the data. Related literature that cited a data set is at least as important as it is in literature search systems. It might be even more important as the question if a research question has already been answered (a use case proposed by the participants) based on a certain data set could be answered by looking at the citing. The participants reported that searching for such publications is rather cumbersome, so far. Furthermore, in case the studies were performed in different countries and not all data sets are available in the presented catalogue, they would appreciate having a link to where to find the corresponding data sets.

6.4 Additional functionalities

In addition to the questions about their search behaviour in the DBK and the more general questions about data set retrieval, we asked the participants about their opinion on a few new features for the DBK. The answers are summarised in Table 1.

Would you like to have ...	Yes	No	Undecided
a list of the Top 10 Downloads?	27%	62%	11%
the information of how often each individual data set was downloaded?	19%	62%	19%
a comment function?	30%	53%	17%
a user account?	55%	45%	-
a list that shows data sets downloaded by researches who also downloaded the currently observed data set?	82%	18%	-
an email-reminder asking for publications that you might have written based on the downloaded data set?	94%	6%	-

Table 1. Answers to the questions about new feature ideas for the DBK.

The features that we have prepared for judgement, were motivated by the experiences with our literature search engine. Most answers are unremarkable, except that both user accounts and email reminders have such a high acceptance rate. Most interviewees strongly demanded free access for as many things as possible and did not mind receiving emails.

6.5 Perceived Usability

It is important to mention that the comments regarding usability are quite different from participants of the lab study and the telephone interviews. In the lab study literally everyone complained about the inconvenience of the search functionality. In the telephone interviews, only one person seemed to be able to remember the inconvenient search. The general sentiment was that the DBK is an extremely useful tool. The users appreciate the access to the data sets as well as the high-quality metadata so much that any intricateness in the search functionality seems to be of marginal importance. Nevertheless, for starting the search process most of them expect a Google-like input field with Boolean search, autocomplete function, search term suggestion and an auto-correct on spelling mistakes as a bare minimum.

7 Open Challenges

From the general answers above, some challenges arise that were mentioned multiple times by the participants. In this section, we present a few selected issues, as they seem rather typical for data set retrieval.

7.1 Search within the metadata

Metadata of data sets can be quite extensive, so much that the line between metadata and digital objects in their own right is hard to draw. There are documentations like questionnaires and codebooks, but also method reports and even publications, which are linked because they describe the genesis of the data set. So far, these documents are often displayed together with a data set, but they are not included in the retrieval process of the DBK. However, at least 20% of the participants of both studies mentioned that they would prefer a search opportunity for questionnaires, and 15% would like to have this for codebooks.

This is a complicated issue. Variable labels and questions are both highly repetitive and much more common than the more relevant data sets. When providing data sets, questionnaires and codebooks are all together in one result list, the data sets thus recede into the background. This is a known problem for systems that offer such services. It works for catalogues with few studies, but the currently used visualisation approaches do not scale well. Probably, a complete new retrieval approach has to be chosen on another level.

7.2 Categorisation/ Grouping of data sets

The most relevant data sets in the Social Sciences belong to comprehensive survey programmes in which the data collections are repeated periodically and often in several countries. For each individual collection, one or more new data sets are created and henceforth exist as separate records in the data catalogue. Although the DBK has a group structure and provides a group search function to address this, these functions

were not used during the lab study. For example, the most popular study archived in the DBK is the Eurobarometer (it is also the most common query). Related to it are 621 data sets and 23 groups. Here most relationships are not clearly defined or sometimes even unknown. This complicates the creation of a suitable representation of the stored data. In addition, the way in which survey programmes publish their data is highly heterogeneous, making a standardised display even harder. The challenges of implementing such a system are outlined in [13]. Some of the participants suggested a hierarchical representation within the result list. So that they can easily skip results from survey programmes they are not interested in, or, on the other hand, so that they get an overview of results that belong together.

7.3 Interlinking between data sets and publications

The users are very interested to know how often and by whom a data set has been cited. Whenever possible, links to publications related to a data set should be included into the data set retrieval system. Both self-reported as well as automatically extracted citation [1] could significantly improve interlinking and, thus, improve the user experience of data set retrieval systems. Since these publication lists can become quite large, breaking the citation down to the specific variables was suggested, but this is again information that is not always available.

Another critical point is the timeliness of such publication lists. A few of the interviewees worry that the list in the data retrieval system is not as complete and current as a list that Google scholar could provide them. They suggested a link to Google scholar with predefined search terms including the DOI as an alternative and/or supplement to any list that might be compiled through self-reporting or text mining. All links should be clearly split to differentiate between secondary analysis and sources for the study. Many studies base some of their methodology on pre-existing research, e.g. they use scales and questions from databases or published research.

7.4 Registration and Personalisation

All forms of registration and login are explicitly frowned upon by 19% of the participants. But this is a general issue and not limited to data set retrieval. Echoing similar principles in digital library systems, all metadata should be and actually is free, only the content is restricted, oftentimes for legal reasons. This implies that all documentations (including codebooks and questionnaires) should be freely accessible, as users tend to only register when they can be sure that the data set is relevant for them.

Inspecting the actual data set for quality and relevance is the last resort step, users avoid as long as possible. Not only is it time-consuming to go through registration and possibly legal requirements, without proper documentation, the data set is very likely unusable anyway. Yet, even with documentation, opening a data set for the first time often holds some surprises, ranging from software incompatibilities to mismatches between documentation and actual data. In this context, one participant even suggested providing an example data set with only a few fake data to get an impression what he/she has to expect and therefore to minimise the downloading effort.

Some sort of community function would allow users to help other users, but this is seen critical by almost half of the interviewees. They fear unqualified rumours to be given attention. A helpline towards qualified personal, or ideally the data set providers, would be much preferred. As the time user spend with analysing data sets is often quite long they are highly interested in notifications about changes in their downloaded data sets.

8 Conclusion

In this paper, we study the requirements users have for a data set retrieval system in the Social Sciences and compare those to the well-known requirements in Digital Libraries. Based on two user studies with altogether 53 participants, we could find similarities, but also differences. The need to know who cited a publication of interest respectively who published an article based on a data set of interest can be seen as the main common requirement users have on both retrieval systems. On the other hand, the underlying difference is: **Choosing a data set is a much more important decision for a researcher than choosing a piece of literature.** Therefore, they are willing to spend much more time on the selection process. As a consequence,

- Metadata quality and quantity is most important. Data without description is useless. Often, decision criteria for relevance are hidden inside lengthy documentation. Compared to the state-of-the-art, users wish for more metadata, in particular better structure and visualisation of the relationship between data sets, more links to additional resources, both on the context of the data set and its reception by the community.
- Researchers are dedicated to the research data they choose, even long after they have downloaded it. They appreciate updates and news, as well as suggestions for what might be new relevant data sets for them. Many are also willing to spend time on improving the metadata when so prompted. For example, they are willing to provide secondary literature.
- As for literature, users expect an optimal support while entering their search terms with search term suggestions, autocomplete function, etc. Furthermore, since users visit the catalogue so rarely it is important that a function is as self-explanatory as possible, or it will not be used.

We believe that the differences are mainly caused by the different importance papers and data sets have for the researchers. Research data sets are much more decisive for research activity than any literature is. Many empirical scientists spend their whole academic career analysing one essential study. Any time spent on choosing the correct data set is therefore time well-spent.

Typical information retrieval UIs, in contrast, are optimised to improve convenience, often understood as implementing time-saving mechanisms. The user of data set retrieval systems gives priority to the quality and completeness of the metadata as well as the data sets. While they prefer indeed a more comfortable search input field, they are less interested in getting fast results. This finding offers new possibilities for

all the ideas in information retrieval that failed at the convenience barrier and spawn new research ideas in an environment where time does not matter (much).

9 References

1. Boland, K., Ritze, D., Eckert, K., Mathiak, B.: Identifying references to datasets in publications." In: Proc of the TDPL (2012)
2. Borlund, P., Ingwersen, P.: The development of a method for the evaluation of interactive information retrieval systems. In: *Journal of documentation* 53.3, 225-250 (1997)
3. Borlund, P. "Experimental components for the evaluation of interactive information retrieval systems." *Journal of documentation* 56.1 (2000): 71-90.
4. Connaway, L. S., Dickey, T. J., & Radford, M. L.: If it is too inconvenient I'm not going after it: Convenience as a critical factor in information-seeking behaviors. In: *Library & Information Science Research*, 33.3, 179-190 (2011)
5. Case, D. O. (Ed.): Looking for information: A survey of research on information seeking, needs and behavior. In: Emerald Group Publishing (2012)
6. DeRidder, J. L., Matheny, K.G.: What Do Researchers Need? In: *D-Lib Magazine* 20.7/8 (2014).
7. Duff, W. M., Johnson, C. A.: Accidentally found on purpose: Information-seeking behavior of historians in archives. In: *The Library Quarterly*, 472-496, (2002)
8. Ericsson, K. A., Simon, H. A.: Protocol Analysis. Verbal reports as data, Cambridge Massachusetts. (1984)
9. Faniel, I. M., and Zimmerman, A.: Beyond the data deluge: A research agenda for large-scale data sharing and reuse. In: *International Journal of Digital Curation* 6.1 (2011).
10. International Organization for Standardization. ISO 9241-210:2010 Ergonomics of Human System Interaction-Part 210: Human-Centred Design for Interactive Systems (Formerly Known as 13407). (2010).
11. Kelly, D., Belkin, N.J.: Reading time, scrolling and interaction: exploring implicit sources of user preferences for relevance feedback. In: Proc. of the SIGIR, ACM, (2001)
12. Kelly, D.: Methods for evaluating interactive information retrieval systems with users. *Foundations and Trends in Information Retrieval* 3.1—2, 1-224 (2009)
13. Mathiak, B., Boland, K.: Challenges in Matching Dataset Citation Strings to Datasets in Social Science. In: *D-Lib Magazine* 2015, 21(1/2) (2015)
14. Meho, Lokman I., and Helen R. Tibbo. "Modeling the information-seeking behavior of social scientists: Ellis's study revisited." *Journal of the American society for Information Science and Technology* 54.6 (2003): 570-587.
15. Ritze, D., and Boland, K.: "Integration of research data and research data links into library catalogues." In: *DC 13 - The Lisbon Proceedings of the International Conference on Dublin Core and Metadata Applications: papers, project reports and posters for DC-2013*, 35-40, (2013)
16. Rogers, Y., Sharp, H., Preece, J.: Interaction design: beyond human-computer interaction. John Wiley & Sons, (2011)
17. Savolainen, R.: Time as a context of information seeking. In: *Library & Information Science Research* 28.1, 110-127, (2006).
18. Shen, Y.: Information seeking in academic research: A study of the sociology faculty at the University of Wisconsin-Madison. *Information technology and libraries* 26.1, 4-13 (2013)
19. Wright, S. J. et al.: Using Data Curation Profiles to Design the Datastar Dataset Registry. In: *D-Lib Magazine* 19.7, 2, (2013)