SD 2016-06 (GLOBAL - Tainted Data)
ISSN: 2059-075X // Sovereign Data Vol. 2 No. 6 (June 2016)

EDITOR: Michael A. Innes
AUTHOR: Tim Stevens

Tainted data and user entitlement to exploit

Non-consensual disclosures of classified and proprietary data represent a wealth of openly available but ethically risky source material for researchers.

It is nearly ten years since WikiLeaks launched as a platform for the dissemination of otherwise restricted data. Since then, there has been a steady stream of major non-consensual disclosures via WikiLeaks, detailing US military operations in Afghanistan and Iraq, US State Department diplomatic cables, and many other document caches revealing practices and processes that states would rather have remained hidden. The Snowden disclosures (2013) and the Panama Papers (2016) have further illustrated that governments and corporate entities are under greater scrutiny than ever before as a result of data leakage and disclosure. Decried as traitors by some, as heroes by others, WikiLeaks and similar entities have sparked furious debates about the balance between transparency and security, and the nature of openness and control in today's global information environment.

Sovereign Data explores the conundrum this presents for specialists working in information-intensive enterprisers, as they contend with ethical and regulatory constraints on the use of non-consensual data.

Governance: Norms of professional practice

In those countries that prize a free press, it is widely accepted that journalists may seek, acquire and use non-consensual data if these activities further the public good. Even though the US government's reaction to, for instance, the actions of Edward Snowden and Chelsea Manning, was overwhelming negative, there has been little suggestion that the press acted illegally in reporting on or assisting the public dissemination of leaked materials. Whereas Manning originally stood trial, *inter alia*, for the capital offence of 'aiding the enemy', prominent publications that participated in this and other leaks have yet to be prosecuted. This constitutional protection does not extend to all jurisdictions. Journalists have been investigated in the UK and elsewhere for publishing information gained from the Snowden case but successful prosecutions have only been brought against the press for more clear-cut breaches of privacy, such as the News of the World 'phone hacking' case in the UK.[i] This has not prevented senior officials from intimidating journalists but the freedom of the press has, in liberal countries, not yet been seriously threatened. In the US, it has been possible to simultaneously argue that while the original leaks were illegal, they also performed a 'public service', as asserted by former US Attorney General, Eric Holder.[ii]

There is a journalistic presumption towards putting information in the public domain if it meets strict criteria of public interest and factual integrity. Once there, this is usually where academics and other researchers encounter it. Their ethical considerations are likely to be personal and local to institutional context. There is a general vocational instinct to publish research, unless of a classified or commercially sensitive nature, but additional norms pertain that constrain this impulse. Academic researchers are ordinarily bound by codes of ethical conduct that constrain research on the basis of the liberal principle of harm avoidance, to both individual human subjects and the wider community. Business ethics generally disbar use of inside information for commercial advantage but leaked data falls outside of this framework. In both cases, the reputation of the university or firm is also a key factor in deciding if and how to use data already in the public domain. Individual academic researchers, in particular, are often confronted with a very personal choice: whether to use ‘tainted data’, or not. Their decision-making processes reveal much about personal integrity and research for the common good, as well as judgements about what counts as public information and what can or should be done with it.[iii]

Infrastructure: Caution and context

Many of the recent revelations have had national security implications and academics working on these issues have resolved into two broad camps.[iv] International Relations scholars have been largely unsurprised by the content of leaks, for instance Cablegate (2010-11), in which quarter of a million US diplomatic cables were published online. Their central debate has been whether disclosure of this information amounts to the ‘violation of a state privilege’, and if such actions require formal sanction.[v] Disagreement on this issue registers amongst political elites too. On the one hand, senior US policy-makers called on the State Department to label WikiLeaks a ‘foreign terrorist organization’.[vi] On the other, Defense Secretary Robert Gates thought the impact on US diplomacy ‘fairly modest’.[vii] The current consensus seems to be that damage to the US from Cablegate has been limited and, in some respects, has shown a ‘human’ side to diplomatic interactions not often revealed to the public.

A second academic grouping considers mass leaks of national security information in a more emancipatory light. The reaction from sociologists and communications scholars to Cablegate, for example, has been tinged with anger and resignation, as well as a mildly triumphant, ‘We Told You So’.[viii] Existing narratives of gross state impropriety and structural inequality are therefore corroborated by disclosures revealing ‘new’ data about state practices, which may then be used to fashion counter-narratives resisting these practices and furthering the public good. Decisions made about whether to use ‘tainted data’ tend, therefore, to be somewhat less rigorous than might be desirable, although this in no way detracts from the positive social impulse that drives them.

In both cases, ethics and politics are intertwined and there is no right or wrong way to approach ‘tainted data’ from this perspective. A more pragmatic solution lies in the methods employed when opportunities to use leaked data-sets arise. As with journalism, no researcher should rely solely on leaked material as an unproblematic source of empirical data. Just as journalists are taught not to base their stories on single sources, so too are academics, in particular, taught the values of caution and context. One criticism of post-Snowden reportage and analysis has been that researchers have been too keen to treat National Security Agency documents as single data-points, without due consideration of contextual factors, such as organisational politics and technical accuracy.[ix] Academics that have used the Snowden disclosures as foundations for empirical studies have been very careful to triangulate these data-points with other material already in the public domain, and to offer only cautious assessments of their wider social and political significance.[x] Moreover, the data-sets may be rather less revelatory than widely imagined, particularly as formal disclosure and transparency mechanisms, particularly in the US, already provide substantial information on and insights into state practices.[xi]

Stakeholders: Uncertainty and insecurity

Most researchers are keen to utilise these data-sets but constrained by professional and institutional ethics. Some have avoided these issues entirely, citing established formal modes and mechanisms of state disclosure and declassification, which for them remain sacrosanct. Disturbingly, some professional journals have *a priori* embargoed use of leaked data, citing legal concerns.[xii] In the absence of specific legal threats this may be considered an unnecessary extension of the principle of caution outlined previously. However, it is indicative of a general uneasiness that using 'tainted' data-sets may expose researchers to unwelcome state attention and possible legal challenge. In the field of national security, for instance, few wish to fall foul of often stringent laws regarding theft and subsequent use of government information, let alone accusations of subversion or treason. Considerations of personal and professional consequences clearly weigh heavily on many researchers' minds, particularly in the political sciences.[xiii]

As suggested above, this does not apply to all researchers and certainly not to all disciplines. Just as Edward Snowden appears to have made an ethical judgement that releasing classified NSA material was, on balance, in the public interest, researchers have to make similar decisions. There is a strong argument that once this material is in the public domain, it is fair game for all to use. The old rules that applied in the time of the Pentagon Papers (1971) no longer apply: the internet means that once in the public domain, something is always in the public domain. It is a one-way release of information and calls to 'return' stolen data look rather naïve in such circumstances.[xiv] Future technologies might be able to seek out and destroy specific data held on networked devices but, given offline storage media and the myriad locations in which data can be stored, this remains speculative at best. The 'nuclear option' of switching off portions of the internet would also not guarantee data deletion, nor would it serve wider societal interests.

This is not an argument that favours technological determinism over ethical considerations surrounding leaking and whistle-blowing, but a simple statement of fact. Legally, however, as the decision of the Library of Congress to block web access to WikiLeaks in 2010 clearly demonstrates, disclosure does not change the classified nature of leaked documents under US law.[xv] The obvious solution in this particular situation is to declassify all leaked documents, although this may dangerously incentivise potential leakers, and is unlikely to happen.

Conclusions

The contemporary importance of mass disclosure of restricted data poses a major challenge to established modes of information control and dissemination. In the past decade, WikiLeaks, Snowden, Manning, the Panama Papers, and many other instances, have shown that institutions are, as famously described by surveillance academic David Lyon, 'leaky containers'.[xvi] Previously discrete public and private sector entities are coming into increased contact with one another and the opportunities for information to flow from one to the other have produced a hybrid crisis over information control and institutional legitimacy. Just as governments and businesses were committing to greater transparency in their affairs, WikiLeaks and its fellow travellers have presented them with a *fait accompli*, in which radical transparency has been forced upon them, mediated by the global internet. Researchers whose professional activities demand they explore the workings of government, economics and diplomacy have found themselves in an awkward, and currently unresolved, situation. How should they negotiate the ethical and legal issues thrown up with this novel state of affairs, whilst continuing to adhere to the highest professional standards? Caution and context are admirable methodological principles to follow but do not exempt researchers from potential legal blowback and state interference. Ultimately, perhaps, to publish or not to publish will remain a matter of professional ethics, as no state will submit to radical transparency, and few researchers will ignore the opportunities presented in such fashion.

NOTES

[i] 'Guardian journalists could face criminal charges over Edward Snowden leaks', *The Telegraph*, 3 December 2013; *R. v. Coulson, Brooks and others* [2013] All ER (D) 287 (Jun).

[ii] 'Snowden leaks illegal but were a "public service", Eric Holder says', *The New York Times*, 31 May 2016.

[iii] E.g., Nathaniel Poor and Roei Davidson, 'Case study: the ethics of using hacked data: Patreon's data hack and academic data standards', Council for Big Data, Ethics, and Society, 2016, http://bdes.datasociety.net/council-output/case-study-the-ethics-of-using-hacked-data-patreons-data-hack-and-academic-data-standards/. [Accessed 8 June 2016.]

[iv] Athina Karatzogianni and Andrew Robinson, 'Digital Prometheus: WikiLeaks, the state-network dichotomy, and the antinomies of academic reason', *International Journal of Communication* 8 (2014): 2704-2717.

[v] Karatzogianni and Robinson, 'Digital Prometheus', p. 2707.

[vi] 'WikiLeaks: Hillary Clinton states WikiLeaks release is "an attack"', *The Telegraph*, 29 November 2010.

[vii] 'Pentagon boss is not sweating WikiLeaks', *Wired*, 30 November 2010, https://www.wired.com/2010/11/pentagon-boss-is-not-sweating-wikileaks/. [Accessed 8 June 2016.]

[viii] Karatzogianni and Robinson, 'Digital Prometheus', p. 2710.

[ix] 'Snowden leaks lack context says security studies professor', *The Register*, 6 January 2015, http://www.theregister.co.uk/2015/01/06/snowden_leaks_hype_and_lack_of_context_bad_for_understanding/. [Accessed 8 June 2016.]

[x] For example, Daniel G. Arce, 'WikiLeaks and the risks to critical foreign dependencies', *International Journal of Critical Infrastructure Protection* 11 (2015): 3-11; Tim Stevens, 'Security and surveillance in virtual worlds: Who is watching the warlocks and why?', *International Political Sociology* 9, no. 3 (2015): 230-247.

[xi] Jon Western, 'We really don't need Wikileaks', *Current Intelligence* 3, no. 1 (2011), http://thesigers.com/analysis/2010/12/2/we-really-dont-need-wikileaks.html. [Accessed 9 June 2016.]

[xii] Gabriel J. Michael, 'Who's afraid of Wikileaks? Missed opportunities in political science research', *Review of Policy Research* 32, no. 2 (2015): 175-199.

[xiii] Michael, 'Who's afraid of WikiLeaks?'.

[xiv] 'James Clapper calls for Snowden and "accomplices" to return NSA documents', *The Guardian*, 29 January 2014.

[xv] 'Why the Library of Congress is blocking WikiLeaks', *Library of Congress Blog*, 3 December 2010, https://blogs.loc.gov/loc/2010/12/why-the-library-of-congress-is-blocking-wikileaks/. [Accessed 8 June 2010.]

[xvi] David Lyon, *Surveillance Society: Monitoring Everyday Life*. Buckingham: Open University Press.