horizon
DIGITAL ECONOMY RESEARCH

The Horizon Digital Economy Research Institute centred at the University of Nottingham was created through grants totalling £30m from the RCUK Digital Economy programme and currently involves over 20 academics and 30 research fellows from many disciplines. Since opening in October 2009 it has engaged in research around the “Lifelong Contextual Footprint” investigating novel and creative uses of personal data, balanced with the need to remain human in the digital age, by addressing social and personal issues such as trust and privacy. Social media has been one of many footprints investigated.

## Horizon response.

***What are the barriers to implementing real time data analysis? Is the new Government data-capability strategy sufficient to overcome these barriers?***

1. ”Social media data offers the possibility of studying social processes as they unfold at the level of populations as an alternative to traditional surveys or interviews.” A true statement for a number of social media systems in existence at this time. However, already many users have migrated from initial services that were open by default and difficult to configure for privacy to new technologies that are declared and implemented to provide more limited and comprehensible sharing (e.g. SnapChat, WhatsApp). The window for this form of analysis may only ever be open for a short time and provide a very limited demographic view.

***What are the ethical concerns of using personal data and how is this data anonymised for research?***

2. While social media data holds a great deal of promise for insight into civil issues, its use is ethically complicated. Government analysis of social media data is not different from academic research, and as such must adhere the same ethical strictures for government action to be legitimate. While there is a temptation to see social media data as fully public and therefore not deserving of protections given to human subjects in other research, this perspective is wrong - the use of any data that is potentially linkable to human beings, thereby making people identifiable, is potentially

privacy-invasive, and must be rigorously examined for its ability to harm them or violate their dignity.

3. Informed consent – Participation in research demands informed consent. Modern privacy thought stresses that, with regard to the use of personal data, context matters (Nissenbaum, 2010; Solove, 2006). The users of social media participate on various websites, fora and applications in the context of those services – for the purposes of sharing, communicating, shopping, entertainment, and so on. Those services were not created for research, and so government use of that data, especially without informed consent, is potentially a violation of the context of the original data collection and of the intent of the human sources.

4. Spectrum of private to public – Social media sites have a wide variety of privacy controls. As social media technology and business models advance, more nuanced and granular controls of data dissemination appear. For example, on Facebook, users' posts can be set to Public, Friends, Friends Except Acquaintances, Custom, and other categories. On Twitter, accounts can be Public or Locked, requiring user consent before her or his tweets are visible to the applicant, and the way tweets are formed determines who sees or doesn't see posts by default. These examples illustrate a spectrum of public to private, and that users' intentions about the appropriateness and sensitivity of their data can be expressed in a variety of ways. The implication of this spectrum is that *social media data must not be considered an undifferentiated mass of public data*, ripe for use without the need to account for the privacy expectations of users. This issue is amplified when one considers that a) users can retract posts or make formerly public posts private, indicating a clear intent for the use of their data, and b) social media sites change their terms of service frequently, often with inadequate notice to users, muddying the already opaque view into users' intentions.

5. Anonymity – A key characteristic of social media is its personal nature. As such, it follows that social media data is highly identifiable. Given this, and the inevitable sensitivity involved in the collection, analysis, comparison, compilation and dissemination of personal data, anonymizing social media data is an essential step in government use of it. The techniques used by the Census for many years would be a good place to start and adapt to this new environment.

6. Vulnerable populations – The inclusiveness of social media networks means that it is very easy to ingest the personal data of vulnerable populations in any research. These populations – including children, those fleeing domestic violence, marginalized social groups, and other at-risk

individuals – must be treated with higher degrees of care and more stringent safety procedures. The danger of potentially sweeping up children's data cannot be overstated. Given the above discussions of consent, anonymity and harm, government must understand that age verification on social media is flawed and ineffective. Existing methods for determining if a user is an adult or child are so weak as to cause any data set drawn from social media to be suspect of containing children's data.

7. Social inequality – It is important to statistically consider the representativeness of social media populations. Government analysis of social media data must account for bias within such data so as to ensure that its research does not rely on and amplify social and economic disparities. The whole of the UK population does not use social media, and so social media analytics should not be used to exacerbate the socioeconomic issues of the less digital parts of the citizenry.

8. Academic research of this form is overseen by ethics review panels that operate under published guidelines, with reviews undertaken by independent experts. It is recommended that government engage privacy scholars, security experts and data scientists external to the stakeholder groups undertaking research, to review research methods and safeguards. Sir David Ormand (Ormand) has identified principles that should apply even when the analysis is in pursuit of state security: 1) There must be sufficient sustainable cause; 2) There must be integrity of motive; 3) The methods used must be proportionate and necessary; 4) There must be right authority validated by external oversight; 5) Recourse to secret intelligence must be a last resort if more open sources can be found.

***What impact is the upcoming EU Data Protection Legislation likely to have on access to social media data for research?***

9. The forthcoming EU General Data Protection Regulation (GDPR) tightens the legal language mandating consent, requiring it to be explicit: "Consent should be given explicitly … enabling a freely given specific and informed indication of the data subject's wishes … ensuring that individuals are aware that they give their consent to the processing of personal data…. Silence or inactivity should … not constitute consent." (Preamble (26)); also: "Where processing is based on the data subject's consent, the controller should have the burden of proving that the data subject has given the consent to the processing operation." (Preamble (32)); and, the GDPR requires the ability for data subjects to meaningfully withdraw their consent at any time (Art. 7(3)).

10. Right to be forgotten – This is, in essence, a person's right to have data about her or him deleted upon request. "Any person should have the right to have personal data concerning them rectified and a 'right to be forgotten'…. In particular, data subjects should have the right that their personal data are erased and no longer processed, where the data are no longer necessary in relation to the purposes for which the data are collected or otherwise processed, [and] where data subjects have withdrawn their consent for processing…" (Preamble (53)); this right also includes "the obligation of the controller which has made the personal data public to inform third parties on the data subject's request to erase any links to, or copy or replication of that personal data." (Preamble (54)); the right has important restrictions, including that the "retention of the data should be allowed where it is necessary for historical, statistical and scientific research purposes, for reasons of public interest in the area of public health, for exercising the right of freedom of expression, when required by law or where there is a reason to restrict the processing of the data instead of erasing them." (Preamble (53)). It is clear that the term "necessary" in the above quote will require a great deal of elaboration and negotiation, and that this ambiguity is particularly salient to government social media research designs.

11. The GDPR mandates that data controllers shall provide easy access to held data in service of portability: "The data subject shall have the right … to obtain from the controller a copy of data undergoing processing in an electronic and structured format which is commonly used and allows for further use by the data subject." (Art. 18(1)). In the case that government social media research contains identifiable records this provision will apply.

12. The GDPR establishes a stronger bias against the potential harm of profiling than previous data protection policy."Every natural person should have the right not to be subject to a measure which is based on profiling by means of automated processing. However, such measure should be allowed when expressly authorised by law, carried out in the course of entering or performance of a contract, or when the data subject has given his consent. In any case, such processing should be subject to suitable safeguards, including specific information of the data subject and the right to obtain human intervention and that such measure should not concern a child." (Preamble (58)). This bias has particular salience for government plans to perform real-time (and thereby automated) analysis of social media data, especially in light of the previous discussion of the challenge of screening out children's data on social networks.

13. Privacy by design – The GDPR requires that personal data processing adhere to 'privacy by design' principles (Art. 23). At the moment, this

means a bias for only collecting minimally necessary amounts of data for a given collection purpose, and storing that data for the minimum necessary time. Further, the data controller "shall ensure that by default personal data are not made accessible to an indefinite number of individuals" (Art. 23(2)). The European Commission has reserved the right to amplify this provision via its delegates and to adopt technical measures to give the provision effect.

### *Is UK legislation surrounding the collection and use of data fit for purpose?*

Privacy advocates have long held that UK legislation regarding personal data has been weak in the following areas (although all of these will need to be reviewed given impending changes due to GDPR):

14. Data breach notification – Currently there is no legal requirement to notify regulators, the public or data subjects in the event of a breach. The Information Commissioner "believes serious breaches should be brought the attention of his Office," and has published guidance on what "serious" means as well as the ICO's possible reaction to a breach (ICO, 2012). Breach notification, explicit penalties and the potential 'naming and shaming' that might result are important features of a strong data protection regime.

15. Location privacy – Granular, large collections of location data are extremely revealing of one's personal activities. A month of location data from someone's mobile phone can answer the following questions: "Did you visit an STI clinic four times recently?", "Did you visit a lawyer several times?", "Were you at a political rally?", "Did you leave the city at night for five days in a row?" (Blumberg and Eckersley, 2009). The privacy-invasive potential of location data is immense, yet there is no primacy given to it in the UK DPA. The DPA has a category for "sensitive personal data," including racial, political, religious and sexual information, which triggers heightened protections for data collection and processing. Location data should be added to this list (Raper, 2010).

16. Consent revocation – The UK DPA contains no provisions for data subjects to revoke their consent to data processing. Consent is not meaningful without an ability to withdraw it.

17. Human Rights – Whatever UK DPA laws says, this still needs to be interpreted with the context of the European Court of Human Rights which may find existing and even future DPA law lacking, for example in policing, the need for processing to be according to clear rules and proportionate.

## References

*Article 29 Data Protection Working Party. (2013, May 13). Advice paper on essential elements of a definition and a provision on profiling within the EU General Data Protection Regulation. Retrieved from http://ec.europa.eu/justice/data-protection/article-29/documentation/other-document/files/2013/20130513_advice-paper-on-profiling_en.pdf*

*Barbaro, M. and Zeller, T. (2006, August 9). A Face Is Exposed for AOL Searcher No. 4417749. New York Times. Retrieved from http://www.nytimes.com/2006/08/09/technology/09aol.html*

*Blumberg, A. and Eckersley, P. (2009). On Locational Privacy, and How to Avoid Losing it forever. Electronic Frontier Foundation. Retrieved from https://www.eff.org/files/eff-locational-privacy.pdf*

*Buchanan, E. and Zimmer, M. (2012). Internet Research Ethics. Stanford Encyclopedia of Philopsophy. Retrieved from http://plato.stanford.edu/entries/ethics-internet-research/*

*Information Commissioner's Office [ICO]. (2012). Notification of data security breaches to the Information Commissioner's Office. Retrieved from http://ico.org.uk/for_organisations/guidance_index/~/media/documents/library/Data_Protection/Practical_application/breach_reporting.ashx*

*Kadushin, C. (2005). Who benefits from network analysis: ethics of social network research. Social Networks, 27(2), 139-153.*

*Mahon, P. (2013). Internet Research and Ethics: Transformative Issues in Nursing Education Research. Journal of Professional Nursing, Article in Press.*

*Nissenbaum, H. (2010). Privacy in Context. Stanford: Stanford University Press.*

*Ormand, Sir David (2010) Securing the State.*

*Raper, J. (2010). Data privacy in geographic information. Association for Geographic Information. Retrieved from http://www.agi.org.uk/storage/foresight/policy/Data%20privacy%20in%20geographic%20information.pdf*

*Solove, D. (2006). A Taxonomy of Privacy. University of Pennsylvania Law Review, 154(3), 477-560.*

*Sweeney, L. (2000) Simple Demographics Often Identify People Uniquely (Data Privacy Working Paper 3). Retrieved from http://dataprivacylab.org/projects/identifiability/paper1.pdf*

*Zimmer, M. (2010). "But the data is already public": on the ethics of research in Facebook. Ethics and Information Technology, 12(4), 313-325.*

*Written on April 2, 2014*