

Bedankt voor het downloaden van dit artikel. De artikelen uit de (online)tijdschriften van Uitgeverij Boom zijn auteursrechtelijk beschermd. U kunt er natuurlijk uit citeren (voorzien van een bronvermelding) maar voor reproductie in welke vorm dan ook moet toestemming aan de uitgever worden gevraagd.

Boom

Behoudens de in of krachtens de Auteurswet van 1912 gestelde uitzonderingen mag niets uit deze uitgave worden verveelvoudigd, opgeslagen in een geautomatiseerd gegevensbestand, of openbaar gemaakt, in enige vorm of op enige wijze, hetzij elektronisch, mechanisch door fotokopieën, opnamen of enig andere manier, zonder voorafgaande schriftelijke toestemming van de uitgever.

Voor zover het maken van kopieën uit deze uitgave is toegestaan op grond van artikelen 16h t/m 16m Auteurswet 1912 jo. Besluit van 27 november 2002, Stb 575, dient men de daarvoor wettelijk verschuldigde vergoeding te voldoen aan de Stichting Reprorecht te Hoofddorp (postbus 3060, 2130 KB, www.reprorecht.nl) of contact op te nemen met de uitgever voor het treffen van een rechtstreekse regeling in de zin van art. 16l, vijfde lid, Auteurswet 1912.

Voor het overnemen van gedeelte(n) uit deze uitgave in bloemlezingen, readers en andere compilatiewerken (artikel 16, Auteurswet 1912) kan men zich wenden tot de Stichting PRO (Stichting Publicatie- en Reproductierechten, postbus 3060, 2130 KB Hoofddorp, www.cedar.nl/pro).

No part of this book may be reproduced in any way whatsoever without the written permission of the publisher.

info@boomamsterdam.nl
www.boomuitgeversamsterdam.nl

EDITORIAL

Qualitative research in the digital humanities

Reinoud Bosch*

In a number of ways, this is an exceptional issue of KWALON. First, because the issue is entirely in English. The choice to do this was made at the insistence of the field. Simply put, it would not have been possible to put this issue together as it is, had it not been in English. Hopefully, this was the right choice to make. In any case, it is an exception: KWALON is a Dutch-language journal and will remain so in the future.

Another exceptional aspect of this issue is that it concerns the humanities. KWALON mostly publishes social science articles, although there is some overlap with the humanities in specialized fields such as discourse analysis, historical approaches, phenomenology, and, of course, hermeneutics. In my (hermeneutic) view, there is no reason why a journal concerned with qualitative methodology should not publish articles concerned with the humanities. After all, qualitative research methodology originated in the humanities long before Dilthey introduced it explicitly into the social sciences (Bosch, 2012). So giving the humanities some special attention in KWALON only appears natural.

A further exceptional aspect of this issue is that it deals with the use of digital tools in the humanities. KWALON has a long history of discussing digital tools in the social sciences, but this mostly concerns computer assisted qualitative data analysis software. A particularly interesting current debate in the digital humanities is rather concerned with the development and application of algorithms, a debate one does not yet often encounter in KWALON. Computer science in general is not foreign to qualitative research: one of the editors of the excellent *The SAGE Handbook of Grounded Theory* (Bryant & Charmaz, 2007) is a professor of informatics (Bryant), and, indeed, there is no reason to think that careful, systematic interpretation would not be required in various fields of computer science dealing with such issues as artificial intelligence, human-computer interaction, semantics, programming languages, and 'big data'. It is precisely the question how 'big data' can be interpreted carefully and systematically that stimulated putting this issue together. Last, but not least, this issue is exceptional because of its contributing authors – all true experts in their respective fields of expertise. Their expertise has made putting together this issue a very pleasant and rewarding experience. Hopefully, reading the issue will be similarly pleasant and rewarding.

* Dr. Ing. Reinoud Bosch is coordinating editor-in-chief of KWALON.
Email: reinoudbosch@hotmail.com.

Some questions regarding qualitative research in the digital humanities

This issue aims to address a number of questions concerning the possibilities and requirements for qualitative research in the digital humanities. First, digitization has made an increasing amount of relevant material easily accessible to qualitative researchers in the humanities. Which questions does this raise for the humanities in general, and for qualitative researchers in particular? Which IT tools are available to qualitative researchers for dealing with this huge increase in easily available information? And how can qualitative researchers use IT in a responsible manner? How can we make sure that all this data is *interpreted* – rather than simply crunched – in an open, honest, and responsible fashion?

Qualitative research is concerned with the interpretation of meaning. It has become accepted, at least in Dutch qualitative research circles, that systematic qualitative research is based on three core methodological components: constant comparison (Weber, 1904; Glaser & Strauss, 1967), analytic induction (Znaniecki, 1934), and theoretical sensitivity (Glaser, 1978). For those with a mixed-methods inclination, this can include results from quantitative research, for example leading to a systematic holistic hermeneutic methodology based on selection, comparison, categorization, theory development, and synthesis (Bosch, 2007, 2012, in press).

There is clear potential for the use of algorithms in this type of research method. Firstly, can algorithms help in the interpretation of meaning implicit in large amounts of available material? This relates to the way in which meaning is accorded during categorization, and the way in which meaning arises out of the overall analysis. Secondly, how can we best find or develop relevant, plausible, and accurate empirical and theoretical material? As Kahneman (2011) has argued, there are two ways to go about this: thinking fast and thinking slow. In thinking fast, simplifying heuristics are used to reach decisions – such as decisions needed in finding or developing relevant, plausible, and accurate theoretical and empirical claims. In thinking slow, a more extensive iterative approach is used. Thinking slow may lead to more accurate results, but the process is lengthier and more complicated than thinking fast. Suggesting that there is room both for thinking fast and thinking slow in qualitative research is in line with Kahneman's views. Just like in thinking slow, qualitative research methodology is iterative. Can algorithms contribute here? Of the utmost importance in the implementation of algorithms is that the researcher should always have the feeling of having control over and an understanding of the research process. Do and can algorithms allow for this, and if so, how?

Over the past decade or so, the use of Computer Assisted Qualitative Data Analysis Software (CAQDAS) has become commonplace in qualitative research. What is the current state of this type of software? Can it handle large amounts of data, and if so, how? Which tools are available in CAQDAS to integrate qualitative and quantitative methods in handling 'big data'? Qualitative research has always required data management planning, repository, and curation. This is not new. But what are the implications of digitization and 'big data'? Which specific issues arise? And, last but not least, how is qualitative research in the digital humanities currently being con-

ducted in practice? What are some of the issues and the possibilities that are encountered? What could the future bring?

An overview of the contributions to this issue

In the opening article, Van Dijck points out that digitization together with the development of IT tools has allowed a wider scope and an increased complexity of research in the humanities – research pursued under the heading of ‘Digital Humanities’. As Van Dijck indicates, doing this type of research requires multi-disciplinary cooperation between humanities scholars and computer and social scientists. In this process, Van Dijck finds four challenges. First, there is a fear of a crowding out of qualitative methods. A second challenge is determining which particular methods to use in specific cases to arrive at the best possible results. The third challenge lies in using common curiosity and an interest in each other’s expertise as a basis for multidisciplinary cooperation aimed at customized interpretation. And the final challenge is concerned with finding a balanced way of curating material and digital sources in a responsible fashion, and clarifying the way in which search algorithms and storage machines inform selections and choices.

What is required, in Van Dijck’s view, is experimentation with large research questions about cultural complexity and cultural change on the basis of computer-mediated access to a wide range of source material. In this process of experimentation, there is more than ever a need for critical interpretation and qualitative analysis. As Van Dijck puts it: “society direly needs the expertise of humanities scholars – their critical insights, analytical acuity, and knowledge of ambiguity and diversity – to make sense of a digital culture that permeates and directs our daily life.”

In the next contribution to this issue, Vossen argues for the importance of making subjective annotations explicit, in order to allow modelling for the purpose of building machines that can interpret language in the same way as humans do. Machines can mimic human interpretations by making such machines associate symbols or signals with annotated interpretations of these symbols. Different human annotations can lead to different machines with different outcomes. On the basis of available annotations, machines can then perform annotations themselves, when they are provided with algorithms to compare features of unlabelled signals with those of labelled ones. A challenge lies in the incompleteness, lack of clarity, and subjectivity of annotations. What is essential, according to Vossen, is not to simplify or reduce the interpretation, but to control the process of interpretation and make the result explicit.

Bosch and Verborgh propose an iterative mixed-methods research cycle as an approach to querying the semantic web – the part of the web that can be automatically processed by machines. They argue that the logical steps in this iterative research cycle could in principle be codified into algorithms. To give an indication of what this could look like in practice, they present practical steps that have been taken in accordance with an iterative approach to querying the semantic web – in the form of a dynamic iterator pipeline that has been developed for efficient and

effective queries on the semantic web. Developing the logic of the iterative research cycle could be advanced by providing detailed and systematic answers to the question of how researchers go about answering complex questions by combining information from different sources on the web.

In the next contribution, Friese provides an overview of the state-of-the-art of CAQDAS. Using Kahneman's ideas about fast and slow thinking as a framework, Friese points out that most automated analysis options of CAQDAS programs are based on fast thinking – some providing the possibility for researcher input leading to algorithm adjustment. Friese also notes an increase in the number of mapping tools and tools based on the quantification of qualitative data. In her evaluation, Friese finds that for decent results, automatic coding currently still needs to be followed up by close reading of the data. Currently available CAQDAS programs cannot yet handle large amounts of data, and automated theme recognition and sentiment coding do not yet work satisfactorily. Friese argues in favor of using specialized tools for the analysis of large data sets, combined with import and export options into and from CAQDAS programs. What is needed, in Friese's view, is to find a way to integrate close reading and small data with distant reading and large data.

Van Horik discusses data management planning and trusted data repositories, needed for the testing of research outcomes and the reuse of data for new research. A data management plan must include comprehensive information about the data, such as data types, the metadata standards used, the policies and facilities for access and data sharing, and the plans for data archiving and preserving. Special attention needs to be paid to license issues, the processing of data sets, and proprietary database formats. Trusted digital repositories (TDRs) exist that provide reliable, long-term access to managed digital resources to designated user communities. Such TDRs must fulfill a number of requirements, including making data findable on the Internet; conforming to legislation with regard to personal information and intellectual property; making data available in a usable format; guaranteeing the reliability of data over time, and providing facilities for stable and robust references to the data sets.

In the final contribution, Noordegraaf notes a gap between the affordances of the increased digitization of sources and the development of analytical tools, and their application in the field of media studies. Digital data and computational tools are being used in textual analysis of media content, and in social and economic historical research into the distribution and reception of media content. But the availability of comprehensive, reliable, standardized, and copyright-free datasets, audiovisual sources, and textual materials is still rather limited. Moreover, most computational tools do not facilitate a critical engagement with culture. According to Noordegraaf, the potential for broadening the scope and complexity of media studies research can be reaped if media scholars invest in digital literacy and methodological awareness, and collaborate with computer scientists and heritage institutions to increase the transparency of analysis tools.

Where do we stand?

Together, then, the contributors to this issue see great opportunities for qualitative research in the digital humanities. Research can be widened in scope and increased in complexity, human interpretations can be mimicked, qualitative research steps can be codified, and trusted digital repositories can offer support. But a number of challenges are recognized: the risk that qualitative methods are crowded out by quantitative ones; the incompleteness, ambiguity, and subjectivity of annotations; the continuing need for close reading; the limitations to processing and automated coding capacities of software; license and data set processing issues; proprietary database formats; a lack of application and transparency of tools; a limited availability of comprehensive, reliable, standardized, and copyright-free datasets and audiovisual sources and texts; and a lack of tools facilitating a critical engagement with culture.

The authors mention several steps that are needed to enhance the potential of qualitative research in the digital humanities. Methods, annotations, and algorithms need to be improved, clarified, and made explicit; and there is a need for respectful multidisciplinary cooperation and experimentation with large research questions. Rather than simplifying the process of interpretation, it needs to be made explicit and open to control. The importance of close reading should not be neglected. So, according to the contributors, qualitative research in the digital humanities holds great promise if the right steps are taken to respond to the challenges ahead.

But with all this attention to digitization and IT tools, one thing should not be forgotten: Life is organic; computers are made of dead inorganic material only 'brought to life' by what human beings do with them. If we want to understand our fellow human beings and their meaningful experiences and creations, we should never forget the importance of personal inter-human contact and interpretation.

References

- Bosch, R. (2007). Pragmatism and the practical relevance of truth. *Foundations of Science*, 12(3), 189-201.
- Bosch, R. (2012). *Wetenschapsfilosofie voor kwalitatief onderzoek*. The Hague: Boom Lemma uitgevers.
- Bosch, R. (in press). *Power: A conceptual analysis*. The Hague: Eleven International Publishing.
- Bryant, A. & Charmaz, K. (Eds.). (2007). *The SAGE Handbook of Grounded Theory*. Los Angeles: SAGE.
- Glaser, B.G. (1978). *Theoretical Sensitivity: Advances in the methodology of grounded theory*. Mill Valley, CA: Sociology Press.
- Glaser, B.G. & Strauss, A.L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Chicago: Aldine Publishing.
- Kahneman, D. (2011). *Thinking, fast and slow*. New York: Farrar, Straus & Giroux.
- Weber, M. (1904). Die 'Objektivität' sozialwissenschaftlicher und sozialpolitischer Erkenntnis. *Archiv für Sozialwissenschaft und Sozialpolitik*, 19(1), 22-87.
- Znaniecki, F. (1934). *The method of sociology*. New York: Farrar & Rinehart.