Digital Research and Methods For All (Researchers)

Loïc Riom (University of Geneva), Julien Ruey (University of Geneva), Sami Coll (Université du Québec à Montréal), and Mathilde Bourrier (University of Geneva)

Who has never used Google to find information about his research project? More or less consciously and deliberately, Internet is already deeply rooted in our practices as researchers. Digital methods can be used for a wide spectrum of research questions, whether they are digital related or not. Today, almost any research on any subject can benefit from using them, whether relying on quantitative or qualitative methods.

Exploring the wide spectrum of researches using digital-based methods

Web-based methods have already become popular, because they enable researchers to build large samples at a relatively low cost (Snee *et al.* 2016). This is especially the case when it comes to collecting data with "classic" quantitative methods (e. g. a survey). Moreover, it also offers new venues to collect data. In particular, social media offers a rich quantity of data, allowing for example to study listening practices (Berkers 2012) or political mobilization during social movements (e. g. during UK riots, Beguerisse-Díaz *et al.* 2014 and Occupy, Thorson *et al.* 2013). For instance, Grandjean (2016) made use of Twitter to map researchers working in the field of digital humanities. The Digital world can be used as a fieldwork too. Golub (2010), for example, conducted participatory observation as a player of the online game *The World of Warcraft* to study how specific knowledge is produced. Social media is not the only source of social data. Recently a research team collected mobile phone geolocalization data and metadata from *Open Street Map* to study street activity of six Italian cities (De Nadai *et al.* 2016).

More generally Internet offers a massive archive on various subjects for content analysis (Ackland 2013; Rogers 2013). Lee and Peterson (2004) studied the *Alt-country* amateur's scene through *Postcard Tow*, a forum devoted to this musical genre. Thelwall, Wilkinson and Uppal (2010) gathered thousands of comments on *MySpace* to study gender differences regarding emotional communication. Balleys and Coll (2015) studied teenagers' behavior on social networks to shed light on the way they build their social prestige amongst peers, whether it be offline or online. Beside textual data, video or photo-sharing platforms such as *Youtube*, as well as *Instagram* or *Flickr* offer large amount of available audiovisual content. For example, Horsti (2016) collected videos on *Youtube* to study the production and the diffusion of collective memory of illegal migration in Europe. In addition, the way Internet – as a culture artifact (Hine 2000) – is shaped can tell us a lot about societies, even beyond the digital sphere. For instance, Zimmermann (2015) made a comparison between *Facebook* and *Happy Network* – a former popular social media in China – to investigate how digital technologies are differently used in different cultural settings.

As these examples show, online-gathered data can be divided into two categories. The first category regroups works where primary data is produced through "nethnography" or online surveys. The second category includes works that use data already produced for other purposes and, in most cases, which are available and free. Both of these categories of data production attest the potential of digital methods whether it be to study the digital world or not. They offer a great opportunity to enrich sociological research at little cost. However, the reliability of methods and the quality of data gathered must be questioned. For example, it must be examined to what extent collected data for other purposes carries unacknowledged biases.

Questioning digital methods and tailoring algorithms to the needs of social scientists

For more than a decade, several books and articles have been published about the development and the use of digital methods (see for instance Hine 2000; Mason *et al.* 2005; Ackland 2013; Rogers 2013; Snee *et al.* 2016). Authors discuss methodological issues such as selection biases that online recruitment can induce, especially when it comes to social media. They introduce researchers to the use of practical tools that can help handle digitally generated data when conducting online interviews (Ackland 2013). Also, they offer a general introduction to the Web and most of them point out to the necessity to understand how Internet works, in order to have the necessary critical view on the generated data (Hine 2000; Ackland 2013; Snee *et al.* 2016). Furthermore, collecting data on the Internet raises questions ranging from informed consent to participant anonymity via the distinction between private and public sphere (e. g. when collecting data on social media or forums) (Beaulieu 2004; Garcia, *et al.* 2009; Ackland 2013). In this regard, the Association of Internet Researcher (AoIR) has established a code of ethics for research on Internet since 2002[1].

The potential of computer algorithms to analyze data should also be taken into account. However, they have mostly been developed by private companies for marketing purposes. *Google* and *Facebook* are probably the ones, which develop the most sophisticated algorithms. Yet, these algorithms are still opaque (Cardon 2015; Pasquale 2015). Consequently, there is a need to pioneer and sponsor the development of algorithms specifically oriented to the benefit of social sciences. Hence, sociologists have a role to play not only in the way data is collected, but also in how it can be produced and analyzed.

Introducing digital methods to social science education

Such development of digital methods should be taken into account as quickly as possible within our education programs (Bachelors, Masters and PhDs). Indeed, already many of our students already have to deal research on the Internet during their investigations. Thus, we should be able to give them the right tools to conduct their research. There is an urgent need for both a better understanding of the "digital world" and an expertise on digital methods. This should include a wide range of courses from applied statistics to programming and data visualization. Social sciences' institutes should not neglect the development of those skills.

Some universities, for example the University of Uppsala, in Sweden, and the University of Sheffield in the UK, have already introduced Master's programs called "Digital media and society." In Switzerland, the University of Lausanne has just introduced a Master's program in digital humanities. These programs are designed both to introduce students to the evaluation of social shifts due to digital technologies and to train them in the practice of inquiry based on online data. If we look more closely at the courses they provide, they are deeply rooted in the already existing research practices of social sciences. Following the steps of this existing programs, it can be recommended the following to be incorporated into the sociological curricula:

- An introduction on the main key concepts of the social and semantic web (e. g. key words, tags, hashtags, links, data identifiers, etc.)
- A presentation of quantitative and qualitative digital methods to gather data online, along with related research ethics and data protection.
- An initiation to technical tools such as web-scraping, visualizing and mapping software, web archives, social networks' APIs (Application Programming Interfaces) and coding techniques.

1 Available at http://aoir.org/ethics/

Thus, applying digital methods does not mean giving up on older methods, like ethnography or interviews, but carrying on the necessary adaptation and development of our research tools. Moreover, testing new ways of inquiry is also an interesting opportunity to reinforce the tradition of critical thinking when it comes to research design, and to stimulate what Mills (1959) named our "sociological imagination."

In sum, digital methods clearly raise a lot of important questions. This paper does not have the pretense of being exhaustive, but rather a way identify the main challenges regarding digital methods. Digital methods are not completely new, contrary to what some of us may think, and they call to be integrated in the everyday work of researchers as soon as possible. They need to be demystified as they are more accessible than we commonly believe. Furthermore, since digital giants such as *Google* or *Facebook* already claim to be able to produce more relevant research than academic researchers, it is important to step in and be part of the game (Boyd and Crawford 2012).

References

Ackland, Robert (2013): *Web social science: Concepts, data and tools for social scientists in the digital age*. London, Thousand Oaks and New Delhi: Sage.

Beaulieu, Anne (2004): Mediating Ethnography: Objectivity and the Making of Ethnographies of the Internet, *Social Epistemology* 18(2–3), pp. 139–159.

Beguerisse-Díaz, Mariano, Guillermo Garduño-Hernández, Borislav Vangelov, Sophia N. Yaliraki, and Mauricio Barahona (2014): Interest communities and flow roles in directed networks: the Twitter network of the UK riots. *Journal of The Royal Society Interface* 11(101), DOI: 10.1098/rsif.2014.0940.

Berkers, Pauwke (2012): Gendered Scrobbling: Listening Behaviour of Young Adults on Last.fm. *Interactions: Studies in Communication & Culture* 2(3), pp. 279-296, doi:10.1386/iscc.2.3.279_1.

Boyd, Danah and Kate Crawford (2012): Critical Questions for Big Data. *Information, Communication & Society* 15(5), 662-679, doi:10.1080/1369118X.2012.678878.

Cardon, Dominique (2015): À quoi rêvent les algorithmes. Nos vies à l'heure des *big data*, Paris: Seuil.

De Nadai, Marco, Jacopo Staiano, Roberto Larcher, Nicu Sebe, Daniele Quercia, and Bruno Lepri (2016): The Death and Life of Great Italian Cities: A Mobile Phone Data Perspective. In International World Wide Web Conferences Steering Committee (Eds.). *Proceedings of the 25th International Conference on World Wide Web*. pp. 413–423, http://dl.acm.org/citation.cfm?id=2883084.

Garcia, Angela Cora, Alecea I. Standlee, Jennifer Bechkoff, and Yan Cui (2009): Ethnographic Approaches to the Internet and Computer-Mediated Communication, *Journal of Contemporary Ethnography* 38(1), pp. 52–78.

Golub, Alex (2010): Being in the World (of Warcraft): Raiding, realism, and knowledge production in a massively multiplayer online game. *Anthropological Quarterly* 83(1), pp. 17–45.

Grandjean, Martin (2016): A social network analysis of Twitter: Mapping the digital humanities community. *Cogent Arts & Humanities* 3(1), doi: 10.1080/23311983.2016.1171458.

Hine, Christine (2000): *Virtual ethnography*. London: Sage.

Horsti, Karina (2016): Communicative memory of irregular migration: The re-circulation of news images on YouTube. *Memory Studies*. doi: 10.1177/1750698016640614.

Lee, Steve S. and Richard A. Peterson (2004): Internet-based virtual music scenes: The case of P2 in alt. country music. In Andy Bennett and Richard A. Peterson. *Music Scenes: Local Translocal and Virtual*. Nashville: Vanderbilt University Press, pp. 187–204.

Mason, Bruce, Amanda Coffey, and Paul Anthony Atkinson (2005): *Qualitative Research and Hypermedia*. London, Thousand Oaks and New Delhi: Sage.

Mills, Charles W. (1959): *The Sociological Imagination.* Oxford: Oxford University Press.

Pasquale, Frank (2015). *The Black Box Society: The Secret Algorithms That Control Money and Information*. Cambridge: Harvard University Press.

Rogers, Richard (2013): *Digital methods*. Cambridge & Londres: MIT Press.

Snee, Helene, Christine Hine, Yvette Morey, Steven Roberts, and Hayley Watson (2016): *Digital methods for social science: An interdisciplinary guide to research innovation*. New York: Palgrave Macmillan.

Thelwall, Mike, David Wilkinson, and Sukhvinder Uppal (2010): Data mining emotion in social network communication: Gender differences in MySpace. *Journal of the American Society for Information Science and Technology* 61(1), pp. 190–199.

Thorson, Kjerstin, Kevin Driscoll, Brian Ekdale, Stephanie Edgerly, Liana Gamber Thompson, Andrew Schrock, Lana Swartz, Emily K. Vraga, and Chris Wells (2013): YouTube, Twitter and the Occupy movement: Connecting content and circulation practices. *Information, Communication & Society* 16(3), pp. 421–451.

Zimmermann, Basile (2015): *Waves and Forms: Electronic Music Devices and Computer Encodings in China*. Cambridge & Londres: MIT Press.