Section A

Institution: The University of Manchester

Unit of Assessment: 21 (Sociology)

Title of case study: Ensuring more effective and safer data sharing practices through the Anonymisation Decision Making Framework

Period when the underpinning research was undertaken: 2002 – 2020

Details of staff conducting the underpinning research from the submitting unit:

Name:	**Role(s) (e.g. job title):**	**Period(s) employed by submitting HEI:**
Mark Elliot	Professor of Data Science	1996 – present
Elaine Mackey	Research Information Governance Manager; Research Associate	2018 – 2020; 2011 – 2018

Period when the claimed impact occurred: 2014 - 2020

Is this case study continued from a case study submitted in 2014? No

Section B

1. Summary of the impact

Research at the University of Manchester into data anonymisation (conducted with partners including the Office of National Statistics, the UK Information Commissioners Office and the Open Data Institute) and the subsequent development of the *Anonymisation Decision-Making Framework* (ADF), has:

1. informed data anonymisation policies and practice within organisations (including Government departments and agencies, businesses, financial services and research and development institutions) who have engaged with the research and the ADF; and
2. ensured the confidentiality of individual data subjects is protected whilst data is used for the optimum economic and social benefit, enabling the sharing of data with confidence, and compliance with data protection regulations.

2. Underpinning research

Extensive research on anonymisation led by Elliot since 2002, has culminated in a new approach to data anonymisation called *functional anonymisation* which is described fully in [1]. The key organising concept is that the question of whether data are anonymised or not (and therefore personal or not) cannot be decided by simply considering the data alone. Risk lies in the relationship between data and the *data environment,* defined in [2,3] as: *“the set of formal and informal structures, processes, mechanisms and agents that either: act on data; provide interpretable context for those data; or define, control and/or interact with those data.”*

Research and engagement with external organisations have focused on the development of a methodology for the delivery of functional anonymisation - the *anonymisation decision-making framework (ADF)* [4], which is the result of a constant and ongoing interplay between original research in anonymisation methodology and the development of best practice. The new methodology builds on earlier statistical and computational work in the field of *statistical disclosure control* research supported by a number of grants from the Office of National Statistics (ONS), the ESRC and the EPSRC between 2002 and 2006. Elliot’s earlier research demonstrated that anonymisation is necessarily an interdisciplinary activity which combines legal analysis [5,6], statistics and computer science but also incorporates elements of criminology and management science. His more recent research from 2004 - 2014 incorporates socio-technical work on *data environment analysis* [i,ii]. The interplay between research and impact is most strongly represented in output [6], an academic article derived from the report from a consultancy project carried out with ONS. The report utilised the ADF’s comprehensiveness principle to assist ONS in its decision making about open data resources and involved applied research at the intersection of legal and technical analysis. The academic article represents the most thorough development of motivated intruder testing (a branch of ethical hacking where data rather than security systems are attacked) published to date.

In sum, the ADF [4] which has resulted from Elliot's programme of research, sets out the methodological basis for a system of thinking on data anonymisation policies and practice which enables clear decision-making by individuals and organisations.

3. References to the research

[1] **Elliot. M. J**., O'Hara, K., Raab, C., Dibben, C., **Mackey, E.,** Gowans, H., O'Keefe, C., Purdam K. and McCullagh, K. (2018). 'Functional anonymisation: Personal data and the data environment', *Computer Law & Security Review,* 34(2), 204-221. https://doi.org/10.1016/j.clsr.2018.02.001

[2] **Elliot, M. J**., Lomax, S., **Mackey, E.** and Purdam, K. (2010). 'Data Environment Analysis and the Key Variable Mapping System', In Domingo-Ferrer, J. and Magkos, E. (eds.) *Privacy in Statistical Databases.* Springer: Berlin. https://doi.org/10.1007/978-3-642-15838-4_13

[3] **Elliot, M. J**,. Smith, D., **Mackey, E.** and Purdam, K. (2011). 'Key Variable Mapping System II' *Proceedings of UNECE work session on Statistical Confidentiality, Tarragona*, October 2011.

[4] **Elliot, M.J**., **Mackey, E**., O'Hara, K. M. and Tudor, C. (2016). *The Anonymisation Decision Making Framework*. UKAN publications; Manchester

[5] Mourby, M., Wallace, S., **Mackey, E., Elliot, M. J**., Gowans, H., Bell, J., Smith, H., Aidinlis S., and Kaye. J. (2018). 'Are 'pseudonymised' data always personal data? Implications of the GDPR for administrative data research in the UK', *Computer Law & Security Review,* 34(2), 222-233. https://doi.org/10.1016/j.clsr.2018.01.002

[6] **Elliot, M. J**., **Mackey, E.,** O'Shea S., Tudor, C. and Spicer, K. (2016). 'End user licence to open government data? A simulated penetration attack on two social survey datasets', *Journal of Official Statistics*, 32(2), 329–348. https://doi.org/10.1515/JOS-2016-0019

Related Grants:

[i] 2008 - 2011 Elliot M, & Purdam K; Data Environment Analysis Service ONS GBP101,007

[ii] 2012 - 2013 Elliot M, Purdam K, & Mackey E; EUL - OGL Phase 1 ONS GBP9,800; 2013 - 2014 Elliot M, Purdam K & Mackey E; EUL - OGL Phase 2 ONS GBP33,499

[iii] 2012 - 2014 Elliot M, Shadbolt, N, O'Hara K & Spicer K. UK Anonymisation Network ICO GBP26,607

[iv] 2014 - 2015 Elliot M, Tennison, J O'Hara K & Spicer K; Building Anonymisation Capacity Cabinet Office GBP64,080

[v] 2016 - 2016 Tudor C. Elliot M & Tennison J; Quality Improvement fund grant GSS GBP15,700

[vi] 2017-2018 Elliot M, & Mackey E; Cross-sector Engagement on Data Anonymisation HEFCE GBP16,300

4. Details of the impact

The ADF has informed and shaped data anonymisation policies and practice within organisations, ranging from Government departments and agencies, to financial services, and research and development institutions. In turn, the adoption of policies informed by the ADF have ensured the confidentiality of individual data subjects is protected, whilst enabling the corresponding data to be used for the optimum economic and social benefit. Use of the ADF enables the sharing of data with both ***confidence*** (meaning that it is able to flow more easily to where it is needed), and ***compliance*** with data protection regulations (meaning it is shared ethically and the privacy of data subjects is better protected).

Pathways to Impact

The UK Anonymisation Network (UKAN – www.ukanon.net) is a cross sector network led by Elliot and Mackey. UKAN delivers work that is positioned at the research and practice interface, and has 600 member organisations from all sectors of the UK economy. Through a series of large focus groups (cross-sectoral, interdisciplinary one-day workshops) run by UKAN over eighteen months in 2013 and 2014, workshop participants were given one aim: to answer the question "what is anonymisation?" The ADF is essentially the answer to that question and necessarily covers technical, legal, and ethical elements of the problem. One of the key outcomes has been harmonising the legal and technical language in this area.

The key dissemination vehicle is the ADF book [4]. In 2017, UKAN worked with Australian academics and their information commissioner's office to produce an edition of the book customised for Australian Law. In 2020 the second edition of the UK ADF was published. In total, across all versions, the book has been downloaded over 18,000 times (not including downloads where 3rd parties have made the ADF available via their own websites). The UKAN website also provides open access to the research outputs (www.ukanon.net). In addition, UKAN works with organisations and individuals to apply the ADF directly, running a clinic service through which any organisation receives a free hour of discussion based on completing the ADF template. UKAN also runs one-day workshops with organisations. These are intensive sessions where collaboration with the client organisation is focussed on building a model of the specific problem that they are dealing with. So far, this has been used with GOVTECH (Singapore), Roche, the Department for Education, HESA and Flatiron (US/UK). UKAN also provides paid-for consultancy services inputting on specific technical anonymisation issues faced by the focal organisation.

Impact

The ADF has improved the practice of businesses and organisations across all sectors [A]. For example, the Chief Methodologist at Privacy Analytics (a Canadian company providing software system solutions for managing data confidentiality) outlines the impact of the ADF on its practices: "*As an organization that works internationally across all sectors, we have used elements of the ADF to inform our own methods and practices. We reference the ADF in most of our work, as a fundamental piece of guidance in the practice of responsibly using and sharing data"* [A]. The Research and Data Governance Lead at retailTRUST (a third sector organisation providing mental health support services to workers in the retail sector) also states that they used the ADF, *"to inform our design of a process within our CRM to strip records of identifiable data. The ADF also helped to challenge more effectively the blanket statements made by some suppliers' about their complete anonymisation of shared data"* [A].

Consultation with various organisations in the development stages of the ADF has yielded substantial benefits to the organisations concerned, illustrated by the following examples.

1. Office for National Statistics (ONS): In 2014, the ONS approached UKAN to help resolve the question of how to continue allowing academics to access its census and survey datasets while moving away from bespoke end user license agreements, to embrace open access in the form of an open government data license. ONS Head of Statistical Disclosure Control Methodology, Keith Spicer, states that, "*the ADF was very relevant to the decisions that we needed to make as it - and its underpinning concept of functional anonymisation - was precisely designed to assess the risk of moving data from one environment to another"* [B]. Collaborative work between UKAN and the ONS demonstrated that open publication of the data under the initial end user license agreement carried a significantly heightened confidentiality risk. The work combined legal and technical analyses, and was structured by (and fed into the refinement of) the ADF. The results were published as an academic paper in 2016 [6], jointly authored with Manchester and ONS staff. As a result of this work ONS **"***decided not to release the data in an open environment - so the impact was the avoidance of* *a potentially disastrous breach of data protection law*" [B]. Such a breach "*would carry sanctions of up to a two year prison sentence for individuals releasing personal information. For the wider ONS, public trust in our statistics is fundamental and such a breach (and no doubt its negative publicity) could have had serious consequences for public compliance in surveys and other collections. In turn, this would have lessened the quality of our statistical outputs for the public good and affected the esteem in which they are held*" [B].

2. Roche Pharmaceuticals: Roche approached UKAN for use of their clinic service in 2016. In common with other pharmaceuticals, Roche is dealing with a regulatory requirement to share detailed information on its clinical trials with the European Medicines Agency (who then publish that information) in combination with legal requirements to maintain data subject confidentiality. As well as concerns about its data subjects' privacy, Roche is concerned about damage to drug development if participation rates in clinical trials drop in response to a breach [C], not least because under the General Data Protection Regulation it could be fined up to 4% of its GBP44,000,000,000 global turnover for breaches arising from negligence.

To solve this problem, a team of 10 Roche staff (from the UK and Swiss HQs) came to Manchester for an ADF workshop hosted by Elliot and Mackey. The purpose was to use the framework to undertake an end-to-end mapping of the multinational company's data situation. The outcome was that Roche commissioned UKAN to carry out a two-phase consultancy project with the specific task of equipping it with a more principled methodology for assessing its confidentiality risks. Phase 1 (2017) carried out an audit of the data situation and developed a set of risk scenarios. Phase 2 (2018-19) has developed a risk classification tool for clinical study reports. Katherine Tucker, Data Privacy Technical Lead at Roche Products Ltd. the UK arm of Roche Pharmaceuticals, states that "*UKAN reassured us that we were right to be concerned, providing confirmation that our somewhat cautious approach was justified"* [C]. Use of the ADF *"has enabled Roche to further develop its strategies and thinking on anonymization topics and ultimately continue to responsibly share clinical trial data and documents with researchers and the general public"* [C].

3. The UK Information Commissioners Office (ICO): From UKAN's inception, it has worked in partnership with the ICO in a mutually beneficial relationship; the ICO consults UKAN on technical anonymisation matters and UKAN in turn refers to the ICO on regulatory matters. As the ICO's Paul Comerford says: "*Alignment between the ADF and ICO guidance is regarded as important by both sides. External stakeholders will closely follow ICO guidance to ensure they are compliant with the GDPR, therefore, the ADF will provide a valuable practical guide to implementation. …The ADF serves as a practical, best practice guide to anonymisation*" [D].

The importance of this relationship is further underlined in the recommendations and policies of external organisations, which reference both the ADF and ICO guidance as valuable resources. For example, in its response to the National Data Guardian for Health and Social Care's Review of Data Security, Consent and Opt-Outs in 2016, the Wellcome Trust *"…commend the* [ADF] *to the Department of Health for a clear, comprehensive account of what needs to happen to data and the environment in which is it used in order for it to be considered 'anonymised' information. This a valuable resource alongside the ICO's Anonymisation Code of Practice for providing technical detail on anonymization."* [E]. The MRC's guidance on using information about people in research states that *"We believe that managing the risk of (re)identification is fundamental to balancing privacy and conducting research, and we highlight resources which describe how to do this (Information Commissioner's Office Anonymisation Code of Practice and UK Anonymisation Network decision-making framework)"* [F]. Similarly, City University of London's policy on data retention instructs that *"All personal information must be encoded or anonymised as far as is possible and consistent with the needs of the study, and as early as possible after collection.*

Guidance is available on the Information Commissioner's Office website and via the UK Anonymisation Network (UKAN) and their Anonymisation Decision-Making Framework" [G].

4. Flatiron Health UK Limited

Flatiron approached UKAN in 2019 to carry out a data situation audit – one of the key tools developed to implement the ADF. This was in fact the first application of the second edition of the ADF, the results of which fed into the fine tuning of the new framework. Flatiron's business consists of structuring the identifiable health records of cancer patients into high-quality, longitudinal, tumour-type specific datasets, which are then shared with third-party oncology researchers. The long-term benefits of such data sharing in terms of accelerating improvements of cancer care are well established but attempts to do so in the UK have floundered, due to a mixture of ill thought-out approaches to anonymisation, poor public engagement, and the difficulties of resourcing the pre-processing necessary to render the relevant datasets suitable for research purposes. As Lauren Brown, Director of Business and Product Development at Flatiron states, "*use of the ADF allowed us to map the problem sufficiently so that we could progress. This led to a piece of novel work called a data situation audit which was an in depth and thorough analysis of our intended processing of the cancer care data. The output of this audit – an analysis, and targeted set of recommendations – has formed the foundation of our approach to anonymisation in the UK*" [H]. Flatiron's well-established process enables it to overcome the technical processing issues and UKANs role was to support it in developing its approach to anonymization, which is both GDPR compliant and ethical [H].

The influence of UKAN and the ADF goes beyond the framework's direct usage, with other organisations advocating its approach. For example, in 2016 the House of Commons Science and Technology Committee recognised the importance of anonymisation, and the work of UKAN. In its report on "The Big Data Dilemma", the Committee recommended that *"The Government should set out its anonymisation strategy for big data in its upcoming Digital Strategy, including a clear funding commitment, a plan to engage industry with the work of the UK Anonymisation Network and core anonymisation priorities"* [I].

5. Sources to corroborate the impact

[A] Web sources and survey results evidencing the use and onward dissemination of the ADF (January 2021).

[B] Testimonial from Head of Statistical Disclosure Control Methodology, Office for National Statistics. Received 25 November 2020.

[C] Testimonial from Data Privacy Technical Lead, Roche Products Ltd. Received 14 October 2020.

[D] Testimonial from Principal Technology Adviser, UK Information Commissioners Office. Received 26 November 2020.

[E] Wellcome Trust (2016). Response to the National Data Guardian for Health and Social Care's Review of Data Security, Consent and Opt-Outs. Available at https://wellcome.org/sites/default/files/NDG-review-data-secrity-consent-Sep16.pdf

[F] Medical Research Council (2017). Guidance on using information about people in research. Available at https://mrc.ukri.org/documents/pdf/using-information-about-people-in-health-research-2017/

[G] City University of London's policy on data retention. Available at https://www.city.ac.uk/__data/assets/pdf_file/0008/417707/Storage-and-destruction-of-research-records.pdf

[H] Testimonial from Director, Business and Product Development, Flatiron UK. Received 17 November 2020.

[I] House of Commons Science and Technology Committee (2016). *The big data dilemma*. *Fourth Report of Session 2015–16*. Available at https://publications.parliament.uk/pa/cm201516/cmselect/cmsctech/468/468.pdf