

<p><b>Institution: The University of Manchester</b></p>
<p><b>Unit of Assessment: 23 (Sociology)</b></p>
<p><b>Title of case study:</b>  <b>Impact on the Statistical Confidentiality Practices of Data Stewardship Organisations</b></p>
<p><b>1. Summary of the impact</b></p> <p>Research at the University of Manchester (UoM) has developed new approaches, methods and algorithms to improve the statistical confidentiality practices of data stewardship organisations (DSOs), such as the UK’s Office for National Statistics. The research and its products have had significant impacts on data dissemination practice, both in the UK and internationally, and have been adopted by national statistical agencies, government departments and private companies. The primary beneficiaries of this work are DSOs, who are able to both disseminate useful data products, and protect respondent confidentiality more effectively. Secondary beneficiaries are respondents, whose confidentiality is better protected, and the research community, as without ‘gold standard’ disclosure risk analysis, data holders can be overcautious.</p>
<p><b>2. Underpinning Research</b></p> <p>The underpinning research on disclosure risk analysis was carried out at the Cathie Marsh Centre for Census and Survey Research (CCSR) at UoM (1996-); led by Dr Mark Elliot (Senior Lecturer, School of Social Sciences). The Manchester team also includes: Dr Kingsley Purdam (Research Fellow, 2003-), Dr Elaine Mackey (Research Associate, 2010-), Dr Duncan Smith (2003-2008, now Honorary Research Fellow), and Susan Lomax (Research Assistant, 2006-2009).</p> <p>‘Statistical disclosure’ occurs when, through statistical matching, a population unit is identified within an anonymised dataset and/or new information about them is revealed. This compromises the privacy of the individuals within the data, breaks data protection legislation and results in reputational and other damage for those responsible for the data. Statistical Disclosure Control (SDC) concerns the prevention of such events. <b>Research at UoM has helped to remodel the way that disclosure risk analysis is carried out, delivering innovations in both statistical methods, and in the whole framing of the problem.</b></p> <p>The distinctive nature of the research contribution flows from the insight that the SDC field was <i>data-centric</i> (focusing on the statistical properties of the data to be disseminated) when it needed to be <i>intruder-centric</i>, focusing on the means, motives and opportunities (MMO) of individuals or organisations who might wish to attack the data. This led to a body of work <b>developing the notion of scenario analysis</b> [E]. By creating pragmatically based MMO models for any given dataset it is possible to look at what the intruder would be trying to achieve, and then generate metrics which measure the risk of them achieving it. This framework is now widely regarded by DSOs as an important part of disclosure risk assessment. This scenario-based framework led in turn to significant <b>methodological innovations</b>:</p> <ul style="list-style-type: none"> <li>• <b>Development of the Data Intrusion Simulation (DIS) measure of risk for microdata [D] and of SUIM (the Special Unique Identification Method) [C].</b> The two metrics were <b>embodied in software called SUDA</b>, developed in collaboration with colleagues in computer science, with EPSRC funding. The software has been adapted for use by several statistical agencies in the UK, the US, Australia, New Zealand and Singapore, in determining outputs from census and surveys.</li> <li>• Parallel to this, <b>the research has extended the notion of scenario analysis</b> [E]. This extended framework proposes that disclosure risk arises from the relationship between the data and its environment (taking into account the additional, external information that might be available to a would-be intruder), raising the issue of how to capture/measure the environmental facet of the risk. This <b>led to the framework of Data Environment Analysis (DEA)</b>, a set of methods for measuring the environmental aspects of the risk including ‘key variable mapping’ [A]. Until this work, assessment of disclosure risk was only measuring one side of the problem. The new framework now enables these additional risks to be formally</li> </ul>

principled. On the basis of this work, the Office for National Statistics (ONS) commissioned and funded the Data Environment Analysis Service, (DEAS, 2008-2012 at UoM), to guide the specifications of the 2011 Census outputs (2012 Annual Report, available upon request).

**3. References to the research** (all references available upon request)

- [A] (2011) Elliot, M., Lomax, S., Mackey, E. & Purdam, K. "Data Environment Analysis and the Key Variable Mapping System" in Domingo-Ferrer, J. & Magkos, E. (eds.) *Privacy in Statistical Databases* (Springer: Berlin)
- [B] (2008) Smith, D. & Elliot, M. "A Measure of Disclosure Risk for Tables of Counts" *Transactions in Data Privacy* 1(1) 34-52 (REF 2014)
- [C] (2002) Elliot, M. J., Manning A. M. & Ford, R. W. "A Computational Algorithm for Handling the Special Uniques Problem". *International Journal of Uncertainty, Fuzziness and Knowledge Based Systems* 5(10) 493-509
- [D] (2002) Skinner, C. J. & Elliot, M. J. "A Measure of Disclosure Risk for Microdata" *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 64(4) 855-867 (RAE 2008)
- [E] (1999) Elliot M. & Dale, A. "Scenarios of Attack: The Data Intruder's Perspective on Statistical Disclosure Risk" *Netherlands Official Statistics* (Spring) 6-10

**Grants:** Over the past 13 years, Elliot has been involved in over £1.58M (Manchester Share) of project funding, from: ESRC, EPSRC, MRC, EU and ONS. Key grants include:

- (2013-18) Administrative Data Service (£6.18m; Manchester £518K, Co-Director)
- (2008-13) ESRC: 'Administrative Data Liaison Service' (£850K; Manchester £100K, Co-I)
- (2012-13) Office for National Statistics (ONS): EUL->OGL (£30K, PI)
- (2005-8) MRC: 'Clinical E-science Framework II' (£4.5m. Manchester £2.55m, Co-I)
- (2003-7) EPSRC: 'High Performance Computing and Statistical Disclosure Control' (£260K, Co-I)
- (2001-3) EU: 'Computational Aspects of Statistical Confidentiality' (Total ~2.2M ECU; Manchester 172K ECU, Partner)

**4. Details of the impact**

**The key impact of this body of research is the reduction of likelihood of personal information disclosures from research data. The immediate beneficiaries are national and international data stewardship organisations (DSOs),** who have responsibility for disseminating data whilst at the same time maintaining confidentiality. The reach of this impact has increased in recent years from national statistical agencies (UK and overseas) to include government departments – such as the Department for Work and Pensions (DWP) and the Department for Communities and Local Government (DCLG) who are under increasing pressure to release administrative data as part of a wider 'open data' agenda.

**Impact has been delivered through collaborative research partnerships with practitioners and professional services in data stewardship.** This research, consultancy and advisory work in relation to data management and confidentiality has helped to shape professional best practice in the field. The body of work described here is impact-embedded research. An example of this is [B], which describes an algorithm developed whilst analysing the risk inherent in the Neighbourhood Statistics – work carried out for ONS several years earlier. Overall, impact has been delivered through a variety of pathways:

- 1. Impact through DSOs adopting methods into their practice.** The Canadian firm Privacy Analytics Inc. has utilised [D]: *"as part of the general heuristic for re-identification risk assessment in our software, [with Elliot's] work on scenarios of attack (1999) and SAR data (2001, 2007) helping shape the methodology that is being used by the company. Privacy Analytics, Inc.'s solutions have been leveraged by approximately 70 organizations in the US and Canada to control their disclosure risk, including: the American Society of Clinical Oncology, Heritage Provider Network, Icahn School of Medicine at Mount Sinai, Department of Preventative Medicine, State of Louisiana Department of Health and Hospitals, Alberta Healthcare, Cancer Care Ontario, Children's Hospital of Eastern Ontario, College of Family*

*Physicians of Canada, Public Health Agency of Canada, TELUS Health, and Vancouver Coastal Health*” [1]. Likewise, after reading articles and presentations of the research at meetings, the Australian Bureau of Statistics noted that: *“Your work helped shape ABS’s current thinking on confidentiality protection for microdata... [and has] assisted the ABS to produce high quality statistics while maintaining provider trust”* [2]. Specifically on [D]: Statistics Canada say: *“The Skinner-Elliot measure, adapted for survey weights by Skinner & Carter, is often used by our surveys to provide an indication of the risk inherent in certain combinations of their variables. The measure has been incorporated in our in-house disclosure risk generalized program”* [3].

On documents [C] and [E] ONS say: *“Previous work carried out by Dr Elliot continues to have an impact on the way we conduct our assessment of risk at ONS [and] ONS occasionally use the Skinner and Elliot (2002) metric to assess microdata releases which come to us for approval via our Microdata Release Panel and in particular a significant contribution consistently used is Dr Elliot’s concept of special uniqueness involving the identification of uniques that stand out regardless of geography. Many of his research ideas have been fundamental in informing GSS Disclosure Control Policy for Social Survey Microdata”* [4]. Further afield, the US Census Bureau contends that the research has *“had a significant impact on the way in which disclosure control is carried out in our organisation”*, and that they have adopted the concept of special uniques in their reidentification studies which: *“has helped us to identify public use microdata records that are at risk of disclosure. We are then able to slightly distort those records so that we can release them while preserving the statistical properties of the data”* [5].

2. Based on the principles developed in [A] and [E], Elliot set up the **Data Environment Analysis Service** (DEAS) with funding from ONS, to guide the specifications of the 2011 Census outputs. This informed ONS’s own disclosure risk analyses and policy, with their Head of Statistical Disclosure confirming that it: *“greatly assisted ONS in performing day-to-day disclosure risk assessments of our microdata and knowing under what conditions they should be released. It also informed the specification of the Sample of Anonymised Records (microdata from the 2011 Census)”* [4].
3. **The provision of consultancy reports to DSOs, evaluating the disclosure risks from specific datasets (to be disseminated) or analysing the risk impact of particular policy decisions.** Commissioned on the basis of the expertise and track record of Dr Elliot and his team, these evaluations (‘Report on the Disclosure Risk associated with Fire and Rescue Incident Datasets’, ‘Report on the Disclosure Risk associated with Supporting People Datasets’, ‘Disclosure Risk Audit of the Smart Steps Data Service’) directly apply the research reported above. For example, the report to *Telefonica* (‘Disclosure Risk Audit of the Smart Steps Data Service’) highlighted the potential risks of providing information at multiple units of geography, and the need to suppress small counts in remote areas to ensure individuals cannot be identified through reverse engineering [6]. These reports have enabled DSOs to make better decisions about the data in question, resulting in safe and usable data. For example, DCLG confirms that: *“[the] work showed that the database which I provided was far from anonymised in practice”* and in turn this: *“...showed that we needed to proceed with access to the database in a secure setting and/or under data access conditions”* [7]. Similarly, ONS note that: *“The outcome of [‘EUL to OGD: A Risk Analysis’] will have significant implications on the way ONS will provide data to its customers in the future, and helps ONS to meet government strategic aims in proactively supporting the Open Data Agenda”* [4].
4. **The provision of dedicated SUDA software** to DSOs to help them carry out of their own disclosure risk analyses [8]. Several DSOs, such as the Australian Bureau of Statistics and the Singapore Department of Statistics, have incorporated the software into their data dissemination practice. For example, the Australian Bureau of Statistics confirms that: *“Since 2007 SUDA has been a part of the standard assessment process for all CURFS [Confidentialised Unit Record Files] released by the ABS. In particular SUDA is used to help identify risky records. Generally, as a result of the SUDA analysis, a number of records on each CURF are given additional confidentiality treatment to mitigate the identification risk”* [2].

**5. The provision of certificated safe researcher training** delivered via the Administrative Data Liaison Service (ADLS), a service set up to support research using administrative data ([www.adls.ac.uk](http://www.adls.ac.uk)). The primary goal of the training is to educate participants in dealing with data safely, and is attended by DSO staff as well as researchers. The course is part of the Safe Researcher Training Programme and is endorsed by major administrative data holding organisations in the UK and the Information Commissioner's Office. By the end of 2013 this course will have run 14 times, each attended by 20-50 participants. The research is embedded throughout the course and is the basis of the most important module of the course. **The course is an approved method for access to administrative data held by the Scottish government and NHS Scotland.**

**6. The development of the UK Anonymisation Network (UKAN), set up to establish best practice in anonymisation.** Led by Elliot and funded by the UK Information Commissioner, UKAN offers practical advice and information to anyone who handles personal data and needs to share it ([www.ukanon.net](http://www.ukanon.net)). The network provides a range of resources to private and public sector data holders in the UK, including **anonymisation clinics** (where individual DSOs attend to receive guidance on specific data anonymisation problems), and a web site providing **case studies of disclosure control** in practice, including several derived from UoM research. In a joint press release, the UK Information Commissioner, said: "*The work of UKAN will help build on the recommendations laid down in the ICO's data protection code of practice on managing the risks related to anonymisation which we published last year*" [9].

The information economy requires the free flow of data, but privacy breaches lead to a breakdown of trust, respondent non-co-operation and DSO over-cautiousness. Armed with the tools and methods developed from this research, DSOs can proceed with greater confidence, and make more informed data release decisions. As ONS state: "*The long-term impact of the reports above and ongoing collaborative work with Dr Elliot has meant that ONS can demonstrate a rigorous and considered approach to data dissemination given a fuller understanding of the potential disclosure risk. This helps ONS to achieve its vision to produce high quality trusted statistics that meet user needs.*" [4] Similarly, the US Census Bureau agrees that: "*The broader impact of this work is to help ensure that we can disseminate high quality data products whilst maintaining the confidentiality of respondents*" [5].

In sum, without this work delivery of the open data agenda would be significantly hampered.

#### **5. Sources to corroborate the impact** (all claims referenced in the text)

- [1] Testimonial from Chief Executive Officer, Privacy Analytics, Inc. (29<sup>th</sup> August 2013)
- [2] Testimonial from Chief Methodologist, Australian Bureau of Statistics (27<sup>th</sup> February 2013)
- [3] (2013) Report and Covering Document from Statistics Canada (18<sup>th</sup> February)
- [4] Testimonial from Head of Statistical Disclosure Control Branch, Office for National Statistics (10<sup>th</sup> April 2013) & (2013) Elliot, M. & Mackey, E. 'EUL to OGD: A Risk Analysis (phase 1) Report' (August)
- [5] Testimonial from Chair, Disclosure Review Board, US Census Bureau (6<sup>th</sup> June 2013)
- [6] (2013) Elliot, M. 'Disclosure Risk Audit of the Smart Steps Data Service For *Telefonica*'
- [7] Testimonial from Head of Fire Statistics, Department of Communities and Local Government (31<sup>st</sup> May 2013)
- [8] (2004) CAPRI/UoM 'Special Uniques Detection Algorithm User Guide' (September)
- [9] ICO/UKAN Press Release