Full reference: Reid, L., and Timan, T (2017) What are digital methods, where have they come from, and how/why are they important? *Working Paper, Centre for Housing Research,* University of St Andrews.

What are digital methods, where have they come from, and how/why are they important?

Introduction

In this blog, the first of a series to come from our digital methods workshop, we attempt to highlight how developments in digital methods offer new opportunities for social science, and research on household sustainability in particular. We believe that many social scientists are unfamiliar with developments in digital methods and hope to encourage greater appreciation of them. Indeed, we conclude our piece with some reasons why social scientists should engage with debates around digital methods and consider using them. Before doing so, we provide some background about what digital methods are and how they have evolved.

1. Digital Methods: what are they?

Digital methods (DM hereafter) are a strand of research methods used to, in general, research the Web. Think for instance of research to understand how Facebook operates as a network[1] or to explore how Twitter may influence elections[2]. DM refer to methods of research that reside in the digital, that allow the researcher to investigate phenomena that are natively digital (i.e. phenomena that are 'born' of the internet and which does not exist offline). Somewhere between qualitative and quantitative branches of research methodology, DM are a set of roughly defined yet heterogeneous methods and tools to capture, collect and analyse 'things' that happen online. They stem from a need in social science to deal with web 2.0, social network sites[3], online news, user generated content and online expressions of society in a scientifically rigorous manner. Since the field is rather young, and deals with a rapidly changing subject of study (the 'digital'; the 'online') definitions, delineations and epistemologies are not set in stone and are under scrutiny. Specifically, the added value of, or difference between DM and 'traditional' offline social science methods, is a focus of debate.

Rogers[4], defines the difference between DM and other social science methods that might also be using digital means, and this is where things get messy: Rogers distinguishes between virtual methods and DM stating that the former means *digitized*

[1] Rieder, B (2013) Studying Facebook via Data Extraction: The Netvizz Application Proceeding WebSci '13 Proceedings of the 5th Annual ACM Web Science Conference France — May 02 - 04, 2013 Pages 346-355 Paris,.ACM New York, NY, USA

[2] See Elmer, G (2012) Live research: Twittering an election debate *New Media Society* 2013 15: 18 DOI: 10.1177/1461444812457328

[3] See Ellison, N. B. (2007). Social network sites: Definition, history, and scholarship. Journal of Computer- Mediated Communication, 13(1), 210-230.

[4] Rogers, R. (2013). Digital methods. MIT press.

versions of phenomena that existed before the Internet, where the latter is solely aimed at phenomenathat have no precedent pre-Internet. To give an example, an online survey would be considered a virtual method, whereas Twitter-analysis would be considered a DM. The point made by Rogers is that the Web, although ever evolving, has its own specific epistemology[5], and needs its own research (methods) because virtual methods are insufficient to grasp digital cultures and digital objects. Not only does the Web exist of digital objects, but it relies on an architecture that is different from that of the 'real' world - different platforms that these digital objects reside in, operating systems, programming languages, system architecture and devices that are involved in the creation and maintenance of these objects[6]. , Additionally, the logics of the Web are of interest. Think of ordering devices such as search engines that devise how search results are depicted and in which order, or social media sites that pre-select and 'feed' news according to preferences (so-called echo chambers). Google[7] has had a major influence on the logics of the Web and in its wake, the kind and type of digital objects to be studied[8]. Moreover, Google itself runs all sorts of analysis on its users, not only to learn from them and to improve their services, but also to pursue all sorts of (third-party-requested) analytics for other purposes (marketing, mainly). Compared to the capacity and speed of analytics that Google can perform, versus the capabilities and resources of an individual scientist, the question for Internet researchers has become "what would Google do?"[9]

The role of DM research is then to develop an alternative discourse than that of Silicon-Valley-driven marketing analytics and to develop methods of research and tools that can delve into questions related to digital phenomena[10]. Moreover, a question that precedes and shapes DM, is how to interpret the digital, the Web and all that comes from it. Whilst in the early stages of Internet research, the divide between digital and physical, online and offline was a demarcation of worlds, Rogers (among others) now proposes that:

> "[…] a new era in Internet research, which no longer concerns itself with the divide between the real and the virtual. It concerns a shift in the kinds of questions put to the study of the Internet. The Internet is employed as a site of research for far more

[5] Rogers, R. (2009). The end of the virtual: Digital methods (Vol. 339). Amsterdam University Press. See also Rogers, R. (2000). Introduction to the Practice of Web Epistemology. Preferred Placement, 11-23.
[6] In that respect, one largely under-studied topic is the archive of the Web and the role of companies and industry standards in storing collective and individual histories
[7] http://www.google.com - or ALPHABET, recently.
[8] See Vaidhyanathan, S. (2012). The Googlization of everything:(and why we should worry). Univ of California Press.
[9] See Rogers (2009) Inaugural lecture – available at http://govcom.org/publications/full_list/oratie_Rogers_2009_preprint.pdf. See also Anderson, (2008) in Wired https://www.wired.com/2008/06/pb-theory/ in which the end of social science is heralded. In this piece, he predicts that due to the coming of big data analytics, there would be no role anymore for social science that Google or other large-part data driven companies cannot or did not already answer - the data gathered by these companies would be far more fine-grained and rich and vast, and the possibilities for analysis far more sophisticated than any social scientist could ever dream of. This rather deductive reasoning in combination with an almost sacred belief in the objectivity and 'truth' of data is scary, and missed completely the point of the politics of data, of infrastructure, and historical questions of data and its context. See also boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. Information, communication & society, 15(5), 662-679.
[10] See https://wiki.digitalmethods.net/Dmi/ToolDatabase for such a tool collection

> than just online culture. The issue no longer is how much of society and culture is online, but rather how to diagnose cultural change and societal conditions using the Internet. The conceptual point of departure for the research program is the recognition that the Internet is not only an object of study, but also a source."

The latter point is important when wanting to transfer and utilise DM outside of the scope of Internet research and in other disciplines, like geography, political science, or information sciences. A challenge faced by all disciplines researching online life, is what these DM mean scientifically, and what kinds of claims can be made about this *onlife*[11] when referring to online accounts of phenomena. One argument in favor of the value of studying the digital as a realm on its own, is that the online/offline, virtual/real, digital/physical are increasingly intertwined, rendering a separation between these binaries increasingly fruitless. Why this entanglement happened is illuminated by providing a very brief history of how we got here…

2. A super-short history of the Web

In historicising the relatively short existence of the Internet and with it, Internet research, a rough division can be made between: 1) Web 1.0; and, 2) the upcoming era of connected PCs, Web 2.0, User-Generated Content and an interactive Web. The former may be described as a pre-interactive Internet, the latter representing a shift towards an increase in mobility and quantity of 'smart' and connected devices. Without going into too much Human-Computer Interaction history[12], the key feature of Web 1.0 was the ability to link to other pages on one's page (hyperlinks). However, that was about it regarding interactivity; adding content as a visitor was rare and complicated, as was creating more intricate and complex websites. Web 2.0 roughly demarcates the development of an interactive Web, in which User-Generated Content (UGC) would be central. With the proliferation of smaller and mobile computational devices, as well as widespread adoption of the Internet as a place for and of communication, innovations were made to online platforms and the underlying software platforms that run the social Web. With Google becoming the dominant search-engine, and social networking sites on the rise, the pinnacle of this era might be the moment Times magazine coined the 'user' as person of the year[13].

The interactive Web brought online services of and for all stakeholders and sectors in society, from governments and NGOs, large companies and small businesses, to schools and hospitals, TV, music and journalism; all utilising similar online means for their own ends. This convergence of communication and information media at large also had (and has) its effects on science. First, due to the "computational turn"[14], there is hardly any science that does not rely on computers[15] and digital tools to do its work, secondly the

[11] See for example the onlife manifesto drawn up by the EC: https://ec.europa.eu/digital-single-market/en/onlife-manifesto
[12] DiSalvo, C. (2014). The need for design history in HCI. *interactions*, *21*(6), 20-21.
[13] http://content.time.com/time/magazine/article/0,9171,1570810,00.html
[14] Berry, D. (2011). The computational turn: Thinking about the digital humanities. *Culture Machine*, *12*.
[15] In some sciences at least partially, for most sciences completely

Internet has brought about new forms of social science that are rapidly branching out from (digital) culture and media studies, to interaction design research, online network analysis, and platform - and software studies. Consequently, this brought about new questions and new forms of doing research, of which DM is just one strand. In addition, the move from textual searches via keywords in a browser-bar to the logic of social network sites that feed content –move media making and User-Generated Content to a more passive form of media consumption[16], something that is also challenging for DM. How these developments influence the Internet and its connected research, thus remains to be seen.

The logics (and attractiveness) of DM rely on the premise of being able to capture an entire network (n=all), thereby departing from sampling strategies developed in classical quantitative social sciences, that via statistical methods claim representation of a society at large, or sub-parts of that society. Due to the accessibility of digital expressions, be they blogposts, links, Tweets, or Facebook-likes, once the researcher can analyse one of those expressions, s/he can access and analyse all. By way of example, one can research an election-debate by capturing ('scraping' in Internet research lingo[17]) *all* Tweets related to that election[18]. Or, based in pre-social network site analyses, DM can map controversies that have emerged on the Web or offline, have a counterpart in different media online[19]. Putting Actor-Network Theory into practice[20], DM allows, for instance, to see how the battle over climate change is fought online by exploring who links to whom[21], or alternatively, how border-claims are confirmed or denied by popular mapping platforms[22]. In short, not only does DM allow for research questions that pre-existed the Web to be pursued (be it in a slightly different shape), but with more analytical power[23], DM allows for empirical-based (sometimes data-driven) research. This makes DM suitable and interesting for social science disciplines outside of Internet research.

3: Of what interest to social science?

Clearly, DM are fast-evolving, but they do offer opportunities and risks for social science, and at the very least, social scientists should become more familiar with these.. To begin

[16] Lee, C. S., & Ma, L. (2012). News sharing in social media: The effect of gratifications and prior experience. *Computers in Human Behavior*, *28*(2), 331-339.

[17] Munzert, S., Rubba, C., Meißner, P., & Nyhuis, D. (2014). *Automated data collection with R: A practical guide to web scraping and text mining*. John Wiley & Sons.

[18] See also, online network analysis explained Hanneman & Riddle (http://faculty.ucr.edu/~hanneman/nettext/). It goes beyond the scope of this small piece to explain network analysis, and its accompanies Bayesian statistics, it is, however important to mention here because a lot of assumption in DM are based o this type of analysis, which in itself has a different pro's and con's and 'camps' within math and computer science on its merits

[19] Marres, N. (2015). Why map issues? On controversy analysis as a digital method. *Science, Technology, & Human Values*, *40*(5), 655-686.

[20] See Law, J. (1999). After ANT: complexity, naming and topology. *The Sociological Review*, *47*(S1), 1-14. or Latour, B. (1999). On recalling ANT. *The Sociological Review*, *47*(S1), 15-25.

[21] See for an example project https://wiki.digitalmethods.net/Dmi/SkepticsIssues

[22] Bier, J. (2014). *Mapping Israel, mapping Palestine: how segregated landscapes shape scientific knowledge* (Doctoral dissertation, Maastricht University).

[23] Simply because it is now easier for one researcher to access and delve into the entire corpus or network of a certain topic and it is more easy/ faster to query it.

the final, and concluding section of this piece, some critiques and benefits of DM are outlined.

As with any research method, DM have limitations, specifically a reliance on computer science and data science-methods to identify and make sense of data. Moreover, serious questions have been raised around issues[24] such as data ethics and informed consent, live influence and testing on millions of users, and privacy and data protection. There are also unresolved questions about the context of data and how that may be understood (or not) when using DM. Specifically, whether DM can accurately capture and explain the broader social environment within which phenomena take place. DM are also not immune to challenges posed in 'traditional' social science, that of representation (e.g. can we truly decontextualise users/consumer/citizens) and of positionality (particularly in the context of social media sites) – if, of course, issues of representation can ever really be fully understood. Additionally, exploration of how DM deals with time is another area for consideration since long-term studies are often difficult to achieve (although the same can be said for 'traditional' social science methods).

Despite these challenges, DM does offer some potential for social science to explore existing phenomena differently, to shape how DM are undertaken, and to gain access to new forms of data to help theorize and explain our social world. Within DM there will likely always be a need for social scientists to help develop the questions that need asked of DM and to offer (alternative) narratives or explanations to complement those which can merely be described by data-driven quantitative methods. For social science, and for studies of household sustainability more specifically, the ease with which one might adopt DM (e.g. largely free from financial costs as opposed to travel, incentives and transcription costs for traditional interviews) and their usefulness at least in the initial stages of a research project to explore a topic, are no doubt attractive. Moreover, they offer potential to accelerate and reduce effort of collecting new data. Perhaps most interestingly, DM would help social scientists capitalize on the increasingly 'smart' aspects of everyday life to more easily access the mundane/private realms that are of so much interest – for instance exploring data from online forums that often capture intimate and very rich accounts of domestic experience.

To end this piece, we identify two critical = areas of DM that we encourage social scientists to consider.

a) <u>Digital objects: classifications and strategies</u>

The first step of any analysis is to decide what it is one is looking at, and to be critical of the origins of the data: how did I get to the data? Can I rely on search engine classifications to prioritise and rank results? How do I analyse the data? Menchen-Trevino offers an insightful strategy to classify digital data types and connected analyzing

[24] Michael Zimmer ''But the data is already public'': on the ethics of research in Facebook" - see also Boyd, D., & Crawford, K. (2011, September). Six provocations for big data. In *A decade in internet time: Symposium on the dynamics of the internet and society* (Vol. 21). Oxford: Oxford Internet Institute.

strategies, separating data into horizontal and vertical forms. Horizontal datasets are "organised around a specific type of trace, for example, search terms, web browsing log files, tweets, hashtags, likes, or friend and follower ties". Vertical data are "those that extend beyond one digital tool and are organised around research participants. In this type of dataset, individuals give permission for researchers to collect their digital traces, sometimes across a variety of digital services". This differentiation can help in framing the research questions and the types of findings and conclusion one can draw from a dataset of digital traces.

A second consideration is of the level of analysis. The issue with DM is that all sorts of places and spaces on the Web and social media are used to base research methods on *and* at the same time, are the subject of research. For example, if one uses a Twitter-scraper to capture tweets related to a new energy policy, one is utilising the methods and the platform of a digital service, often relying on their limits, logics and ways of archiving. In short, one is relying on the internal *methods* of a digital service, while at the same time capturing the *content* of that service. It is all too easy and tempting to use the many tools offered online to do online research, however, a researcher must understand how these tools work, and what they do (and don't do).

b) Decontextualisation of data

A downside of the promise of widely available online data is that the researcher does not know the context of this data. Too often in the use of DM and quantitative media studies researchers are blindly pulling decontextualised data through analysis software to then present the outcomes as unquestionable facts or findings (because, n= all, so it must be true). Interacting with data scientists and data designers and looking at their practices often reveals that cleaning up datasets takes up most of the work, not least because datasets are often incomplete. Even if complete datasets can be found, there are often gaps in the entries, and more importantly, there are gaps in metadata - in notes or other forms of information that can help contextualise what is created, by whom, when, in which conditions, how representative the data is in relation to the phenomenon under research etc. By dismissing questions of representation and context and without any 'real-world' validation, the outcomes of DM research must be taken with a pinch of salt, due to the rapidly changing digital realm it moves itself in ("moving-target" research). Other, practical issues arise because of the temporality of online platforms and services - often once it reaches a popularity peak, its main monetary value lies in the network of users and their accompanied data, resulting in a closing down of the service to research (Twitter is a good example).

These issues are clearly ones for consideration by social scientists and whilst we would encourage all to keep a focus on the research question, rather than allowing DM itself to lead, we do believe that DM may add analytical strength to uncover novel strands of research and advance our understanding of social phenomena.