Methods Futures Briefing #003

Digital Footprint Management

By Robert Meckin[1] and Mark Elliot

This Methods Futures Briefing focuses on digital footprint management. It first explains what digital footprint management is. The following section describes possible developments. The briefing closes with an outline of potential implications for social research methods of such developments.

What is digital footprint management?

In essence, digital footprint management tools and methods aim to give data subjects more understanding about who holds and processes data about them, and control over what happens to that data, who it is shared with and how long for. Such tools can be regarded as a natural implication of data protection laws such the GDPR and embodies an approach to informational privacy that brings it closer to a norm of autonomous (as opposed to paternalistic) data protection.

Digital footprint management covers a range of technologies and practices. At the most basic level, this includes such things as privacy settings on browsers, platforms, and websites, the use of guest logins for online purchases, and anonymous browsers such as Tor[2].

More developmental examples include **personal data stores**[3], which act as single point of truth for the data about a person that is curated and controlled by the person themselves; and **digital footprint erasers**, services or software that find information about a person and cause it to be deleted.

A prominent exemplar of his idea is **Self-sovereign identity** (SSI) allowing individuals to control the identifying information they provide. The technology most often suggested as appropriate for SSI is blockchain, where public-key infrastructure secures the information cryptographically. Identifying information is encrypted and stored in a **distributed ledger system** such as the **blockchain**, where its use can be monitored and controlled by the individual, who only gives it out when consenting to its use. Personal data does not need to be perpetually held by organisations; rather, they can hold anonymised digital identifiers which can be authenticated using the transparency and immutability of blockchain.

Work on the automatic processing and classification of terms and conditions (e.g. Sadeh et al 2013, Braun et all 2023) is another plank in this. There have been several attempts to develop this as a solution to the tensions raised by check box privacy policy consent processes dating back to the Platform for Privacy Preferences of the late 1990s (see e.g., Grimm and Rossnagel 2000). Ultimately these have failed through a combination of the lack of enforceability and the poor usability of tools for users who have become accustomed to instant accessibility. However, an increase in the power of natural language processing will likely empower users as privacy policies become in effect transparent reducing the power of organisations to obfuscate their actions through opaque policies (See e.g., Hosseini et al., 2021; Adhikari et al., 2022).

Future developments

This area is a fast-moving mix of several different technologies, so futures work here is even more perilous than with most new tech spaces. However, it seems likely that further developments in the potential of digital footprint management are likely to be facilitated by the current wave of advances in AI.

---

[1] Contact author: Robert.meckin@manchester.ac.uk

[2] The last two are not specifically about giving control to the user in their specific operation but their availability makes options of being less visible to second and third parties.

[3] For extended definitions and discussion of the privacy implications of each of the emboldened terms see Elliot et al 2024.

For example, privacy avatars which 'sit alongside' a person online, manage person's data store and negotiate with other entities regarding the exchange of or access to data, according to the person's privacy preferences – perhaps constructed as their personal privacy policy - are likely to be a technological possibility soon. Search capacities of digital footprint erasers are likely increase – turning the notion of a digital trace on its head. So, the eraser is not just searching in a brute force way for instances of my personal information but for echoes of such data too (for example in inferences made about me).

Blockchain will continue to mature with the advent of blockchain-as-a-service, improvements in user interfaces and interchain interoperability being several likely developments. This in concert with AI will enable the possibility near-flawless personal authentication systems reducing the need for organisations to hold data about individuals.

Critical issues for social research

The potential impacts of digital footprint management on research are significant. If we imagine a world in which individuals are in full control of their personal data, then secondary data analysis is likely to become a thing of the past. All data analysis will become primary. Data might be analysed in situ with data subjects (or their avatars) giving **just-in-time consent**. Data will then be accessed live as needed but not stored by the researcher or indeed - in principle at least - by any other party. There are societal level questions about whether statutory right of access might apply in some instances, but these are out of scope for this briefing!

Issues of **data quality and integrity will change** although not disappear completely. On the one hand if a distributed ledger system is used to store an individual's personal life record then issues like recall bias will become non-problems; the trace of the record will be permanent and although changes can be made these will effectively in the form of timestamped updates, so if your research question relates to the past the correct information for any given time point will be available if the individual chooses to share it.

However, new issues will arise, the question - as the physical and digital interact more deeply and even merge – of what someone's identity is becomes critical. The possibility of humans having **multiple identities** without any one being primary throws up fundamental epistemological questions for the data driven social sciences. Such concerns have a long tradition in some areas of social science and psychology (see e.g. Kurzweilly 2019, Burke, 2020). However, this is rarely embraced in data focused empirical work where the notion of singular identity has been important in for example tracing entities over time in longitudinal analysis.

An important sidebar is that such developments have significant educational, political, legal, economic and infrastructural implications. A world in which individuals were in control of the data about them also implies that those individuals would take on a new level of responsibility for that information. What is the status in such a world of individuals who lack capacity (or incentive) to take that responsibility? What does this imply for a digital economy where the capacity of organisations to monetise customer information is priced into the costs of goods and services? For the general populace such a world implies an education system focused on informational citizenship and a model of privacy which is about individual empowerment rather than the simple "right to be let alone" which was the first modern era attempt to define privacy (Warren and Brandeis 1890).

This in turn implies an information literate populace. It is difficult to anticipate what impact this would have on response/participation rates. The reduction of burden may mitigate the increasing reluctance of citizens to participate but citizens control of their own information may lead to a change in the psychology of participation. This would likely require a different relationship between the researcher and the researched in social science, with participatory research perhaps become normalised.

In principle this world could throw into sharp relief longstanding sociological question about the relationship between structure and agency and this can only affect how we think about methodology too. If an individual is truly the sovereign of their own digital domain, then the 'story of them' surely might supplant or perhaps fully embed mere longitudinal data.

References

Adhikari, A., Das, S., & Dewri, R. (2022). Privacy Policy Analysis with Sentence Classification. 2022 *19th Annual International Conference on Privacy, Security & Trust*, 1-10.

Braun, D. and Matthes, F., (2021). NLP for consumer protection: Battling illegal clauses in German terms and conditions in online shopping. *In Proceedings of the 1st Workshop on NLP for Positive Impact*. 93-99. https://doi.org/10.18653/v1/2021.nlp4posimpact-1.10.

Burke P. Identity. *In*: Kivisto P, ed. *The Cambridge Handbook of Social Theory*. Cambridge University Press; 2020:63-78. https://doi.org/10.1017%2F9781316677452.005.

Alessi, M., Camillò, A., Giangreco, E., Matera, M., Pino, S., & Storelli, D. (2018). Make Users Own Their Data: A Decentralized Personal Data Store Prototype Based on Ethereum and IPFS. *In: Proceedings of 3rd International Conference on Smart and Sustainable Technologies*, 1-7.

Elliot, M., Mandalari, A-M, Mourby, M., and O'Hara, K., (2024, forthcoming). *Dictionary of Privacy, Data Protection and Information Security*, London: Elgar. https://www.e-elgar.com/shop/gbp/dictionary-of-privacy-data-protection-and-information-security-9781035300914.html.

Grimm, R. and Rossnagel, A., (2000). Can P3P help to protect privacy worldwide? *In: MULTIMEDIA '00: Proceedings of the 2000 ACM Workshops on Multimedia*, New York: ACM, 157-160, https://doi.org/10.1145/357744.357917.

Howarth, J. (2024). 10 Important Blockchain Trends (2024-2027). https://explodingtopics.com/blog/blockchain-trends.

Kurzwelly, J., (2019). "Being German, Paraguayan and Germanino: Exploring the Relation Between Social and Personal Identity". Identity. 19 (2): 144–156. https://doi.org/10.1080/15283488.2019.1604348.

Hosseini, H., Degeling, M., Utz, C., & Hupperich, T., (2021). Unifying Privacy Policy Detection. *In: Proceedings on Privacy Enhancing Technologies*, 2021, 480 - 499. https://doi.org/10.2478/popets-2021-0081.

Mansour, E., Sambra, A.V., Hawke, S., Zereba, M., Capadisli, S. Ghanem, A., Aboulnaga, A. and Berners-Lee, T., (2016). A demonstration of the solid platform for social web applications. In: *Proceedings of the 25th international conference companion on world wide web* 223-226.

Papadopoulou, E., Stobart, A., Taylor, N.K. and Williams, M.H., (2015). Enabling data subjects to remain data owners. In: *Agent and Multi-Agent Systems: Technologies and Applications: 9th KES International Conference, KES-AMSTA 2015 Sorrento, Italy, June 2015, Proceedings* 239-248. Springer International Publishing.

Preukschat, A. and Reed, D., eds., (2021). *Self-sovereign Identity*. Shelter Island, NY: Manning Publications.

Sadeh, N., Acquisti, A., Breaux, T.D., Cranor, L.F., McDonald, A.M., Reidenberg, J.R., Smith, N.A., Liu, F., Russell, N.C., Schaub, F. and Wilson, S., (2013). The usable privacy policy project. *Technical Report, CMU-ISR-13-119*. Carnegie Mellon University.http://reports-archive.adm.cs.cmu.edu/anon/anon/home/ftp/isr2013/CMU-ISR-13-119.pdf.

Warren, S.D. and Brandeis, L.D., 1890. The Right to Privacy. Harvard Law Review, 4(5),193-220. https://heinonline.org/HOL/Page?handle=hein.journals/hlr4&div=31&g_sent=1&casa_token=&collection=journals.

The authors would like to thank Kieron O'Hara for his constructive comments on a draft. Any errors, omissions and decision-making are the authors' responsibility.

If you would like to contribute a Methods Futures Briefing to the series, or would like to give feedback, please get in touch by emailing Robert.meckin@manchester.ac.uk.

National Centre for Research Methods
Social Sciences
University of Southampton
Southampton, SO17 1BJ
United Kingdom.

Web http://www.ncrm.ac.uk
Email info@ncrm.ac.uk
Tel +44 23 8059 4539
Twitter @NCRMUK