

After cyberspace: Data-rich media online

Richard Rogers

Abstract

The article takes up the question of the distinctiveness of the Web as site of social and cultural research. First, it seeks to situate analytical associations between the Internet and ideas of cyberspace and the virtual. It seeks to demonstrate the current conceptual opportunities available for cyberspace in security studies and the virtual in game studies. It subsequently makes a plea for a shift in focus for research away from the Internet as bracketed realm. How to employ the Internet for research into more than online culture only? Subsequently, it asks, what opportunities are available for research that takes up the Web as source? In the event there are currently competing programs that seek to introduce the Web as well as other digital media as data sets to be studied for purposes unrelated to cyberculture or similar. After a brief synopsis of the debate surrounding the Web as data set, the contribution made here is an underlying media theory that seeks to treat the Internet as a specific medium in the sense of the methods it offers. Thus instead of digitizing and bringing online existing method from the humanities and social sciences, the proposal is to follow the methods in the medium, and repurpose them for rather traditional social and cultural research purposes.

Key words: Cyberspace, virtual reality, media theory, internet, methodology

Internet-related research: Beyond the study of online culture only

The argument made here is an overall case for taking the Internet far more seriously than we have in the past, specifically in terms of what it has to offer for social and cultural research. The first step is to dispense with the ideas of cyberspace and the virtual as primary points of departure for Internet-related research, or in fact reposition those terms to reflect the conceptual opportunities they currently offer. Cyberspace, with its origins in science fiction literature and its legacy in cybercultural studies, most recently has become a specific realm of inquiry in Internet security studies, with the U.S. military, for example, recently creating a 'Cyber Command', and the U.S. Air Force rephrasing its mission as 'fly, fight and win in air, space and cyberspace' (U.S. Air Force, 2009). Similarly, the virtual, a term with a rich theoretical history, refers less to the Internet generally than it does most poignantly to virtual worlds such as Second Life and game environments such as World of Warcraft (Shields, 2003; van Doorn, 2009). Studies of the virtual, as in those of specific types of online

worlds and environments, would thus become a subset of Internet-related research just as cyberspace studies refer to the niche areas of cyber-war together with cyber-espionage and cyber-crime (Information Warfare Monitor, 2009).

Second, we may wish to reconsider the primacy of treating the Web as a site for the study of amateur production practices and user-generated content, for we are re-running the 'online quality' debates. Arguably the Web has seen recurrences of such debates, the first in the 1990s on the value of information online, where the Web was treated as a rumor mill and as breeding ground for conspiracy theory (Marres and Rogers, 2000). In the mid 2000s, the second such debate referred to the quality of content, where the Web became a free amateur content space threatening the professional (Keen, 2007; Thelwall and Hasler, 2007). I would like to argue that the Web, which in a sense arrived on the scene as 'empty' infrastructure, continues to pose problems for content analysis. It disappoints those in search of traditional markers of quality generally, or an underlying interpretive apparatus more specifically (Galloway, 2004; Hayles, 2004). Especially with the decline of surfing and with it hypertext as literary theory underpinning it, the Web has lost some of its hermeneutic productivity (Elmer, 2001). I would like to put forward that the Web nowadays invites more of the stance popular culture and television researchers radically put forward decades ago regarding their relatively new object of study – that one can read and diagnose cultural concerns from the medium. The question, however, becomes the means and techniques by which to do so. As I come to, instead of content analysis, here I will put forward 'digital methods' as such means.¹

Third, the argument recognizes that the Internet has reputational issues for researchers accustomed to thinking of it as cyberspace and virtual realm, as domain of rumors and self-publication, as well as, most recently, a site of messy data. The quality of information debate that was followed by the quality of (amateur) content debate has become the quality of data debate. Initial concerns had to do with incomplete Web archives. Subsequently, multiple dates on Webpages and search engine indexing practices were unable to provide accurate results for date range queries; longitudinal analysis, a marker of quality research, was doomed (Hellsten et al., 2006). Questions now arise about the robustness of so-called user-generated data such as social bookmarks, tags, comments, likes and shares (Thelwall et al., 2005). There have been specific reactions to this mess online. One is the perceived need for large, exhaustive data sets for the e-social sciences and the e-humanities (Borgman, 2009). In this account of the work to be done, it becomes desirable for social researchers, in a

¹ Apart from content analysis, another candidate approach would be user studies.

computational turn for the field, to collect and maintain large-scale data sets (Lazer et al., 2008). Another reaction to the online jumble could be called Google envy. Here the task would be to approximate the apparatus and the data collected and analyzed by Google, Facebook, etc. This particular way forward has a critical rejoinder. Are the only viable online data sets those collected and maintained by the large corporations? A third approach worth mentioning eschews the Web entirely as data collection site, and prefers digitized data of print and other media born offline, which are long-range and complete (Manovich, 2008).

The question of alternatives arises. What would be the 'little science' approach to Web data analysis? The notion of digital methods has been developed so as to offer one way forward (Rogers, 2009). Indeed, digital methods put forward a series of techniques that use data online to perform a kind of diagnostics. In the following, I would like to open the discussion of what is different about Internet and specifically Web thought these days that would allow research to move beyond associations with the cyber, the virtual, the rumor mill and the free amateur content space. In the account of the Web research that follows, it is first recognized that the Web has 'natively digital objects' (the hyperlink, the archived webpage, the search engine query, the search engine result, etc.) that are routinely employed in algorithms to recommend information, books, tunes, friends, etc. online. Digital methods are techniques that learn from how dominant Web devices treat this particular type of data, the digital objects. Yet instead of serving recommendations (ranked source lists and potential friends 'to friend', as online applications do), the digital methods, for example, trace association (through hyperlink mapping), show changing issue commitments (through time-lapsed photography of an archived website), and display national cultural preference (through special query design and reinterpretation of search engine results from local domain Googles). Digital methods reconsider the objects of study online, paying less attention to stand-ins such as the avatar than to the data in a user's profiles on a social networking site. In doing so, digital methods put forward techniques that strive to enable social and cultural research – the study of association, commitment, concern, resonance, preference and so forth – through online data analysis. The larger question is whether the findings made with online data subsequently require grounding with more traditional methods (and offline data). However, once we begin to ask those sorts of questions, the idea of studying the Internet as bracketed realm begins to recede from the picture. In all the undertaking is to introduce a research practice that moves Internet-related research beyond the study of online culture only.

Cyberspace as conflation of the Internet and virtual reality

The beginnings of Internet-related research that refers to the terms, virtual and cyberspace, may be located in the brief timeframe between the mid-1980s and the early 1990s, with its zenith circa 1991-1992. Indeed, prior to Web browsers and the growth of the World Wide Web, virtual reality and the Internet were often treated together as one space, at least in the writings of such dominant figures of the time as John Perry Barlow, Howard Rheingold, John Walker, Jaron Lanier and (in the U.K.) Roger Stone. In 1991 John Perry Barlow, the co-founder of the Electronic Frontier Foundation, the digital rights advocacy organization, described cyberspace in the *Communications of the ACM* as such: 'While it is familiar to most people as the location of a long-distance telephone conversation, it is also the repository for all digital or electronically transferred information.... [In there] 'everyone is as virtual as the shadows in Plato's cave' (Barlow, 1991: 19). Barlow unified virtual reality (VR) and the Internet into the single notion of cyberspace, saying he preferred the science fiction term from a celebrated cyberpunk novel over his own, the datasphere. Indeed, to VR developers Jaron Lanier and John Walker (and less so Scott Fisher), cyberspace was the preferred term for the place or space that came into being through the hooking up of the head-mounted display, the data glove and the rendering engine (Walker, 1990). Howard Rheingold, whose *Virtual Reality* book also was published in 1991, writes of cyberspace when discussing Fisher's NASA Ames VR set-up, Lanier's VPL gear as well as Walker and Autodesk's briefcase version (1991). Michael Benedikt's celebrated volume, *Cyberspace: First Steps* (1991), is similarly attuned to the association between virtual reality and the Internet – a connection which perhaps ends with John Perry Barlow's 1996 'Declaration of Independence from Cyberspace,' a seminal document in itself that refers to an Internet threatened by state regulation and thus no longer a virtual reality. The cybercultural scholar Steve Bukatman has recounted the history of cyberspace as the conflation of virtual reality and the Internet, albeit for different purposes (2000). By dating the association, the point that I would like to make concerns the significance of the work performed to undo it and the legacy it has left behind – especially those of the U.K. Virtual Society? Programme (1997-2001). The Virtual Society? programme, through a series of empirical studies, debunked many of the popular claims made of cyberspace and the virtual, and in a sense also grounded the virtual by establishing its complementary as opposed to substitutive relationship with the real (Woolgar, 2002). Virtual community, perhaps the most well-known use of the term apart from virtual reality, was shown to be reliant on knowing and interacting with people in one's town and city. One's identity was not so much the matter of play and experimentation, but rather was hard won, both online and offline. Whilst the research provided correctives to the larger claims about the virtual, it paradoxically offered the field of Internet research a term

that has stuck. 'Virtual' became an adjective for types of Internet-related study, including virtual ethnography and virtual methods (Hine, 2000; Hine, 2005).

I would like to argue that these days the objects in 'virtual studies' in Internet-related research are online instantiations, or disembodied counterparts, be they identifies, bodies or practices. They may be summarized through a glance at the results of a Google Images query for 'virtual,' where there are largely two sets of images, one concerning virtual reality and the other virtual worlds technology and gaming, including the avatars themselves, and the skills and insights cultivated to operate cleverly in these spaces (practices).² (See Figure 1.) Thus one notes the NASA Ames image of virtual reality from the late 1980s, as well as two other images of head-mounted displays. Also displayed are online individuals – virtual women – as well as settings from TV news studios to a meeting place in Second Life. Certain sources make light of the connection between virtual and reality – also specifically separating the virtual worlds, games and TV from any association with the Internet, in a move that I would argue is typical for the field of Internet-related research. For example, the humor blog, 2030, writing about the future in what it calls 'the really early edition,' reports: 'March 31, 2029 – American adults spend an average of 12 hours a day divorced from reality, immersing themselves in 3D television, virtual environments and holographic imagery, according to a new study' (2009). The virtual becomes a realm for visual cultural studies as opposed to Internet-related research.

[Insert Figure 1: Virtual as returned by Google Images, 27 June 2010.
Source: images.google.com.]

[Insert Figure 2: Cyberspace as returned by Google Images, 27 June 2010.
Source: images.google.com.]

Cyberspace may be similarly drawn into focus as no longer a catch-all term for online space but rather as a particular idea of it. The leading images of cyberspace returned by Google Images depict a rather specific subset of associations, which I would like to call 'network spatialities' – all of which mystify rather than clarify an Internet ready for social and cultural

² The use of Google Images as proxy for current, dominant representations for a subject matter relies on techniques, developed across science and industry, for identifying and ranking the subject matter of images. Whilst the literature discusses the limitations of current technology in techniques of disambiguation, it is precisely the technology's capacity to catch many images of cyberspace (owing to textual reference) and rank them (owing to frequency) that is of interest here. Thus the Image search preserves the potential open-endedness of cyberspace representation, and also shows which ones enjoy favor (Jing et al., 2007; Jing and Baluja, 2008).

research, at least in the manner sketched above (see Figure 2). Cyberspace imagery in network spatial terms has three separate compositions. One is the layered space which connotes dimensions; that is, cyberspace adds a dimension to reality. Another network space image refers to the corridor, which conjures a mind space. (Other images of an outline of the head or mind have a similar point of reference.) A third relies on imagery of a vortex, or a virtual funnel, perhaps combined with fibre optics cables. Here the reference is to communication and information exchange, and perhaps the principles of Internet infrastructure (packet-switching and end-to-end principle) that prompted a series of early hyperbolic pronouncements; the Internet treats censorship as a malfunction and routes around it, if one thinks with the packets following different routes and reconvening at the destination (Boyle, 1997). In all the three network spatialities organize cyberspace (powerfully) as a substance, culture and carrier, in classic medium terms, or indeed as realm, thought construct and special infrastructure, with particular transformative qualities. As such cyberspace comes without so many of the elements it subsequently has developed – interfaces, political economies, advertisements, and everyday patterns of usages (navigating from Google to Facebook and back), to name a few. However, there is one image of a toy soldier between the keys on a computer keyboard, which is a reference to cyberwarfare, as mentioned in the opening. The arcane practices of cyberwar seem to be one of the few associations cyberspace can currently maintain. Indeed leading work into cyberwar and cyberespionage plays upon the mysteries of cyberspace. The malware that infected the computers in the offices of the Dalai Lama and elsewhere were run, as the researchers called it, through a ‘ghostnet’ (Information War Monitor, 2009).

The Internet (or Web) as problem for content analysis

McLuhan is sometimes recalled as a seminal media theorist for his insistence on studying the medium instead of its content. In setting up his position he employed such terms (in reference to mindsets of analysts) as ‘typographic cultural bias,’ ‘typographic trance of the West,’ and ‘typographic spell’ (McLuhan, 1964). The content distracts the analyst in media studies for its habitual dominance over other potentially more fruitful research practices. From the substance offered in forums and blogs, to YouTube videos and tweets in Twitter, the value of the content has remained elusive for many researchers. Value in each of those software platforms has been framed for the potentials each has for the underdog in media settings crucial for the health of a democratic society: forums for average citizens to take part in public debate, blogs for average writers to contribute to discourse in a new public sphere, YouTube as video publishing platform for average videomakers diversifying content and subject matter in heretofore mass media, and Twitter for everyone’s status updates (as opposed to only those usually found on radio scanners,

such as emergency vehicles, taxicabs and similar). All give voice, dislodge gatekeepers, and reorganize audience and attention so as to create reputation, awareness as well as celebrity, until the content analysts arrive, and undertake studies similar in spirit to those which debunked the claims made about the virtual and cyberspace. Concluding after an in-depth study of the forum discussions on the issue of immigrants in the Netherlands, social researcher Tamara Witschge writes: 'It becomes clear that the quality of the [forum] discussion leaves a lot to be desired' (Witschge, 2007: 242) Writing about the contribution made by the Web to the debate surrounding the murder of the Dutch critic, Theo van Gogh, media researcher Thomas Poell writes: '[T]he majority of the forum and blog discussions did not in any way match the normative criteria of the Habermasian public sphere' (Poell, 2008: 220-221). These are two isolated Dutch studies that find resonance with others regarding the qualities of the forums and the blogosphere, but the point I would like to make concerns the weight the Web continually must bear for its contents, and analysts' finding that it is unable to do so (Dean, 2002). Uploaded home movies such as 'Charlie bit my finger – again!', with its 270 million views on YouTube (through mid January 2011), is among the top ten YouTube videos of all time. The 'Annoying Orange' is another type of YouTube success story, as it eventually serialized; there is now a channel of videos of the talking mouth superimposed on an orange as well as other vegetables, with scores of episodes, and a total number of views nearing 500 million. One can subscribe and eventually contribute additional statistics to the videos, and also do business with the maker. Here we have a story of the gradual professionalization and produsage of the new medium (Bruns, 2008). One could continue with the study of the effects of online amateur content not only on the industry but on the amateurs themselves, where one of the more dominant lines of inquiry concerns how continued publishing in the medium gradually professionalizes the amateur (Burgess and Green, 2009). However, as amateur video gains audience, their makers upload more, prompting a variety of debates concerning the effects of micro-celebrity on content-making, including the use of children in video. 'Is it wrong to take your kids viral?' is the title of a magazine article referring to the efforts of a graduate student and single dad, who, after the success of the video of a duet with his young daughter, made more video and created a Facebook fan page, with a link to 'One day I'm Gonna Whistle' t-shirts with his daughter's scribbles on them, offered for purchase so the father can pay off his student loans (Williams, 2011). Critique of web content commodifying kids followed the more typical discussions of corrupting influences of (media) attention, or the need for far more widespread media training. Thus discussion of amateur content moves beyond supposed low quality itself to the effects of open access by amateurs to audience. It is also the vast scale of amateur content production – with audience effects – that is underappreciated.

The Web as data set

Above I discussed how the Web initially arrived on the scene 'empty,' as infrastructure awaiting content. 'Under construction' sites or pages may be regarded as sources of nostalgia these days, like other aesthetics of the 1990s, such as starry blue nights as website backdrops, or 'random site' links which invite surfers to surf on to territories to be discovered, and to jumpcut to another hyperspace (Espenschied and Lialina, 2009). This Web 0.1 is our cyberspace, the precursor to what is now becoming historicized as Web 1.0 (info-Web) and Web 2.0 (social Web). Above I also mentioned how each of these Webs had particular quality debates associated with them. Whether it is associated with fandom, porn and aliens, with imposters, conspiracy and self-publishers, or with amateur production practices that may professionalize or potentially corrupt its makers – the Internet has had serious reputational issues. Thus it is likely, as stated above, that any introduction of a new object of online study would be met with similar scepticism. Nowadays to view the Web as data sets for social and cultural research is to be confronted with a variety of issues about messy data. The webometrician Mike Thelwall summarized the challenges of employing the Web as data set as follows:

One [issue] is the messiness of Web data and the need for data cleansing heuristics. The uncontrolled Web creates numerous problems in the interpretation of results (...). Indeed, a skeptical researcher could claim the obstacles (...) are so great that all Web analyses lack value. [O]ne response to this (...) is to demonstrate that Web data correlate significantly with some non-Web data in order to prove that the Web data are not wholly random. (Thelwall et al., 2005: 81)

Here the general reputation problem about quality online is transformed, initially, into the question of how to clean up the data, since there is a lack of uniformity in how users fill in forms, fields, boxes, bars and other text entry spaces. In a sense the (unedited) Web is viewed as one large 'free text' space. There are misspellings. There are too few conventions. For example, different tags are used for the same content. To Thelwall, this state of affairs makes many researchers simply renounce the Web as source, unless data sets come whole (all transactions in Second Life) and one studies online culture only, or unless there is the introduction of offline data for comparative purposes. The offline would serve as reassuring calibration or as a baseline. Indeed, perhaps the most well-known use of Web data for social research is Google Flu Trends. It is a project by the non-profit arm of Google, the Google Foundation or Google.org. Search query logs are analyzed, whereby a keyword list (or dictionary) is employed to find flu and flu-related symptoms, as well as

where people are searching for them. The locations of the users querying these flu keywords provide indications of the current locations of flu (Ginsberg et al., 2008). When Google.org first released Flu Trends the project appeared to base its findings on the prevalence of flu on the search engine queries only. However, the question of the baseline was answered with the comparison of Google's findings from the Web with those of the U.S. Center for Disease Control, whose data are based on hospital and other traditional medical reporting practices. It was found that Google was charting instances of flu some 7 days prior to the official records, thus in a sense predicting the occurrence on the ground on the basis of search online. The Web here became an anticipatory medium, a medium of trend indications, not only as the name of this Google project suggests, but also in keeping with how search engine data have been packaged and presented in previous projects by other companies, including Yahoo! Buzz and Alexa Movers and Shakers.

Can the Web provide more than mere indications, later to be grounded by traditional data (and data collection)? Of interest in this regard is the set of graphics published in the *New York Times* the day prior to the 2009 Thanksgiving holiday, the national feast in the U.S. (Severson, 2009). The states on the U.S. map were differently shaded depending on the number of queries from each state made for particular recipes at allrecipes.com, the country's most popular site of its kind. Thus certain states had users of the recipe website query macaroni and cheese far more than other states. The same held for a long list of foods, seemingly queried the day before Thanksgiving so as to be prepared on the holiday itself: sweet potato pie, corn casserole, green beans, turkey brine, yams, etc. The Web data produced a geography of taste, leading to the usual questions of the quality of the Web data, and additionally to questions of the baseline. To compare the findings made with the online data to some baseline, does one conduct a telephone survey, or obtain nationwide supermarket data? Given deeper insight in the demographics and statistics available from allrecipes.com, one would be in the position to weigh the desirability and likelihood of performing this research by other means, lest it become highly costly or unmanageable (Weber and Castillo, 2010). In other words, when does the online provide the more salutary means for conducting social and cultural research? Under which conditions are findings grounded only online?

Online groundedness and digital methods

The term online groundedness, coined initially as digital groundedness, was developed in reaction to reporting by investigative journalists in the Netherlands. As mentioned above, the question of societal reactions to the murder of social commentator and critic, Theo van Gogh, has been framed in terms of the larger immigration debate, and in terms of the muslimization of society. Right-wing populist politicians as Pim Fortuyn

and later Geert Wilders have represented a strain of politics that raise questions about the 'hardening' of Dutch culture more generally. Is it becoming more extremist? The investigative reporters took up the question, and undertook in-depth research on right-wing and extremist right-wing groups. The investigation did not assume the form of 'going native' or embedding themselves in particular groups (Buford, 1991). They also did not go to the library or the social history institute, perusing ephemera or other materials penned by the groups. Rather, the journalists used Web data. They studied the evolution of the tone of language on right-wing Websites over time, using the Wayback Machine of the Internet archive (Dohmen, 2007). They found that over time the language on the right-wing sites gradually began to approximate that on the right-wing extremist sites, thereby prompting the cautious conclusion that Dutch culture is 'hardening'. For those accustomed to thinking of the Internet as cyberspace, virtual realm, rumor mill or free amateur content space, employing its data for social research would seem improbable. It also could lead to skepticism about the results generally. Contrariwise, it also could be considered an invitation to think through the use of archived websites as object of study, as others also have considered recently (Brügger, 2009; Dougherty et al. 2010). Thought of relation to research project on extremism above, archived websites became source material for more than the history of the Web or for evidentiary purposes in lawsuits (Brügger, 2010; Howell, 2006; Grimmelman, 2008/2009). Indeed they transform the Web into primary source material in order to research societal concern.

In the remainder, I would like to treat certain of the other questions of social and cultural research that may be asked (more experimentally) with Web data. How to make use of natively digital objects, or the Web data regularly employed by dominant devices online? I would like to begin with what could be considered a fundamental unit of the Web (for research purposes), the hyperlink (Barabasi, 2003; Huberman, 2003; Watts, 2003). The hyperlink has been crucial to early Web theory, inherited from literary theory, whereby surfers through choosing their path through linked documents online would author a trail, or even a story (Landow, 1992; Elmer, 2001). One of the other lineages of study concerns the hyperlink, together with the click or hit, as the core of the development of search engine algorithms that rank sources, and thereby author not paths or trails as much overall source authority online (Brin and Page, 1998; Hindman, 2009). Indeed, search engines arguably have taken over from other social markers (such as name recognition and graphical design style) as reputation authors, or at least reliability makers, matching user expectations with desired results, as I will return to (Van Couvering, 2007).

How else to use links, still following the lead of search engines? One may trace associations between websites (link mapping), and thereby show a politics of association – how reputation is the source of link-making, but also the reason why links are not reciprocated. One case in point is a mapping of the Armenian non-governmental space online, whereby the Armenian organizations reach out through hyperlinks to the inter-governmental and international organizations (chiefly, the United Nations), and no U.N. bodies link back (see Figure 1). The clustering that one often sees on Web maps, generally, is to be viewed, at least in this specific example, as classic politics of association, with the one set of actors networking and aspiring for association, and the other set of actors concerned that a link would show endorsement, so they do not reciprocate. Of course tracing association – a foundation of sociology – is not normally done through links online (Latour, 2005). Here instead of asking how one would harden these findings through introducing the baseline of interviews or social network analysis style surveys, it is of interest to note how such maps may be used by the actors themselves in presentations at pre-summit meetings and similar settings where NGOs and intergovernmental organizations meet. The map serves as proof or an indication of the need to network more strategically, and to make new ties (Elmer, 2006; Waddell, 2011).

[Insert Figure 3: Armenian E-Government Issue Network Map showing two distinctive clusters of interlinking between site types, Armenian and international organizations, with Armenian sites linking to the international organizations and the latter not reciprocating. Map from Issuecrawler.net, Govcom.org Foundation, July 2004. Crawl by Audrey Selian.]

From hyperlinks, I wish to move from the natively digital object gradually to larger units, discussing work on the Website and on the search engine. I have discussed how to repurpose the study of social networking platforms, the spheres (Websphere, blogosphere, newsphere), as well as national webs elsewhere, though I will touch on the study of national concerns through reinterpreting local domain Google results below (Rogers, 2009).

As mentioned above the Website (as archived object) is of interest for its capacity to retell its history when loaded and played in a kind of time-lapsed photography technique. Using the Wayback Machine of the Internet Archive, one captures and loads the archived pages of, for example, the homepage of the U.S. White House (whitehouse.gov) into a movie, and watches the history of the transition from the Clinton to the Bush administration, and finally to the Obama administration, including, in the far left column, the rise and fall of social issues or White House commitments (Man, 2010). To wit, by 10 January 2002, some 4 months

after the September 11th attacks, all 'issues' on the White House homepage contained the term 'security': 'National Security,' 'Homeland Security,' 'Economic Security' and 'More Issues' (Whitehouse.gov, 2002; Man, 2010).

The search engine, in this case Google, also may be the object of a single-site history, in the style of time-lapsed photography, where the homepage was studied for its subtle changes over time, as above (Govcom.org, 2008). Whilst to the naked eye Google appears to have changed only modestly over the past ten years, of note are the tabs above the search box, and especially which of the various search services, such as Images, Videos, Maps, News, etc., have achieved the front-page tab status, and which ones have been relegated to behind the more or even more buttons. The striking finding from the history of the Google homepage is the gradual demise of the Directory tab, and service, which appeared on the homepage in 2000, and maintained its place until March 2004, when it was dropped behind the more button, and in 2006, to even more. By the end of 2008, it was no longer accessible via the homepage; one needed to search Google in order to find the Google directory. The decline of the directory stands in for the overall fall of the human-edited Web, and the rise of the back-end and the algorithm as the primary determinant of the order of content.

Google and other search engines may be critiqued for their lack of commitment to upholding the human-edited Web over the algorithmically ranked Web only, and the critique may be made forcefully through a single-site history, in the rendering described above. Another critique concerns how engines systematically exclude sites. It is a thought carried over from the first systematic studies of search engine crawling and indexing in the late 1990s, where it was found that engines indexed a much smaller percentage of the known Web than imagined (Lawrence and Giles, 1998). If engines indexed only some 30% of the Web, the rest is 'dark'. This dark Web was thought of as an exclusionary one, and the language employed aestheticized it. Linkless and thus 'orphan' Websites were never found by engines, and could be considered buried (Introna and Nissenbaum, 2000). Another set of critiques concerns the power of the engine to reorganize significance, relevance and reliability, and thus the value of sources. Search engine users also contribute to the value of the real estate of the top results by looking at fewer and fewer results over the years (Jansen and Spink, 2006). There are other critiques concerning the growing behemoth and perhaps hegemon that is Google, summarized in the notion of Googlization (Vaidhyathan, 2011).

Societal search

Google, however, may be of use for other research purposes than critique. Google is of interest for social research purposes in that it is a data

collecting and epistemological machine with crawlers that fetch and index vast amounts of material and algorithms that rank on the basis of relevance. Relevance is a somewhat controversial term as it elides the difference between matching query with the quality source, with that of matching query with popular source (Van Couvering, 2007). Put differently, search engines are increasingly meeting the expectations of the users, and defining 'relevant' results accordingly. Search engine algorithms manage to give the user what he or she wants by relying not only links between Websites as indicators of significance. Rather, the results that are clicked by users are boosted in subsequent interactions with the engine. The question is, may one make use of the rankings of sources favored by users and consider results as showing a form of societal resonance? In this particular research practice, first, one moves from glancing at engine results, and clicking on one of them, to reading the returns. For a particular query, for example RFID (or radio frequency identification tags), what kinds of actors are in the top results? Does the actor composition indicate that a particular social issue is mature, polarized, or unoccupied by regulators (for example)? More to the point, may one interpret the rankings as showing a hierarchy of engagement, and by whom? Such a question assumes that when one reads Google results, one views primarily societal as opposed to engine dynamics.

I would like to put forward that this question is different from the idea, said to be propagated by search engine companies and industry trade press, that results are either democratic or organic (Hindman, 2008). Both descriptions of engine results are worthy of critique, for they point to questionable explanations for engine result sense-making. In the event, they are not democratic, it could be argued, for websites are not treated equally. Those with more inlinks are more influential in granting authority to other sites, and websites in so-called spammy neighborhoods are not allowed to vote with their links. The dominant engine algorithm results in what scholars call 'googlearchy' (Hindman et al., 2003). Naturalizing engine returns – calling them organic – equally would belittle the achievement of search technology, and the entire artifice that has been created to produce relevant results (Latour, 2005).

Instead, it is proposed to explore the extent to which search results may be called societal. In a sense it is a research project that is in keeping with the evocative idea of a search engine as 'database of intentions,' which the search researcher Alex Halavais, following John Battelle, calls a 'fine-grained map of our interests' (Battelle, 2005; Halavais, 2009: 30). One's intentions, however, normally would be explored through search query log analysis (as in Google Flu Trends above). One of the more telling examples of the use of engine query logs is the 2006 documentary film, *I Love Alaska* by Lernert & Sander, which in 15 episodes presents an intimate character

portrait of a woman on the basis of her daily engine queries. The object here is to move to a much larger scale, inquiring into whether engine results are reflective of national preference. In order to explore such a question, colleagues and I queried a series of local domain Googles (Google dot country code top level domain) for the term 'rights,' translated into the languages of the various countries. The query design was one that was devised so that a variety of rights types were returned, but that the order of the rights mentioned could be different per country and thus telling. In which order were which types of rights (human rights, children's rights, animal rights, air passengers' rights, cyclists' rights, etc.) mentioned per country in the results for the query? Whilst preliminary and experimental, we nevertheless were able to explore the larger question of whether Google could tell us which issues are held dear per country (see Figure 4). For example, via Google.fi, we learned that the 'right to roam' is a top right type for Finland.

Far more conventional has been the work on hierarchies of commitment and concern which may be detected by querying an organization's website for all the issues it claims to work on, and ranking the issues according to the number of mentions on the website (see Figure 5).³ Thus Greenpeace International's main issues were extracted (from their issue list on their homepage), and a google.com site query of greenpeace.org was made for each issue, resulting in one measure of a hierarchy of issue commitment. In the examples above, Google's epistemology of relevance as well as its depth of indexing of websites (respectively) are repurposed for societal and organizational issue hierarchy research.

[Insert Figure 4: Rights Types: Hierarchy of rights resonance per country. Results of queries of local domain Googles for 'rights' in the respective country languages, with the rights types ranked according to their appearance in the results.]

[Insert Figure 5: Greenpeace International's hierarchy of issue commitment, according to the number of mentions on its Website of each of its issues. Output from the Lippmannian Device by Govcom.org, October 2009.]

Conclusion

The digital methods project, whilst much larger and more varied than described above, may be summarized as in keeping with a turn away from the Web as virtual, in the terms described above, but also in terms of the Web methods, called 'virtual methods'. I have argued elsewhere that there

³ One page equals one mention.

is an ontological distinction between the natively digital and the digitized, that is between that of the medium, and that which has migrated to it (Rogers, 2009). Such a distinction also may be made for method. There are those of the medium, as folksonomy and the ever-evolving PageRank, and those that have been digitized, as online surveys. I would like to argue that virtual methods may be seen as digitized methods adapted to the medium, e.g., creating an online survey. There are medium effects that require detailed attention to existing method, and the term 'virtual' is employed to indicate a sensitivity to the difference the medium makes. Virtual methods are often used to study the difference the medium makes in other ways, too. For example, how does searching for health information alter the patient-doctor relationship (if at all)? One may ask that of a particular cohort of patients by sending a link of an online survey to a discussion or mailing list, and waiting for responses. Often the size of the list is unknown, and thus is the response rate. Indications may result, instead of grounded findings, thereby diluting the impact of the analysis. These particular questions, as I mentioned above, give pause to researchers skeptical of the online, and the virtual. Digital methods do not aim to provide redress or remedy to the issues facing the movement of the humanities and the social sciences online. That is better left to the e-humanities and the e-social sciences. Rather, with digital methods I would like to propose means to make use of the data routinely generated online.

References

Barabasi, A.-L. (2003) *Linked*. New York: Plume.

Battelle, J. (2005) *The Search*. New York: Portfolio.

Bukatman, S. (2000) 'Terminal Penetration', pp. 149-174 in Bell, D. and Kennedy, B.M. (eds.) *The Cybercultures Reader*. New York: Routledge.

Barlow, J.P. (1991) 'Coming into the Country', *Communications of the ACM* 34(3): 19-21.

Barlow, J.P. (1996) 'A Declaration of the Independence of Cyberspace.' URL (accessed December, 2010): http://w2.eff.org/Censorship/Internet_censorship_bills/barlow_0296.declaration

Benedikt, Michael (ed.) (1991) *Cyberspace: First Steps*. Cambridge, MA: MIT Press.

Borgman, C. (2009) 'The Digital Future is Now: A Call to Action for the Humanities', *Digital Humanities Quarterly* 3(4). URL (accessed December,

2010):

<http://www.digitalhumanities.org/dhq/vol/3/4/000077/000077.html>.

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

Brin, S. and Page, L. (1998) 'The Anatomy of a Large-Scale Hypertextual Web Search Engine', Seventh International World-Wide Web Conference (WWW 1998), April 14-18, Brisbane, Australia.

Brügger, N. (2009) 'Website history and the website as an object of study', *New Media & Society* 11(1-2): 115-132.

Brügger, N. (ed.) (2010) *Web History*. New York: Peter Lang.

Bruns, A. (2008) *Blogs, Wikipedia, Second Life, and Beyond: From Production to Prodsusage*. New York: Peter Lang.

Buford, B. (1991) *Among the Thugs: The Experience, and the Seduction, of Crowd Violence*. London: Secker and Warburg.

Burgess, J. and Green, J. (2009) *YouTube: Online video and participatory culture*. Cambridge: Polity Press.

Dean, J. (2002). *Publicity's Secret*. Ithaca, NY: Cornell University Press.

Dohmen, J. (2007) 'Opkomst en ondergang van extreemrechtse sites', *NRC Handelsblad*. URL (accessed January, 2010): http://www.nrc.nl/binnenland/article1831689.ece/Opkomst_en_ongang_van_extremrechtse_sites.

Dougherty, M., Meyer, E., Madsen, C., Van den Heuvel, C., Thomas, A. and Wyatt, S. (2010) 'Researcher Engagement with Web Archives: State of the Art', Joint Information Systems Committee Report. URL (accessed December 2010): <http://ssrn.com/abstract=1714997>.

Elmer, G. (2001) 'Hypertext on the Web: The Beginning and End of Web Path-ology', *Space and Culture* 10: 5-14.

Elmer, G. (2006) Re-tooling the Network: Parsing the Links and Codes of the Web World', *Convergence* 12(1): 9-19.

Espenschied, D. and Lialina, O. (2009) *Digital Folklore*. Stuttgart: Merz Academy.

- Future Update (2009) 'Americans Spend Most of Day Detached From Reality, Study Shows', blog, URL (consulted December 2010): <http://futureupdate.wordpress.com/2009/03/30/americans-spend-most-of-day-detached-from-reality-study-shows/>
- Galloway, A. (2004). *Protocol: How Control Exists after Decentralization*. Cambridge, MA: MIT Press.
- Ginsberg, J., Mohebbi, M. H., Patel, K. S., Brammer, L., Smolinski, M. S. and Brilliant, L. (2009) 'Detecting influenza epidemics using search engine query data', *Nature* 457(7232): 1012–1014.
- Govcom.org (2008). 'Google and the Politics of Tabs', Video. URL (accessed January, 2009): <http://movies.issuecrawler.net/google.html>.
- Grimmelman, J. (2008/2009) 'The Google Dilemma,' *New York Law School Law Review* 53: 939-950.
- Halavais, A. (2009) *Search Engine Society*. London: Polity.
- Hayles, K. (2004) 'Print Is Flat, Code Is Deep: The Importance of Media-Specific Analysis', *Poetics Today* 25(1): 67-90.
- Hellsten, I., Leydesdorff, L., & Wouters, P. (2006) 'Multiple presents: How search engines rewrite the past', *New Media & Society* 8(6): 901-924.
- Hindman, M., Tsioutsoulis, K. and Johnson, J. (2003) "Googlearchy': How a Few Heavily-Linked Sites Dominate Politics on the Web', Annual Meeting of the Midwest Political Science Association, Chicago, IL.
- M. Hindeman (2008). *The Myth of Digital Democracy*. Princeton: Princeton University Press.
- Howell, B.A. (2006). 'Proving web history: How to use the Internet archive.' *Journal of Internet Law* 9(8): 3-9.
- Huberman, B. (2003) *The Laws of the Web*. Cambridge, MA: MIT Press.
- Information Warfare Monitor (2009) 'Tracking Ghostnet: Investigating a cyberespionage network', Ottawa: SecDev Group.
- Introna, L. and Nissenbaum, H. (2000) 'Shaping the Web: Why the Politics of Search Engines Matters', *The Information Society* 16(3): 1-17.

Jansen, B. and Spink, A. (2006) 'How are we searching the World Wide Web? A comparison of nine search engine transaction logs', *Information Processing and Management* 42: 248–263.

Jing, Y. and Baluja, S. (2008) 'PageRank for Product Image Search' *WWW 2008: Proceedings of the 17th International World Wide Web Conference*. New York: ACM, 307–315.

Jing, Y., Baluja, S. and Rowley, H. (2007) 'Canonical Image Selection from the Web', *Proceedings of the 6th International Conference on Image and Video Retrieval*, New York: ACM, 280-287.

Keen, A. (2007) *The Cult of the Amateur*. New York: Doubleday-Currency.

Latour, B. (2005) *Reassembling the social: An introduction to actor-network-theory*. New York: Oxford University Press.

Lawrence, S., and Giles, C. L. (1998) 'Searching the World Wide Web', *Science* 280(5360): 98–100.

Lazer, D. et al. (2009) 'Computational Social Science', *Science* 323(5915): 721-723.

Man, P. (2010). 'Whitehouse.gov: Rhetorical use of spreading fear,' YouTube video. URL (accessed November 2010): <http://www.youtube.com/watch?v=5XSTPY4NuQc>.

Manovich, L. (2008) 'Cultural Analytics', White paper. University of California San Diego: Software Studies Initiative.

Marres, N. and Rogers, R. (2000) 'Depluralising the Web, repluralising public debate: The case of the GM food debate on the Web', pp. 113-136 in Rogers, R. (ed.) *Preferred placement: Knowledge politics on the Web*. Maastricht: Jan van Eyck Editions.

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

Poell, T. (2009) 'Conceptualizing Forums and Blogs as Public Sphere', pp. 239-251 in Lammes, S., Lehmann, A.-S., Raessens, J. and Schäfer, M.T. (eds.), *Digital Material - Tracing New Media in Everyday Life and Technology*. Amsterdam: Amsterdam University Press.

Rogers, R. (2009) *The End of the Virtual: Digital Methods*. Amsterdam: Amsterdam University Press.

Rogers, R., Jansen, F., Stevenson, M. and Weltevrede, E. (2009), 'Mapping Democracy', *Global Informaton Society Watch 2009*, Association for Progressive Communications and Hivos, 47-57.

Severson, K. (2009) 'Butterballs or Cheese Balls, an Online Barometer', *New York Times*. 25 November. URL (accessed November, 2009): <http://www.nytimes.com/2009/11/26/dining/26search.html>.

Shields, R. (2003) *The Virtual*. London: Routledge.

Thelwall, M. and Hasler, L. (2007) 'Blog search engines', *Online Information Review* 31(4): 467-479.

Thelwall, M., Vaughan L. and Björneborn, L. (2005) 'Webometrics', pp. 81-135 in Cronin, B. (ed.), *Annual Review of Information Science and Technology*, 39, Medford, NJ: Information Today.

U.S. Air Force (2009). 'Our Mission', U.S. Air Force Website. URL (accessed January 2011): <http://www.af.mil/main/welcome.asp>.

Vaidhyanathan, S. (2011) *The Googlization of Everything*. Berkeley: University of California Press.

Van Couvering, E. (2007) 'Is relevance relevant? Market, science, and war: Discourses of search engine quality', *Journal of Computer-Mediated Communication* 12(3), article 6. <http://jcmc.indiana.edu/vol12/issue3/vancouvering.html>

Van Doorn, N. (2009) *Digital Spaces, Material Traces*. PhD Disseration, University of Amsterdam.

Waddell, S. (2011) *Global Action Networks*. New York: Palgrave Macmillan.

Walker, J. (1990) 'Through the looking glass', pp. 439-448 in Laurel, B. (ed.) *The Art of Human-Computer Interface Design*. Reading, MA: Addison-Wesley.

Watts, D. (2003) *Six Degrees*. New York: W.W. Norton.

Weber, I. and Castillo, C. (2010) 'The Demographics of Web Search', SIGIR'10, New York: ACM, forthcoming.

Williams, M.E. (2011) 'Is it wrong to take your kids viral?' *Salon*. URL (consulted January 2011):

http://www.salon.com/life/going_viral/index.html?story=/mwt/feature/2011/01/12/jorge_navarez_family_viral

Witschge, T. (2007) *(In)difference online: The openness of public discussion on immigration*. PhD Dissertation, University of Amsterdam.

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

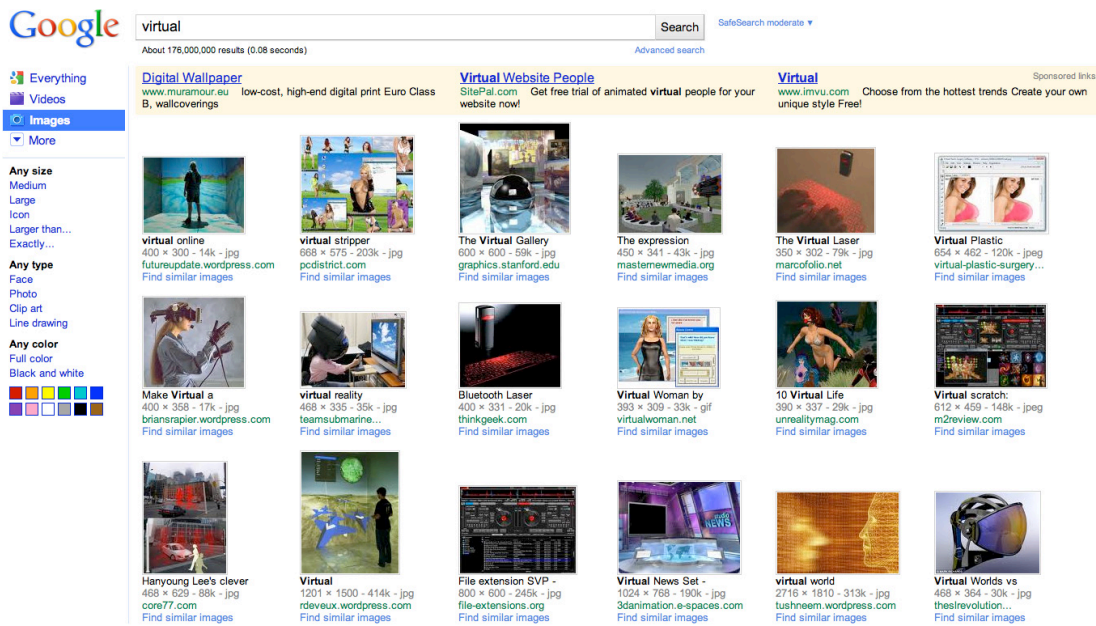
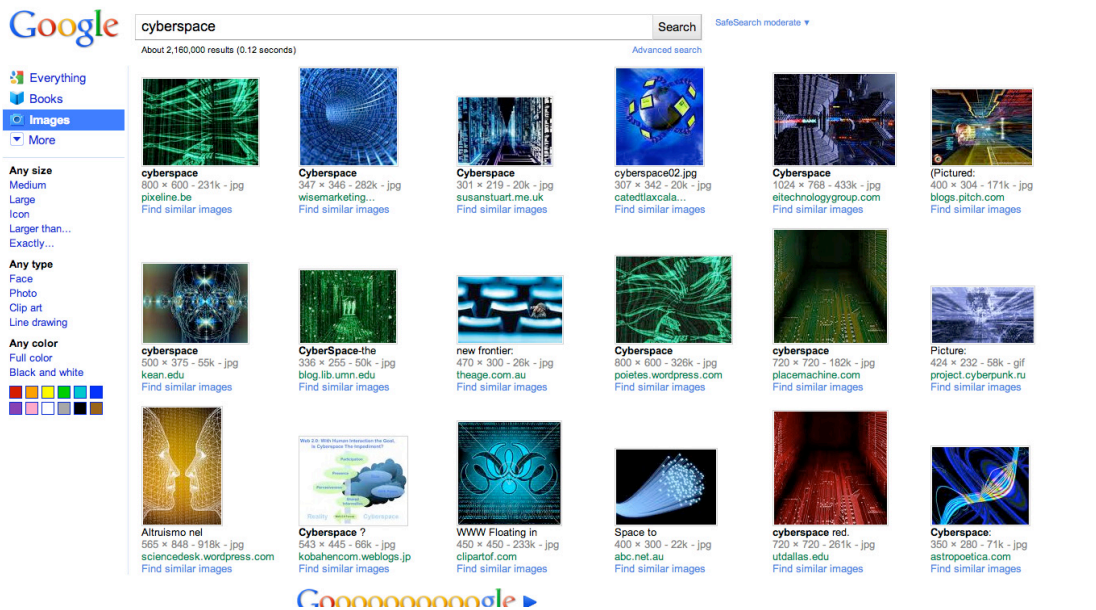


Figure 1: Virtual as returned by Google Images, 27 June 2010. Source: images.google.com.



Insert Figure 2: Cyberspace as returned by Google Images, 27 June 2010. Source: images.google.com.

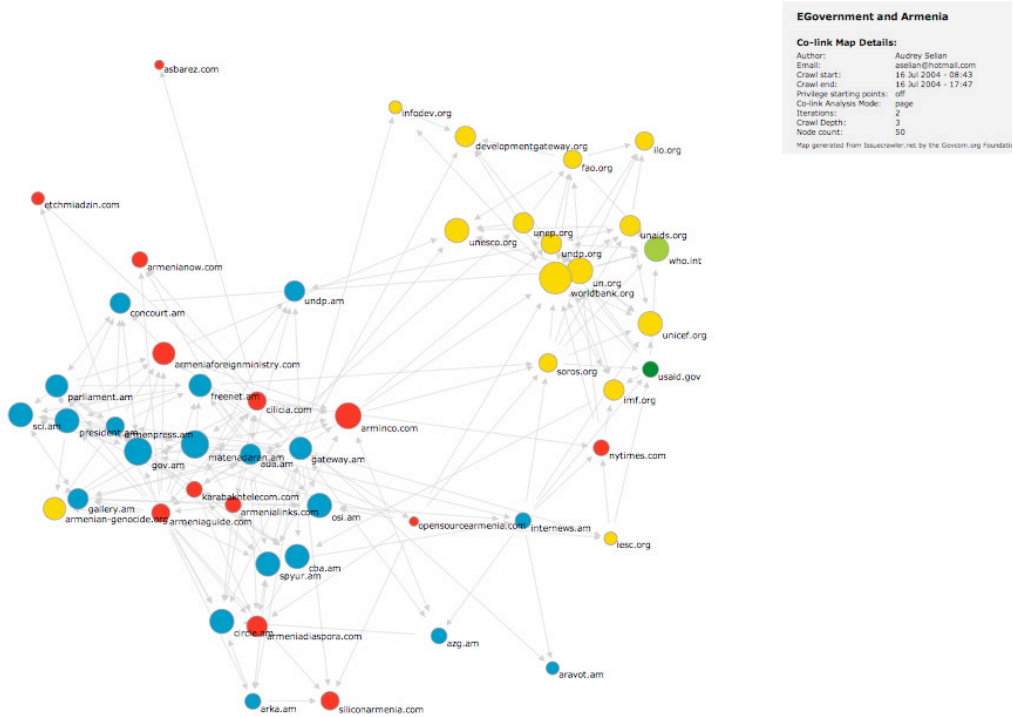


Figure 3: Armenian E-Government Issue Network Map showing two distinctive clusers of interlinking between site types, Armenian and international organizations, with Armenian sites linking to the international organizations and the latter not reciprocating. Map from Issuecrawler.net, Govcom.org Foundation, July 2004. Crawl by Audrey Selian.

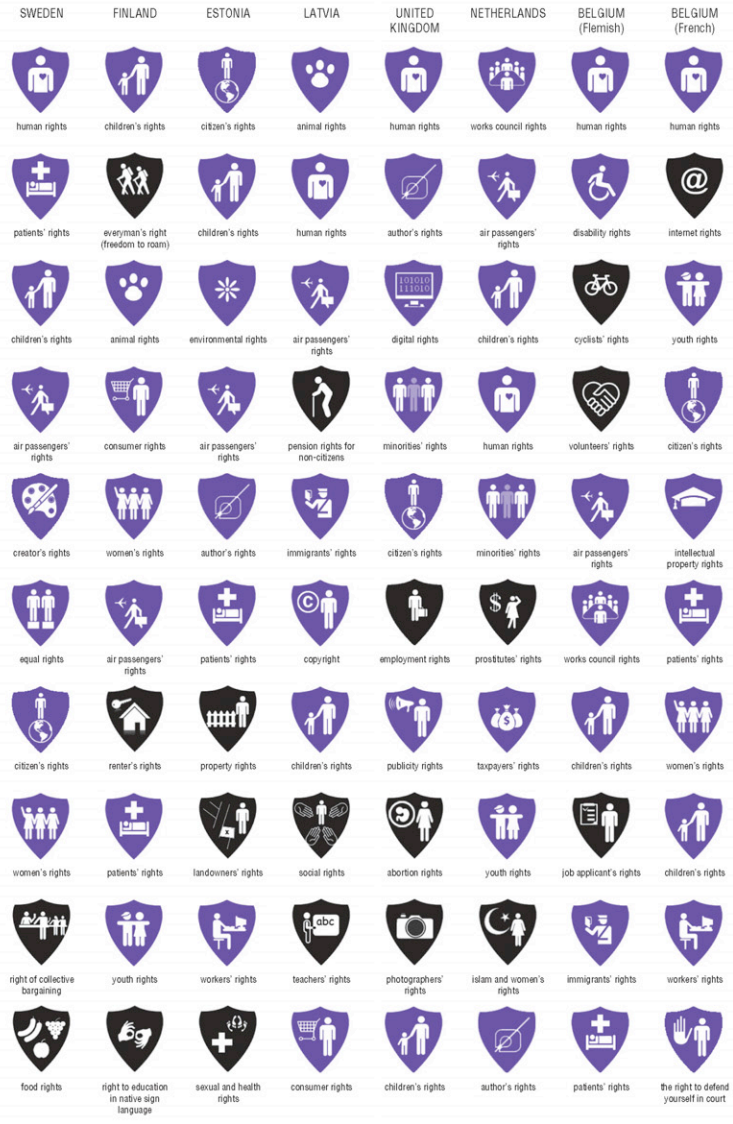


Figure 4: Rights Types - Hierarchy of rights resonance per country. Results of queries of local domain Googles for 'rights' in the respective country languages, with the rights types ranked according to their appearance in the results. Source: Rogers et al. 2009.

"genetic engineering" (918) "climate change" (894) disarmament (884) "nuclear power" (883) "ancient forests" (855) oceans (847) "sustainable trade" (782) "toxic chemicals" (649)

Figure 5: Greenpeace International's hierarchy of issue commitment, according to the number of mentions on its Website of each of its issues. Output from the Lippmannian Device by Govcom.org, October 2009.