CORA Cork Open Research Archive
Cartlann Taighde Oscailte Chorcaí

Title	The privacy dilemma for official statistics in a big data world
Authors	MacFeely, Stephen
Publication date	2018-05-08
Original Citation	MacFeely, S. (2018) 'The Privacy Dilemma for Official Statistics in a Big Data World', Open Access Biostatistics & Bioinformatic, 2(1) (3 pp). doi: 10.31031/OABB.2018.02.000526
Type of publication	Article (peer-reviewed)
Link to publisher's version	10.31031/OABB.2018.02.000526
Rights	© 2018 All rights are reserved by Steve Mac Feely. This is an open access article distributed under the terms of the Creative Commons Attribution License - https://creativecommons.org/licenses/by-nc/4.0/
Download date	2026-10-12 06:56:55
Item downloaded from	https://hdl.handle.net/10468/9506

UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

CRIMSON PUBLISHERS
Wings to the Research

Open Access
Biostatistics & Bioinformatics

ISSN 2578-0247

Mini Review

The Privacy Dilemma for Official Statistics in a Big Data World

Steve Mac Feely*

Department of Statistics and Information, University College Cork, Ireland

***Corresponding author:** Steve Mac Feely, Head of Statistics and Information, United Conference on Trade and Development, Switzerland Adjunct Professor, Centre for Policy Studies, University College Cork, Cork, Ireland.

Submission: May 08, 2018; **Published:** June 15, 2018

To Use or Not to Use - That is the Question?

Over recent years the potential of big data for government, for business, for society has excited much comment, debate and even evangelism. Described as the 'new science' with all the answers [1] or a paradigm destroying phenomena of enormous potential [2] big data are all the rage. Official statisticians, already with a long history of using non-survey data, which are often very large in terms of volume, must decide whether big data is really something new and useful or just hype. On the one hand, some argue that big data needs to be seen as an entirely new ecosystem comprising new data, new tools and methods [3]. Whereas others argue to the contrary that big data is just hype and that big data are just Data [4]. In deciding whether big data can be useful for official statistics, National Statistics Offices (NSOs) must keep the protection of confidential data at the top of their decision making tree.

The Importance of Confidentiality

For official statistics, safeguarding the confidentiality of individual data is sacrosanct and is enshrined in Principle 6 of the United Nations Fundamental Principles of Official Statistics [5], which states 'Individual data collected by statistical agencies for statistical compilation, whether they refer to natural or legal persons, are to be strictly confidential and used exclusively for statistical purposes.' The UN Handbook of Statistical Organization [6], too 'underscores repeatedly the requirement that the information that statistical agencies collect should remain confidential and inviolate. The Scheveningen Memorandum [7] prepared by the Directors General of NSOs in the European Union identified the need to adapt statistical legislation in order to use big data-both to secure access but also protect privacy. The failure to treat individual information as a trust would prevent the statistical agency from functioning effectively. For a NSO to function, confidentiality of the persons and entities for which it holds individual data must be protected i.e. a guarantee to protect the identities and information supplied by all persons, enterprises or other entities, and guarantee that their data are used for statistical purposes only. In short, everyone who supplies data for statistical purposes does so with the reasonable presumption that their confidentiality will be respected and protected. In most countries, safeguarding confidentiality is enshrined in national statistical legislation. But with the increased volumes of big data being generated, and the potential to match those data, greater attention must be paid to data suppression techniques to ensure confidentiality can be safeguarded.

Is Privacy Really Dead?

The emergence of big data is forcing many challenging questions to be asked, not least with regard to privacy and confidentiality. Mark Zuckerberg, the founder of Face book, famously claimed that the age of privacy is over [8]. Scott McNealy, CEO of Sun Microsystems, too famously asserted that concerns over privacy are a 'red herring' as we 'have zero privacy' [9]. Many disagree and have voiced concerns over the loss of privacy [10,11]. Fry [12] has likened developments with regard to big data and the loss of privacy to the opening of Pandora's Box - what he terms, Pandora 5.0. The introduction in Europe of the new General Data Protection Regulation which comes into effect in 2018, reinforcing citizen's data-protection rights, including among other things the right 'to be forgotten', suggests that privacy is still a real concern [13] at least in some regions of the world. By contrast, in the United States, users who provide information under the 'third-party doctrine' i.e. to utilities, banks, social networks etc. should have 'no reasonable expectation of privacy.'

The Dilemma for Official Statistics

This introduces two new challenges for official statisticians: one technical and one of perception. The technical challenge arises from the availability of large, linkable datasets which present a problem thought to have been solved in traditional statistics-anonymisation. But big data, combined with the enormous computing power available today, it is clear that simply removing personal identifiers and aggregating individual data is not a sufficient safeguard. A paper by Ohm [14] outlining the consequences of failing to adequately anonymise data graphically illustrates why there is no room for complacency. Thus a problem that had been solved in the context of traditional official statistics must now be re-solved, in the context

Copyright © All rights are reserved by Steve Mac Feely.

of a richer and more varied data ecosystem. The changing nature of perception is arguably a trickier problem.

What if Zucker berg and McNealy are correct and future generations are less concerned about privacy? There appears to be some evidence to suggest that they may be correct. It seems there are clear inter-generational differences in opinion vis-a-vis privacy and confidentiality, where those 'born digital' are less concerned about disclosing personal information than older generations [15]. Taplin [16] ponders this, musing 'It very well may be that privacy is a hopelessly outdated notion and that Mark Zuckerberg belief that privacy is no longer a social norm has won the day.' If this is so, what are the implications for official statistics and anonymisation? If other statistical providers, not governed by the UN fundamental principles, take a looser approach to confidentiality and privacy, it may leave official statistics in a relatively anachronistic and disadvantaged position vis-a-vis other data providers. But moving away from or discarding principle 6 of the UN Fundamental Principles for Official Statistics would seem to be a very risky move, given the importance of public trust for NSOs.

A Worthwhile Trade off?

Taplin [16] argues that we trade our privacy with corporations in return for innovation or benefits, 'but it is one thing to forfeit our privacy as individuals to a company that we believe is delivering a needed service and another to open our personal lives to the federal government.' Mac Feely [17] has warned that if the benefits of privacy are insufficiently clear to the public or policy makers, then it leaves official statistics vulnerable, and possibly facing a precarious and bleak future. Rudder [18] highlights this challenge too noting that 'the fundamental question in any discussion of privacy is the trade-off - what you get for losing it.' Like Taplin [16] and Rudder [18] also argues that the trade-off benefit with the private sector is clear-better targeted ads! He argues that 'what we get in return for the government's intrusion is less straightforward.

McNealy too, who seems unconcerned about the lack of privacy in the private sector, takes a very different attitude when it comes to government, saying 'It scares me to death when the NSA or the IRS know things about my personal life and how I vot Every American ought to be very afraid of big government' [9]. Curiously, while there is a real fear of government Big Brother, there appears to be few concerns regarding the emergence of a corporate Big Brother. A challenge for official statistics is how to put clear blue water between the NSO and the other institutions of government from the perspective of data sharing, but highlight the common benefits of official statistics as a public good. To some extent there is ideology at play here, where a neo-liberal agenda is pushing to minimize the role of the public sector, but it also illustrates the challenge facing national governments and their agencies where their contribution to the wellbeing of economies and societies is poorly understood.

Conclusion

Big data, if they can be harnessed properly, would appear to offer some tantalizing opportunities - not least improved timeliness and the chance to better align public and official statistics with policy needs. The possibilities of matching different digital data sets may allow us to dramatically improve our understanding of complex, cross-cutting issues, such as, the impacts of life style on health. Advances, such as, the Internet of Things and biometrics will all surely present opportunities to compile new and useful statistics. As yet, the implications of this 'big data bang' for statistics is not immediately clear, but one can envisage a whole host of new ways to measure and understand the human condition. In relative terms, big data are still new. At the turn of the century, Scott Cook, the CEO of Intuit mused we're still in the first minutes of the first day of the Internet revolution [19]. Almost two decades later we are probably only in the first hours. Many norms and standards are yet to evolve. But it does not take a huge leap of imagination to foresee that in the not too distant future, the misuse of big data will be at the heart of a serious human rights abuse scandal. Official statistics must take the ethical dimension seriously. Just because something can be measured doesn't mean it should be. Norms and cultural values regarding privacy may be changing, but in assessing whether and how to use big data, NSOs and international organizations must carefully consider the human rights of citizens in this digital age.

References

1. https://whatsthebigdata.com/2012/06/22/big-data-quotes-of-the-week-10/
2. Davidowitz S (2017) Everybody lies-What the internet can tell us about who we really are. Bloomsbury, London, UK.
3. Letouzé E, Jütting J (2015) Official statistics big data and human development. Data Pop Alliance, White Paper Series, North America.
4. Thamm A (2017) Big Data is dead.
5. United Nations (2014) Resolution adopted by the general assembly on 29 January 2014 fundamental principles of official statistics. General Assembly, USA.
6. United Nations (2003) Handbook of statistical organization-3rd edition: The operation and organization of a statistical agency. Department of Economic and Social Affairs Statistics Division Studies in Methods. United Nations, New York, USA.
7. European Commission (2013) Scheveningen memorandum on: big data and official statistics. European Statistical System Committee, Belgium.
8. Kirkpatrick M (2010) Face books zuckerberg says the age of Privacy is Over. Readwrite, Belgium.
9. Noyes K (2015) Scott McNealy on privacy: you still don't have any. PC World, UK.
10. Payton T, Claypoole T (2015) Privacy in the age of big data: recognising the threats defending your rights and protecting your family, Rowman & Littlefield, Lanham, USA.
11. Pearson E (2013) Growing up digital: presentation to the oss statistics system seminar big data and statistics. Wellington, New Zealand.
12. Fry S (2017) The way ahead. Hay Festival, Hay-on-Wye, UK.
13. European Parliament (2016) 'Regulation (EU) of the European parliament and of the council of the protection of natural persons with regard to the processing of personal data and on the free movement of such data. European commission, Belgium.
14. Ohm P (2010) Broken promises of privacy: Responding to the surprising failure of anonymisation. UCLA Law Review, USA, pp. 1701-1777.

How to cite this article: Steve M F. The Privacy Dilemma for Official Statistics in a Big Data World. Open Acc Biostat Bioinform. 2(1). OABB.000526. 2018.
DOI: 10.31031/OABB.2018.02.000526

15. European Commission (2011) Attitudes on data protection and electronic identity in the European Union. Special Eurobarometer, TNS Opinion and Social, UK.

16. Taplin J (2017) Move fast and break things-how face book, Google and Amazon cornered culture and undermined democracy. Little Brown and Company, New York, USA.

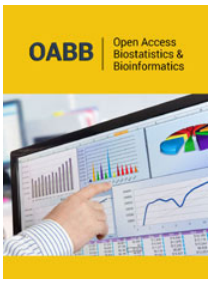
17. MacFeely S (2016) The continuing evolution of official statistics: some challenges and opportunities. Journal of Official Statistics 32(4): 789-810.

18. Rudder C (2014) Dataclysm: What our online lives tell us about our offline selves. 4th Estate, London, UK.

19. Levington S (2000) Internet entrepreneurs are upbeat despite market's rough ride. The New York Times, USA.

Creative Commons Attribution 4.0 International License

For possible submissions Click Here Submit Article



Open Access Biostatistics & Bioinformatics

Benefits of Publishing with us

- High-level peer review and editorial services
- Freely accessible online immediately upon publication
- Authors retain the copyright to their work
- Licensing it under a Creative Commons license
- Visibility through different online platforms