What's the Deal with the Web/Blogs/the Next Big Technology: A Key Role for Information Science in e-Social Science Research?

Mike Thelwall[1] and Paul Wouters[2]

[1] School of Computing and Information Technology, University of Wolverhampton,
Wulfruna Street, Wolverhampton WV1 1SB, UK
m.thelwall@wlv.ac.uk
[2] Networked Research and Digital Information - Nerdi, NIWI-KNAW,
Joan Muyskenweg 25, PO Box 95110, 1090 HC Amsterdam,
The Netherlands
paul.wouters@niwi.knaw.nl

Abstract. Since many nations have provided substantial funding for new e-social science and humanities investigations, there is now an opportunity for information scientists to adopt an enabling role for this new kind of research. Logically, a more information-centred environment should be more conducive to information science and to information scientists taking part in other types of research. In this article it is argued that information scientists can play a valuable role by evaluating new information *sources* in a meta-disciplinary context, developing tools and methods to analyse the data and, crucially, contributing to the prediction of the kinds of research questions that the data may usefully help address. It is argued that this is both an essential service for social science research and one that information science is uniquely placed to provide. A timely response to this challenge may also generate novel research problems within information science itself.

1 Introduction

There have been so many changes to our lives enabled by computing technologies that it seems inadequate to refer to a single computer/information/communication revolution. Governments have recognised the potential advances that the technologies may facilitate in many ways, including the funding of large e-science programmes and infrastructures. Recently, the term e-science has been adopted for initiatives that have taken advantage of Grid infrastructures for shared computing power [1]. Following e-science, and often explicitly seeking to imitate it, funding has been provided for e-social science and humanities research (e.g., the UK's ESRC National Centre for e-Social Science[1]; the Netherlands' Virtual Knowledge Studio for the Humanities and Social Sciences[2]; and the USA's ACSL-sponsored Commission on

---

[1] http://www.ncess.ac.uk/
[2] http://www.virtualknowledgestudio.nl

Cyberinfrastructure for the Humanities & Social Sciences[3]). We use the phrase e-social science research to describe research that is enabled by the new electronic technologies (e.g., digital media creation; Grid-enabled social science research, e.g. [2]) or studies their social impacts. Logically, the new funding should give challenges and opportunities to those who study information in various ways, and information scientists in particular. Yet there does not seem to be an explicit debate over this issue within the journals of information science, perhaps because practitioners are already struggling to cope with managing continuous changes in core activities such as information retrieval, effective (mainly digital) search strategies, and library services.

One of the by-products of the digital revolutions and the Internet has been the creation of huge informal repositories of public access, easily discoverable information including the web and newsgroups. An advantage of Internet sources is that it is often possible to use tools to automatically retrieve and process information in large quantities. A second advantage is that large sections of the information come from genres that have previously been inaccessible to researchers in any quantity, and may therefore help to address research questions that have been previously left unanswered. This may even create new research areas. There is a precedent for this in the creation of the field of bibliometrics largely in response to the availability of the Institute for Scientific Information's databases [3,4].

The central thesis of this paper is that there is an opportunity for information science as a discipline to take on the role of assessing new *information sources* for use in social science research, including the development of appropriate *methods*. There is a theoretical basis for this in the novel roles that information is starting to play in scientific and scholarly research. This has been captured in the notion of the "informational turn" in research [5,6]. We argue that information science is uniquely positioned to most effectively develop each new potential data source from a social science perspective. First, however, we review three case studies of large-scale Internet phenomena/information sources that have attracted research interest: web links, newsgroups, and blogs; contrasting computer science, information science and social sciences approaches. We also discuss current e-social science initiatives and examine their possible roles in developing new data sources.

2   New Data Sources

2.1   Web Links

Link analysis has produced significant findings that underpin the potential for the academic Web to be used as an information source, a potential that was recognised once commercial search engines introduced facilities that could be used for link counting [7,8]. Results have shown that university Web site links are influenced by a combination of geographic [9] and research-related [10] factors, confirming that significant patterns can be mined from this kind of link data. Mapping techniques have also been developed to visualise the flow of information between national educational systems [11]. Colinks have been used to map patterns of interlinking between universities in Europe [12]. In contrast to the above university-wide

---

[3] http://www.acls.org/cyberinfrastructure/cyber_charge.htm

studies, some link analysis research has analysed the Web pages relating to a single subject within a country (e.g., [13-17]).

In the social sciences, Social Network Analysis [18] has been the perspective from which some hyperlink analyses have been developed, particularly in communication and Internet studies. Often, the existence of a hyperlink is taken as an indicator of either a social connection or a communication channel. Links have been used to map online connections between the web sites of political movements [19], to track individual issues [20,21], and as indicators of business connections [22]. Linking has also been investigated for its own sake: as a phenomenon that is important in its own right, although as part of wider investigations into web use [23,24].

Computer scientists tend to use links as the basis of new algorithms or to improve the functioning of existing ones. The name web structure mining is often given to computer science link analysis. A high profile example is Google's link-based PageRank algorithm, designed to help identify the most authoritative pages in the web [25]. Another common computer science use of links is as the raw data for automatically-generated web site navigation aids [26]. There are also many descriptive link analyses, although these tend to be justified in terms of future benefits for improved algorithm design. Examples include a topological analysis of the link structure of the web, as crawled by AltaVista [27], and investigations into the relationship between links and text in web pages [28,29].

2.2   Newsgroup Postings

Newsgroups are themed discussion lists that allow anyone to contribute although some are moderated. They are sources of relatively informal opinions even though contributions probably typically vary in tone between the relatively informal communications such as personal e-mails and relatively more formal documents such as personal home pages (which are more visible to general users). For social sciences research, they provide a non-intrusive, though possibly ethically problematic [30], source of information about the informal opinions of a section of the population. Intuitively, they may help to address many social science research questions, especially because of their topic-structured nature. For example, researchers into sport attitudes may visit the sporting groups. Intuitively also, newsgroups participants are likely to come from a small segment of society: the more IT literate and perhaps the more expressive or opinionated.

Information scientists have analysed newsgroups from the perspective of demonstrating that bibliometric laws apply to them [31] and analyzing their influence on scholarly communication and the invisible college [32]. The former has a more information centred approach, concerned with mathematical modelling of the data. Caldas [32], in contrast, addresses a more social sciences type question focussing on the impact of technology upon scholarly communication, a recurrent theme in information science (e.g., [33]). Outside of information science, social scientists have analysed newsgroups from various disciplinary perspectives: as an example of computer mediated communication in communication science by researchers seeking to investigate social aspects of its use (e.g., [34]), also an information science theme [35].

Computer science newsgroup research tends to be characterised by the development of algorithms to extract particular types of information, such as which articles are likely

to be of interest to an individual user. One typical and highly cited example is a paper describing a new approach for constructing newsreader software [36].

2.3 Blog Postings

Web logs or blogs, are a newer development than the web itself. They are based around software that makes it easy for non-technical web users to maintain a frequently updated collection of online information, which is typically presented in the form of an online log or diary. The software automatically moves old postings into an archive, accessible by following links from the main page, and formats the postings in an attractive style. Many other features make blogs easy to use and integrate blogs with each other, such as permalinks (link to a particular post, even after it has been archived), trackback (to find who links to a blog post), and blogrolls (collections of links to recommended blogs).

Blogs were a hot topic at the 2004 Association of Internet Researchers conference, a multidisciplinary conference mainly reflecting sociology, media and cultural studies, and social sciences [37,38]. The excitement centred around the potential for public glimpses into the private lives of individuals on a scale not previously possible [39]. For example, one presentation examined anti-war blogs to see how mainstream mass media could be challenged online [40]. Bloggers are probably a wider social group than newsgroup posters because a lower technical competence is needed. Most likely the IT skills of the average blogger are typically significantly lower than that of newsgroup posters. A disadvantage is the lack of an 'official' topic organisation of blogs, although blogrolls make an unofficial substitute. At the 2004 ASIST conference, representing here the information science angle, there were two blog discussions: "Blogs for information dissemination and knowledge management" and "Beyond the sandbox: Wikis and blogs that get work done" and a paper on the value of blogs for information dissemination [41]. The themes here are information dissemination and communication. Information science seems to be a latecomer to blog research, with no papers at all at any previous ASIST conference, nor any blog papers published in JASIST before 2005.

Computer scientists have quietly found one very interesting blog application: using blogs as a source of information about the public reaction to large-scale marketing campaigns. IBM's WebFountain project continually monitors thousands of blogs, creating a large, real time database of public opinions [42]. The underlying belief is that even if only a very small percentage of blogs happen to give any reaction to an advertising campaign, as long as there are a few responses then these can give instant feedback to advertisers. This is a commercial rather than scientific application, but has been reported in an academic forum, centring on quantitative descriptions of aspects of blog evolution such as time series for word usage in blogs. Perhaps the commercial applications found by IBM for social context information explain the interesting mix of computer science and other approaches found in the blog workshop within the traditionally computer science-dominated World Wide Web series of conferences (http://www.blogpulse.com/www2004-workshop.html). The computing presentations in 2004 included a descriptive, structure-seeking contribution (Implicit structure and the dynamic of blogspace, by Adar and Zhang) and an initiative similar to IBM's (BlogPulse: Automated Trend Discovery for weblogs, by Glance, Hurst &

Tomokiyo). The non-computer science contributions reported on where bloggers live (Mapping the blogosphere in America, by Lin & Halavais), and the social and psychological background to blogging (Blogs as "Protected Space", by Gumbrecht), as well as one aiming at social context (How can we measure the influence of the blogosphere? by Gill).

2.4 Overview of Data Sources and Research Areas

The above examples illustrate the approaches that different research areas take to new Internet information formats. Table 1 summarizes the discussion. The categories are generalizations and there will be exceptions. This is unavoidable because fields are ultimately defined by human factors rather than solely by objective considerations of content. In particular, journal and conference reviewers may accept work that might objectively be considered to be part of a different field than the journal/conference, particularly if the authors are associated with the normal field of the journal or if the referees are unaware of more directly relevant fields. Moreover, computer science is a very large field that incorporates a range of disciplinary backgrounds and overlaps (e.g. with psychology), and there are intersections and grey areas between different fields.

Table 1. Research into new data sources

Source	Social Science	Information Science	Computer Science
Web links	Descriptive analysis of user communities; descriptive web mapping; analysing linking as a social phenomenon; using links as indicators for social relations	Mapping scholarly communication	Algorithms to build web navigation and information retrieval tools; descriptive modelling
Newsgroups	Social organisation of newsgroups; newsgroups as a new communication form	Modelling of information in newsgroups.	Algorithms to help individuals to discover of relevant postings
Blogs	Social context of creators and users; new communication forms; new insights into existing social issues.	Blogs for communication and information dissemination	Algorithms to extract social information; descriptive modelling
Overall	Social impact and context, tools to analyse society	Scholarly communication, libraries, information dissemination	Algorithm development, descriptive modelling

The discussions so far have primarily treated research areas as separate entities. In practice, however, for decades there has been a trend towards problem solving research involving increasingly large and increasingly interdisciplinary teams of investigators [43]. Nevertheless, disciplines are still important both as focal points for training and research [44,45]. It seems clear that some types of social science problems will have the computational complexity to require computer scientists to be engaged as part of interdisciplinary research teams, but this does not seem to be a dominant paradigm in social sciences research yet, perhaps because of its relatively individualistic nature [44,45].

3   E-Social Science Research Initiatives

Significant funding has been allocated by many national governments to various forms of e-social science research. The UK has taken the lead with its ESRC National Centre for e-Social Science. This programme "aims to stimulate the uptake and use by social scientists, of new and emerging Grid-enabled computing and data infrastructure, both in quantitative and qualitative research"[4]. In its initial stage, which was set to start in April 2005, the ESRC's e-social science strategy is made up of three components: a training and awareness programme, pilot demonstrator projects and the National Centre for e-Social Science (NCeSS). NCeSS has a distributed structure, comprising a co-ordinating hub, based at the University of Manchester in collaboration with the UK Data Archive at the University of Essex, and a set of research-based nodes distributed across the UK. The research has a two-fold goal. It will focus either on the application of Grid technologies to generate new solutions to social science research problems, or on the social shaping and socio-economic impact of e-science. At the time of writing, it was not yet clear how the balance between these two aims would be struck.

In the Netherlands, the Royal Netherlands Academy of Arts and Sciences has taken the lead together with the Dutch national science foundation (NWO). In the course of 2005, a new research centre was set to start, the Virtual Knowledge Studio for the Humanities and Social Sciences[5]. This will be accompanied by a reorganisation of existing data archives for the social sciences and humanities in a new initiative called DANS: Digital Archive and Networked Services. This data archive, comparable to the UK Data Archive[6], should also harness R&D to further develop the technical expertise needed to archive social science and humanities data sets, including non-textual sources. The new Virtual Knowledge Studio has a multi-tiered goal. It aims 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; and to contribute to a better understanding of the dynamics of knowledge creation [46].

---

[4] http://www.esrc.ac.uk/esrccontent/researchfunding/esciencecentre.asp
[5] http://www.virtualknowledgestudio.nl
[6] http://www.data-archive.ac.uk/home/

In the US, the natural sciences and engineering are still central to initiatives in "cyberinfrastructures", which is the key concept in e-science initiatives. The closest parallel to the British and Dutch initiatives is the national Commission on Cyberinfrastructure for the Humanities & Social Sciences, which was initiated by the American Council of Learned Societies[7]. The Commission was set to report in early 2005, and was charged to "describe and analyze the current state of humanities and social science cyberinfrastructure; articulate the requirements and the potential contributions of the humanities and the social sciences in developing a cyberinfrastructure for information, teaching, and research; and recommend areas of emphasis and coordination for the various agencies and institutions, public and private, that contribute to the development of this cyberinfrastructure"[8]. It is not yet clear to what extent this will lead to substantial initiatives in e-social science research in the US.

4 A New Information Science Approach

Recall that the central thesis of this paper is that there is an opportunity for information science as a discipline to take on the role of *assessing new information sources* for use in social science research, including the development of appropriate *methods*. Traditional boundaries between information sharing and social communication have become blurred in the new Internet based information environments. Moreover, the distinctions between communication and collaboration are often more difficult to uphold in digital collaboratories (i.e., electronic environments for collaborative research between geographically distant partners: [47]). This is the case, for example, in collaborative annotation tools that are developed by scholars in literary research. An example of this is the Dutch e-laborate project[9]. The emergence of digital information, embedded in information and communication technologies, has enabled a radical lowering of the costs related to many types of information dissemination. At the same time, new research technologies have affected the process of data generation itself. They have enabled new types of experiments (e.g., sequencing technologies in bioinformatics), measurements (e.g., statistical pattern recognition in astronomy), imaging (e.g., body scanning in medical sciences) and data visualisation (e.g. using modelling software to visualise complex protein structures). These have in their turn vastly increased the level of data production in research. Where this happens, scientific research is becoming more dependent on information and communication technologies [48]. In the social sciences and humanities, this development leads to specific configurations. Because the implications of *e*-research for the humanities and social sciences are still far from clear, these configurations have not yet stabilised. 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

---

[7] http://www.acls.org/
[8] http://www.acls.org/cyberinfrastructure/cyber_charge.htm
[9] www.e-laborate.nl

dominance of particular computational approaches in the world of *e*-science. Our claim is that precisely because the informatisation of academic research in the social sciences and humanities has not yet been blackboxed in stable socio-technical configurations (see [45]), information scientists can play an exploratory and supportive role. We also expect that this will generate research puzzles for the information scientists that are different from the ones that have emerged in the context of the natural sciences and engineering. This is related to both the data sets and the research questions in humanities and social science research. In the remainder of this paper, we focus on the characteristics of the data sets and the assessment of the methods to analyse these data.

Assessing New Information Sources for Social Science Research. The Internet is characterised by the rapid evolution of new communication forms. Aside from e-mail and newsgroups, this is true for the web alone. To give some examples of revolutionary parts of the web: commercial search engines are themselves a mass media [49]; blogs have been discussed; chat rooms are a new communication form [35]; and the web has spawned many new document genres [50]. Perhaps each newly identified genre or publishing/communication form should be assessed to see what new insights it can give into aspects of human behaviour. For example, Herring [51] claims that each new communication technology gives rise to research into how it helps to shape social practices.

Whilst many social scientists wish to and do explore emerging Internet phenomena enthusiastically, there is an obvious danger that in the rush to explore and exploit, many researchers in different fields will develop largely similar techniques. An example is Hyperlink Network Analysis [52], derived from Social Networks Analysis, and Link Analysis, derived from bibliometrics [53]. Whilst each subject specialist can ask whether the new source can provide new answers, for their concerns and develop new methods, the information scientist can step back and attempt to identify the research areas for which the data source would be appropriate, and develop a generic methodology and tools that each subject could adopt or adapt. In practice, of course, social sciences research fields and methods are too diverse for any researcher to be able to match new data sources to research areas in a comprehensive way. It would also be impossible and undesirable to ask social scientists to avoid new data sources until someone else has first assessed them. In practice, however, it may be possible for information scientists to combine the roles of data evaluator/method developer with brokering organically growing social science methods so that the information science approach incorporates the best of the best of a range of social science initiatives and also plays the role of disseminating the developed methods to a wider social science audience, perhaps through books, chapter in social science methods books and presentations at social sciences methodologies conferences.

Creating Programs to Mine New Information Sources. Given that future new information sources will be almost exclusively digital and predominantly Internet-based, part of the task of developing tools to effectively exploit them is likely to involve developing computer programs to automatically process large volumes of online data. A previous example of this is a suite of link analysis programs [53]. Such development requires programming skills or the services of a programmer (e.g., in a collaborative project). Information science as a field does seem to contain many

programmers, so this is not necessarily an obstacle, and information science must be one of the better-furnished fields in this regard amongst the social sciences. Computer science is an alternative choice as the centre of new information tools development, however, as the home of computer programming. It has played an increasing role in many research areas as they harness its data processing capabilities to their needs (e.g., computational linguistics, bioinformatics, medical visualisation, data mining). Disciplinary contributions to each field, in addition to processing power, seem to be in relatively portable data processing algorithms, often highly mathematical ones. Computational linguistics, for example employs the generic Expectation Maximisation and Viterbi Search algorithms [54]. It does not seem to contribute developed methodologies for extracting meaning from data types, in the same way that information science has contributed Author Cocitation Analysis [55], Link analysis [53] and citation analysis [56]. Perhaps this distinction is partly phenomenological: very application-oriented research by computer scientists may not be labelled as computer science research. Nevertheless, it seems that what computer scientists excel at is producing software and systems to match certain goals rather than producing computer centred methodologies for tackling research questions.

The blog case study above shows that computer scientists can develop tools for the extraction of social information from data source, when there is a market for them. It seems to be the case that computer scientists do not tend to independently create tools designed for research outside of computer science, so this is a role that information science may fill. In fact computer scientists have a rich tradition of giving software for free, exemplified by the open source community, and these are sometimes used in others' research (e.g., [57]) but this is typically not the computer scientists' primary purpose. There are also some examples of research tools developed by computer scientists and used by social scientists including the Pajek and VisOne network drawing software packages. But there are also many examples (more, we believe) of programs created and used by social scientists. Examples in the field of network analysis include UCINET, NetMiner, SocSciBot and, in bibliometrics, BibExcel.

Developing Methods Suitable for New Information Sources. Information scientists already contribute to research in other disciplines by applying information science methods to analyze specific data sources. For example, bibliometrics researchers with the expertise to analyse the Institute for Scientific Information's data can contribute to subject-specific investigations in any academic field. Others analyse the same data but orient towards applying the techniques to make advances in theory outside of information science. For example, Leydesdorff has developed a suite of tools for the co-word analysis of bibliographic records, investigating questions typically grounded in Luhmann's sociology of communicative systems [58,59].

The development of methods to analyse a particular type of data is not peculiar to information science; arguably research fields are defined by their methods and practices rather than just the content of their object of study [60, p.8-12] so the whole of science is much more methods-oriented than is apparent from a surface perspective. Statisticians perhaps deserve a special mention as specialists in information analysis methods, but there are other information methods specialists such as data miners. Nevertheless, because methods are so central to disciplines, it is unlikely that one generic research method for a new data source would be appropriate to a wide range of different fields. Hence any method developed by

information scientists for social scientists would have to be *customisable* so that individual researchers could adapt it to suit local needs.

5   Conclusion

We have argued that information science, as a field, is particularly well placed to assess new information sources for their use in social sciences research, including large scale and hybrid or fuzzy data sets. This includes developing appropriate general methods and programs in addition to identifying the types of research questions that the data may support. In practice, since information science is a large and varied field, this argument will apply to a minority of researchers, perhaps those with access to programming skills and an interest in social sciences research. The payoff for our field is an increased profile within the social sciences. One lesson that we will need to learn, however, is the need to react faster to new developments, because the blog examples suggest that other fields have been quicker to respond.

References

1. NCeSS, *What is the Grid?* 2004: p. http://www.ncess.ac.uk/grid/index.shtml.
2. NCeSS, *Frequently asked questions.* 2004: p. http://www.ncess.ac.uk/resources/faqs/.
3. Garfield, E., *Citation indexing for studying science.* Nature, 1970. 227(669-671).
4. Wouters, P., *The creation of the SCI*, in *Proceedings of the 1998 Conference on the History and Heritage of Science Information Systems*, M.E. Bowden, T.B. Hahn, and R.V. Williams, Editors. 1999. p. 127-136.
5. Wouters, P. *Cyberscience: The informational turn in science*. in *Lecture at the Free University*. 2000. Amsterdam.
6. Beaulieu, A., *From brainbank to database: the informational turn in the study of the brain.* Studies in History and Philosophy of Biological and Biomedical Sciences., 2004.
7. Ingwersen, P., *The calculation of Web Impact Factors.* Journal of Documentation, 1998. 54(2): p. 236-243.
8. Rodríguez i Gairín, J.M., *Valorando el impacto de la información en Internet: AltaVista, el "Citation Index" de la Red.* Revista Española de Documentación Científica, 1997. 20(2): p. 175-181.
9. Thelwall, M., *Evidence for the existence of geographic trends in university web site interlinking.* Journal of Documentation, 2002. 58(5): p. 563-574.

10. Thelwall, M. and G. Harries, *Do better scholars' Web publications have significantly higher online impact?* Journal of American Society for Information Science and Technology, 2004. 55(2): p. 149-159.

11. Thelwall, M. and A.G. Smith, *A study of interlinking between Asia-Pacific university web sites.* Scientometrics, 2002. 55(3): p. 335-348.

12. Polanco, X., et al., *Clustering and mapping Web sites for displaying implicit associations and visualising networks.* 2001, University of Patras.

13. Chen, C., et al., *How did university departments interweave the Web: A study of connectivity and underlying factors.* Interacting With Computers, 1998. 10(4): p. 353-373.

14. Thomas, O. and P. Willet, *Webometric analysis of departments of Librarianship and information science.* Journal of Information Science, 2000. 26(6): p. 421-428.

15. Chu, H., S. He, and M. Thelwall, *Library and Information Science Schools in Canada and USA: A Webometric Perspective.* Journal of Education for Library and Information Science, 2002. 43(2): p. 110-125.

16. Li, X., et al., *The relationship between the WIFs or Inlinks of computer science departments in UK and their RAE ratings or research productivities in 2001.* Scientometrics, 2003. 57(2): p. 239-255.

17. Tang, R. and M. Thelwall, *US academic departmental web-site interlinking in the United States disciplinary differences.* Library and Information Science Research, 2003. 25(4): p. 437-458.

18. Wasserman, S. and K. Faust, *Social network analysis: Methods and applications.* 1994, Cambridge, NY: Cambridge University Press.

19. Garrido, M. and A. Halavais, *Mapping networks of support for the Zapatista movement: Applying Social Network Analysis to study contemporary social movements*, in *Cyberactivism: Online activism in theory and practice*, M. McCaughey and M. Ayers, Editors. 2003, Routledge: London. p. 165-184.

20. Rogers, R., *Operating issue networks on the Web.* Science as Culture, 2002. 11(2): p. 191-214.

21. Rogers, R., *Information Politics on the Web*. 2004, Massachusetts: MIT Press.

22. Park, H.W., G.A. Barnett, and I. Nam, *Hyperlink affiliation network structure of top web sites: Examing affiliates with hyperlink in Korea.* Journal of American Society for Information Science and Technology, 2002. 53(7): p. 592-601.

23. Foot, K.A., et al., *Analyzing linking practices: Candidate sites in the 2002 US electoral web sphere.* Journal of Computer Mediated Communication, 2003. 8(4): p. http://www.ascusc.org/jcmc/vol8/issue4/foot.html.

24. Hine, C., *Virtual Ethnography.* 2000, London: Sage.

25. Brin, S. and L. Page, *The anatomy of a large scale hypertextual Web search engine.* Computer Networks and ISDN Systems, 1998. 30(1-7): p. 107-117.

26. Wheeldon, R. and M. Levene. *The best trail algorithm for assisted navigation of Web sites.* in *1st Latin American Web Congress (LA-WEB 2003)*. 2003. Sanitago, Chile: IEEE Computer Society.

27. Broder, A., et al., *Graph Structure in the Web.* Journal of Computer Networks, 2000. 33(1-6): p. 309-320.

28. Chakrabarti, S., et al., *The structure of broad topics on the Web*. 2002, WWW2002.

29. Menczer, F., *Lexical and semantic clustering by web links.* Journal of the American Society for Information Science and Technology, 2004. 55(14): p. 1261-1269.

30. Ess, C. and A.E.W. Committee, *Ethical decision-making and Internet research. Recommendations from the aoir ethics working committee.* 2002.

31. Bar-Ilan, J., *The 'Mad Cow Disease', Usenet Newsgroups and bibliometric laws.* Scientometrics, 1997. 39(1): p. 29-55.
32. Caldas, A., *Are newsgroups extending 'invisible colleges' into the digital infrastructure of science?* Economics of Innovation and New Technology, 2003. 12(1): p. 43-60.
33. Kling, R., *The Internet and unrefereed scholarly publishing.* Annual Review of Information Science and Technology, 2004. 38: p. 591-631.
34. MacKinnon, R.C., *Searching for the Leviathan in Usenet*, in *Cybersociety: Computer-mediated communication and community*, S. Jones, Editor. 1995, Sage: Thousand Oaks, CA. p. 112-137.
35. Herring, S.C., *Computer-mediated communication on the Internet.* Annual Review of Information Science and Technology, 2002. 36: p. 109-168.
36. Konstan, J., et al., *GroupLens: Applying collaborative filtering to usenet news.* Communications of the ACM, 1997. 40(3): p. 77-87.
37. Rall, D.N., *Exploring the breadth of disciplinary backgrounds in internet scholars participating in AoIR meetings, 2000-2003.* Proceedings of AoIR 5.0, 2004: p. http://gsb.haifa.ac.il/~sheizaf/AOIR5/399.html.
38. Silver, D., *Internet/cyberculture/digital culture/new media/fill-in-the-blank studies.* New Media & Society, 2004. 6(1): p. 55-64.
39. Schaap, F. *Multimodal interactions and singular selves: Dutch weblogs and home pages in the context of everyday life.* in *AoIR 5.0.* 2004. Brighton, UK.
40. Gorgura, H. *The war on the terror consensus: Anti-war blogs as an online sphere of dissensus.* in *AoIR 5.0.* 2004. Brighton, UK.
41. Bar-Ilan, J., *Blogarians – A new breed of librarians*, in *Proceedings of the American Society for Information Science & Technology.* 2004.
42. Gruhl, D., et al. *Information diffusion through Blogspace.* in *WWW2004.* 2004. New York, http://www.www2004.org/proceedings/docs/1p491.pdf.
43. Gibbons, M., et al., *The New Production of Knowledge.* 1994, London, UK: Sage.
44. Whitley, R., *The Intellectual and Social Organization of the Sciences.* 2 ed. 2000, Oxford: Oxford University Press.
45. Fuchs, S., *The Professional Quest for Truth: A Social Theory of Science and Knowledge.* 1992, Albany, NY: SUNY Press.
46. Wouters, P., *The Virtual Knowledge Studio for the Humanities and Social Sciences.* 2004, The Royal Netherlands Academy of Arts and Sciences: Amsterdam.
47. Finholt, T., *Collaboratories.* Annual Review of Information Science and Technology, 2002. 36: p. 73-107.
48. Lenoir, T. *Shaping Biomedicine as an Information Science.* 1998.
49. Van Couvering, E. *New media? The political economy of Internet search engines.* in *Annual Conference of the International Association of Media & Communications Researchers.* 2004. Porto Alegre, Brazil.
50. Rehm, G. *Towards automatic web genre identification.* in *the 35th Hawaii International Conference on System Sciences.* 2002.
51. Herring, S.C., *Slouching toward the ordinary: Current trends in Computer-Mediated Communication.* New Media & Society, 2004. 6(1): p. 26-36.
52. Park, H.W., *Hyperlink network analysis: A new method for the study of social structure on the web.* Connections, 2003. 25(1): p. 49-61.
53. Thelwall, M., *Link Analysis: An Information Science Approach.* 2004, San Diego: Academic Press.
54. Mitkov, R., *The Oxford handbook of computational linguistics.* 2003, Oxford: Oxford University Press.

55. White, H.D. and K.W. McCain, *Visualizing a discipline: An author co-citation analysis of information science, 1972-1995.* Journal of the American Society for Information Science, 1998. 49(4): p. 327-355.

56. Garfield, E., *Citation indexing: its theory and applications in science, technology and the humanities.* 1979, New York: Wiley Interscience.

57. Goodrum, A.A., et al., *Scholarly publishing in the Internet age: a citation analysis of computer science literature.* Information Processing and Management, 2001. 37(5): p. 661-676.

58. Hellsten, I. and L. Leydesdorff, *Measuring the meaning of words in contexts: An automated analysis of controversies about 'Monarch butterflies,' 'Frankenfoods,' and 'stem cells.'* in preparation, 2005: p. http://users.fmg.uva.nl/lleydesdorff/meaning/measuring%20meaning.pdf.

59. Leydesdorff, L., *Words and co-words as indicators of intellectual organization.* Research Policy, 1989. 18: p. 209-223.

60. Hyland, K., *Disciplinary discourses: social interactions in academic writing.* 2000, Harlow: Longman.