Text prepared for the Inaugural Speech, Chair, New Media & Digital Culture, University of Amsterdam 8 May 2009

The End of the Virtual – Digital Methods

Richard Rogers

Situating Digital Methods in Internet research

Arguably, there is an ontological distinction between the natively digital and the digitized, that is, the objects, content, devices and environments that are "born" in the new medium, as opposed to those that have "migrated" to it. Should the current methods of study change, however slightly or wholesale, given the focus on objects and content of the *medium*? The research program put forward here thereby engages with "virtual methods" that import standard methods from the social sciences and the humanities. That is, the distinction between the natively digital and the digitized also could apply to current research methods. What kind of Internet research may be performed with methods that have been digitized (such as online surveys and directories) vis-à-vis those that are natively digital (such as recommendation systems and folksonomy)?

Second, I propose that Internet research may be put to new uses, given an emphasis on natively digital methods as opposed to the digitized. I will strive to shift the attention from the opportunities afforded by transforming ink into bits, and instead inquire into how research *with* the Internet may move beyond the study of online culture only. How to capture and analyze hyperlinks, tags, search engine results, archived Websites, and other digital objects? How may one learn from how online devices (e.g., engines and recommendation systems) make use of the objects, and how may such uses be repurposed for social and cultural research? Ultimately, I propose a research practice that grounds claims about cultural change and societal conditions in online dynamics, introducing the term "online groundedness." The overall aim is to rework method for Internet research, developing a novel strand of study, digital methods.

To date the methods employed in Internet research have served the purpose of critiquing the persistent idea of the Internet as a virtual realm apart. Such thinking arose from the discourse surrounding virtual reality in the late 1980s and early 1990s, and the Internet came to stand for a virtual realm, with opportunities for a redefinition of consciousness, identity, corporality, community, citizenry and (social movement) politics.[1] Indeed, in 1999 in one of the first efforts to synthesize Internet research, the communications scholar, Steve Jones invited researchers to move beyond the perspective of the Internet as a realm apart, and opened the discussion of method.[2] How would social scientists study the Internet, if they were not to rely on the approaches associated with it to date: human-computer interaction,

[1] Barlow, 1996; Benedict, 1991; Dibbell, 1998; Rheingold, 1991; Rheingold; 1993; Shaviro, 2008; Stone, 1995; Turkle, 1995.
[2] Jones, 1999.

social psychology and cybercultural studies?[3] In their ground-breaking work on Internet usage in Trinidad and Tobago, the ethnographers, Daniel Miller and Don Slater, challenged the idea of cyberspace as a realm apart where all "inhabiting" it experienced its identity-transforming affordances, no matter one's location.[4] Slater and Miller grounded the Internet, arguing that Trinis appropriated the medium, making it fit their own cultural practices. Whilst a case study, the overall thrust of the research was its potential for generalizability. If Trinis were using the Internet to stage Trini culture, the expectation is that other cultures are doing the same.

The important Virtual Society? program (1997-2002) marked another turning point in Internet research, with its debunking of the transformative capacities of cyberspace through multiple empirical studies about Internet users. The program ultimately formulated five 'rules of virtuality'.[5] In what is now the classic digital divide critique, researchers argued that the use of new media is based on one's situation (access issues), and the fears and risks are unequally divided (skills issues). With respect to the relationship between the real and the virtual, virtual interactions supplement rather than substitute for the 'real,' and stimulate more real interaction, as opposed to isolation and desolation. Finally, the research found that identities are grounded in both the online as well as the off-line. Significantly, the program settled on approaches that have been characterized as virtual methods, with an instrumentarium for studying users. Surveys, interviews, observation and participant-observation became the preferred methods of inquiry. In the humanities, subsequent user studies – concentrating on the amateur, the fan, and the 'produser' – also are grappling with the real and virtual divide, seeking to demonstrate and critique the reputational status of online culture.[6] The argument put forward here is that virtual methods and user studies in the social sciences and the humanities have shifted the attention away from the *data* of the medium, and the opportunities for study of far more than online culture.

How may one rethink user studies with data that are (routinely) collected by software? User studies to date have relied on accounts that privilege observation, interviews and surveys, owing, in one reading, to the difference in armatures between social scientific and humanities computing, on the one hand, and the large commercial companies, with their remarkable data collection achievements, on the other. In a sense, Google, Amazon and many other dominant Web devices are already conducting user studies, however much the term is not used. User inputs (preferences, search history, purchase history, location) are captured and analyzed so as to tailor results. Taking a lead from such work, the new media theorist, Lev Manovich, has called for a methodological turn in Internet research, at least in the sense of data collection. With "cultural analytics," named after Google Analytics, the proposal is to build massive collection, storage and analytical facilities for humanities computing.[7] One manner to describe the methodological turn is its marked departure from the reliance on (negotiated) access to commercial data sets, e.g., AOL's set of users' search engine queries, Linden Lab's set of the activities of millions of users in Second Life or Sony's for Everquest,

[3] Hine, 2000.
[4] Slater & Miller, 2000.
[5] Woolgar, 2002.
[6] Jenkins, 2006; Keen, 2007; Bruns, 2008.
[7] Manovich, 2007.

however valuable the findings have been.[8] Cultural analytics as well as what has been termed computational social science have a "big science" outlook.[9] 'Visualizations should be designed to take full advantage of the largest gigapixel wall-size displays available today.'[10]

In a sense the research programs are the scientific community's answer to the question, what would Google do? The programs could be situated in the larger context of the extent and effects of "googlization". To date the googlization critique, which originated in the reaction to the search engine company's entrance into the library (the Google Books project), has examined the growing "creep" of Google, its business model as well as its aesthetics across information and knowledge industries.[11] Especially library science scholars concern themselves with the changing locus of access to information and knowledge (from public shelves and stacks to commercial servers). "Google effects" also may be couched in terms of the supplanting of surfing and browsing by search. They also may be studied in terms of the demise of the expert editor, and the rise of the back-end algorithm, themes to which I return. Here, however, the point is that they also may be studied in terms of models for research – ones that seek to replicate the scale of data collection as well as analysis.

The proposal I am putting forward is more modest, yet still in keeping with what are termed approaches to user studies that are registrational. Online devices and software installed on the computer (e.g., browsers) register users' everyday usage. Browser histories would become a means to study use. The larger contention is that data collection, in the methodological turn described above, could benefit from thinking about how computing may have techniques which can be repurposed for research. Thus the proposal is to consider first and foremost the availability of computing *techniques*.

I would like to put forward a new era in Internet research, which no longer concerns itself with the divide between the real and the virtual. It concerns a shift in the kinds of questions put to the study of the Internet. The Internet is employed as a site of research for far more than *just* online culture. The issue no longer is how much of society and culture is online, but rather how to diagnose cultural change and societal conditions with the Internet. The conceptual point of departure for the research program is the recognition that the Internet is not only an object of study, but also a source. Knowledge claims may be made on the basis of data collected and analyzed by devices such as search engines. One of the more remarkable examples is Google Flu Trends, a non-commercial (Google.org) project launched in 2008, which anticipates local outbreaks of influenza by counting search engine queries for flu, flu symptoms and related terms, and 'geo-locates' the places where the queries have been made. It thereby challenges existing methods of data collection (emergency room reports), and reopens the discussion of the Web as anticipatory medium, closer to the ground than one expects.[12]

---

[8] Contractor, 2009.
[9] Lazer et al., 2009.
[10] Manovich, 2008.
[11] Jeanneney, 2007; Vaidhyanathan, 2007; Rogers, 2009.
[12] Rogers, 2003.

Where did the 'grounded Web,' and its associated geo-locative research practice, originate? The 'end of cyberspace' as a placeless space (in the terms of Manual Castells) may be located in the technical outcomes of the famous Yahoo lawsuit, brought by two non-governmental organizations in France in 2000.[13] At the time French Web users were able to access the Nazi memorabilia pages on Yahoo.com in the United States, and two French non-governmental organizations desired that the pages be inaccessible – in France. IP-to-geo (address location) technology was developed specifically to channel content nationally; when one types google.com into a browser in France, now google.fr is returned by default. This 'grounding' of the Web has been implemented by such major content-organizing projects, as Google, Microsoft Live and YouTube; online television is served geographically, too.

Diagnostic work as Google Flu Trends, whereby claims about societal conditions are made on the basis of captured Internet practices, leads to new theoretical notions. For the third period of Internet research, the digital methods program introduces the term *online groundedness*, in an effort to conceptualize research that follows the medium, captures its dynamics and makes grounded claims about cultural and societal change. Indeed, the broader theoretical goal of digital methods is to think through anew the relationship between the Web and the ground. Like the ethnographers who came before them, for the U.K. Virtual Society? program, one needed to visit the ground in order to study the Web. Here the research program complicates the order in which one's findings are grounded.[14] For example, journalism has methodological needs now that the Internet has become a significant meta-source, where the question normally concerns the trustworthiness of a source. Snowballing from source to source was once a social network approach to information-checking, to speak in terms of method. Who else should I speak to? That is the question at the conclusion of the interview, if trust has been built. The relationship between 'who I should speak to' and 'who else do you link to' is asymmetrical for journalism, but the latter is the question asked by search engines when recommending information. How to think through the difference between source recommendations from verbal and online links? Is search the beginning of the quest for information that ends with some grounded interview reality beyond the net, whereby we maintain the divide between some real and some virtual? Or is that too simplistic? Our ideal source set divide (real and virtual, grounded or googled) raises the question of what is next. What do we 'look up' upon conclusion of the interview to check the reality? The Internet may not be changing the hierarchy of sources for some (e.g., the restrictions on citing Wikipedia in certain educational settings), but it may well be changing the order of checking, and the relationship of the Web to the ground.

I developed the notion of online groundedness after reading a study performed by the Dutch newspaper, the *NRC Handelsblad*. The investigation into right-wing as well as hate groups in the Netherlands inquired into whether the language used was becoming more extremist over time, perhaps indicating a 'hardening' of right-wing and hate culture more generally. Significantly, the investigators elected to use the Internet Archive, over an embedded researcher (going native), or the pamphlets, flyers and other ephemera at the Social History Institute.[15] They located and analyzed the changes in tone over time on right-

[13] Castells, 1996; Goldsmith & Wu, 2006; Rogers, 2008.
[14] Marres & Rogers, 2008.
[15] NRC Handelsblad, 2007.

wing as well as extremist sites, finding that right-wing sites were increasingly employing more extremist language. Thus the findings made about culture were grounded through an analysis of Websites. Most significantly, the online became the baseline against which one may judge a societal condition.

Follow the Medium: The Digital Methods Research Program

Why follow the medium? A starting point is the recognition that Internet research is often faced with unstable objects of study. The instability is often discussed in terms of the ephemerality of Websites and other digital media, and the complexities associated with *fixing* them, to borrow a term from photography. How to make them permanent so that they can be studied with care? Web archiving is continually faced with the dilemma of capturing Websites on the one hand, and maintaining their liveliness on the other. In one approach, vintage hardware and software are maintained so as to keep the media 'undead.' In another, also practiced in game environments, the ephemerality issue is addressed through simulation/emulation, which keeps the nostalgic software, as Atari games, running on current hardware. The ephemerality issue, however, is much larger than the issues of preservation. The Internet researcher is often overtaken by events of the medium, such as software updates that 'scoop' one's research.

As a research practice, following the medium, as opposed to striving to fix it, may also be discussed in a term borrowed from journalism and the sociology of science – "scooping." Being the first to publish is to 'get the scoop.' 'Being scooped' refers to someone else having published the findings first. The sociologist of science, Michael Lynch, has applied this term to the situation in which one's research subjects come to the same or similar conclusions as the researchers, and go on record with their findings first. The result is that the "[research subjects] reconfigure the field in which we previously thought our study would have been situated".[16] In Internet research, 'being scooped' is common. Industry analysts, watchdogs and bloggers routinely coin terms (e.g., googlization) and come to conclusions that shape ongoing academic work. I would like to argue, however, that scooping is also done by the objects, which are continually reconfigured. For example, Facebook, the social networking site, has been considered a case of a 'walled garden,' a relatively closed community system, where by default only 'friends' can view information and activities of other friends. The 'walled garden' is a series of concentric circles: a user must have an account to gain access, must friend people to view their profiles and must change privacy default settings to let friends of friends view one's own profile. Maximum exposure is opening profiles to friends of friends. In March of 2009, Facebook changed a setting; users may make their profile open to all other users with accounts, as opposed to just friends, or friends of friends, in its previous configuration. Which types of research would be 'scooped' by Facebook's flipping of a switch? Which would benefit? Facebook serves as one notable example of the sudden reconfiguration of a research object, which is common to the medium.

More theoretically, following the medium is a particular form of medium-specific research. Medium specificity is not only how one sub-divides disciplinary commitments in media

[16] Lynch, 1997.

studies according to the primary objects of study: film, radio, television, etc. It is also a particular plea to take seriously ontological distinctiveness, though the means by which the ontologies are built differ. To the literary scholar and media theorist, Marshall McLuhan, media are specific in how they engage the senses.[17] Depth, resolution and other aesthetic properties have effects on how actively or passively one processes media. One is filled by media, or one fills it in. To the cultural theorist, Raymond Williams, medium specificity lies elsewhere. Media are specific in the forms they assume – forms that are shaped by the dominant actors to serve interests.[18] For example, the creation of 'flow,' the term for how television sequences programming so as to keep viewers watching, serves viewer ratings and advertising. Thus, to Williams, media are not a priori distinctive from one another, but can be made so. To Katherine Hayles, media have characteristics in their materiality; book specifies, whilst text does not.[19] Her proposal for 'media-specific analysis' is a comparative media studies program, which takes materially instantiated characteristics of media (e.g., hypertext in digital media), and enquires into their (simulated) presence in other media (e.g., print). One could take other media traits and study them across media. For example, as Alexander Galloway has argued, flow is present not only in radio and television, but also on the Web, where dead links disrupt surfing.[20]

Hayle's point of departure may be seen in Mathew Fuller's work on Microsoft Word and Adobe Photoshop, which studies how particular software constrains or enables text.[21] To Fuller a Microsoft document or a Photoshop image are specific outputs of software, distinctive from some document or some image. An accompanying research program would study the effects of (software) features, as Lev Manovich also points to in his work on the specificity of computer media. With these media Manovich's ontology moves beyond the outputs of media (Hayle's hypertextual print, Fuller's Word document and Photoshop image).[22] Computer media are metamedia in that they incorporate prior media forms, which is in keeping with the remediation thesis put forward by Jay David Bolter and Richard Grusin.[23] But, to Manovich, computer media not only refashion the outputs of other media; they also embed their forms of *production*.

The medium specificity put forward here lies not so much in McLuhan's sense engagement, Williams' socially shaped forms, Hayles's materiality, or other theorists' properties and features. Rather it is situated in method. Previously I described such work as 'Web epistemology'.[24] On the Web information, knowledge and sociality are organized by recommender systems – algorithms and scripts that prepare and serve up orders of URLs, media files, friends, etc. In a sense Manovich has shifted the discussion in this direction, both with the focus on forms of production (method in a craft sense) as well as with the methodological turn associated with the cultural analytics initiative. I would like to take this

[17] McLuhan, 1964.
[18] Williams, 1974.
[19] Hayles, 2004.
[20] Galloway, 2004.
[21] Fuller, 2003.
[22] Manovich, 2008.
[23] Bolter & Grusin, 1999.
[24] Rogers, 2004.

turn further, and propose that the underinterrogated methods of the Web also are worthy of study, both in and of themselves as well as in the effects of their spread to other media, e.g., TV shows recommended to Tivo users on the basis of their profiles.

The initial work in the area of Web epistemology was in the context of the politics of search engines. [25] It sought to consider the means by which sources are adjudicated by search engines. Why, in March of 2003, were the U.S. White House, the Central Intelligence Agency, the Federal Bureau of Investigation, the right-of-center Heritage Foundation and leading news organizations such as CNN the top returns for the query "terrorism"? In a sense the answer lies in how hyperlinks are handled. Hyperlinks, however, are but one digital object, to which may be added: the thread, tag, PageRank, wikipedia edit, robots.txt, post, comment, trackback, pingback, IP address, URL, whois, timestamp, permalink, social bookmark and profile. In no particular order, the list goes on. The proposal is to study how these objects are handled, specifically, in the medium, and learn from medium method.

In the following, I would like to introduce a series of medium objects, devices, spaces as well as platforms, first touching briefly on how they are often studied with digitized methods and conceptual points of departure from without the medium. Subsequently, I would like to discuss the difference it makes to research if one were to follow the medium – by learning from and reapplying how digital objects are treated by devices, how Websites are archived, how search engines order information and how geo-IP location technology serves content nationally or linguistically. What kinds of research can be performed through hyperlink analysis, repurposing insights from dominant algorithms? How to work with the Internet archive for social research? Why capture histories of Websites? How may search engine results be studied so as to display changing hierarchies of credibility, and the differences in source reliance between the Web, the news and blogosphere? Can geo-IP address location technology be reworked so as to profile countries and cultures? How may the study of social networking sites reveal cultural tastes and preferences? How are software robots changing how quality content is maintained on Wikipedia? What would a research bot do? Thus, from the micro to the macro, I treat the hyperlink, Website, search engine and spheres (including the Websphere, blogosphere, newssphere, etc.), and the Web (or Webs, including national ones). I finally turn to social networking sites as well as Wikipedia, and seek to learn from these profiling and bot cultures (respectively), and rethink how to deploy them analytically. The overall purpose of following the medium is to reorient Internet research to consider the Internet as a source of data, method and technique.

The Link

How is the hyperlink often studied? There are at least two dominant approaches to studying hyperlinks: hypertext literary theory and social network theory, including small world and path theory.[26] To literary theorists of hypertext, sets of hyperlinks form a multitude of distinct pathways through text. The surfer, or clicking text navigator, may be said to author a

[25] Introna & Nissenbaum, 2000.
[26] Landow, 1994; Watts, 1999; Park & Thewall, 2003

story by choosing routes (multiple clicks) through the text.[27] Thus the new means of authorship as well as the story told through link navigation are of interest. For small world theorists, the links that form paths show distance between actors. Social network analysts use pathway thought, and zoom in on how the ties, uni-directional or bi-directional, position actors.[28] There is a special vocabulary that has been developed to characterize an actor's position, especially an actor's centrality, within a network. For example, an actor is 'highly between' if there is a high probability that other actors must pass through him to reach each other.

How do search engines treat links? Theirs arguably is a scientometric (and associational sociology) approach. As with social network analysis, the interest is in actor positioning, but not necessarily in terms of distance from one another, or the means by which an actor may be reached through networking. Rather, ties are reputational indicators, and may be said to define actor standing. Additionally, the approach does not assume that the ties between actors are friendly, or otherwise have utility, in the sense of providing empowering pathways, or clues for successful networking.

Here I would like to follow how engines treat links as markers of impact and reputation. How may an actor's reputation be characterized by the types of hyperlinks given and received? Actors can be profiled not only through the quantity of links received, as well as the quantity they received from others which themselves have received many links, in the basic search engine algorithm. Actors may also be profiled by examining which links they give and receive in particular.[29] In previous research colleagues and I found linking tendencies among domain types, i.e., governments tend to link to other governmental sites only, non-governmental sites tend to link to a variety of sites, occasionally including critics. Corporate Websites tend not to link, with the exception of collectives of them – industry trade sites and industry "front groups" do link. Academic and educational sites typically link to partners and initiatives they have created. Taken together these linking proclivities of organization types show an everyday "politics of association".[30] For example, in work colleagues and I conducted initially in 1999, we found that Greenpeace linked to governmental sites, and government did not link back. Novartis, the multinational corporation, linked to Greenpeace, and Greenpeace did not link back. When characterizing an actor according to inlinks and outlinks, one notices whether there is some divergence from the norms, and more generally whether particular links that are received may be telling for an actor's reputation. A non-governmental organization receiving a link from a governmental site could be construed as a reputation booster, for example.[31]

---

[27] Elmer, 2001.
[28] Krebs, 2002.
[29] cf. Beaulieu, 2005.
[30] Marres & Rogers, 2000; Rogers, 2002.
[31] The Issue Crawler software, with particular allied tools, has been developed specifically to perform such hyperlink analysis. Websites are crawled, and links are gathered and stored. The crawler-analytical modules are adaptations from scientometrics (co-link analysis) and social networking analysis (snowball). Once a network is located with the Issue Crawler, individual actors may be profiled, using the actor profiler tool. The actor profiler shows, in a graphic, the inlinks and outlinks of the top ten network actors. The other technique for actor

Apart from capturing the micro-politics of hyperlinks, analysis of links also may be put to use in more sophisticated sampling work. Here the distinction between digitized and natively digital method stands out in greater relief. The Open Net Initiative at the University of Toronto conducts Internet censorship research by building lists of Websites (from online directories such as the Open Directory Project and Yahoo). The researchers subsequently check whether the sites are blocked in a variety of countries. It is important work that sheds light on the scope as well as technical infrastructure of state Internet censorship practices worldwide.[32] In the analytical practice, sites are grouped by category: famous bloggers, government sites, human rights sites, humor, women's rights, etc.; there are approximately 40 categories. Thus censorship patterns may be researched by site type across countries.

The entire list of Websites checked per country (some 3,000) is a sample, covering of course only the smallest fraction of all Websites as well as those of a particular subject category. How would one sample Websites in a method that follows the medium, learning from how search engines work (link analysis) and repurposing it for social research? Colleagues and I contributed to the Open Net Initiative work by employing a method that crawls all the Websites in a particular category, captures the hyperlinks from the sites, and determines additional key sites (by co-link analysis) that are not on the lists. I dubbed the method 'dynamic URL sampling', in an effort to highlight the difference between manual URL-list compilation, and more automated techniques of finding significant URLs. Once the new sites are found, they are checked for connection stats (through proxies initially, and later perhaps from machines located in the countries in question), in order to determine whether they are blocked. In the research project on 'social, political and religious' Websites in Iran, researchers and I crawled all the sites in that ONI category, and through hyperlink analysis, found some 30 previously unknown blocked sites. Significantly, the research was also a page-level analysis (as opposed to host only), with one notable finding being that Iran was not blocking the BBC news front page (as ONI had found), but only its Persian-language page. The difference between the two methods of gathering lists of Websites for analysis – manual directory-style work and dynamic URL sampling – shows the contribution of medium-specific method.

The Website

Up until now, investigations into Websites have been dominated by user and 'eyeball studies,' where attempts at a navigation poetics are met with such sobering ideas as 'don't make me think'.[33] Many of the methods for studying Websites are located over the shoulder, where one observes navigation or the use of a search engine, and later conducts interviews with the subjects. In what one may term classic registrational approaches, a popular technique is eye-tracking. Sites load and eyes move to the upper left of the screen, otherwise known as the golden triangle. The resulting heat maps provide site redesign cues. For

---

profiling relies on a scraper that would capture all outlinks from a site, and a scraper of a search engine, the Yahoo inlink ripper, which provides a list of the links made to a Website.

[32] Diebert et al., 2006.

[33] Krug, 2000; Dunne, 2005.

example, Google.com has moved its services from above the search box (tabs) to the top left corner of the page (menu). Another dominant strand of Website studies lies in feature analysis, where sites are compared and contrasted on the basis of levels of interactivity, capacities for user feedback, etc.[34] The questions concern whether a particular package of features result in more users, and more attention. In this tradition, most notably in the 9/11 special collection, Websites are often archived for further study. Thus much of the work lies in the archiving of sites prior to the analysis. One of the crucial tasks ahead is further reflection upon the means by which Websites are captured and stored, so as to make available the data upon which findings are based. Thus the digital methods research program engages specifically with the Website as archived object, made accessible, most readily, through the Internet Archive's Wayback Machine. It asks, which types of studies of Websites are enabled and constrained by the Wayback Machine?

In order to answer that question, the work first deconstructs, or unpacks, the Internet Archive and its Wayback Machine. In which senses does the Internet Archive as an object, formed by the archiving process, embed particular preferences for how it is used, and for the type of research performed with it? Indeed, the Web archiving scholar, Niels Brügger, has written: '[U]nlike other well-known media, the Internet does not simply exist in a form suited to being archived, but rather is first formed as an object of study in the archiving, and it is formed differently depending on who does the archiving, when, and for what purpose.'[35] That the object of study is constructed by the means by which it is tamed and captured by method and technique is a classic point from the sociology and philosophy of science and elsewhere.[36] Thus the initial research questions are, which methods of research are privileged by the specific form assumed by the Web archive, and which are precluded? For example, when one uses the Internet Archive (archive.org), what stands out for everyday Web users accustomed to search engines, is not so much the achievement of the existence of an archived Internet. Rather, the user is struck by how the Internet is archived, and, particularly, how it is queried. One queries a URL, as opposed to key words, and one receives a list of stored pages associated with the URL from the past. In effect, the Internet Archive, through the interface of the Wayback Machine, has organized the story of the Web into the histories of single Websites.

Which types of research approaches are favored by the current organization of Websites by the Internet Archive? With the Wayback Machine, one can study the evolution of a single page (or multiple pages) over time, for example, by reading or collecting snapshots from the dates that a page has been indexed. How can such an arrangement of historical sites be put to use? Previously I mentioned the investigative reporting work done by the *NRC Handelsblad* in their analysis of the rise of extremist language in the Netherlands. The journalists read some hundred Websites from the Internet archive, some dating back a decade. It is work that should be built upon, methodologically as well as technically. One could scrape the pages of the right-wing and extremist sites from the Internet Archive, place the text (and images) in a database, and systematically query it for the presence of particular

[34] Foot & Schneider, 2006.
[35] Brügger, 2005, p. 1.
[36] Latour & Woolgar, 1986; Knorr-Cetina, 1999; Walker, 2005.

keywords over time. As the *NRC Handelsblad* did, one could determine changes in societal conditions through archived Website analysis of particular sets of sites.

How else to perform research with the Internet Archive? The digital methods program has developed means to capture the history of sites by taking snapshots and assembling them into a movie, in the style of time-lapsed photography.[37] As a demonstration of how to use the Internet archive for capturing such evolutionary histories, colleagues and I took snapshots of the frontpages of Google from 1998 up to the end of 2007. The analysis concerned the subtle changes made to the interface, in particular the tabs. We found that the directory project, the organization of the Web by topic undertaken by human editors, has been in decline. After its placement on the frontpage of Google in 2001, it was moved in 2004 under the 'more' button, and in 2006 under 'even more.' By late 2007, with the removal of the 'even more' option, one had to search Google in order to find its directory.[38] The larger issue of the demise of the human editor, read in this case from the evolution of Google's interface, has far-reaching implications for how knowledge is collected and ordered. Indeed, after examining Google, researchers and I turned to Yahoo, the original Web directory, and found that there, too, the directory had been replaced by the back-end algorithm. In examining the outputs of a query in the directory, we also learned that at Yahoo the results are no longer ordered alphabetically, in the egalitarian-style of information and source ordering, inherited from encyclopedias. Yahoo is listing its directory sources according to *popularity*, in the well-known style of recommendation systems more generally.

Are the histories of search engines, captured from their interface evolutions, indicating changes in how information and knowledge are ordered more generally? A comparative media studies approach would be useful, with one of the more poignant cases being the online newspaper. With the *New York Times*, for example, articles are still placed on the front page and in sections, but are also listed by 'most emailed' and 'most blogged', providing a medium-specific recommender system for navigating the news. The impact of recommender systems – the dominant means on the Web by which information and knowledge are ordered – may also be studied through user expectations. Are users increasingly expecting Web-like orderings at archives, libraries, tourist information centers and other sites of knowledge and information queries?

The Search Engines & the Spheres

The study of search engines was jolted by the now infamous AOL search engine data release in 2006, where 500,000 users' searches over three months were put online, with frightening and often salacious press accounts about the level of intimate detail revealed about searchers, even if their histories are anonymized (no names) and decoupled from geography (no IP address). One may interpret the findings from the AOL case as a shift in how one considers online presence, if that remains the proper term. A person may be 'googled', and his or her self-authored presence often appears at or towards the top of the returns. Generally

[37] Screen-capturing software has been employed previously for the analysis of Wikipedia pages, showing the evolution of entries and thus how Wikipedians build knowledge.
[38] The 'even more' button returned to the interface of Google.com in 2008.

speaking, what others have written about a person would appear lower down in the rankings. However, with search engine queries stored, a third set of traces could come to define an individual. It opens up policy questions. How long may an engine company keep search histories? Thus search engines are being studied in the legal arena, especially in terms of how data retention laws may be applied to search.

Previously I mentioned another strand in search engine studies, summed up in the term 'googlization.' It is a political economy style critique that considers how Google's free-service-for-profile model may be spreading across industries and (software) cultures. I have covered the critique elsewhere, striving to propose a research agenda for googlization scholars which includes front-end and back-end googlization. Front-end googlization would include the study of the information politics of the interface (including the demise of the human-edited directory). Back-end googlization concerns the rise of the algorithm that recommends sources hierarchically, instead of alphabetically, as mentioned above. The significance of studying the new information hierarchies of search engines also should be viewed in light of user studies. A small percentage of users set preferences to more than 10 results per page; typically they do not look past the first page of results; and they increasingly click the results appearing towards the top.[39] Thus the power of search engines lies in the combination of its ranking practices (source inclusion in the top results) together with the users' apparent "respect" for the orderings (not looking further). Google's model also relies on registrational interactivity, where a user's preferences as well as history are registered, stored and employed, increasingly, to serve tailored results. Prior to the Web and search engine algorithms and recommendation systems, interactivity was 'consultational,' with pre-loaded information that would be 'called up'.[40] A query would return the same information for all users at any given time. Now the results are dynamically generated based on one's registered preferences, history and location.

The different orders of sources and things served by engines are under-studied, largely because they are not stored, and made available for research, apart from the AOL data release, or other negotiated agreements with search engine companies. Where Google is concerned, the company once made available an API (application programming interface) that allowed for data collection. A limited number of queries could be made per day, and the results repurposed. Researchers relying on the API were scooped by Google when it discontinued or 'deprecated' the service in late 2006. With its reintroduction in a different form in 2009, Google emphasized, however, that automated queries and the permanent storage of results are against the terms of service. How to study search engine results under such conditions? Colleagues and I scrape Google, and put up a notice appreciating Google's forbearance.[41]

What may be found in Google's search engine results? As I have remarked, search engines, a crucial point of entry to the Web, are epistemological machines in the sense that they crawl, index, cache and ultimately order content. Previously I described the Web, and particularly a

[39] Spink & Jansen, 2004.
[40] Jensen, 1999.
[41] The notice appears on the credits page of the Issue Dramaturg, http://issuedramaturg.issuecrawler.net/.

search engine-based Web, as a potential collision space for alternative accounts of reality.[42] The phrasing built upon the work of the sociologist, C. Wright Mills, who characterized the purpose of social research as 'no less than to present conflicting definitions of reality itself'.[43] Are engines placing alternative accounts of reality side by side, or do the results align with the official and the mainstream? Storing and analyzing search engine results could answer such questions. Such has been the purpose of the software project called the Issue Dramaturg, so called for the potential drama on display within the top results, whereby sites may climb to or suddenly fall from the top. It is important to point out that top engine placements are highly sought after; organizations make use of search engine optimization techniques so as to boost site visibility. There are white hat and black hat techniques, that is, those accepted by engines and those that prompt engines to delist Websites from results until there is compliance again with engine etiquette.

In the Issue Dramaturg project, colleagues and I have stored Google search engines results for the query, 9/11, as well as other keywords for two purposes. The one is to enquire into source hierarchies, as described above. Which sources are privileged? Which are "winning" the competition to be the top sources returned for particular queries? The other purpose has been to chart particular sources, in the approach to engine studies that I have termed 'source distance'. For the query 9/11, how far from the top of the engine returns are such significant actors in 9/11 accounts as the New York City government and the *New York Times*? Are such sources prominent, or do they appear side by side with sources that challenge more official and familiar views? Thus, apart from the New York City government and the *New York Times* another actor that we have monitored is the 9/11 truth movement (911truth.org). For months between March and September 2007, the 9/11 truth movement's site appeared in the top five results for the query 9/11, and the other two were well below result fifty. In mid-September 2007, around the anniversary of the event, there was drama. 911truth.org fell precipitously to result two hundred, and subsequently out of the top one thousand, that is, the maximum number of results served by Google. Colleagues and I believe that it is one of the first fully documented cases of the apparent removal of a Website in Google – from a top five placement for six months to a sub-one thousand ranking.[44] The case leads to questions of search engine result stability and volatility, and opens up an area of study.

However dominant it may be, there are more search engines than Google's Web search. What is less appreciated perhaps is that there are other dominant engines per section or sphere of the Web. For the blogosphere, there is Technorati, for the newssphere Google News, and the tagosphere or social bookmarking space, Delicious. Indeed, thinking of the Web in terms of spheres refers initially to the name of one of the most well-known, the blogosphere, as well as to scholarship that seeks to define another, the 'Web sphere'.[45] The sphere in blogosphere refers in spirit to the public sphere; it also may be thought of in terms of the geometrical form, where all points on the surface are the same distance from the center or core. One could think about such an equidistance measure as an egalitarian ideal, where every blog, or even every source of information, is knowable by the core, and vice

[42] Rogers, 2004.
[43] C. Wright Mills, 1971, p. 212; Rogers & Marres, 2002.
[44] Rogers, 2009.
[45] Foot & Schneider, 2002; Schneider & Foot, 2002.

versa. On the Web, it has been found, however, that certain sources are central. They receive the vast majority of links as well as hits. Following such principles as the rich get richer (aka Pareto power law distributions), the sites receiving attention tend to garner only more. The distance between the center and other nodes may only to grow, with the ideal of a sphere being a fiction, however much a useful one. I would like to put forward an approach that takes up the question of distance from core to periphery, and operationalize it as the measure of differences in rankings between sources per sphere. Spherical analysis is a digital method for measuring and learning from the distance between sources in different spheres on the Web.

Conceptually, a sphere is considered to be a device demarcated source set, i.e., the pure PageRank of all sources on the Web (most influential sites by inlink count), or indeed analogous pageranks of all sources calculated by the dominant engines per sphere, i.e., Technorati, Google News and Delicious. Thus, to study a sphere, we propose first to allow the engines to demarcate it. In sphere analysis one considers which sources are most influential, not only overall but per query. Cross-spherical analysis compares the sources returned by each sphere for the same query. It can therefore be seen as comparative ranking research. Most importantly, with cross-spherical analysis, one may think through the consequences of each engine's treatment of links, freshness, tags, etc. Do particular sources tend to be in the core of one sphere, and not in others? What does comparisons between sources, and source distances, across the spheres tell us about the quality of the new media? What do they tell us about current informational commitments in particular cultures?

In a preliminary analysis, colleagues and I studied which animals are most associated with climate change on the (English-language) Web, in the news and in the blogosphere. We found that the Web has the most diverse set of animals associated with climate change. The news favored the polar bear, and the blogosphere amplified, or made more prominent, the selection in the newssphere. Here we cautiously concluded that the Web may be less prone to the creation of media icons than the news, which has implications for studies of media that take as their point of departure a publicity culture. The blogosphere, moreover, appeared parasitic on the news as opposed to an alternative to it.

The Webs

As mentioned above, Internet research has been haunted by the virtual/real divide. One of the reasons for such a divide pertains to the technical arrangements of the Internet, and how they became associated with a virtual realm, cyberspace. Indeed, there was meant to be something distinctive about cyberspace, technologically.[46] The protocols and principles, particularly packet switching and the end-to-end principle, initially filled in the notion of cyberspace as a realm free from physical constraints. The Internet's technical indifference to the geographical location of its users spawned ideas not only of placeless-ness. In its architecture, it also supposedly made for a space untethered from the nation-states, and their divergent ways of treating flows of information. One recalls the famous quotation attributed to John Gilmore, co-founder with John Perry Barlow of the Electronic Frontier Foundation.

[46] Chun, 2006.

'The Internet treats censorship as a malfunction, and routes around it'.[47] Geography, however, was built in to cyberspace from the beginning, if one considers the locations of the original thirteen root servers, the unequal distributions of traffic flows per country as well as the allotment of IP addresses in ranges, which later enabled the application of geo-IP address location technology to serve advertising and copyright needs. Geo-IP technology as well as other technical means that locate (aka locative technology) also may be put to use for research that takes the Internet as a site of study, and inquires into what may be learned about societal conditions across countries. In the digital methods research program, colleagues and I have dubbed such work national Web studies.

Above I discussed the research by the British ethnographers who grounded cyberspace through empirical work on how Caribbean Internet users appropriated the medium to fit their own cultural practices. This is of course national Web studies, however with observational methods (from outside of the medium). To study the Web, nationally, one also may inquire into the data that are routinely collected, for example by large enterprises as Alexa's top sites by country (according to traffic). Which sites are visited most frequently per country, and what does site visitation say about a country's informational culture? Alexa pioneered registrational data collection with its toolbar, which users would install in their browsers. The toolbar provided statistics about the Website one had loaded in the browser, such as its freshness. All the Websites that the user loaded, or surfed, also would be logged, and the logged URLs would be compared with the URLs already in the Alexa database. Those URLs not in the database would be crawled, and fetched. Thus was born the Internet Archive.

The Internet Archive (1996 -) was developed during the period of Internet history, if I may, that one could term cyberspace. (I have developed periodizations of Internet history elsewhere, and will not further elaborate here.[48]) To illustrate the design and thought between the Internet Archive, and the national Web archives that are sprouting up in many countries, it may be pointed out that the Internet Archive was built for surfing – an Internet usage type that arguably has given way to search.[49] At the Wayback Machine of the Internet Archive, type in a single URL, view available pages, and browse them. If one reaches an external link, the Internet Archive looks up the page closest in date to the site one is exiting, and loads it. If no site exists in the Internet Archive, it connects to the live Website. It is the continuity of flow, from Website to Website, that is preserved.[50] National Web archives, on the other hand, have ceased to think of the Web in terms of cyberspace. Instead their respective purposes are to preserve national Webs. For the purposes of contributing method to Internet research, the initial question is, how would one demarcate a national Web?

At the National Library in the Netherlands, for example, the approach is similar to that of the Internet censorship researchers, discussed above. It is a digitized method, that is, a directory model, where an expert chooses significant sites, based on editorial criteria. These sites are continually archived with technology originally developed in the Internet Archive

[47] Boyle, 1997.
[48] Rogers, 2008.
[49] Shirky, 2005.
[50] Galloway, 2004.

project. At the time of writing, approximately one hundred national Websites are archived in the Netherlands – a far cry from what is saved at the Internet Archive.[51] In accounting for the difference in approaches and outcomes of the two projects, I would like to observe that the end of the virtual, and the end of cyberspace, have not been kind to Web archiving; the return of the nation-state and the application of certain policy regimes (especially copyright) have slowed efforts dramatically. Would digital methods aid in redressing the situation? I would like to invite national Web archivists to consider a registrational approach, e.g., the Alexa model adapted for a national context. The results may be salutary.

Social Networking Sites & Post-demographics

'We define social networking websites here as sites where users can create a profile and connect that profile to other profiles for the purposes of making an explicit personal network.'[52] Thus begins the study of American teenage use of such sites as MySpace and Facebook, conducted for the Pew Internet & American Life Project. Surveys were made. 91% of the respondents use the sites to 'manage friendships'; less than a quarter use the sites to 'flirt'. Other leading research into social networking sites considers such issues as presenting oneself and managing one's status online, the different 'social classes' of users of MySpace and Facebook and the relationship between real-life friends and 'friended' friends.[53] Another set of work, often from software-making arenas, concerns how to make use of the copious amounts of data contained in online profiles, especially interests and tastes. I would like to dub this latter work 'post-demographics.' Post-demographics could be thought of as the study of the data in social networking platforms, and, in particular, how profiling is, or may be, performed. Of particular interest here are the potential outcomes of building tools on top of profiling platforms. What kinds of findings may be made from mashing up the data, or what may be termed meta-profiling?

Conceptually, with the 'post' prefixed to demographics, the idea is to stand in contrast to how the study of demographics organizes groups, markets and voters in a sociological sense. It also marks a theoretical shift from how demographics have been used 'bio-politically' (to govern bodies) to how post-demographics are employed 'info-politically,' to steer or recommend certain information to certain people.[54] The term post-demographics also invites new methods for the study of social networks, where of interest are not the traditional demographics of race, ethnicity, age, income, and educational level – or derivations thereof such as class – but rather of tastes, interests, favorites, groups, accepted invitations, installed apps and other information that comprises an online profile and its accompanying baggage. Demographers normally would analyze official records (births, deaths, marriages) and survey populations, with census-taking being the most well known of those undertakings. Profilers,

[51] By comparison, in current national Web archiving efforts at the National Library in the Netherlands, the total collection of national Websites on offer is a fraction of the entire national Web, with approximately one hundred Dutch Websites archived of the three million in existence (in the .nl domain only as of 2008). See Weltevrede, 2009.
[52] Lenhart & Madden, 2007.
[53] Boyd & Ellison, 2007.
[54] Foucault, 1998; Rogers, 2004.

contrariwise, have users input data themselves in platforms that create and maintain social relations. They capture and make use of information from users of online platforms.

Perhaps another means of distinguishing between the two types of thought and practice is with reference to the idea of digital natives, those growing up with online environments, and unaware of life prior to the Internet, especially with the use of manual systems that came before it, like a library card catalogue.[55] The category of digital natives, however, takes a 'generational' view, and in that sense is a traditional demographic way of thinking. The post-demographic project would be less interested in new digital divides (digital natives versus non-natives) and the narratives that emerge around them (e.g., moral panics), but rather in how profilers recommend information, cultural products, events or other people (friends) to users, owing to common tastes, locations, travel destinations and more. There is no end to what *could* be recommended, if the data are rich and stored. How to study the data?

With 'post-demographics,' the proposal is to make a contribution to Internet research by learning from those profilers and researchers that both collect as well as harvest (or scrape) social networking sites' data for further analysis or software-making, such as mash-ups.[56] How do social networking sites make available their data for profilers? Under the developers' menu item at Facebook, for example, one logs in and views the fields available in the API (or application programming interface). Sample scripts are provided, as in 'get friends of user number x,' where x is yourself. Thus the available scripts generally follow the privacy culture, in the sense that the user decides what the profiler can see. It becomes more interesting to the profiler when many users allow access, by clicking 'I agree' on a third-party application.

Another set of profiling practices are not interested in personal data per se, but rather in tastes and especially taste relationships. One may place many profiling activities in the category of depersonalized data analysis, including Amazon's seminal recommendation system, where it is not highly relevant which person also bought a particular book, but rather that people have done so. Supermarket loyalty cards and the databases storing purchase histories similarly employ depersonalized information analysis, where like Amazon, of interest is the quantity of particular items purchased as well as the purchasing relationships (which chips with which soft drink). Popular products are subsequently boosted. Certain combinations may be shelved together.

Whilst they do not describe themselves as such, of course the most significant post-demographic machines are the social networking platforms themselves, collecting user tastes, and showing them to others, be they other friends, everyday people watchers or profilers. Here I would like to describe briefly one piece of software researchers and I built on top of the large collection device, MySpace, and the kinds of post-demographic analytical practices that result.

---

[55] Prensky, 2001.

[56] Non-users refer to profilers. Of course, profilers also may be users of the platforms, and most probably are, for one's sense of what may be mined, and how it may be analyzed or mashed up, would come from usage, with at least a minimal level of activity.

Elfriendo.com is the outcome of thinking through how to make use of the profiles on the social networking platform, MySpace. At Elfriendo.com, enter a single interest, and the tool creates a new profile on the basis of the profiles of people expressing that single interest. One may also compare the compatibility of interests, i.e., whether one or more interests, tunes, movies, TV shows, books and heroes are compatible with other ones. Is Christianity compatible with Islam, in the sense that those people with one of the respective interests listen to the same music and watch the same television programs? Elfriendo answers those sorts of questions by analyzing sets of friends' profiles, and comparing interests across them. Thus a movie, TV show, etc. has an aggregate profile, made up of other interests. (To wit, Eminem, the rapper, appears in both the Christianity and Islam aggregate profiles, in early February 2009.) One also may perform a semblance of post-demographic research with the tool, gaining an appreciation of relational taste analysis with a social networking site, more generally.[57]

It is instructive to state that MySpace is more permissive and less of a walled garden than Facebook, in that it allows the profiler to view a user's friends (and his/her friends' profiles), without you having friended anybody. Thus, one can view all of Barack Obama's friends, and their profiles. Here, in an example, one queries Elfriendo for Barack Obama as well as John McCain, and the profiles of their respective sets of friends are analyzed. The software counts the items listed by the friends under interests, music, movies, TV shows, books and heroes. What does this relational taste counting practice yield? The results provide distinctive pictures of the supporters of the two presidential candidates campaigning in 2008. The compatibility level between the interests of the friends of the two candidates is generally low. The two groups share few interests. (The tastes of the candidates' friends are not compatible for movies, music, books and heroes, though for TV shows the compatibility is 16%. There seem to be particular media profiles for each set of candidate's friends, where those of Obama for example watch the Daily Show, and those of McCain watch Family Guy, Top Chef and America's Next Top Model. Both sets of friends watch Lost. The findings may be discussed in terms of voter post-demographics in the sense that the descriptions of voter profiles are based on media tastes and preferences as opposed to educational levels, income and other standard indicators.

Wikipedia & Networked Content

To date the approaches to the study of Wikipedia have followed from certain qualities of the online encyclopedia, all of which appear counter-intuitive at first glance. One example is that Wikipedia is authored by so-called amateurs, yet is surprisingly encyclopedia-like, not only in form but in accuracy.[58] The major debate concerning the quality of Wikipedia vis-à-vis *Encyclopedia Britannica* has raised questions relevant to digital methods, in that the Web-enabled collective editing model has challenged the digitized work of a set of experts. However, research has found that there is only a tiny ratio of editors to users in Web 2.0

[57] One gains only a sense of how analysis may be performed, and the kinds of findings that may be made, because Elfriendo captures only the top 100 profiles, thus providing only an indication, as opposed to a grounded finding from a proper sampling procedure.
[58] Giles, 2005.

platforms, including Wikipedia. This is otherwise known as the myth of user-generated content.[59] Wikipedia co-founder, Jimbo Wales, has often remarked that the dedicated community is indeed relatively small, at just over 500 members. Thus the small cadre of Wikipedia editors could be considered a new elite, leading to exercises in relativizing the alleged differences between amateurs and experts, such as through a study of the demographics of Wikipedians.[60] Another example of a counter-intuitive aspect of Wikipedia is that the editors are unpaid, yet committed and highly vigilant. The vigilance of the crowd, as it is termed, is something of a mythical feature of a quality-producing Web, until one considers how vigilance is performed. Who is making the edits? One approach to the question lies in the Wikiscanner project (2007-), developed by Virgil Griffith studying at the California Institute of Technology. The Wikiscanner outs anonymous editors by looking up the IP address of the editor and checking it against a database with the IP address locations (geoIP technology). Wikipedia quality is ensured, to Griffith, by scandalizing editors making self-serving changes, such as a member of the Dutch Royal Family, who embellished an entry and made the front-page of the newspaper after a journalist used the tool.

How else are vandals kept at bay on Wikipedia, including those experimenters and researchers making erroneous changes to an entry, or creating a new fictional one, in order to keep open the debate about quality?[61] Colleagues and I have contributed to work about the quality of Wikipedia by introducing the term, networked content.[62] It refers to content held together by human authors and non-human tenders, including bots and alert software that revert edits or notify Wikipedians of changes made. Indeed, when looking at the statistics available on Wikipedia on the number of edits per Wikipedian user, it is remarkable to note that the bots are by far the top users. The contention, which is being researched in the digital methods program, is that the bots and the alert software are the significant agents of vigilance.

From the Wikiscanner project and the bots statistics related above, it is worth emphasizing that Wikipedia is a compendium of network activities and events, each logged and made available as large data sets. Wikipedia also has in-built reflection or reflexivity, as it shows the process by which an entry has come into being, something missing from encyclopedias and most other *finished* work more generally. One could study the process by which an entry matures; the materials are largely the revision history of an entry, but also its discussion page, perhaps its dispute history, its lock-downs and re-openings. Another approach to utilizing the data of Wikipedia would rely on the edit logs of one or more entries, and repurpose the Wikiscanner's technical insights by looking up where they have been made. "The places of edits" show subject matter concerns and expertise by organization and by country.

[59] Swartz, 2006.
[60] van Dijck, 2009.
[61] Chesney, 2006; Read, 2006; Magnus, 2008.
[62] Niederer, 2009.

Conclusion. The End of the Virtual – Grounding Claims Online

My aim is to set into motion a transformation in how and why one performs research with the Internet. The first step is to move the discussion away from the limitations of the virtual (how much culture and society are online) to the limitations of current method (how to study culture and society, and ground findings with the Internet).

I would like to conclude with a brief discussion of these limitations in Internet research as well as a proposal for renewal. First, the end of cyberspace and its placeless-ness, and the end of the virtual as a realm apart, are lamentable for particular research approaches and other projects. In a sense the real/virtual divide served specific research practices.[63] Previously I mentioned that Internet archiving thrived in cyberspace, and more recently it suffers without it. Where cyberspace once enabled the idea of massive Website archiving, the grounded Web and the national Webs are shrinking the collections.

Indeed, I have argued that one may learn from the methods employed in the medium, moving the discussion of medium specific theory from ontology (properties and features) to epistemology (method). The Internet, and the Web more specifically, have their ontological objects, such as the link and the tag. Web epistemology, among other things, is the study of how these natively digital objects are handled by devices. The insights from such a study lead to important methodological distinctions, as well as insights about the purpose of Internet research. Where the methodological distinction is concerned, one may view current Internet methods as those that follow the medium (and the dominant techniques employed in authoring and ordering information, knowledge and sociality) and ones that remediate or digitize existing method. The difference in method may have significant outcomes. One reason for the fallowing of the Web archiving efforts may lie in the choice of a digitized method (editorial selection) over a digital one (registrational data collection), such as that employed in the original Internet Archive project, where sites surfed by users were recorded. Indeed, I have employed the term digital methods so that researchers may consider the value and the outcomes of one approach over another. As a case in point, the choice of dynamic URL sampling over the editorial model could be beneficial to Internet censorship research, as I discussed.

Third, and finally, I have argued that the Internet is a site of research for far more than online culture and its users. With the end of the virtual/real divide, however useful, the Internet may be rethought as a source of data about society and culture. Collecting it and analyzing it for social and cultural research requires not only a new outlook about the Internet, but method, too, to ground the findings. Grounding claims in the online is a major shift in the purpose of Internet research, in the sense that one is not so much researching the Internet, and its users, as studying culture and society *with the Internet*. I hope you join me in this urgent project.

---

[63] The end of cyberspace also has not been helpful for projects relying on the classic Internet feature of the anonymous user. For example, organizations and governments ban employees from editing Wikipedia at work, for the edits may be traced to locations and made into subjects of scandal.

References

Barlow, J. P., 'A Declaration of the Independence of Cyberspace,' Davos, Switzerland, 1996, http://homes.eff.org/~barlow/Declaration-Final.html (accessed 28 January 2009).

Beaulieu, A., 'Sociable Hyperlinks: An Ethnographic Approach to Connectivity', in: C. Hine (ed), *Virtual Methods: Issues in Social Research on the Internet*. Berg, Oxford, 2005, pp. 183-197.

Benedict, M., 'Cyberspace: Some Proposals', in: M. Benedict (ed.), *Cyberspace – First Steps*. MIT Press, Cambridge, MA, 1991, pp. 119–224.

Bolter, J. D. and R. Grusin, *Remediation: Understanding New Media*. MIT Press, Cambridge, MA, 1999.

Boyd, D. and N. Ellison, 'Social network sites: Definition, history, and scholarship,' in: *Journal of Computer-Mediated Communication*, 13(1), 2007.

Boyle, J., 'Foucault in Cyberspace', in: *Univ. Cincinnati Law Review*, 66, 1997, pp. 177-205.

Brügger, N., *Archiving Websites: General Considerations and Strategies*. Centre for Internet Research, Aarhus, 2005.

Bruns, A., *Blogs, Wikipedia, Second Life, and Beyond: From Production to Produsage*. Peter Lang, New Yor, 2008.

Castells, M., *The Information Age: Economy, Society and Culture - The Rise of the Network Society*. Blackwell, Malden, MA, 1996.

Chesney, T., 'An empirical examination of Wikipedia's credibility', in: *First Monday*, 11(11), 2006.

Chun, W., *Control and Freedom: Power and Paranoia in the Age of Fiber*. MIT Press, Cambridge, MA, 2006.

Contractor, N., 'Digital Traces: An Exploratorium for Understanding and Enabling Social Networks', presentation at the annual meeting of the American Association for the Advancement of Science (AAAS), 2009.

Dibbell, J., *My Tiny Life: Crime and Passion in a Virtual World*. Henry Holt, New York, 1998.

Diebert, R., J. Palfrey, R. Rohozinski, and J. Zittrain (eds.), *Access Denied: The practice and policy of global Internet filtering*. MIT Press, Cambridge, MA, 2008.

van Dijck, J., 'Users Like You: Theorizing Agency in User-Generated Content', in: *Media, Culture and Society*, 31(1), 2009, pp. 41-58.

Dunne, A., *Hertzian Tales: Electronic Products, Aesthetic Experience, and Critical Design*. MIT Press, Cambridge, MA, 2005.

Elmer, G., 'Hypertext on the Web: The Beginnings and Ends of Web Path-ology', in: *Space and Culture*, 10, 2001, pp. 1-14.

Elmer, G., *Profiling Machines*. MIT Press, Cambridge, MA, 2004.

Foot, K. and S. Schneider, 'Online Action in Campaign 2000: An Exploratory Analysis of the U.S. Political Web Sphere', in: *Journal of Broadcast and Electronic Media*, 46(2), 2002, pp. 222-244.

Foot, K. and S. Schneider, *Web Campaigning*. Cambridge, MA: MIT Press.

Foucault, M., *The History of Sexuality Vol.1: The Will to Knowledge*. Penguin, London, 1998.

Fuller, M., *Behind the Blip: Essays on the Culture of Software*. Autonomedia, Brooklyn, 2003.

Galloway, A., *Protocol: How Control Exists After Decentralization*. MIT Press, Cambridge, MA, 2004.

Giles, J., 'Internet encyclopedias go head to head', in: *Nature*, 438, 2005, pp. 900-901.

Goldsmith, J. and T. Wu, *Who Controls the Internet? Illusions of a Borderless World*. Oxford, New York, 2006.

Hayles, K., 'Print Is Flat, Code Is Deep: The Importance of Media-Specific Analysis', *Poetics Today*, 25(1), 2004, pp. 67-90.

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

Hine, C. (ed.), *Virtual Methods: Issues in Social Research on the Internet*. Berg, Oxford, 2005.

Introna, L. and H. Nissenbaum, 'Shaping the Web: Why the Politics of Search Engines Matters', *The Information Society*, 16(3), 2000, pp. 1-17.

Jeanneney, J.-N., *Google and the Myth of Universal Knowledge*. University of Chicago Press, Chicago, 2007.

Jenkins, H., *Convergence Culture: Where Old and New Media Collide*. NYU Press, New York, 2006.

Jensen, J., 'Interactivity: Tracking a New Concept in Media and Communication Studies', in: P. Mayer (ed.), *Computer Media and Communication*. Oxford University Press, Oxford, 1999, pp. 160-188.

Jones, S., 'Studying the Net: Intricacies and Issues.' in: S. Jones (ed.), *Doing Internet Research: Critical Issues and Methods for Examining the Net*. Sage, London, 1999, pp. 1-28.

Keen, A., *The Cult of the Amateur: How Today's Internet is Killing Our Culture*. Nicholas Brealey, London, 2007.

Knorr-Cetina, K., *Epistemic Cultures*. Harvard University Press, Cambridge, MA, 1999.

Krebs, V., 'Mapping Networks of Terrorist Cells', in: *Connections*, 24(3), 2002, 43-52.

Krug, S., *Don't Make Me Think! A Common Sense Approach to Web Usability*. New Riders, Indianapolis, IN, 2000.

Landow, G., *Hyper/Text/Theory*. Johns Hopkins University Press, Baltimore, MD, 1994.

Latour, B. and S. Woolgar, *Laboratory Life*. Princeton University Press, Princeton, NJ, 1986.

Lazer, D. et al., 'Computational Social Science', in: *Science*, 323, 2009, pp. 721-723.

Lenhart, A. and M. Madden, 'Social Networking Websites and Teens', Pew Internet Project Data Memo, Pew Internet & American Life Project, Washington, DC, 2007.

Lynch, M., 'A sociology of knowledge machine'. in: *Ethnographic Studies*, 2, 1997, pp. 16-38.

Magnus, P.D., 'Early response to false claims in Wikipedia', in: *First Monday*, 13(9), 2008.

Manovich, L., 'Cultural Analytics.' unpublished ms., 2007, http://www.manovich.net/cultural_analytics.pdf (accessed 28 January 2009).

Manovich, L., *Software Takes Command*. unpublished ms., 2008, http://www.manovich.net/ (accessed 10 April 2009).

Marres, N. and R. Rogers, 'Depluralising the Web, Repluralising Public Debate. The GM Food Debate on the Web,' in: R. Rogers (ed.), *Preferred Placement*. Jan van Eyck Editions, Maastricht, 2000, pp. 113-135.

Marres, N. and R. Rogers, 'Subsuming the Ground: How Local Realities of the Ferghana Valley, Narmada Dams and BTC Pipeline are put to use on the Web', *Economy & Society*, 37(2), 2008, pp. 251-281.

McLuhan, M., *Understanding Media: The Extensions of Man*. McGraw Hill, New York, 1964.

Miller, D. and D. Slater, *The Internet: An Ethnographic Approach*. Berg, Oxford, 2000.

Mills, C. Wright, *The Sociological Imagination*. Penguin, Harmondsworth, 1971.

Niederer, S., 'Wikipedia and the Composition of the Crowd,' unpublished ms., 2009.

NRC Handelsblad. 28 August 2007.

Park, H. and M. Thewall, 'Hyperlink Analyses of the World Wide Web: A Review', in: *Journal of Computer-Mediated Communication*, 8(4), 2003.

Prensky, M., 'Digital Natives, Digital Immigrants', *On the Horizon*. 9(5), 2001.

Read, B., 'Can Wikipedia Ever Make the Grade?' *Chronicle of Higher Education*, 53(10), 2006, p. A31.

Rheingold, H., *Virtual Reality: Exploring the Brave New Technologies*. Summit, New York, 1991.

Rheingold, H., *The Virtual Community: Homesteading on the Electronic Frontier*. Addison-Wesley, Reading, MA, 1993.

Rogers, R., 'Operating Issue Networks on the Web,' in: *Science as Culture*, 11(2), 2002, pp. 191-214.

Rogers and Marres, N., 'French scandals on the Web, and on the streets: A small experiment in stretching the limits of reported reality', in: *Asian Journal of Social Science*, 30(2), 2002, pp. 339-353.

Rogers, R., 'The Viagra Files: The Web as Anticipatory Medium', in: *Prometheus*, 21(2), 2003, pp. 195-212.

Rogers, R., *Information Politics on the Web*. MIT Press, Cambridge, MA, 2004.

Rogers, R., 'The Politics of Web Space,' unpublished ms., 2008.

Rogers, R., 'The Googlization Question, and the Inculpable Engine', in: Stalder, F. and K. Becker (eds.), *Deep Search: The Politics of Search Engines*. Edison, NJ: Transaction Publishers, 2009.

Schneider, S. and K. Foot, K., 'Online structure for political action: Exploring presidential Web sites from the 2000 American election', *Javnost*, 9(2), 2002, pp. 43-60.

Shaviro, S., 'Money for Nothing: Virtual Worlds and Virtual Economies', in: M. Ipe (ed.), *Virtual Worlds*. The Icfai University Press, Hyderabad, 2008, pp. 53-67.

Shirky, C., 'Ontology is Overrated: Categories, Links, and Tags', The Writings of Clay Shirky, 2005, http://www.shirky.com/writings/ontology_overrated.html (accessed 28 January 2009).

Spink, A. and B.J. Jansen, *Web Search: Public Searching on the Web*. Kluwer, Dordrecht, 2004.

Stone, A.R., *The War of Desire and Technology at the Close of the Mechanical Age*. MIT Press, Cambridge, MA, 1995.

Sunstein, C., *Infotopia: How Many Minds Produce Knowledge*. Oxford University Press, New York, 2006.

Swartz, A., 'Who writes Wikipedia?' Raw Thoughts blog entry, 4 September 2006, http://www.aaronsw.com/weblog/whowriteswikipedia/ (accessed 22 August 2008).

Turkle, S., *Life on the Screen: Identity in the Age of the Internet*. Simon & Schuster, New York, 1995.

Vaidhyanathan, S., 'Where is this book going?' The Googlization of Everything Blog, 25 September 2007, http://www.googlizationofeverything.com/2007/09/where_is_this_book_going.php (accessed 22 December 2008).

Walker, J., 'Feral Hypertext: When Hypertext Literature Escapes Control', *Proceedings of the Sixteenth ACM conference on Hypertext and Hypermedia*, 6-9 September 2005, Salzburg, Austria, pp. 46-53.

Watts, D., *Small Worlds*. Princeton University Press, Princeton, 1999.

Weltevrede, E., *Thinking Nationally with the Web: A Medium-Specific Approach to the National Turn in Web Archiving*. M.A thesis, University of Amsterdam, 2009.

Williams, R., *Television: Technology and Cultural Form*. Fontana, London, 1974.

Woolgar, S., 'Five Rules of Virtuality', in: S. Woolgar (ed.), *Virtual Society? Technology, Cyberbole, Reality*. Oxford University Press, Oxford, 2002, pp. 1-22.