

The Virtual Knowledge Studio for the Humanities and Social Sciences

Paul Wouters

Networked Research and Digital Information (Nerdi), NIWI-KNAW, The Royal Netherlands Academy of Arts and Sciences, PO Box 95110, Amsterdam, Netherlands

Email: paul.wouters@niwi.knaw.nl

Abstract

This paper presents the goals and theoretical underpinning of a new national programme in *e*-social science in the Netherlands. Recent transformations in communication and information exchange have created new opportunities for researchers in the humanities and social sciences. It is not self-evident, however, in what ways scholars can best use these possibilities while maintaining and further developing their specific roles in academia and society. This is the rationale of a new national programme in the Netherlands, *The Virtual Knowledge Studio for the Humanities and Social Sciences*, hosted by the Royal Netherlands Academy of Arts and Sciences. It aims to support researchers in the social sciences and humanities in the creation of new scholarly practices, termed *e*-research, as well as in their reflection on *e*-research in relation to the development of their fields. A core feature of the Virtual Knowledge Studio is the integration of design and analysis in a close cooperation between social scientists, humanities researchers, information technology experts and information scientists. This integrated approach should provide insight in the way *e*-research can contribute to new research questions and methods in the humanities and social sciences.

Introduction

This paper presents the goals and theoretical underpinning of a new national programme in *e*-social science in the Netherlands. *e*-Science is generally defined as the combination of three different developments: the sharing of computational resources, distributed access to massive datasets, and the use of digital platforms for collaboration and communication (Hey & Trefethen, 2002; Nentwich, 2003). The precise definition and terminology varies between the UK, the US and the Netherlands, which inter alia illustrates the local nature of this type of global developments. Nevertheless, these three elements are generally recurring in *e*-science projects and programmes. In the Dutch initiatives of *e*-science, the *e* stands not in the

first place for “electronic” but for “enhancement”¹. The core idea of the *e-science* movement (most of it is still promise rather than practice) is that knowledge production will be enhanced by the combination of pooled human expertise, data and sources, and computational and visualisation tools. In this paper, the notion of *e-research* rather than *e-science* is used to indicate that it is not a matter of importing *e-science* ways of working into the social sciences and humanities. The humanities and social sciences will develop their own specific ways of integrating the use of networked information and communication technologies (Bijker & Peperkamp, 2002; Bijker et al., 2003; Boonstra, Breure, & Doorn, 2004; Kircz, 2004). This does not have to mean that the difference with natural sciences will become less important. Hence, the generic term *e-research* is preferable over the notion of *e-science*.

Programme mission

Recent transformations in communication and information exchange have created new opportunities for researchers in the humanities and social sciences. It is not self-evident, however, in what ways scholars can best use these possibilities while maintaining and further developing their specific roles in academia and society. This is the rationale of a new national programme in the Netherlands, *The Virtual Knowledge Studio for the Humanities and Social Sciences*, hosted by the Royal Netherlands Academy of Arts and Sciences. It aims to support researchers in the social sciences and humanities in the creation of new scholarly practices, termed *e-research*, as well as in their reflection on *e-research* in relation to the development of their fields.

A core feature of the *Virtual Knowledge Studio* is the integration of design and analysis in a close cooperation between social scientists, humanities researchers, information technology experts and information scientists. The programme is integrated at all levels: its goal and mission, its research projects, its internal peer review, its funding acquisition, its collaboration and publication policies and its data management. This does not mean that disciplinary differences will become invisible. On the contrary, we rather expect that these distinctions will be productive and contribute to that creative tension which is the hallmark of innovative research and scholarship. This integrated approach should thereby provide insight in the way *e-research* can contribute to new research questions and methods in the humanities and social sciences.

The *Virtual Knowledge Studio* has the following goals:

- to contribute to the design and conceptualisation of novel scholarly practices in the humanities and social sciences
- to support scholars in their experimental play with new ways of doing research and emerging forms of collaboration and communication
- to facilitate the travel of new methods, practices, resources and techniques across different disciplines
- to contribute to a better understanding of the dynamics of knowledge creation.

¹ See <http://www.wtcw.nl/nl/projecten/eScience.pdf>.

These goals are intimately connected, one cannot be reached without the other. Moreover, they are not only relevant to the humanities and social sciences but to the development of - and reflection on – present-day knowledge societies.

A systematic and critical interrogation of the potential of *e*-research paradigms and methodologies for the humanities and social sciences has been hampered by disciplinary boundaries between fields, by a relative lack of resources and research infrastructures, and by the dominance of particular computational approaches in the world of *e*-science. The Studio will address these problems by:

- demonstrating and exploring the potential of additional, non-computational as well as computational, ways of doing *e*-research
- making disciplinary boundaries more permeable for new scholarly practices
- pooling resources that are available to the scholarly communities in the Netherlands and abroad.

The humanities and social sciences are not a unified set of knowledge practices, as is well known. Methodologies and techniques do not travel easily from one field to another. This has a direct bearing upon the development of *e*-research tools, practices, infrastructures and institutions. For example, in fields like philosophy, art history or literature studies, scholars even may not be aware of *e*-science as relevant to them at all. The equipment of many academic scholars with the tools and “play space” they need to independently assess the merits of *e*-research has moreover been hindered. Lack of funding has decreased the space for advanced instrumentation and support staff. Moreover, ICT has been standardised within the paradigm of office automation and therefore lacks many features that would be useful for scholarly work. In so far as ICT has been tailored to research needs, it has been based on computational research and often assumes mathematical and programming skills on the part of the researcher. In many fields scholars have different needs, such as the representation of ill-defined data, analysis-oriented visualisation of manuscripts and multimedia sources, and specific source-oriented analytical tools. These needs are often not met by standard computational and mathematical analytic methods.

The meeting of *e*-research and the academic scholar is moreover problematic because it is far from clear whether the present needs of the scholar can be met by *e*-research at all. Important fields in the humanities and social sciences are characterised by a huge epistemic diversity; by specific, sometimes person-bound, roles of the researcher; by the lack of consensus about the research agenda in a host of specialties; by a relatively low-tech research environment (often aggravated by the scarcity of university funding); by the specificity of writing and reading as features of knowledge creation; and by a historically grounded and relatively large share of solitary research practices (Becher 1989; Whitley 2000). In all these dimensions, many fields seem ill-suited to become enthusiastic adopters of the *e*-science paradigm as it now stands. If *e*-research should make sense to a variety of specialties in the humanities and social sciences, new non-computational and computational paradigms of *e*-research need to be developed.

The Studio will therefore orient itself to a critical interrogation of the very notion of *e*-research, by taking seriously the intellectual and social characteristics of the humanities and social sciences and the implications of these characteristics for the hermeneutics of *e*-research as a prospective intellectual and technical horizon. At the

same time, existing knowledge practices in the humanities and social sciences should not be taken for granted. There is ample space for enhancement indeed.

Research problem

The creation of knowledge as a social and cultural process is the research object of the Virtual Knowledge Studio. In the context of *e*-research, it is tempting to focus on the digital technologies. Such a singular approach would ignore, however, the epistemic dimension of knowledge practices. Scholarly and scientific developments are intellectually codified, so that they are relatively resistant to external steering (Van den Daele, Krohn and Weingart 1977; Weingart 1974; Whitley, 2000), but not to technological change (Joerges and Shinn 2001). The study of the interaction between users, designers and technological artefacts is the topic of a number of well-established fields of research: humanities computing, social informatics, technology studies, human computer interaction, computer supported cooperative work, and innovation studies. And although the researchers at the Studio will be able to draw upon this knowledge, its research agenda should not be limited to these types of user-oriented paradigms. The central problem in the development of *e*-research is not the technology, nor the role of the user in technological environments, but the cultural and historical specificity of knowledge production in the new technologically mediated contexts. This is the reason the Studio puts knowledge creation centre stage.

The central research question of the Studio is *how it is possible to develop novel ways of knowledge creation in the humanities and social sciences by utilizing and adapting e-research concepts, instruments and ways of working*. This includes the epistemic and cultural effects of *e*-research on the humanities and social sciences. In terms of *e*-research: are the social sciences and humanities susceptible to enhancement and what would enhancement mean for the nature and role of academic scholarship?

This approach contrasts with three more limited approaches:

- It is not in the first place a matter of changing the humanities and social sciences to make them fit a particular model of large-scale, data oriented research
- It is not only a matter of creating information technologies that better fit the needs of scholars in the humanities and social sciences
- It is not simply a matter of responding to the perceived needs of scholars with new tools, tailor-made software and research instruments.

All three aspects are part of the problem of *e*-research, and will provide important building blocks of the Studio, but none of them are the central issue.

Studio design

The central research question will be tackled by developing analysis and design in three intellectual dimensions. The first dimension consists of the topical research themes, the second dimension is the development of novel methodologies, and the third dimension is theory development in the humanities and social sciences to further specify the research question and its ramifications.

The first dimension of the Studio is the *research theme* to which a particular research project belongs. In the first five years of the Studio, three themes will be studied:

- Data and Digital Information: the role of data, digital information and data standards in scholarly research
- Networked Research: novel forms of collaboration and communication in the humanities and social sciences
- Virtual Institutions: the emergence and dynamics of new institutional arrangements in *e*-research.

The second dimension is methodological innovation of the study of *e*-research. Of course, this should be relevant to other researchers in the humanities and social sciences as well. The Studio focuses on those methodologies that (1) are not yet well covered by methodologists in social sciences and humanities at the universities, and (2) are particularly relevant for the study of scientific and scholarly knowledge practices. Three *methodological foci* will be given priority in the first three years of the Studio:

- Virtual Ethnography
- Web Archiving for scholarly research
- Simulation in *e*-research.

The third dimension is the intellectual frame of reference for the research at the Studio. The empirical research in the Studio will address a number of theoretical questions that pertain to the development of *e*-research in the humanities and social sciences. Although we will focus our theoretical work, the Studio will maintain theoretical plurality in its approach of *e*-research phenomena. This is necessary since we cannot expect that *e*-research, being a part of culture and of social reality, can ever be interpreted as a theoretically unified reality (Wallerstein et al., 1996). To better understand the dynamics and meaning of *e*-research practices and infrastructures therefore means to converse in different theoretical traditions. We need to be able to create “trading zones” (Galison, 1997; Galison & Stump, 1996) to make local connections between different theories. The most important questions are:

- does *e*-research lead us to redefine how we can understand the development of scholarly cultures?
- how can we explain and understand diversity of mediated knowledge practices, for example across disciplines and specialties?
- what roles do digital epistemic objects play in knowledge creation, and how can we use them to reformulate informatics problems in the humanities and social sciences?
- do we need to rethink the conceptualisation of scientific labour and markets to understand the dynamics of *e*-research practices?
- does the extra connectivity of *e*-research lead to new forms of complex relationships in social structures and does this lead to new understandings of complex systems?
- does the operationalisation of concepts of agency, institution, textuality and infrastructure need to be revised in order to study mediation in *e*-research?

To realise its dual mission of increasing our understanding of *e*-humanities and *e*-social science, and of supporting scholars to make use of *e*-research, the Studio has two interrelated modules: the Analytic Centre (AC) and the Construction Platform

(CP). These facilitate long-term research based on a clear intellectual agenda (AC) combined with flexible short-term projects created in response to the changing needs of researchers at universities and research institutes (CP). For this reason, all Studio research projects will have a complex blend of curiosity-driven and application oriented goals (Ang & Cassity, 2004). All projects in the CP result from, and are led by, partnerships with external research groups. Whereas the CP helps create new epistemic objects and practices in the humanities and social sciences, both inside and outside of the Studio, the AC studies this process. To facilitate this, the AC is responsible for the creation and maintenance of the Studio's inhouse knowledge database.

For the scholars who are the client-partners of the Studio, the design work must lead to useful insights in the form of concrete deliverables, such as new protocols, best practice manuals, new software tools, perspectives on new analytical techniques to answer old questions, and new research questions in their fields. For the researchers at the Studio, this design work is also a mode of enquiry into the process of knowledge creation. In other words, the Construction Platform is a field laboratory in which different scholarly practices and configurations are tried out and assessed on their consequences. This will lead to a better understanding of the characteristics of knowledge creation as a cultural and social process. The researchers in the Analytic Centre have a special responsibility to link up the results of the CP to the scientific and scholarly literature in the fields of information science, science & technology studies, and communication sciences. To facilitate the management of this type of research, the AC reviews the research in the Studio on its contribution to basic knowledge about the process of knowledge creation. The CP will specifically examine the utility of the Studio research for scholars in the humanities and social sciences based in universities and research institutes in the Netherlands and abroad.

Observation, with all the advanced observation tools available in social and cultural analysis, is central to the Studio. This may involve participant observation of prototypes of new infrastructures (such as collaboratories or Grid computing for social science), but may also entail the systematic observation of mundane processes in research in the social sciences and humanities. This is important to counteract the danger of bias in favour of "the new new thing" (Lewis, 2000; Woolgar, 2002). We can only put the promise and practice of *e*-research into perspective by taking distance from the claims and critically interrogate both the promise and the practice (cf. Wouters & Schröder, 2003). This also holds for the innovative projects that are conducted within the Studio itself. Since these are oriented to the exploration of new modes of inquiry, they run the danger of biasing the novel over the traditional. Reflexive self-observation in different forms is therefore an important element of the research cycle in the Studio.

Research themes

Data and Digital Information

Digital information and data play complex roles in research in the humanities and social sciences (Arzberger et al., 2004; Boonstra et al., 2004; SWR, 2003). This creates particular challenges for the application of *e*-research methods and techniques, especially if complex and fuzzy data sets are involved (eg. visual data, music,

complex texts). The increased availability of digital resources, data and collections, partly the result of digitisation of cultural heritage and of administrative databases, promises to facilitate more possibilities for comparative research. There may be more scope for interdisciplinary research that is based on the combination of data from very different types of sources. Questions that until recently could only be dealt with in a speculative way may now be approached by data-oriented empirical research. Re-use of data may become more prominent (SWR, 2003). The capacity to process and visualise huge datasets is moreover expected to create additional opportunities for empirical research with the help of new computational research methods. In short, both in the humanities and in the social sciences new objects of research, which we call “epistemic objects” (Rheinberger, 1997), will emerge. This development is parallel to the creation of new experimental arrangements in *e*-science.

The research in this theme will address the question what the characteristics of these new epistemic objects will and should have, and how they may reconfigure scholarly research. What type of questions will be foregrounded and which questions may become less central? Which assumptions are built into the new epistemic objects and how may they influence the boundaries between scientific specialties? We will also pay attention to the specificity of qualitative data. They are often more fuzzy and less easy to standardise. This also influences the development of research traditions to share qualitative data for comparative (re)-analysis (Wouters & Schröder, 2003).

To provide a sharper focus on the particularities of data handling in the social sciences and humanities (Boonstra et al., 2004; Hockey, 2000; SWR, 2003), the research in this theme will maintain a firm comparative perspective with the natural and technical sciences. This will also enable the Studio researchers to be alert to new developments in data science and technology. For example, in those fields that have undertaken major digitisation projects, how does *e*-research change the way data is conceptualised, handled and shared? And how do disciplinary communities organise their work around digitised data, eg. do practices become standardised or do field differences persist? In this respect, the comparison of the development of data initiatives in the humanities with ‘data grids’ in the social sciences seems relevant.

The data theme will also pay specific attention to the issue of data sharing and data sharing policies. This research is based on the completed Nardi projects on data sharing (Arzberger et al., 2004; Beaulieu, 2003; Wouters, 2000; Wouters & Schröder, 2003). The emergence of *e*-research creates specific tensions for data sharing, partly because it may no longer be clear who has control over the data sets. Increased attention to data sharing, also in the framework of the organisation of new data archives in the social sciences and humanities, may create tensions with established research practices and routines that are often not oriented to data sharing. The Studio will therefore not only study data sharing but also resistance to data sharing.

The last decade has witnessed an increase in quantitative methods using Web data and in sophisticated quantitative analyses of the structure of the Web and the internet (Adamic, 1999; Albert & Barabasi, 2002; Ebeling & Feistel, 1990; Scharnhorst, 2003; Watts, 1999). This has even led to the establishment of a new field in the information sciences, “webometrics” (Almind & Ingwersen, 1997; Björneborn & Ingwersen, 2001; Boudourides, Sigrist, & Alevizos, 1999; Rousseau, 1997). Web data can be used to analyse the internet and the Web as a complex information space in which

communication patterns emerge and self-organise (Leydesdorff, 2002). Webometrics can also be used to study the change of institutional structures (by means of hyperlink analysis) and the emergence of new institutional structures and infrastructures. Changes in scientific production and communication can be studied in so far as they can be represented in Web based indicators. We expect that webometrics will also contribute to our understanding of the emergence of new forms of Web based scientific communication and collaboration, such as related to *e*-journals, collaboratories, online databases, file sharing and collaborative simulations. Indicators developed on the basis of Web data can have both an evaluative and descriptive role. In this collaboratory, they should primarily provide insights in the nature of knowledge production in *e*-research.

The research in this theme builds further on recent European research projects in webometrics, in particular on WISER² and EICSTES³. It will extend the research questions in these projects toward a “reflexive webometrics”. It aims to develop novel methods for automated data gathering (with open source web crawlers, commercial software, Web page annotation schemes, and search engine tools) and to contribute to the development of professional standards to observe the dynamic Web. We expect that this will lead to analytic tools that can be used by other researchers in the social sciences and humanities without the need for additional programming expertise (M. Thelwall, 2001, 2002, 2005). We expect that these methods will be particularly successful if they are intimately related to qualitative and quantitative content analysis of Web phenomena. For instance, hyperlink network analysis has shown interesting topological features in graph theory. It is, however, still far from clear how these graph theoretical structures can be interpreted. An important aspect of future research in webometrics will be the development of dynamic observation based on the self-organizing and fluid nature of the web as a medium. New insights of complexity theory into the description of complex structures will have to be taken into account in this research.

Networked Research

e-Research is not only about data, it is also about collaboration. It is expected to facilitate new forms of large-scale collaboration and more collaboration across the boundaries of disciplines and specialties (Berman, Fox, & Hey, 2003; Finholt, 2002; Walsh & Maloney, 2002). Many *e*-science pilots are moreover the fruit of intense cooperation between academia and industry.

The humanities and social sciences are a particularly interesting area to study the development of scientific and scholarly collaboration because the variation of forms of collaboration and non-collaboration is so huge (Fry, 2003). Virtually every possible configuration is practiced in one field or another. The spectrum goes from the traditional, lone scholar working in a decidedly low-tech environment to the tight industrially organised group in which each PhD student and postdoc solves a particular problem. This means that comparative fieldwork in the humanities and social sciences is very rewarding. The same is true for the forms in which scholars and researchers choose to communicate their work and results to larger audiences. The Studio research in this theme will focus on the way the new media interact with

² <http://www.webindicators.org/>
³ <http://www.eicstes.org/>

forms of collaboration and communication. It moreover aims to support them with building new forms of collaboration (eg. collaboratories) and communication (eg. new Web site conceptions).

A key issue concerns the ways the dynamics of collaboration are affected by mediation by new media and digital networks. How does the technological possibility intersect with traditional human need for communication? The implications of collaborative work for the resulting knowledge products will also be studied. Are forms of knowledge affected by the way they need to be communicated? Which types of intellectual work seem amenable to virtualisation and digitisation? How does audience variation across disciplines shape collaborative practices and the integration of *e*-research? Does the organisation of research change, as units within a field become more dependent or more specialised? An interesting question is in what ways the dynamics of very large-scale collaboration differs from more modest networks and how this affects the development of scientific collaboration in the humanities and social sciences. An important question is also how *e*-research shapes the traditional boundary between informal and formal communication across fields. Answers to these questions affect the way we think about the design of tools for collaboration. This is for example relevant to the construction of collaborative analytic instruments.

Within this theme, the use of the Web as means of representation and collaboration will be a specific focus. This will intersect with the work in the methodological focus on Web archiving. The creation and dynamics of collaboratories will be a priority within this theme. The collaboratories of the Studio themselves will be monitored in order to draw lessons about their dynamics.

Virtual Institutions

In *e*-research, digital infrastructures and emergent institutions play a crucial role (Bowker & Star, 1999). Collaboratories, research infrastructures or the lack thereof, digital libraries, digital repositories and collections, and new venues for scholarly publication directly influence the extent to which scholars in the humanities and social sciences can effectively make use of new research possibilities. Given the recent emergence of *e*-research, the consequences of the accompanying institutional rearrangement are not yet well understood. It is therefore relevant to understand the specificities of institution building in the humanities and social sciences. The theme Virtual Institutions will explore which institutional arrangements are conducive to the humanities and social sciences.

Important questions are how textual infrastructures affect the textual practices of researchers and scholars. Does it make a difference that digital infrastructures are also forms of writing? Standardisation and ordering of these infrastructures, such as computer interoperability or database standards, have a tremendous impact on the work of scholars. To what extent can they influence these processes if they are implemented at a higher level of organisation (such as the university or a data repository)? For example, how does infrastructure sustain various levels of formalisation and circulation of knowledge and information?

In this theme, specific attention will be paid to the systems of accountability in universities and research institutes. How universities and research institutes have organised their systems of quality control and accountability may have a profound

effect on knowledge creation because of its impact on the criteria of scientific and scholarly quality and integrity. Does *e*-research go together with new ways of assessing research performance and output? In what ways do new research practices create problems for existing peer review and visitation procedures? How will individual careers be judged in very large-scale collaborative research institutes and networks? How do the forms of knowledge evolve in *e*-research and are particular practices hampered by the way researchers are being assessed? And in what ways are internet based information systems being used by the institutions of accountability?

Methodological foci

Virtual Ethnography

Virtual ethnography is a recent development in the area of anthropology, science & technology studies, and internet research (Beaulieu, 2004, forthcoming 2005; Hine, 2000, forthcoming 2005; Howard, 2002; Mason, 1996). It extends the notions of field and ethnographic observation from the exclusive study of co-present and face to face interactions, to a focus on mediated and distributed ones (Hine, forthcoming 2005). It combines two related but distinct ideas. First, virtual ethnography tries to create virtual counterparts of the basic ethnographic concepts and interrogates whether they can be applied to mediated interaction. Second, virtual ethnography aims to change the notion of the fieldsite from a localised space into a network of interlinked mediated settings. In this, it is related to the ethnography of networks (Newman, 1998). Virtual ethnography maintains a number of values of traditional ethnographic work. It aims to sustain practices of “thick description” (Geertz, 1983), and to achieve this by paying attention to the perspective of the actors themselves (Slater, 2002; Ward, 1999). This makes virtual ethnography distinct from Web site content analysis or Webometric studies, although it may make use of the same qualitative and quantitative techniques to locate networks and Web sites (Scharnhorst, 2003), or to understand media forms.

The key research question in the Virtual Ethnography Focus is: how can ethnography be pursued in mediated settings? Research in this focus will establish which aspects of ethnographic research are challenged in particular in the shift from face to face interaction to mediated digitised interaction. This should make clear how ethnography can be conceived as a flexible practice, while remaining recognisable as a specific methodology.

More specific questions that will be dealt with are which new concepts of “field” or “research site” are needed for virtual ethnography, how virtual elements can be integrated in traditional fieldwork, and which new ethical issues arise in the practice of virtual ethnography (Fox, 2000). In the field of internet research, a fair amount of work has already been devoted to the challenge to crucial notions of research ethics in the context of mediated social science research, and the Studio research will elaborate on these guidelines and reflections in the course of doing research.

Webarchiving

The focus Webarchiving will develop a new methodology for systematic, longitudinal analysis of the Web sites that are produced in the sciences and humanities. It is clear that Web and internet data are central in many ways in *e*-research, also in the

humanities and social sciences. To the extent that researchers and scholars are creating more and more presentations on the Web (the Web is the dominant interface with the internet in academia), Web data are clearly important to understand the development of these fields. However, presently there is no clear way how to make Web data available for scholarly research. Libraries and archives are only beginning to develop concepts that enable the medium- and long-term archiving of Web sites. They are confronted, among others, with the problem that the concept of the document is not adequate for this area. The Studio will not take it on to crawl and archive the Web itself: this must be the responsibility of libraries and archives. It will however in cooperation with the WebArchivist organisation⁴, the Internet Preservation Consortium⁵, the Internet Archive⁶, and the nascent European Internet Archive, develop methods and techniques to conceptualise Web archives in such a way that they can produce datasets for social science and humanities research.

In this effort, the emphasis will be on Web archiving of the Dutch and European academic Web. This work will build on a current project at NIWI-KNAW on digitisation of two KNAW institutes and expand the scope of this effort in order to be able to analyse the whole of “the academic Web” for analysis. A “Web sphere” is conceptualised in this research as a linked set of dynamically defined digital resources spanning multiple Web sites that together demarcate a specific type of content or action in online structures (Foot & Schneider, 2002). Presently, there are still many methodological problems that need to be solved before one can analyse the dynamics of digitisation as represented in the Web sites produced by scientific and scholarly research. This research aims to solve these problems by using a suite of software tools, developed by WebArchivist.org, based on lessons learned through collaborative projects with the US Library of Congress and the Internet Archive. In collaboration with the internet researchers involved in developing this tool, at the University of Washington (Seattle) and the Institute of Technology of the State University of New York (Utica), this method will be adapted to the study of scientific Web sites.

Simulation

The extended use of data visualisation technologies and virtual reality techniques in simulation research methods is often seen as one of the hallmarks of *e-science* (Berman et al., 2003). Computer based simulation and modelling has become a standard repertoire in the natural and technical sciences and is increasingly used in the life sciences and medicine (Banks, 1998; Fishwick, 1995). In a number of social sciences and humanities, models have been a standard tool for decades (eg. economics, sociology, archeology) (Burenhult, 2002; Gilbert & Troitzsch, 1999; Schweitzer, 2002). The use of simulation is however a more recent phenomenon in these fields. Moreover, the value of simulating and modelling as a research method is often not undisputed.

The aim of the research in the methodological focus Simulation in the Studio is to develop further expertise in simulations and systematic reflection on the heuristic value of modelling and simulation for theory building in the social sciences and humanities. The research will not focus primarily on the creation of new models since

⁴ <http://www.webarchivist.org/>

⁵ The author is member of the scientific committee of the IPPC.

⁶ <http://www.archive.org/>

there is already a large variety of models and simulation techniques available. Instead, the respecification of general models for research questions in the social sciences and humanities will be central. Both agent-based and network-oriented models and simulations will be included in the research agenda. The study of the heuristic and epistemic value of modelling and simulating as a research strategy is intrinsic part of this respecification. For example, from the perspective of social theory one might wish to give agents in large systems as many individual traits as possible. However, this easily leads to an exponential increase of the degrees of freedom of the model. This raises the question how low-dimensional approaches can be combined with individualisation of agents in multi-agent models and simulations.

The research in this methodological focus will focus on evolutionary modelling (including the modelling of innovation), the application of evolutionary strategies as heuristic concept and as mathematical tools, and the diffusion of simulation into the humanities (Ratto & Scharnhorst, 2004; Scharnhorst, 1998; Scharnhorst, 2001). We expect that this might also lead to a toolbox of simulation principles in combination with principles for the reflection on the conceptual implications of such models (which are often not explicitly discussed in the literature). The simulation will moreover be developed in close collaboration with domain-specific experts in the social sciences and humanities. This may also produce interactive simulations that can be used in different contexts in research and teaching.

Acknowledgement

This work has been developed in close collaboration with Anne Beaulieu, Jenny Fry, Ina Hellsten, Matt Ratto, Andrea Scharnhorst, and Katie Vann. I would also like to thank the scholars who were interviewed in the preparation of this new research programme, as well as the members of the committee who reviewed it. Full details on the Virtual Knowledge Studio can be found at <http://www.virtualknowledgestudio.org>.

References

- Adamic, L. (1999). The small world Web. In *Proceedings of the 3rd European Conference on Digital Libraries. Lecture notes in Computer Science*. (pp. 443-452). New York: Springer.
- Albert, R., & Barabasi, A. L. (2002). Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74(1), 47-97.
- Almind, T., & Ingwersen, P. (1997). Informetric Analyses on the World Wide Web: Methodological Approaches to "Webometrics". *Journal of Documentation*, 53, 404-426.
- Ang, I., & Cassity, E. (2004). *Attraction of Strangers. Partnerships in Humanities Research*. Sydney: Australian Academy of the Humanities.
- Arzberger, P., Schröder, P., Beaulieu, A., Bowker, G., Casey, K., Laaksonen, L., Moorman, D., Uhlir, P., Wouters, P. (2004). SCIENCE AND GOVERNMENT: An International Framework to Promote Access to Data. *Science*, 303 (5665), 1777-1778.
- Banks, J. (Ed.). (1998). *Handbook of Simulation*. New York et al.: John Wiley & Sons.

- Beaulieu, A. (2003). Research woes and new data flows. In P. Wouters & P. Schröder (Eds.), *Promise and practice in data sharing* (pp. 96). Amsterdam: NIWI-KNAW.
- Beaulieu, A. (2004). Mediating ethnography: objectivity and the making of ethnographies of the internet. *Social Epistemology*, 18(2-3), 139-164.
- Beaulieu, A. (forthcoming 2005). Sociable Hyperlinks: an ethnographic approach to connectivity. In C. Hine (Ed.), *Virtual Methods: Issues in Social Research on the Internet*.
- Berman, F., Fox, G., & Hey, T. (2003). The Grid: past, present, future. In F. Berman, G. Fox & T. Hey (Eds.), *Grid Computing. Making the Global Infrastructure a Reality* (pp. 9-50). Chichester, West-Sussex, UK: John Wiley & Sons.
- Bijker, W., & Peperkamp, B. (2002). *Geëngageerde geesteswetenschappen. Perspectieven op cultuurveranderingen in een digitaliserend tijdperk* (No. Achtergrondstudie nr. 27). Den Haag: Adviesraad voor het Wetenschaps- en Technologiebeleid.
- Bijker, W., Schurer, K., Stronks, E., Uszkoreit, H., Wittenburg, P., & Woolgar, S. (2003). *Building the KNAW International Research Institute on e-Science Studies in the Humanities and Social Sciences (IRISS)*. Amsterdam: Royal Netherlands Academy of Arts and Sciences.
- Björneborn, L., & Ingwersen, P. (2001). Perspectives of webometrics. *Scientometrics*, 50(1), 65-82.
- Boonstra, O., Breure, L., & Doorn, P. (2004). Past, Present and Future of Historical Information Science.
- Boudourides, M. A., Sigrist, B., & Alevizos, P. D. (1999, 1999). *Webometrics and the Self-Organization of the European Information Society*. Retrieved August, 12
- Bowker, G. C., & Star, S. L. (1999). *Sorting Things Out: Classification and its Consequences*. Cambridge, MA: MIT Press.
- Burenhult, G. (Ed.). (2002). *Archaeological Informatics: Pushing The Envelope CAA2001*. Oxford: Archeopress.
- Ebeling, W., & Feistel, R. (1990). *Evolution of Complex Systems*. Dordrecht: Kluwer.
- Finholt, T. A. (2002). Collaboratories. In B. Cronin & D. Shaw (Eds.), *Annual Review of Information Science and Technology* (Vol. 36). Medford, NJ: Information Today.
- Fishwick, P. A. (1995). *Simulation, model design and execution*. Upper Saddle River: Prentice-Hall.
- Foot, K. A., & Schneider, S. M. (2002). Online Action in Campaign 2000: An Exploratory Analysis of the U.S. Political Web Sphere. *Journal of Broadcasting and Electronic Media*, 46(2), 222-244.
- Fox, R. G. (2000). Hearing where we're coming from - Ethically and professionally. In *Ethics and Anthropology* (Vol. 925, pp. 1-8).
- Fry, J. (2003). *The cultural shaping of scholarly communication within academic specialisms*. University of Brighton, Brighton.
- Galison, P. (1997). *Image and logic : a material culture of microphysics*. Chicago and London: The University of Chicago Press.
- Galison, P., & Stump, D. J. (1996). The Disunity of Science. Boundaries, Contexts, and Power. In P. Galison & D. J. Stump (Eds.), *The Disunity of Science. Boundaries, Contexts, and Power*. Stanford, CA: Stanford University Press.
- Geertz, C. (1983). *Local Knowledge*. New York: Basic Books.
- Gilbert, N., & Troitzsch, K. G. (1999). *Simulation for the Social Scientist*. Buckingham, Philadelphia: Open University Press.

- Hey, T., & Trefethen, A. E. (2002). The UK e-Science Core Programme and the Grid. *Future Generation Computer Systems*, 18(8), 1017-1031.
- Hine, C. (2000). *Virtual ethnography*. London: Sage.
- Hine, C. (Ed.). (forthcoming 2005). *Virtual Methods*: Berg.
- Hockey, S. (2000). *Electronic texts in the Humanities: principles and practices*. Oxford: Oxford University Press.
- Howard, P. (2002). Network ethnography and the hypermedia organization: New organizations, new media, new methods. *New Media & Society*, 4(4), 551-575.
- Kircz, J. (2004). *E-based humanities and e-humanities on a SURF platform*. Utrecht: Stichting SURF.
- Lewis, M. (2000). *The fnew new thing: A Silicon Valley Story*. New York: Norton.
- Leydesdorff, L. (2002). The Communication Turn in the Theory of Social Systems. *Systems Research and Behavioral Science*, 19(2), 129-136.
- Mason. (1996). Moving Toward Virtual Ethnography. *American Folklore Society Newsletter*, 25(2), 4-6.
- Nentwich, M. (2003). *Cyberscience. Research in the Age of the Internet*. Vienna: Austrian Academy of Sciences Press.
- Newman, S. (1998). Here, there and nowhere at all: distribution, negotiation and virtuality in postmodern ethnography and engineering. *Knowledge and Society*, 11, 235-267.
- Ratto, M., & Scharnhorst, A. (2004). *From numeric to metaphoric simulations – attempts to draw a simulation landscape. Paper given at Epistemological Perspectives on Simulation . A Cross-Disciplinary Workshop. Koblenz, Germany, July 1-2, 2004*. Unpublished manuscript.
- Rheinberger, H.-J. (1997). *Toward a History of Epistemic Things: Synthesizing Proteins in the Test Tube*: Stanford University Press.
- Rousseau, R. (1997). Situations: an exploratory study. *Cybermetrics*, 1(1), <http://www.cindoc.csic.es/cybermetrics/articles/v1i1p1.html>.
- Scharnhorst, A. (1998). Citation-Networks, Science Landscapes and Evolutionary Strategies. *Scientometrics*, 43(1), 95-106.
- Scharnhorst, A. (2001). Constructing Knowledge Landscapes within the Framework of Geometrically Oriented Evolutionary Theories. In J. Kriz (Ed.), *Integrative Systems Approaches to Natural and Social Sciences Systems Science 2000* (pp. 505-515). Berlin: Springer.
- Scharnhorst, A. (2003). Complex Networks and the Web: Insights from Nonlinear Physics. *Journal of Computer-Mediated Communication*, 8(4).
- Schweitzer, F. (Ed.). (2002). *Modeling complexity in economic and social systems*. New Jersey: World Scientific.
- Slater, D. (2002). Making things real - Ethics and order on the Internet. *Theory Culture & Society*, 19(5-6), 227-+.
- SWR, S. S. C. (2003). *Networked Data Services: Towards a Future Data Infrastructure for the Social Sciences in the Netherlands*. Amsterdam: KNAW.
- Thelwall, M. (2001). A Web crawler design for data mining. *Journal of Information Science*, 27(5), 319-325.
- Thelwall, M. (2002). Methodologies for Crawler Based Web Surveys. *Internet Research: Electronic Networking and Applications*, 12(2), 124-138.
- Thelwall, M. (2005). *Link Analysis: An Information Science Approach*. San Diego: Academic Press.

- Wallerstein, I., Calestous, J., Fox Keller, E., Kocka, J., Lecourt, D., Mudimbe, V., et al. (1996). *Open the Social Sciences*. Stanford: Stanford University Press.
- Walsh, J. P., & Maloney, N. G. (2002). Computer network use, collaboration structures and productivity. In P. Hinds & S. Kiesler (Eds.), *Distributed work* (pp. 433-458). Cambridge, MA: MIT Press.
- Ward, K. J. (1999). Cyber-ethnography and the emergence of the virtually new community. *Journal of Information Technology*, 14(1), 95-105.
- Watts, D. J. (1999). *Small Worlds: The dynamics of networks between order and randomness*. Princeton, New Jersey: Princeton University Press.
- Whitley, R. (2000). *The intellectual and social organization of the sciences*. Oxford: Oxford University Press.
- Woolgar, S. (2002). *Virtual Society? Technology, Cyberbole, Reality*. Oxford: Oxford University Press.
- Wouters, P. (2000). *Exploring the future of information and communication in Research*. Paper presented at the AVRIL NIWI Research Programme 2000-2004, , Amsterdam, the Netherlands.
- Wouters, P., & Schröder, P. (2003). *Promise and Practice in Data Sharing*. Amsterdam: NIWI-KNAW.