Post-demographic Machines

Richard Rogers

Post-demographics?

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 (Boyd & Ellison, 2007). 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, including two described below. What kinds of findings may be made from mashing up the data, or what may be termed meta-profiling? Elfriendo.com is an application that profiles a set of friends. It allows one to compare the tastes of a set of friends to those of another, using MySpace data. Which TV shows are most referenced by those who have friended Barack Obama? How do they differ from those shows as well as books, music and movies from John McCain's 'friends' online? (The small case study was performed prior to the U.S. presidential elections in November, 2008.)

The second example of post-demographic work described here is the Leaky Garden Project (leakygarden.net), which furnishes a list of online services a particular user has subscribed to. One 'profiles' an individual (username) from the accounts taken out in Web 2.0 applications. Subsequently one sees the amount and also the details of the username's activity per platform, if, that is, the user's traces have been indexed by the major search engine, Google. These are 'leaks' in the so-called walled gardens, a term I return to.

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 (Foucault, 1998; Rogers, 2004). 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. As with Elfriendo and the Leaky Garden Project, the question concerns, which approaches and methods may be brought to bear in order to create new derivations from profile information, apart from niches and other, more specific products of behavioral marketing (Turow, 2006)?

Post-demographics is preferred over post-demography, as it recognizes popular usage of the notion of a 'demographic,' referring to a segment or niche that may be targeted or polled. Crucially the notion attempts to capture the difference between how 'demographers' and, say, 'profilers' collect as well as use data. Demographers normally would analyze official records (births, deaths, marriages) and survey populations, with census taking being the most well known of those undertakings. Profilers, 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 (Prensky, 2001). 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.

Social Networking Sites as Object of Post-demographic Study

"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" (Lenhart & Madden, 2007). Thus begins the study of American teenage use of such sites as MySpace and Facebook, conducted for the Pew Internet & American Life Project. 91% of the respondents use the sites to 'manage friendships'; less than a quarter use the sites to 'flirt'. Leaving behind surveys of user experiences for a moment, what is not as well known is what 'non-users' do with social network sites, with the occasional exception, such as the enquiry into how spammers leverage MySpace (Zinman & Donath, 2007). Non-users are those who do not manage friendships or flirt, but still visit the sites and read the profiles. They also may be interested in the data sets, and in automated means of capturing them, such as making use of the APIs, or screen-scraping the pages. With 'post-demographics,' the proposal is to make a contribution to the non-user studies -- 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.[1]

[1] 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,

How could one characterize the difference between the databases of online platforms and the databases of old (and new) that profile people to 'sort' them (Gandy, 1993)? Database philosophers were once deeply concerned about mandatory fields and field character limits -- the number of letters and numbers that would fit on each line in the electronic or hard copy form. The paucity of fields and the limited space available for an entry would impoverish the self, similar to how bureaucracy transformed individuals into numbers (Poster, 1991). People could not describe themselves fittingly in a few fields and characters.

Other critiques of early database profiling practices pointed out that the 'anomaly' was the most significant output of analysis. Certain people (in the sense of data constructs) would stand out from the rest, owing to their lack of statistical normalcy. In a cultural theory sense, the database became the site to derive the other.

What may be derived from the new databases? More otherness? Now, with online platforms, there are longer character limits, more fields, and far greater agency to author oneself, or as one scholar aptly put it, "to type oneself into being" (Sunden, 2003). 'Other,' that last heading available on the form, standing for difference, or taxonomic indeterminacy, has been replaced, generally speaking, by 'more.' For example, the user is invited to 'write note,' a freestyle field that provides opportunities for further self-definition and self-presentation. Now that the database is reaching out, providing you with more space to be yourself, questions may be posed. What does your form-filling say about you? Do you fill in the defaults only? Do you have many empty fields? What do your interests, and those of your friends, tell the profiler?

---

and how it may be analyzed or mashed up, would come from usage, with at least a minimal level of activity.

From a post-demographics perspective, the profile, together with the entities in orbit around it, lies at the core of research. Profilers are interested in what to do with all the 'interests' and 'favorites'.

You are media

What surrounds the profile? Generally, it has been observed that the Web, or at least a part of it, has new 'glue,' or 'plasma' in the Latourian sense (Latour, 2005). Where once hyperlinks tied sites together, now the social networking sphere is viewed as less of a hypertext than a hyper-object space. From this perspective, the Web is more social than informational. The network has profiles as its nodes, with links between friends as well as social objects, not to mention 'social' third-party applications, socially derived recommendations as well as adverts (Knorr Cetina, 2001; Engeström, 2005). An initial question is how sociality is organized.

For one's profile, the user is invited to fill in certain personal information and list favorites. The fields for age, gender and location are still present; yet profiles invite the post-demographic, with requests for media listings, as favorite movies, music, TV shows, books, etc. It also asks for and stores media files, as pictures, clips and tunes. Once the profile has been completed (for the time being), the social linking begins. One 'friends' (the new verb), shares, joins groups and accepts invitations for events.

Sociality breeds more of it. The more social you are, the more prominent you become, in a presence sense. That is, your own activity boosts you on other (friends') pages, be it a tweet, wall writing, or comment, which may appear as running entries on other (friends') pages (Facebook). The platforms continually encourage more activity, inviting commentary on everything posted, and recommending to you more friends (who are friends of friends). With all the ties being made, and all the activity being logged, the opportunities for analysis, especially for social network researchers and profilers, appear to be boundless.

There are of course constraints. Certain of these concern the issues involved in harvesting the data, and making derivations. Which social networking sites are scrapable, and to which extent? When, and under which conditions, is it acceptable to harvest data? Apart from data collection, at issue is also data usage. The depersonalization of the data would be helpful in particular ethical discussions of social network site analysis, however much celebrated cases have shown “why 'anonymous' data sometimes isn't” (Schneier, 2007). There are norms for data usage, the most basic of which is user consent. When signing up, the user makes an agreement with the platform, and there are terms of use for both parties, as well as a service privacy policy. Of crucial importance however is the blurring of the line as to who is the primary agent of ensuring privacy. Arguably, on social networking sites, the user is assuming more and more responsibility for privacy, in the settings chosen. Whilst the services have thought through the default settings, the user is the one who lets his or her guard down, if you will, by changing the profile viewing setting from friends only, to friends of friends, which is the maximum exposure level inside Facebook.

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,

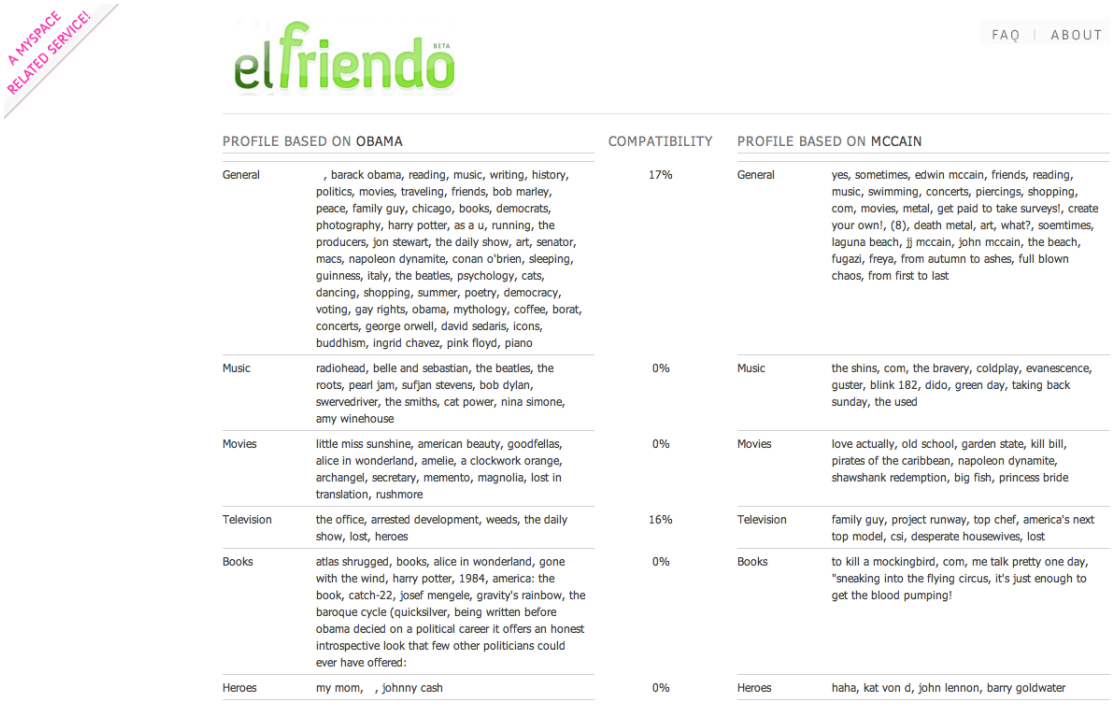
including Amazon's seminal recommendation system, where it is not highly relevant which person also bought a particular book, but rather that people have done so. Supermarket loyalty cards and the databases storing purchase histories similarly employ depersonalized information analysis, where like Amazon, 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.

Post-demographic Machines

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 however I would like to describe briefly two pieces of software built on top of machines, in the post-demographic analytical spirit, and the kinds of research practices that result.

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? 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.[2]



PROFILE BASED ON OBAMA		COMPATIBILITY	PROFILE BASED ON MCCAIN	
General	, barack obama, reading, music, writing, history, politics, movies, traveling, friends, bob marley, peace, family guy, chicago, books, democrats, photography, harry potter, as a u, running, the producers, jon stewart, the daily show, art, senator, macs, napoleon dynamite, conan o'brien, sleeping, guinness, italy, the beatles, psychology, cats, dancing, shopping, summer, poetry, democracy, voting, gay rights, obama, mythology, coffee, borat, concerts, george orwell, david sedaris, icons, buddhism, ingrid chavez, pink floyd, piano	17%	General	yes, sometimes, edwin mccain, friends, reading, music, swimming, concerts, piercings, shopping, com, movies, metal, get paid to take surveys!, create your own!, (8), death metal, art, what?, soemtimes, laguna beach, jj mccain, john mccain, the beach, fugazi, freya, from autumn to ashes, full blown chaos, from first to last
Music	radiohead, belle and sebastian, the beatles, the roots, pearl jam, sufjan stevens, bob dylan, swervedriver, the smiths, cat power, nina simone, amy winehouse	0%	Music	the shins, com, the bravery, coldplay, evanescence, guster, blink 182, dido, green day, taking back sunday, the used
Movies	little miss sunshine, american beauty, goodfellas, alice in wonderland, amelie, a clockwork orange, archangel, secretary, memento, magnolia, lost in translation, rushmore	0%	Movies	love actually, old school, garden state, kill bill, pirates of the caribbean, napoleon dynamite, shawshank redemption, big fish, princess bride
Television	the office, arrested development, weeds, the daily show, lost, heroes	16%	Television	family guy, project runway, top chef, america's next top model, csi, desperate housewives, lost
Books	atlas shrugged, books, alice in wonderland, gone with the wind, harry potter, 1984, america: the book, catch-22, josef mengele, gravity's rainbow, the baroque cycle (quicksilver, being written before obama decied on a political career it offers an honest introspective look that few other politicians could ever have offered:	0%	Books	to kill a mockingbird, com, me talk pretty one day, "sneaking into the flying circus, it's just enough to get the blood pumping!
Heroes	my mom, , johnny cash	0%	Heroes	haha, kat von d, john lennon, barry goldwater

Figure One: The interests of Barack Obama's and John McCain's MySpace friends, 10 September 2008. Elfriendo.com, Govcom.org Foundation, Amsterdam, 2008.

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 the example, one queries Elfriendo for Barack Obama as well as John McCain, and the profiles of their respective sets of friends

[2] 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.

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%. See Figure One.) 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 Leaky Garden Project

"Social networks require a degree of exclusion to work properly" (Shirky, 2003). Whilst commonly associated with certain social network sites, the term walled garden also refers to a business practice, notably in the software and hardware industries, where one firm's formats are incompatible with another's, thereby keeping the consumer 'locked in' (Arthur, 1989). Mobile phone rechargers come to mind, where Nokia's does not fit a Motorola phone, and vice versa. One of the arguments used in favor of lock-in is that dedicated hardware ensures the proper functioning of the technology. AT&T, with its historical slogan of "one company, one system, universal service," made this argument repeatedly, in efforts to disallow 'foreign,' or third party products and services, to run on the phone system, until the MCI lawsuit, and subsequent anti-trust work, finally unwound the Ma Bell monopoly in the 1970s and 1980s. With social networking sites, the notion of a walled garden cannot be applied as effortlessly. Social networking sites, especially Facebook, encourage third-party applications, in the new media style, with the realization that not only user's content, but also user's applications

increase the value as well as levels of participation. This is the classic argument concerning the inversion of the 'value chain' in online games as well as in the entire Web 2.0 industry, summed up in the idea that the more who use it, and contribute to it, the better and more valuable it becomes (Shirky, 2008). (Like the now famous graphic by Bruce Clay that shows the dependencies between search engines, in a kind of data eco-system approach, see in Figure Two a rendition of the flows between leading 2.0 services, Facebook, Flickr and Twitter (Clay, n.d.).)

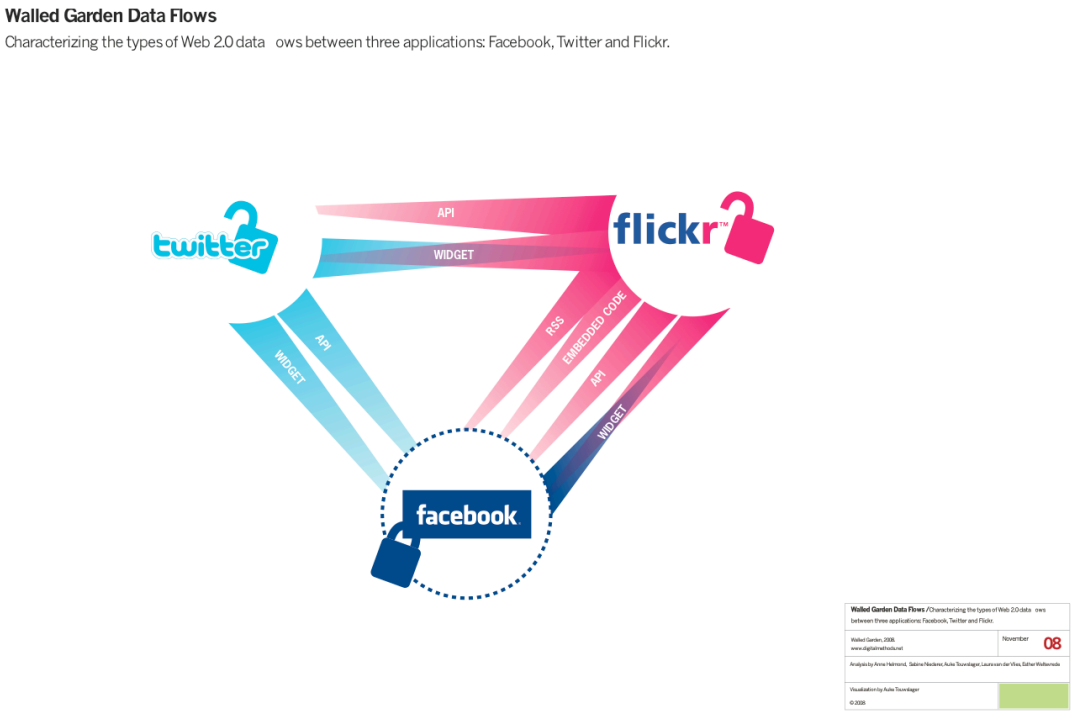


Figure Two: Walled Garden Data Flows. Digital Methods Initiative, Amsterdam, 2008.

Here the question concerns, just how walled are these gardens? Apart from examining the data flows between applications, as above, the question of the permeability and penetrability of the platforms also may be approached by examining whether and to what extent each is indexed by search engines. In order to do so, leakygarden.net sits atop a machine

that checks the availability of a particular username across a growing list of Web 2.0 applications. Usernamecheck.com is a useful service. When considering a new username, you may wish to know if and where it is taken, across the broader landscape of platforms. Here usernamecheck.com is repurposed, and in the first instance made into a profiling machine. Type in a username and check which services a person uses. Here the project researchers observed that generally speaking people seem to have two usernames, an alias as well as the real name (first and last name) as one word. Thus one may need to perform two queries for a fuller picture. Subsequently, leakygarden.net looks up references to the username. Does Google return pages from that username per platform? In all, the Leaky Garden Project shows which 'walled gardens' leak, and which are watertight (see Figure Three).

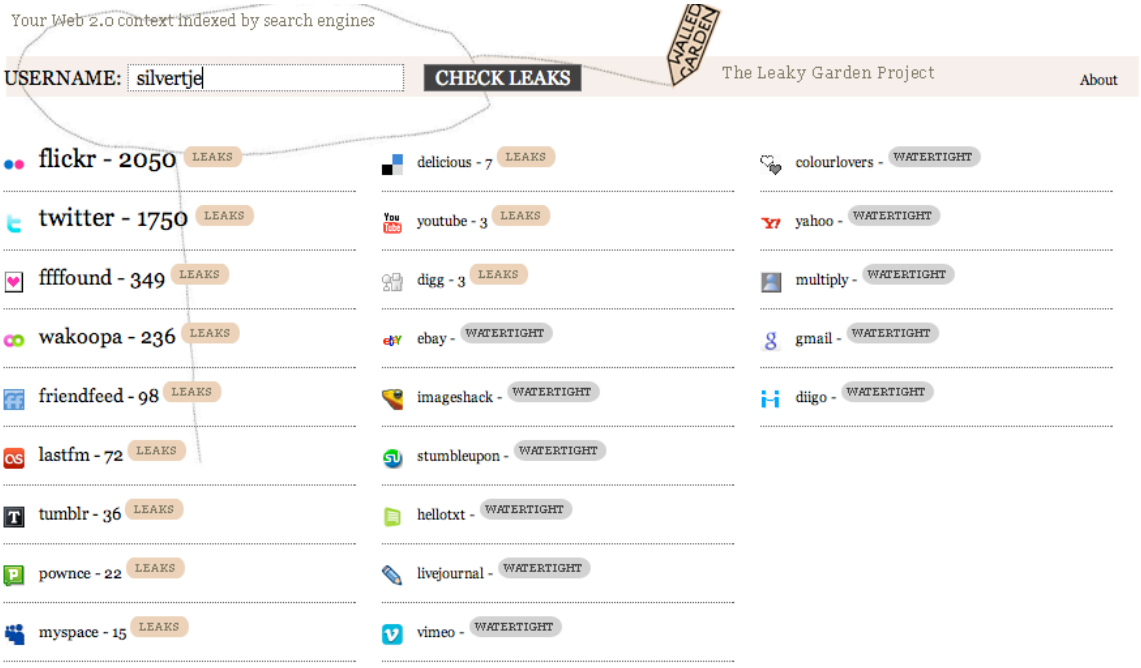


Figure Three: Username service subscription profile of silvertje (Anne Helmond), including the ‘leaks,’ or the amount of silvertje references per service, indexed by Google. Leakygarden.net, Govcom.org Foundation and the Digital Methods Initiative, Amsterdam, 2008.

Conclusion: What Would Nielsen Do?

Two methods dominate old media-style 'audience' research, the hand-written diary of a TV viewer or radio listener and the automated meter, registering how long a TV or radio channel is on, per household or household member. The diary technique is still in use, with the Nielsen company sending out a survey pack to its randomly selected families four times per year to record viewing habits during the so-called 'sweeps weeks'. Each person surveyed provides demographics, and a list of the shows they watch. Advertising is subsequently targeted to a TV show's demographic, with soap operas being the classic case of ads tied to a type of show. Because of survey effects, i.e., people changing their viewing habits owing to their need to keep a diary and fit a profile, an automated technique may be preferred (Stabile, 1995). In the United States, such recording devices were first employed for radio listeners, with the introduction in the 1940s of the Nielsen audimeter, which registered which frequency a radio was tuned to, and for how long (McLuhan, 1951). The results were useful for advertisers, and remain so. Of the initial study performed with the audimeter in 1942, *Time Magazine* wrote: "When the star of one of radio's most popular nighttime shows said 'Good night,' listening dropped sharply. The sponsor's closing commercial was heard by only a fraction of the program's audience" (Time Magazine, 1943). Nielsen's automated television ratings began in the 1950s, and were taken to the next level with the black box known as the Storage Instantaneous Audimeter, which captured TV viewing of each set in the household, sending data back to headquarters daily through a phone line. "People meters" have been employed since the 1980s, where each member of the household has his/her own button on the remote control. Behind the button, in the database, are the user's age and gender, and the meter on top of the television is tagged with a location.

TV shows are rated through a point system, with one point given per percentage of all households watching. Advertising rates are subsequently

expressed in cost per point. A show has an expected rating (based on history) as well as an actual rating. Of interest to the advertisers is the 'post-buy' calculation of actual audience reach, that is, whether their advert actually had the expected audience types and numbers. Was the advert a good buy?

Figure Four: Top 50 Brands of Hyvers, August 2007. Digital Methods Initiative, Amsterdam, 2007.

Should post-demographics emulate the Nielsen machines and metrics? Are there post-demographic equivalents to the machines and their metrics? Indeed, one may transfer the counting method from TV audience research to social networking sites, using the available interest fields as well as basic demographic data (gender, age and location). Thus one may tally references to a particular interest across an entire social networking platform, as colleagues and I did for Hyves in the Netherlands in 2007 (see Figure Four). (No demographic data were used in the example.)

Among the types of favorites at Hyves are brands, and Hyvers, as the users are called, fill in that field, albeit often without the care and diligence that would be demanded of a Nielsen family member.

Examples of ‘non-cooperative’ Hyvers' brands field (to 6 August 2007):

My Style is My Brand
ben geen merkentype
Houd er niet van ge(brand)merkt te worden
ik ben niet zo van de merken
I don't spend much time thinking about brands
Daar doe ik dus ff lekker niet aan mee he
Ik merk het
geen zin in aanvinken

How to tidy the data and make ratings? What would Nielsen do? One could strive to transfer the audience research technique to the new medium. Perhaps particular Hyvers would agree to become Nielsen social networkers, and provide meticulous up-to-date profiles. The fields would be monitored by Nielsen for changes in interests and tastes, and ratings could be provided with a point system, where fans are the equivalents of viewers.

As unlikely as the proposal may sound, it points up the larger question of whether and when to import standard methods of study onto the new medium. It also raises the question of the distinctive uses to be put to post-demographics.

References

Arthur, W. Brian (1989), "Competing Technologies, Increasing Returns and Lock-in by Historical Events," *Economic Journal*. 99: 106-131.

Boyd, Danah and Ellison, Nicole (2007). "Social network sites: Definition, history, and scholarship." *Journal of Computer-Mediated Communication*. 13(1), http://jcmc.indiana.edu/vol13/issue1/boyd.ellison.html

Clay, Bruce (n.d.). "Search Engine Relationships Chart." Bruceclay.com, http://www.bruceclay.com/searchenginerelationshipchart.htm.

Engeström, Jyri (2005). "Why some social network services work and others don't. Or: the case for object-centered sociality." Blog posting. Zengestrom.com, http://www.zengestrom.com/blog/2005/04/why_some_social.html

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

Gandy, Oscar (1993). *The Panoptic Sort: A Political Economy of Personal Information*. Boulder, CO: Westview Press.

Knorr Cetina, Karin (2001). "Objectual Practice." In: Schatzki, Theodor R., Knorr Cetina, Karin and von Savigny, Eike (eds.), *The Practice Turn in Contemporary Theory*. London: Routledge, 175-188.

Latour, Bruno (2005). *Reassembling the Social: An Introduction to Actor-Network Theory*. Oxford: OUP.

Lenhart, Amanda and Madden, Mary (2007). "Social Networking Websites and Teens." Pew Internet Project Data Memo. Washington, DC: Pew

Internet & American Life Project, http://www.pewinternet.org/pdfs/PIP_SNS_Data_Memo_Jan_2007.pdf.

McLuhan, Marshall (1951). *The Mechanical Bride*. New York: Vanguard.

Poster, Mark (1991). *The Mode of Information*. Chicago: University of Chicago Press.

Prensky, Marc (2001). "Digital Natives, Digital Immigrants." *On the Horizon*. 9(5), http://pre2005.flexiblelearning.net.au/projects/resources/Digital_Natives_Digital_Immigrants.pdf.

Rogers, Richard (2004). *Information Politics on the Web*. Cambridge, MA: MIT Press.

Schneier, Bruce (2007). "Why 'Anonymous' Data Sometimes Isn't." *Wired*. December, http://www.wired.com/politics/security/commentary/securitymatters/2007/12/securitymatters_1213.

Shirky, Clay (2003). People on page: YASNS... Blog posting. *Corante's Many-to-Many*. http://many.corante.com/archives/2003/05/12/people_on_page_yasns.php.

Shirky, Clay (2008). *Here Comes Everybody*. New York: Penguin.

Stabile, Carol (1995). "Resistance, Recuperation and Reflexivity: The Limits of a Paradigm." *Critical Studies in Mass Communication*. 12: 403-422.

Sunden, Jenny (2003). *Material Virtualities: Approaching Online Textual Embodiment*. New York: Peter Lang.

Time Magazine (1943). "Who Listens to What?" XLI(1), 4 January.

Turow, Joseph (2006). *Niche Envy*. Cambridge, MA: MIT Press.

Acknowledgments

Richard Rogers would like to acknowledge the contributions made to earlier versions of the text by the Digital Methods Initiative, including Erik Borra, Sabine Niederer, Michael Stevenson, Marijn de Vries Hoogerwerff and Esther Weltevrede. The Leaky Garden Project software and graphics are the outcome of the Walled Garden event, organized by the Virtual Platform, Amsterdam, November, 2008, with additional contributions by Marieke van Dijk, Andrei Fiore, Anne Helmond, Koen Martens, Auke Touwslager and Laura van der Vlies. The Elfriendo software is the product of the Suggested Fields workshop by the Govcom.org Foundation at the Netherlands Media Art Institute, Amsterdam, January, 2008, for which the term post-demographics was coined.