Humboldt University of Berlin

Institute for Library and Information Science

Seeking Research Software. A Qualitative Study of Humanities Scholars' Information Practices.

Ronny Gey

March 19, 2020

Zusammenfassung

Zusammenfassung

Abstract

Abstract

Contents

List of Figures

List of Tables

1. Introduction

Today, software is a central component of science. Throughout the entire research life cycle, researchers use software tools for data collection, transformation, analysis and presentation as well as for generating hypotheses, managing literature and writing scientific papers (Kethers et al., 2017; Pan et al., 2016; Wolski et al., 2017). Software has changed the way we actually do science. The complexity of the analyses carried out by researchers has increased, as has the amount of data that researchers can process. Software supports the documentation of the research process and enables reproducibility (Dallmeier-Tiessen, 2016; Waltemath and Wolkenhauer, 2016) and accuracy of results.

Due to the increased importance of research software for research (Katz, 2017) and the increase in the sheer number of software, it is all the more important for researchers to identify suitable software and select the one which best fits the research problem, the actual step in the research process, or the research data which has to be processed, and, in consequence, which satisfies the researchers information need (Wilson, 1994). In addition to increased efforts, difficulties in seeking software can also endanger the scientific reproducibility of studies or even lead to multiple developments of software with identical functions instead of reusing existing software (Hucka and Graham, 2018).

Information seeking of researchers is generally of great interest within the field of information, be it information behavior (Ahmadianyazdi and Chandrashekara, 2018; Barrett, 2005; Campbell, 2017; Catalano, 2013; Ellis, 1993; Hemminger et al., 2007; Korobili et al., 2011; Liyana and Noorhidawati, 2017; Rimmer et al., 2006; Rupp-Serrano and Robbins, 2013; Wang et al., 2008, e.g.) or information practices (Bøyum and Aabø, 2015; Bulger et al., 2011; Fry, 2006; Given and Willson, 2018; Roos, 2015, e.g.). However, seeking software is still rather challenging for researchers (Howison and Bullard, 2015). In a recent study, Hucka and Graham (2018) surveyed scientists and engineers from several fields to better understand their approaches and selection criteria for seeking software. They found out that "*finding software suitable for a given purpose remains surprisingly difficult*". Even in the domain of software development, the main challenge for software reuse are difficulties in finding software artifacts as Bauer et al. (2014) revealed in a study on code reuse at Google. Grossman et al. (2009) identified users unawareness of specific software tools. These results are all the more surprising because the participants in the cited studies come from a group with a greater affinity for software (software developers, engineers).

The lack of awareness of specific software tools among researchers has been addressed by several technical solutions. Code aggregators, specialized search engines, programming language package repositories, code and application repositories, research repositories (e.g. Zenodo or Figshare), and curated web lists and catalogues aid users in discovering software (Struck, 2018). Standards and tools for citing software enable re-

searchers to identify, cite and reuse software (Niemeyer et al., 2016; Smith et al., 2016; Soito and Hwang, 2017, e.g.). Research funding agencies and research organizations (Haupt et al., 2018; Katerbow and Feulner, 2018; Scheliga et al., 2019, e.g.) adopt guidelines for the development and use of research software with the aim of increasing the reusability and quality of the software artifacts developed. In turn, the technical solutions presented are also aimed more at a technically experienced audience, often even at software developers directly. For researchers with less experience in the use of software, e.g. from the humanities (Rimmer et al., 2006), seeking software remains a difficult undertaking.

The information-seeking behavior of humanities scholars in general has been addressed widely (Barrett, 2005; Bronstein, 2007; Bronstein and Baruchson-Arbib, 2007; Catalano, 2013; Ellis, 1993; Given and Willson, 2018; Korobili et al., 2011; Liew and Ng, 2006; Rimmer et al., 2006, e.g.). In his pioneering work on Grounded Theory in information-seeking, Ellis (1993) identified patterns of information-seeking of social sciences, sciences, and humanities scholars. In 2005, Barrett (2005) analyzed information-seeking habits of graduate student researchers in the humanities. Korobili2011 examined factors influencing information-seeking behavior at the philosophy faculties. While studies in information behavior draw on the cognitive viewpoint, information practice studies are guided by the ideas of social constructionism and collectivism (Savolainen, 2007; Talja et al., 2005; Talja and McKenzie, 2007). Humanities scholars information-seeking practices have also been addressed in several studies (Benardou et al., 2013; Bulger et al., 2011; Given and Willson, 2018; Palmer and Cragin, 2009). In previous studies, however, the classic research process of humanities scholars has been examined, which is mainly based on literature research. Although the information-seeking in the humanities is also based on software tools, e.g. databases, web-based editions, search engines, or online journals (Barrett, 2005; Rimmer et al., 2006), the search for software itself is rarely discussed. One of these rare examples, however a non-humanities one, is Hepworths et al. (2017) study of journalism professors' information-seeking behavior. While seeking new online tools, journalism professors rely on other journalism professors, followed closely by media-related foundations, media professionals, and conferences.

2. Theory

2.1. Research Software

2.2. Domain Analysis: Humanities/Philology

3. Research Design

Since "*[u]nderstanding the nature of information practices and their relation to the production of scholarship is important for both theoretical and applied work in library and information science (LIS)*" (Palmer and Cragin, 2009, p. 165) this thesis sets out to study information practices of humanities scholars and their seeking for research software to better understand humanists needs and future LIS services (Case, 2008; Cunningham, 2010). With information practices we mean practices of seeking, managing, giving, and using information in context (Palmer and Cragin, 2009). The aim of this work is to investigate the information-seeking practices of early-career philologists when seeking research software. This research focuses on information needs of philologists, their information sources, problems, contradictions and barriers in finding information and their information sharing about research software. Special emphasis will be placed on the respondents' recourse to their own research process and the knowledge and practice structures in their field (Hjørland and Albrechtsen, 1995) which are socially constructed.

RQ1: How do humanists seek for research software?
RQ2: How do domain specific structures shape the information practices of humanists?

I chose an exploratory study design (Rinsdorf, 2013) where the personal realm of experience of each interviewee lies in the center of the analysis. Interviews are the main data gathering technique which are applied in a semi-structured manner, guided by interview guidelines, and implemented in a face-to-face manner (Bryman, 2004) in German language. With the interviews I obtain emotions, thoughts, and intentions of the participants and discover their perspective of the social world (Patton, 2002). I will conduct 4-6 interviews of about 60-90 minutes length. The interview data will be analyzed with a qualitative content analysis to explore qualified hypotheses (Kohlbacher, 2006; Krippendorff, 2012; Mayring, 2000, 2014). It enables the researcher to include textual information and to identify its properties systematically. I will make all data generated during the concept, survey, analysis, and writing phases publicly available on GitHub[1], as long as it meets research ethics standards (e.g. interview audio records and unanonymized interview transcripts will be excluded).

[1] https://github.com/geyslein/Masters_Thesis

3.1. Theory

3.2. Data Gathering

3.3. Data Processing

3.4. Data Analysis

4. Findings

5. Discussion

6. Conclusion

7. Zusammenfassung (German Conclusion)

Bibliography

Ahmadianyazdi, F. and Chandrashekara, M. (2018). Research challenges and delayed gratification in information seeking behavior: a case study of research scholars. *Library Philosophy & Practice*.

Barrett, A. (2005). The information-seeking habits of graduate student researchers in the humanities. *The Journal of Academic Librarianship*, 31(4):324–331.

Bauer, V., Eckhardt, J., Hauptmann, B., and Klimek, M. (2014). An exploratory study on reuse at google. In *Proceedings of the 1st International Workshop on Software Engineering Research and Industrial Practices - SER&IPs 2014*. ACM Press.

Benardou, A., Constantopoulos, P., and Dallas, C. (2013). An approach to analyzing working practices of research communities in the humanities. *International Journal of Humanities and Arts Computing*, 7(1-2):105–127.

Bøyum, I. and Aabø, S. (2015). The information practices of business PhD students. *New Library World*, 116(3/4):187–200.

Bronstein, J. (2007). The role of the research phase in information seeking behaviour of jewish studies scholars: A modification of ellis's behavioural characteristics. *Information Research: An International Electronic Journal*, 12(3):n3.

Bronstein, J. and Baruchson-Arbib, S. (2007). The application of cost—benefit and least effort theories in studies of information seeking behavior of humanities scholars: the case of jewish studies scholars in israel. *Journal of Information Science*, 34(2):131–144.

Bryman, A. (2004). *Social research methods*. Oxford Univ. Press, 2. ed. edition. Includes index. - Bibliography. - Previous ed.: 2001.

Bulger, M. E., Meyer, E. T., la Flor, G. D., Terras, M., Wyatt, S., Jirotka, M., Eccles, K., and Madsen, C. M. (2011). Reinventing research? information practices in the humanities. *SSRN Electronic Journal*.

Campbell, L. (2017). The information-seeking habits of architecture faculty. *College & Research Libraries*, 78(6).

Case, M. M. (2008). Partners in knowledge creation: An expanded role for research libraries in the digital future. *Journal of Library Administration*, 48(2):141–156.

Catalano, A. (2013). Patterns of graduate students' information seeking behavior: a meta-synthesis of the literature. *Journal of Documentation*, 69(2):243–274.

Cunningham, L. (2010). The librarian as digital humanist: The collaborative role of the research library in digital humanities projects. *Faculty of Information Quarterly*, 2(1):1–11.

Dallmeier-Tiessen, S. (2016). Reproduzierbarkeit Und Open Science: Bestandteile Und Erste Erfahrungswerte - Mit Besonderem Augenmerk Auf Software. Online. Presented at Helmholtz Open Science Workshop 2016.

Ellis, D. (1993). Modeling the information-seeking patterns of academic researchers: A grounded theory approach. *The Library Quarterly*, 63(4):469–486.

Fry, J. (2006). Scholarly research and information practices: a domain analytic approach. *Information Processing & Management*, 42(1):299–316.

Given, L. M. and Willson, R. (2018). Information technology and the humanities scholar: Documenting digital research practices. *Journal of the Association for Information Science and Technology*, 69(6):807–819.

Grossman, T., Fitzmaurice, G., and Attar, R. (2009). A survey of software learnability. In *Proceedings of the 27th international conference on Human factors in computing systems - CHI 09*. ACM Press.

Haupt, C., Meinel, M., and Schlauch, T. (2018). The software engineering initiative of dlr: overcome the obstacles and develop sustainable software. In *2018 IEEE/ACM 13th International Workshop on Software Engineering for Science (SE4Science)*, pages 16–19. IEEE.

Hemminger, B. M., Lu, D., Vaughan, K., and Adams, S. J. (2007). Information seeking behavior of academic scientists. *Journal of the American Society for Information Science and Technology*, 58(14):2205–2225.

Hepworth, K., Mensing, D., and Yun, G. W. (2017). Journalism professors' information-seeking behaviors: Finding online tools for teaching. *Journalism & Mass Communication Educator*, 73(3):255–270.

Hjørland, B. and Albrechtsen, H. (1995). Toward a new horizon in information science: Domain-analysis. *Journal of the American Society for Information Science*, 46(6):400–425.

Howison, J. and Bullard, J. (2015). Software in the scientific literature: Problems with seeing, finding, and using software mentioned in the biology literature. *Journal of the Association for Information Science and Technology*, 67(9):2137–2155.

Hucka, M. and Graham, M. J. (2018). Software search is not a science, even among scientists: A survey of how scientists and engineers find software. *Journal of Systems and Software*, 141:171–191.

Katerbow, M. and Feulner, G. (2018). Recommendations on the development, use and provision of Research Software. Technical report, Deutsche Forschungsgemeinschaft.

Katz, D. S. (2017). Software in Research: Underappreciated and Underrewarded. Presented at eResearch Australia.

Kethers, S., Treloar, A., and Wu, M. (2017). Building tools to facilitate data reuse. *International Journal of Digital Curation*, 11(2):1–12.

Kohlbacher, F. (2006). The Use of Qualitative Content Analysis in Case Study Research. *Forum: Qualitative Social Research*, 7(1):23.

Korobili, S., Malliari, A., and Zapounidou, S. (2011). Factors that influence information-seeking behavior: The case of greek graduate students. *The Journal of Academic Librarianship*, 37(2):155–165.

Krippendorff, K. (2012). *Content Analysis: An Introduction to its Methodology*. Sage Publications, 3 edition.

Liew, C. L. and Ng, S. N. (2006). Beyond the notes: A qualitative study of the information-seeking behavior of ethnomusicologists. *The Journal of Academic Librarianship*, 32(1):60–68.

Liyana, S. and Noorhidawati, A. (2017). How graduate students seek for information: Convenience or guaranteed result? *Malaysian Journal of Library & Information Science*, 19(2).

Mayring, P. (2000). Qualitative Content Analysis. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 1(2):1–10.

Mayring, P. (2014). *Qualitative content analysis: theoretical foundation, basic procedures and software solution*. Beltz.

Niemeyer, K. E., Smith, A. M., and Katz, D. S. (2016). The challenge and promise of software citation for credit, identification, discovery, and reuse. *Journal of Data and Information Quality*, 7(4):1–5.

Palmer, C. L. and Cragin, M. H. (2009). Scholarship and disciplinary practices. *Annual Review of Information Science and Technology*, 42(1):163–212.

Pan, X., Yan, E., and Hua, W. (2016). Disciplinary differences of software use and impact in scientific literature. *Scientometrics*, 109(3):1593–1610.

Patton, M. Q. (2002). *Qualitative Research & Evaluation Methods*. SAGE, 3 edition.

Rimmer, J., Warwick, C., Blandford, A., Gow, J., and Buchanan, G. (2006). Humanities scholars' information-seeking behaviour and use of digital resources. *Digital Libraries in the Context of Users' Broader Activities*, page 19.

Rinsdorf, L. (2013). Qualitative methoden. In Umlauf, K., Fühles-Ubach, S., and Seadle, M., editors, *Handbuch Methoden der Bibliotheks-und Informationswissenschaft.*, pages 64–79. DeGruyter Saur, Berlin/Boston.

Roos, A. (2015). Medical scientists' information practices in the research work context. *Health Information & Libraries Journal*, 32(1):23–36.

Rupp-Serrano, K. and Robbins, S. (2013). Information-seeking habits of education faculty. *College & Research Libraries*, 74(2):131–142.

Savolainen, R. (2007). Information behavior and information practice: Reviewing the "umbrella concepts" of information-seeking studies. *The Library Quarterly*, 77(2):109–132.

Scheliga, K., Pampel, H., Konrad, U., Fritzsch, B., Schlauch, T., Nolden, M., Zu Castell, W., Finke, A., Hammitzsch, M., Bertuch, O., and Denker, M. (2019). Dealing with research software: Recommendations for best practices. Technical report, Helmholtz Association.

Smith, A. M., Katz, D. S., and and, K. E. N. (2016). Software citation principles. *PeerJ Computer Science*, 2:e86.

Soito, L. and Hwang, L. J. (2017). Citations for software: Providing identification, access and recognition for research software. *International Journal of Digital Curation*, 11(2):48–63.

Struck, A. (2018). Research Software Discovery: Challenges, Risks And Opportunities. In *IEEE eScience 2018*. Zenodo.

Talja, S. and McKenzie, P. J. (2007). Editors' Introduction: Special Issue on Discursive Approaches to Information Seeking in Context. *The Library Quarterly*, 77(2):97–108.

Talja, S., Tuominen, K., and Savolainen, R. (2005). "Isms" in information science: constructivism, collectivism and constructionism. *Journal of Documentation*, 61(1):79–101.

Waltemath, D. and Wolkenhauer, O. (2016). How Modeling Standards, Software, and Initiatives Support Reproducibility in Systems Biology and Systems Medicine. *IEEE Transactions on Biomedical Engineering*, 63(10):1999–2006.

Wang, P., Dervos, D. A., Zhang, Y., and Wu, L. (2008). Information-seeking behaviors of academic researchers in the internet age: A user study in the united states, china and greece. *Proceedings of the American Society for Information Science and Technology*, 44(1):1–29.

Wilson, T. (1994). Information needs and uses: Fifty years of progress. *Fifty Years of Information Progress: A Journal of Documentation Review*.

Wolski, M., Howard, L., and Richardson, J. (2017). The importance of tools in the data lifecycle. *Digital Library Perspectives*, 33(3):235–252.

Declaration of independence

I hereby declare that I have prepared this paper independently, have not submitted it for testing purposes and have not used any other aids than those specified. All knowingly used text excerpts, quotations or contents of other authors have been explicitly marked as such.

Leipzig, March 19, 2020

Ronny Gey

Appendices

A. Validation of Interview Questions

B. Interview Guide

The contents...

C. Consent Form