

Hans-Christoph Hobohm (Hrsg.)

Informationswissenschaft zwischen virtueller Infrastruktur und materiellen Lebenswelten

**Information Science between
Virtual Infrastructure and Material Lifeworlds**

Unter Mitarbeit von Judith Peffing

Proceedings des 13. Internationalen Symposiums
für Informationswissenschaft (ISI 2013)

Potsdam, 19. bis 22. März 2013

vwh

Verlag Werner Hülsbusch
Fachverlag für Medientechnik und -wirtschaft

A Thematic Analysis to Facilitate Data Curation Relevance Decisions

The case of Research Data in Video Studies

Anke Reinhold¹, Marc Rittberger^{1,2}, Stefan Keil²

¹ German Institute for International Educational Research (DIPF)
Information Center for Education
Schloßstraße 29, D-60486 Frankfurt am Main
{reinhold|rittberger}@dipf.de

² Hochschule Darmstadt University of Applied Sciences
Department for Media, Information Science
Max-Planck-Straße 2, D-64807 Dieburg
thorsten.s.keil@stud.h-da.de

Zusammenfassung

In Infrastruktur-Einrichtungen stellt sich derzeit die Frage nach der Relevanz bestimmter Datentypen für die Kuratierung. Dies gilt besonders für audiovisuelle Forschungsdaten, die einen erheblichen Aufwand hinsichtlich Erschließung, Speicherung und Datenschutz mit sich bringen. Anhand einer thematischen Analyse der FIS-Bildung-Datenbank am Beispiel des Forschungsbereichs Unterrichtsvideografie lässt sich eine Zunahme an Forschungsaktivitäten im Feld, und somit auch der Relevanz der dort anfallenden Daten, feststellen. Zudem belegt die Untersuchung die Anwendbarkeit thematischer Analysen von Fachdatenbanken auf das Monitoring von Forschungsaktivitäten.

In: H.-C. Hobohm (Hrsg.). Informationswissenschaft zwischen virtueller Infrastruktur und materiellen Lebenswelten. Tagungsband des 13. Internationalen Symposiums für Informationswissenschaft (ISI 2013), Potsdam, 19.–22. März 2013. Glückstadt: Verlag Werner Hülsbusch, 198–209.

Abstract

Information service providers and infrastructure providers are increasingly dealing with questions regarding the data curation relevance of research data. The field of video studies in educational research provides a wide range of research data types. In addition, constraints in terms of data collection, storage and data privacy need to be considered. In order to analyze the suitability of video data for curation, a thematic analysis of the bibliographic German Education Index (GEI) was conducted. Demonstrating a continuous increase in research activity in the field, data curation for research data in video studies can be considered as relevant. Also, the approach demonstrates that thematic analyses of scientific databases are applicable to the monitoring of research activities.

1 Introduction

In the near future, information service providers and infrastructure providers might be facing a research ‘data deluge’ (e.g. Hey/Trefethen 2003) as demands for publishing and curating research data become more frequent, e.g. by the Wissenschaftsrat in 2011 or the European Commission in 2010. These demands, motivated by the vision of making research data available for evaluation purposes and secondary analyses, are aimed at the development of institutionalized research data management infrastructures.

However, before making research data available on a large scale, discipline-specific data curation relevance decisions need to be made. Data curation is a time-consuming and costly task, and researchers or institutions cannot curate all data types produced within a community. One option for facilitating this process is to monitor publication activity in data-driven areas of research. This allows conclusions to be drawn regarding the relevance of certain data types for secondary or re-analysis in a field. This article will focus on facilitating data curation relevance decisions for a specific area of educational research, namely video studies in educational research.

2 Data curation in educational research

2.1 Research data in educational research

Educational research examines the theory and practice of education and presents a heterogeneous research landscape with a wide range of methodological approaches. Consequently, many different types of research data are being collected, processed and analyzed. Following the so-called ‘empirical turn’ in the mid-1990s, triggered by disappointing results of German students in large-scale assessments like TIMSS¹ or later PISA², educational research in Germany has become more data-driven.

However, in educational research, data privacy issues need to be considered. Highly sensitive personal data are being collected, e.g. in schools or other educational contexts. Therefore, discipline-specific data curation³ policies need to strike a balance between research demands and data protection particularities. Due to its multi- and interdisciplinary nature which encompasses overlaps with social sciences, psychology, economics and history, the data output in educational research is very diverse, as summarized in Bambey et al. (2012). Typically, a differentiation is made between quantitative and qualitative research data. In the case of quantitative approaches, questionnaires are the main instruments of data acquisition. Qualitative data types include transcripts of interviews or videos as well as contextual information like field notes or research diaries.

For quantitative data in educational research in Germany, data curation infrastructures have already been established, e.g. the Research Data Centre at the Institute for Educational Quality Improvement (IQB), the four research data centres hosted by the Leibniz Institute for the Social Sciences (GESIS)

1 Trends in International Mathematics and Science Study. <http://timssandpirls.bc.edu/> <February 15, 2013>

2 Programme for International Student Assessment; <http://www.oecd.org/pisa/> <February 15, 2013>

3 Data curation activities are defined by the *Center for Informatics Research in Science and Scholarship* at the University of Illinois at Urbana-Champaign as activities which “enable data discovery and retrieval, maintain data quality, add value, and provide for re-use over time” (CIRSS 2012).

with its registration agency *da|ra* or the DaQS⁴ database offering access to questionnaires or scaling documentation.

With regard to qualitative research, promising frameworks are currently being established in the context of the eHumanities⁵ initiative by the Federal Ministry of Education and Research (BMBF) and the Thematic Information Networks Programme⁶ of the German Research Foundation (DFG). However, no centralized data curation infrastructure exists for qualitative data types in educational research. Thus, information service and infrastructure providers need to identify relevant qualitative data types suitable for data curation.

2.2 Video studies in educational research

An interesting case with regard to data curation is the field of video studies.⁷ It deals with the topic of instruction/didactics which has been identified as a popular topic in educational research, as analyzed by Schulzeck (2008: 37). Furthermore, the domain offers a complex data structure by providing a multitude of common data types in educational research, e.g. videos, transcripts, questionnaires or field notes. The field uses video technology to document instruction practices (mostly in schools) with the aim of scientific analysis as well as usage in teacher training. For research purposes, both qualitative and quantitative approaches are common to the analysis of videotaped lessons. International studies, e.g. Pythagoras⁸ or TIMSS, often use quantitative approaches to analyze video data.

Regardless of methodological aspects, curating data in the field might challenge data curation providers in several ways with regard to: 1) the high data storage volume of the videos, 2) the above-mentioned diversity of data types and 3) the problems of data privacy. In addition, as producing video

4 Datenbank zur Qualität von Schule. <http://daqs.fachportal-paedagogik.de/> <February 15, 2013>

5 http://pt-dlr-gsk.de/_media/Bekanntmachung_eHumanities_10_05__2011.pdf <February 15, 2013>

6 http://www.dfg.de/formulare/12_12/12_12.pdf <February 15, 2013>

7 In literature, the term videography which often implies an ethnographic approach to research is also used to describe a video study. Here, both terms are used as synonyms.

8 Unterrichtsqualität und mathematisches Verständnis in verschiedenen Unterrichtskulturen

data is a labor-intensive and expensive task, data sharing and secondary analyses of video-taped instruction might offer important options for researchers and institutions.

As described below, in order to gain certainty about the relevance of research data for data curation in the field of video studies, a thematic analysis of the German Education Index (GEI) was carried out.

3 Thematic analysis of the German Education Index (GEI) in the context of video studies

In contrast to formal bibliometric analyses which mostly focus on authors, their affiliations and geographical indicators, a thematic analysis focuses on specific topics and can serve to analyze publication activity in an area of research (Tunger 2009: 70–75). Consequently, a thematic analysis of the GEI with regard to data curation relevance decisions is aimed at finding out whether there is an increase in publication activity in the field of video studies. Accordingly, indications for a growing need to provide an infrastructure for the domain-specific data types can be formulated.

The GEI is a database hosted by the German Institute for International Educational Research (DIPF). Its content of about 800 000 bibliographic records from educational research and practice is collaboratively set up by around 30 documentation centers. The GEI is suitable for a thematic analysis of the field in Germany for several reasons. First, it covers a broad range of relevant topics in educational research. Second, the GEI contains diverse publication types. According to Dees (2008: 29), the most frequently used publication types in German educational research are ‘excerpts from collections’ (46.7%), followed by journal articles (33.4%), books (14.8%) and other publication types (5.1%).⁹ Therefore, the analysis of citation databases like Web of Science with a focus on journal articles (with regard to educational research in Germany) is unsuitable for a thematic analysis of the field. In this case, it is likely that too many relevant documents would be ignored. Only 60 of 505 identified relevant journals in German educational research

⁹ As Dees (2008: 29) notes, these numbers correspond to other findings from similar studies.

are included in the indexes of the citation database Web of Science as illustrated by Dees (2008: 30).

Third, the GEI incorporates references in different languages with a focus on German. As Dees (2008: 30) points out, educational researchers from Germany seldomly publish in English or other languages. In view of these facts, analysis of reference databases in the international context could lead to an incomplete data set. In summary, the GEI provides an adequate database for a thematic analysis in educational research.

3.1 Setup of the collection¹⁰

In order to develop adequate queries for the thematic analyses, the particularities of the GEI need to be considered, i.e. the heterogeneous indexing due to complexity of the domain and collaborative setup as well as the acquisition process, based upon voluntary data delivery (Dees/Botte 2013: 5). Thus, there are three restrictions to a thematic analysis: 1) the controlled terms of the GEI vocabulary are problematic for selecting documents by content-driven criteria; 2) the observation period needs to be restricted due to a possible lack of data timeliness, and 3) the validity of the results may be weakened by missing data. Still, the GEI serves as an important basis for bibliometric analyses in the field, e.g. (Dees/Botte 2013; Martini/Dees 2013).

Due to the aforementioned limitations of the GEI for a thematic analysis, a co-occurrence based approach was chosen in order to identify descriptors describing the field of video studies. Co-occurrence is defined as a “common occurrence of certain descriptors with proper names of video studies e.g. ‘TIMSS’ or ‘DESI’¹¹”. Therefore, every descriptor that is assigned to a document dealing with a well-known study which uses videography as a research method was examined for relevance in describing the research field. Using a building block approach based on the descriptors which were identified as relevant, two queries which differ by their specification level were developed. It was thus possible to find out how many relevant documents are

10 The authors would like to give credit to Jonathan Geiger, graduate at the Darmstadt University of Applied Sciences, for developing a baseline for the chosen queries of the thematic analysis.

11 DESI (Deutsch-Englisch-Schülerleistungen-International)

delivered by each query. Query A is broader¹² than query B in the selection of relevant terms regarding the topic in question. Both result sets were then filtered for the publication years 1990 until 2011.¹³ As the TIMSS study started in 1995 and can be interpreted as a landmark study, all documents relating to video studies in a five-year period before TIMSS are taken into account.

At this point, query A delivered 371 and query B 155 documents respectively. Afterwards, two raters with expertise in the field of video studies intellectually rated the relevance of the documents. To provide consistency throughout the rating process, relevant documents were applied to match the following criteria: The document must deal with videography of instruction used in research or in teacher trainings and must not be informal (e.g. a leaflet about education programs for teachers). Following these criteria, a precision of 79 per cent could be measured for query A (293 out of 371 documents rated as relevant) and of 94 per cent for query B (146 out of 155 documents rated as relevant).

3.2 Thematic analysis of the GEI

For the analysis a combined set of 301 bibliographic records resulting from query A and B which had been judged as relevant was selected. These bibliographic records were compared to a combined set of documents generated with the same queries but without the limitation to the usage of video techniques in observation of instruction. A benchmark for the increase in publication activity in comparison to a subordinate theme can be defined as described by Tunger (2009: 75).

Figure 1 shows the publication activity of researchers in the field of instruction without any specification of the research method compared to all bibliographic records in the GEI as well as the publication activity in the

12 For example, the broadness of query A in comparison to query B results from selecting truncated terms like “*unterricht” for all kind of school subjects or a truncated search for “*video*” in all fields of a database record. Query B aimed at selecting more category-like terms for a search.

13 Limitations of timeliness and coverage of the GEI should be taken into account for 2008 to 2011. Thus, data should only be interpreted as slight tendencies. Missing data will be added in routine updating procedures.

field of video studies of instruction compared to all bibliographic records dealing with observation of instruction.¹⁴

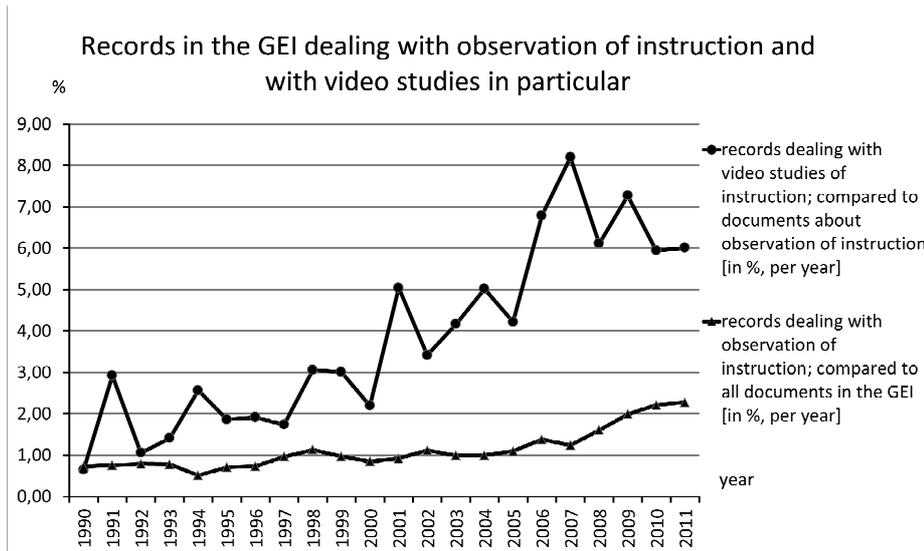


Fig. 1 Development of bibliographic records in the GEI with regard to observation of instruction and video studies of instruction

First, a steady rise in the percentage of records dealing with the observation of instruction compared to all records in the GEI is evident. Second, a parallel increase in the percentage of records indicating the usage of video techniques to gain research data is clearly visible. Here, a more irregular rise with some peaks in 1991, 2001 and 2007 is shown.

In a complementary analysis, the data discussed above were combined with the results of work by Martini/Dees (2013) and Martini/Schmidt-Hertha (2012) in order to confirm a positive trend with regard to publication activity of doctoral and postgraduate students in the field of video studies.¹⁵ These

14 For example, a total of 238 records (from 25961 in the GEI) published in 2001 dealt with observation of instruction (0.92%) whereas 12 records out of those 238 dealt with video studies (5.04%). A rise in activity relating to observation of instruction and video studies could be measured as a total of 564 records (from 28341 in the GEI) published in 2009 dealt with observation of instruction (1.99%) whereas 41 out of those 564 records dealt with video studies (7.27%).

15 The publication of postgraduate theses is not compulsory in Germany. Thus, it can be assumed that not all of such theses have been recorded in the sources selected.

works examined the development of qualification theses in educational research in Germany between 1998 and 2009. Their sources were bibliographic records on theses from the GEI, the catalogue of the German National Library as well as announcements of qualification theses in *Zeitschrift für Pädagogik*, a peer-reviewed journal of high relevance in German educational research. Due to data extraction from multiple sources and its intellectual treatment e.g. in recognizing duplicates, the collection is very exhaustive and of high quality.

For the analysis, a subset of our data – namely the records being marked as qualification theses – was added to a collection of theses dealing with video studies gained from the data of Martini/Dees (2013). The observation period was limited from 1998 to 2009.¹⁶ Altogether 47 theses dealing with video studies of instruction could be identified. As shown in Figure 2, a substantial rise in theses dealing with video studies of instruction since 2003 can be observed.

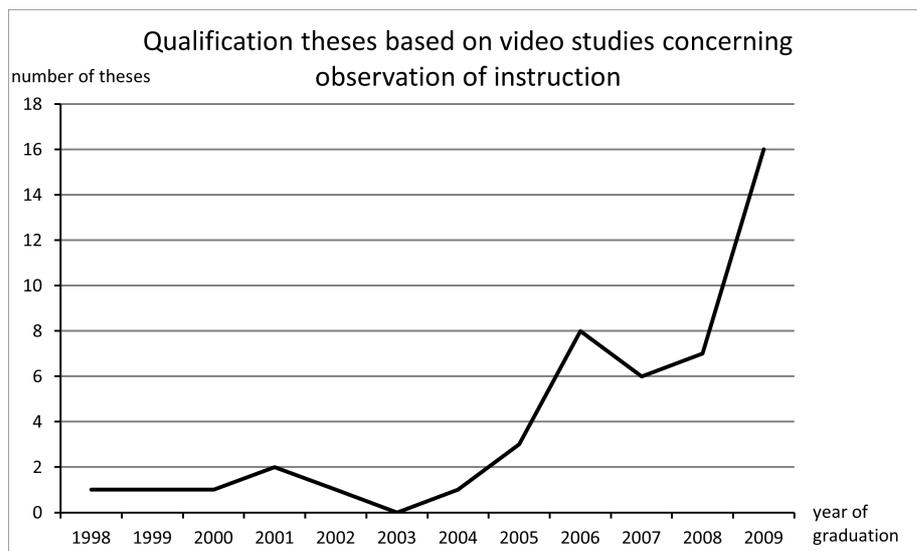


Fig. 2 Qualification theses based on video studies of instruction

In summary, these thematic analyses imply a positive trend with regard to publication activity in the field of video studies in German educational re-

¹⁶ In case a thesis was not published in the year of graduation, the year of graduation was used as a time mark.

search. As follows, results need to be critically discussed. Firstly, with regard to occurrences which might have an impact on publication activity and secondly, with emphasis on data curation relevance decisions.

3.3 Discussion of results

Publication activity in the field of videography of instruction has experienced a positive trend since the early 1990s with a peak in 1991. This is clearly due to technological progress in video technology, e.g. S-VHS and Hi8. However, until 2000 – despite the popularity of the TIMSS study conducted in 1995 – researchers in the field of classroom observation still preferred alternative data collection methods over videography.

Prior to PISA 2000 and other large scale assessments, there was a weaker demand to produce empirical data in educational research. As a consequence, access to the field might have been comparably difficult for researchers interested in video studies. The picture changes after the empirical turn in educational research described in chapter 1. The publication activity in video studies experiences a further growth, with another peak in 2001.

Again from 2005, a substantial and steady rise in video-related publications can be observed, reaching a peak in 2007, which might have its origin in a boost of research activity in the aftermath of the Pythagoras and IPN video studies¹⁷. In the context of both studies, detailed manuals for the practical setting of a video study as well as fundamental work on theoretical categories of analysis were published. The interruption in 2005 is possibly attributable to a stronger awareness of data protection issues. For example, in 2006, the German Educational Research Association (GERA 2006) published guidelines for enhanced data protection in video studies and other qualitative approaches.

The general picture of a growth in significance of publications in the field of video studies is supported by a rise in publication activity with regard to qualification theses. Video studies seem to have become a popular approach for educational purposes. Here, an interesting application area can be identified for secondary analyses.

17 Lehr-Lernprozesse im Physikunterricht

4 Conclusion and outlook

In this article, a method for facilitating data curation relevance decisions in educational research was successfully applied to the area of video studies. As a result, by demonstrating a steady rise in research activity in the field, data curation for research data in video studies can be considered as relevant. This implies a growing need for metadata development and clarification of data protection legislation in the context of audiovisual data in educational research as well as qualitative research in general. In addition, the article offers insights into the scope of research data potentially available for secondary analyses, data sharing and/or data curation in the field. Furthermore, potential accelerators or barriers to data production could be identified.

The methodology provides general implications for query generation and the setup of data collections with regard to thematic analyses. The prior identification of landmark studies in the field enabled the formulation of very precise queries. As the positive trend in research activities in video studies was confirmed by the complementary analysis of doctoral theses – based on a corpus of high quality – a certain validity of the approach can be assumed.

The approach presented could potentially be transferred to other areas of educational research or other disciplines. Certainly, some limitations of the study need to be outlined. Due to the small sample size that might occur frequently in thematic analyses, generalizations of the results need to be treated with caution. Further analyses of processes and information needs with respect to secondary analysis and data sharing are required. Also, user-centered studies are necessary to identify relevant data types, quality criteria, access conditions and data sharing practices in the field. This could enhance data usage practices with a positive effect on research quality as a whole.

References

- Bambey, D.; Reinhold, A.; Rittberger, M. (2012). Pädagogik und Erziehungswissenschaft. In: Neuroth, H.; Strathmann, S.; Oßwald, A.; Scheffel, R.; Klump, J.; Ludwig, J. (Hrsg.). Langzeitarchivierung von Forschungsdaten. Eine Bestandsaufnahme. Boizenburg: Hülsbusch, 111–135.

- CIRSS (2012). Data Curation Education Program. Center for Informatics Research in Science and Scholarship. <http://cirss.lis.illinois.edu/CollMeta/dcep.html> <16.01.2013>.
- Dees, W. (2008). Transparenz und Evaluierbarkeit des erziehungswissenschaftlichen Publikationsaufkommens – eine anwendungsorientierte Studie. In: *Erziehungswissenschaft* 19 (37), 27–32. <http://www.pedocs.de/volltexte/2010/2485/> <16.01.2013>.
- Dees, W.; Botte, A. (2013). *Publikationspraxis in der Bildungsforschung 1998–2007*. Frankfurt a. M.: DIPF. <http://www.pedocs.de/volltexte/2013/7047> <16.01.2013>.
- European Commission (2010). *Riding the wave. How Europe can gain from the rising tide of scientific data. Final report of the high level Expert Group on Scientific Data*. European Union. <http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf> <16.01.2013>.
- GERA (German Educational Research Association) (2006). Anonymisierung von Daten in der qualitativen Forschung. Probleme und Empfehlungen. In: *Erziehungswissenschaft*, 17 (32), 33–34.
- Hey, A. J. G.; Trefethen, A. E. (2003). The Data Deluge: An e-Science Perspective. In: Berman, F.; Fox, G. C.; Hey, A. J. G. (eds.). *Grid Computing – Making the Global Infrastructure a Reality*. New York: Wiley and Sons, 809–824.
- Martini, R.; Dees, W. (2013). *Qualifizierungsarbeiten in der Bildungsforschung 1998–2007*. Frankfurt a. M.: DIPF. <http://www.pedocs.de/volltexte/2013/7048> <16.01.2013>.
- Martini, R.; Schmidt-Hertha, B. (2012). Förderung des wissenschaftlichen Nachwuchses. In: Thole, W.; Faulstich-Wieland, H.; Horn, K.-P.; Weishaupt, H.; Züchner, I. (Hrsg.). *Datenreport Erziehungswissenschaft 2012*. Opladen: Verlag Barbara Budrich (2012), 181–192.
- Schulzeck, U. (2008). *Bildungsforschungsprojekte im Forschungsinformationssystem Sozialwissenschaften (FORIS)*. In: Weishaupt, H. et al. (Hrsg.). *Zur Situation der Bildungsforschung in Deutschland*. Berlin: Bundesministerium für Bildung und Forschung (BMBF), 36–55. http://www.bmbf.de/pub/bildungsforschung_bd_achtundzwanzig.pdf <16.01.2013>.
- Tunger, D. (2009). *Bibliometrische Verfahren und Methoden als Beitrag zu Trendbeobachtung und -erkennung in den Naturwissenschaften*. Dissertation, Universität Regensburg. <http://nbn-resolving.de/urn:nbn:de:bvb:355-opus-12785> <16.01.2013>.
- Wissenschaftsrat (2011). *Recommendations on Research Infrastructures in Humanities and Social Sciences*. Wissenschaftsrat. http://wissenschaftsrat.de/download/archiv/10465-11_engl.pdf <16.01.2013>.