scripted

Volume 12, Issue 2, December 2015

REFLECTIONS ON THE CONCEPT OF OPEN DATA

Edward S. Dove[*]

Abstract

Though "open data" is much discussed as a practice, it is much less discussed as a concept. There is consensus that open data is an emerging global social movement—an Open Data Movement—that encourages a shift in behaviour about performing data-centric tasks, such as governing or researching, to make them more connected and collaborative and thereby improve transparency, accountability, research discovery, knowledge access and knowledge co-production. But just what do we mean by the qualifying word "open"? Open data is understood to mean data resources that are: (1) free for people to access; (2) free from most legal constraints on reuse; and (3) put into formats that maximise interoperability and linkage. This definition, however, fails to fully address all the conceptual and policy issues at play in the Open Data Movement. In this analysis piece, I offer some critical reflections on the Open Data Movement and unpack the meaning of "open" data so as to offer a richer understanding of the concept.

DOI: 10.2966/scrip.120215.154

[*] PhD Candidate, School of Law, University of Edinburgh.

1. Introduction

Data have always been intertwined with humanity — collected, used and shared in some way by individuals, families, communities and governments to make sense of our world and to improve our social and material condition.[1] Yet for much of history, observations of the natural world and of lifeworlds were limited to a small number; data linkage was nearly nonexistent; far-flung data sharing was seen as noble but limited or nonexistent in everyday life; and extracted data were typically static and coarse. Large amounts of time and resources were needed to collect and interpret data, which blunted their impact on various domains, including science, business, government and civil society.[2]

Today, however, thanks to disruptive innovations, especially personal computers, sensors, the Internet and distributed computing, data are everywhere: they are increasingly generated, shared, made dynamic, diversified, scaled up, linked, distributed, digitised and thereby made increasingly accessible. Data also have become big: they are huge in volume, high in velocity of creation and movement, diverse in variety in type, flexible in their ability to add new data fields easily, and scalable in their ability to expand in size rapidly.[3] Consider, for example, the emergence of the "Internet of Things", where sensor-based "wearables" such as clothes and watches can monitor one's health status. Personal health data can be directly uploaded into the cloud, "linked to social networks and potentially broadcast publicly, enabling identification of users and tracking of the behaviour and movements of individuals and crowds."[4]

Information and communication technologies and social media platforms are reconfiguring data assemblages. In so doing, they enable a hitherto neglected (or under-utilitised) *demos* to capture its everyday lived experience and world as data, to interpret those data, and in consequence to affect or transform almost every human endeavour, if not human behaviour.[5] As the expense and resources required to generate data dwindle, and the ability of analytic tools to make sense of data improves, access to and sharing of data have correspondingly increased. "Openness, participation and collaboration" is in many ways the mantra of the 21st century. Governments, keen on maintaining their (democratic) legitimacy, fulfilling mandates for transparency, and meeting expectations for efficient and effective services, are paying heed to this ostensible libertarian and communitarian ethos by opening up their vaults of public data and asking citizens to not only peer inside, but to actively

---

[1] L Gitelman (ed), *Raw Data Is an Oxymoron* (Cambridge, Mass.: MIT Press, 2013). Surveying and mapmaking are classic examples of the generation and use of data to advance society.

[2] R Kitchin, *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences* (London: SAGE Publications, 2014).

[3] V Mayer-Schonberger and K Cukier, *Big Data: A Revolution That Will Transform How We Live, Work and Think* (Boston: Houghton Mifflin Harcourt, 2013).

[4] European Data Protection Supervisor, "Opinion 4/2015, Towards a New Digital Ethics: Data, Dignity and Technology" (2015) available at https://secure.edps.europa.eu/EDPSWEB/webdav/site/mySite/shared/Documents/Consultation/Opinions/2015/15-09-11_Data_Ethics_EN.pdf (accessed 14 Dec 15), at 7.

[5] S Leonelli, "What Difference Does Quantity Make? On the Epistemology of Big Data in Biology" (2014) 1 *Big Data & Society* 1–11.

contribute to the production of innovative delivery of services, often paid out of the public purse.[6]

Institutions and businesses, too, recognise that "freeing the data" facilitates enterprise, improves development and delivery of products (e.g. drugs and devices) and benefits the consumer and the commonweal (e.g. public health). The results range from the mundane to the profound. Science has been one of the domains most affected in the tidal wave of data, which is now according to the late computer scientist Jim Gray experiencing a "Fourth Paradigm" characterised by data-intensive scientific discovery.[7] Data have become a commodity; they are an "oil" or "currency" thoroughly integrated into sharable, open, networked infrastructures that enable societies to rethink, among other things, how knowledge is generated, capital is produced, cities are designed, transportation networks are organised, education is delivered and the human body functions. Those who do not participate in this evolving era of "share and share alike" now run the risk of being labelled as data pariahs, hoarders of the elements necessary to build knowledge and improve humanity.

But just what this phrase "open data" means remains murky. It is clear enough that the Open Data Movement, now a global social movement in its own right, stresses that data should be open to anyone, free to use, reuse and redistribute, provided only, perhaps, that one properly attributes the creator(s) and shares-alike.[8] It is also clear enough that the Open Data Movement encourages a shift in behaviour about performing data-centric tasks, be it governing, researching or something else, to make them more connected and collaborative and thereby improve transparency, accountability, research discovery, knowledge access and knowledge co-production. And yet, as much as has been written in the last few years about data and their various modalities (e.g. big, open, aggregated, linked) and practices, too little thought has been given to teasing out what is meant exactly by the concept "open" and the technoscientific and societal imaginaries of the Open Data Movement. It is clear enough that not *all* data can be open, and not *everyone* is able to have access to data. Open data is complex and inextricably tied to legitimate challenges and institutional concerns, including ownership and property rights, cost, competition, confidentiality and privacy, and data misuse. Query then, the sensible bounds of this term "open". Surely some degree of qualification is needed.

To consider what open data is and what it is not (and should not be) allows us to build a critical ethical frame that enables individuals and communities to define themselves and their imaginaries, and that determines who may access what data on what terms. Key questions must be posed and answered to illuminate significant issues that have bearing on our role as citizens. In this analysis piece, I pose and address these questions. It is hoped that by looking critically at "open" data, we can achieve greater

---

[6] A Zuiderwijk and M Janssen, "Open Data Policies, their Implementation and Impact: A Framework for Comparison" (2014) 31 *Government Information Quarterly* 17–29.

[7] T Hey, S Tansley and K Tolle (eds), *The Fourth Paradigm: Data-Intensive Scientific Discovery* (Redmond, WA: Microsoft Research, 2009).

[8] Share-alike is a copyright licensing term to describe the possible requirement of users of a work to provide the content under the same or similar conditions as the original. See Open Knowledge Foundation, "Open Data Handbook" (2015) available at http://opendatahandbook.org/ (accessed 14 Dec 15).

understanding of open data as both concept and practice, and the extent to which the Open Data Movement raises reasonable concerns about privacy, power and public trust.

2. What is "open"?

The qualifier "open" provides important insight into how the Open Data Movement frames data. One of the core purposes behind the Open Data Movement is not just to enhance participatory democracy through access to an organisation's data; nor is it just to enhance transparency by allowing people to assess an organisation's accountability or value for money. Critically, it is also to facilitate more efficient and effective knowledge production, and to generate economic innovation, social innovation and wealth creation through greater access to data in a convenient and modifiable form at no more than a reasonable reproduction cost. Given access to data about urban traffic patterns and appropriate software (e.g. open source code), for instance, one can determine an optimal commuting route, and in so doing boost productivity and save drivers' time, taxpayers' money and the environment's clean air.

The Open Data Movement has placed great importance on — and has gone to great lengths to ensure the formulation of — mature, collaboratively developed understandings of the openness that is worthy of being advocated for. As science and technology studies scholar Sheila Jasanoff writes in the context of the "open science" wing of the Open Data Movement: "Openness is a treasured attribute of science, but like most good things, even scientific openness has to be purposefully cultivated and judiciously deployed in order to serve its intended functions well."[9] The Open Definition project provides a definition of "open" on the most general level and aims to make "precise the meaning of 'open' with respect to knowledge."[10] According to the Open Definition project: "Knowledge is open if anyone is free to access, use, modify, and share it — subject, at most, to measures that preserve provenance and openness."[11] But a number of other definitions have emerged in more specific circumstances, some of which aim to be legally enforceable, such as in "open publishing", in licensing "free and open source software" (FOSS) and in "open government" standards.

2.1 Beyond mere obtainability: formats, reuse, interoperability, linkage

The Open Data Movement has never restricted itself to simply calling for data that are more easily obtainable. Rather, its advocates demand that any barriers or restrictions on access to and use of data be eliminated to the degree practically possible. The Open Data Handbook, for example, states that open data "is data that can be freely

---

[9] S Jasanoff, "Transparency in Public Science: Purposes, Reasons, Limits" (2006) 69 *Law and Contemporary Problems* 21–45, at 42.

[10] Open Knowledge Foundation, "Open Definition: Version 2.1" (2015) available at http://opendefinition.org/od/2.1/en/ (accessed 13 Dec 15). Note that in this quote, "knowledge" is collapsed into "data", whereas often in the information sciences, definitional distinctions are made between data, information, knowledge and wisdom. See note 2 above, at 9–10.

[11] *Ibid*.

used, re-used and redistributed by anyone – subject only, at most, to the requirement to attribute and sharealike."[12]

We should pause and reflect on this distinction between *mere* obtainability of data, and free access, use, reuse and redistribution of data *by anyone*. By way of comparative illustration, in many jurisdictions, even though law court records are on the public record and are available to the public, they are not "open data" as conceived here. Listening to audio recordings of hearings in most courtrooms, for example, generally requires that a person physically present themselves at a courthouse during its open hours, and have the means to copy the recordings, as they are not facilitated by any staff.[13]

So if mere obtainability of data does not make it "open", what attributes would? The Open Knowledge Foundation maintains three requirements for data to be considered open — requirements that are repeatedly found among open standards developed by the open data community:

- Availability and Access: the data must be available as a whole and at no more than a reasonable reproduction cost, preferably by downloading over the Internet. The data must also be available in a convenient and modifiable form.
- Reuse and Redistribution: the data must be provided under terms that permit reuse and redistribution including the intermixing with other datasets.
- Universal Participation: everyone must be able to use, reuse and redistribute – there should be no discrimination against fields of endeavour or against persons or groups. For example, "non-commercial" restrictions that would prevent "commercial" use, or restrictions of use for certain purposes (e.g. only in education), are not allowed.[14]

These requirements provoke several observations. First, they extend beyond requiring mere obtainability of the data, as in the court records example, touching on practical considerations such as convenience, and placing a premium on features like interoperability (e.g. open machine-readable format) and data integrity. Second, and perhaps surprisingly, open data would seem to be compatible with commercial use of data; in fact, commercial use is explicitly protected (though not promoted) by the definition. In many cases, open data is compatible with copyright, as in the case of FOSS, which is generally copyrighted by its initial author but licensed under any one of several FOSS licenses to allow use by any third party. Third, the standard's demand for "universal participation" is limited to removing active restrictions by third parties on the use of data. Thus, the requirements impose no *positive* obligations on any actor, aside from the burden of making the data available "in a convenient and

[12] See note 8 above.

[13] D Stepniak, "Technology and Public Access to Audio-Visual Coverage and Recordings of Court Proceedings: Implications for Common Law Jurisdictions" (2004) 12 *William & Mary Bill of Rights Journal* 791–823; P Coppock, "Doors to Remain Open during Business Hours: Maintaining the Media's (and Public's) First Amendment Right of Access in the Face of Changing Technology" (2013) 58 *South Dakota Law Review* 319–346.

[14] See note 8 above.

modifiable form." This may be trivial, but depending on the data, it could also impose significant costs for production, curation and distribution.[15]

The absence of positive obligations raises questions of justice. For example, Michael Gurstein reports that following the digitisation and opening of surveys, deeds and land titles in Nova Scotia, surveyors and lawyers used the data to profit by dispossessing poor landholders of their properties.[16] This focus on an "open divide" has led Gurstein to recommend a fundamental shift in the approach of the Open Data Movement, away from the traditional focus on removing restrictions on the free flow, use and reuse of data, and instead toward seeing open data as a "process" of evolving data with the goal of providing appropriate technology to all. The question to ask should not be whether any given data producer has erected barriers to accessing their data, but instead whether consumers of the data have the capacity to make equally "effective use" of that data.[17] To promote this principle, Gurstein advocates attentiveness to "such factors as the cost and availability of Internet access, the language in which the data is presented, the technical or professional requirements for interpreting and making use of the data, and the availability of training in data use and visualization, among others."[18]

Gurstein seems to want to transform the "open" in data from a negative freedom not to have barriers attached to accessing data into a positive right which entitles access to the tools and knowledge to allow people to make practical use of socially important data. In Gurstein's words, it is a shift in focus from the open data "supply" or "access" side, on which it has traditionally been centred, to the "demand" or "user" side.[19]

2.2 Underlying rationales for openness

As already noted, a key purpose of making data "open" is to accelerate or refine knowledge production and promote economic and social innovation. The FOSS movement — the earliest of the strand of what would become the Open Data Movement — first made plain the differing value systems of the adherents it brought together.

Some thought that FOSS created better outcomes. Software developer and open source software advocate Eric Raymond exemplifies this current, especially in his book, *The Cathedral and the Bazaar*, in which he argues that just as scientific inquiry requires a peer-review process to produce quality results and knowledge, so too do transparent, co-operative, egalitarian, or meritocratic software development processes (at the time exemplified by the community that had emerged surrounding the GNU/Linux operating system) produce better software than processes characterised

---

[15] See note 2 above, at 57–60.

[16] M Gurstein, "Open Data: Empowering the Empowered or Effective Data Use for Everyone?" (2011) available at http://firstmonday.org/article/view/3316/2764 (accessed 14 Dec 15).

[17] *Ibid*.

[18] *Ibid*.

[19] *Ibid*.

by secrecy and hierarchy (at the time exemplified by Microsoft). The ethos is encapsulated in Raymond's maxim that "secrecy is the enemy of quality."[20]

At the other end of the debate is software freedom activist and computer programmer Richard Stallman. For Stallman, while FOSS may well lead to a better product, that fact should have no bearing on its social importance. Instead, software must be open as a precondition of human freedom. Free software is seen as a necessary precondition of freedom and creativity not only for programmers — by empowering them to tailor software as they see fit according to their own needs and circumstances — but more importantly, for all software users — who are otherwise deeply vulnerable to the whim of the owner of rights in the software. In this conception, the computer code that allows a technologically advanced society to function must be the common heritage of humanity and belong to all.

In contrast to a traditional sectarian divide, in which the basis is a bitter disagreement over strategy or tactics within a movement that shares common basic principles, the FOSS divide was the reverse: its members relied on fundamentally different underlying motivations that nonetheless led them to adopt the same strategy — to make software and its underlying source code freely available and modifiable without restriction. As a result, although lively debate flourished, the practical effects on the FOSS movement were relatively minor and a variety of FOSS licenses, including the GNU Public License, were adopted by partisans on each side of the debate.

But the specificity of the software development context masked further disagreements in the Open Data Movement that would emerge in other contexts. When faced with barriers imposed by proprietary software, both Stallman and Raymond agree that the solution is to write free and open software that serves the same function and to make it available as an alternative. In most other areas of the Open Data Movement — government, scientific, or creative endeavours, for example — producing functional equivalents is usually either impractical or illegal. In response to this hurdle, the Open Data Movement has had to convince institutions to open their data, either prospectively, retroactively, or both. This goal has been pursued both by legal (e.g. lobbying, court challenges) and illegal means (e.g. leaking classified or confidential information), each with some success.

Many successful attempts have been made to convince local and national governments to open their data as part of the "open government" wing of the Open Data Movement. These initiatives have encompassed everything from stores of ancient land survey data to up-to-the-minute detailed information on the legislative process, including debates and committee hearings. Closely connected to these initiatives have been the emergence of Legal Information Institutes around the world, seeking to open legal information to the public, with attempts being made, at least ostensibly, to achieve even coverage between developed and developing countries.[21]

---

[20] E Raymond, *The Cathedral and the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary* (Cambridge, Mass.: O'Reilly, 2001), at 142.

[21] Legal Information Institutes of the World, "Montreal Declaration on Free Access to Law" (2007) available at http://www.canlii.org/en/info/mtldeclaration.html (accessed 14 Dec 15); D Poulin, "Open Access to Law in Developing Countries" (2004) available at http://firstmonday.org/ojs/index.php/fm/article/view/1193/1113 (accessed 14 Dec 15).

The Open Data Movement has also extended to many areas beyond open government. The Creative Commons organisation has expanded "open access" to creative works, including the articles in this journal. A significant shift in scholarly writing has been underway for over a decade, in which increasing numbers of technical and academic journals and books are publishing works and scientific results as open access.[22]

But merely prospectively opening up scholarly publishing can leave great stores of historical scholarship, much of which is now digitised, in private hands. The inaccessibility of this trove of information allegedly led the late open access advocate Aaron Swartz to systematically download the scholarly content of JSTOR's private service.[23] Although sharing these data may have been illegal, for Swartz "sharing isn't immoral — it's a moral imperative," and he laid out the rationale for engaging in civil disobedience in his manifesto:

> Forcing academics to pay money to read the work of their colleagues? Scanning entire libraries but only allowing the folks at Google to read them? Providing scientific articles to those at elite universities in the First World, but not to children in the Global South? It's outrageous and unacceptable.[24]

A similar logic, which conceives of data sharing as civil disobedience, or at the very least as operating in a liminal space between (immoral) legality and (moral) illegality, has proliferated in other areas. Where creative works are concerned, Peter Sunde, co-founder and ex-spokesperson of the popular BitTorrent file sharing site The Pirate Bay, asserts that in his view "copyright is not needed when it comes to personal use."[25] As for classified or sensitive government data, the principles motivating WikiLeaks' Julian Assange's efforts to open data were unique in that they were relatively unconcerned with the public's right to access the data and instead relied on the effectiveness of making data open to weaken states' abilities to engage in (ostensibly) serious crimes or other misdeeds:

> [I]n a world where leaking is easy, secretive or unjust systems are nonlinearly hit relative to open, just systems. Since unjust systems, by their nature induce opponents, and in many places barely have the upper hand, mass leaking leaves them exquisitely vulnerable to those who seek to replace them with more open forms of governance.

---

[22] J Willinsky, *The Access Principle: The Case for Open Access to Research and Scholarship* (Cambridge, Mass.: MIT Press, 2006).

[23] JSTOR (www.jstor.org), short for "Journal Storage", is a digital library that contains digitised issues of academic journals, as well as books and primary sources.

[24] A Swartz, "Guerrilla Open Access Manifesto" (2008) available at https://archive.org/stream/GuerillaOpenAccessManifesto/Goamjuly2008_djvu.txt (accessed 14 Dec 15).

[25] T Mennecke, "The Pirate Bay Interview" (2008) available at http://www.slyck.com/story1638_The_Pirate_Bay_Interview (accessed 14 Dec 15).

> Only revealed injustice can be answered; for man to do anything intelligent he has to know what's actually going on.[26]

But the most long-lasting and influential example of civil disobedience related to open data are likely to be Edward Snowden's revelations in 2013 of an unlawful, intensive global surveillance program directed by the United States' National Security Agency (NSA) and Five Eyes, an intelligence alliance of Australia, Canada, New Zealand, the United Kingdom and the United States. Although, as in the other examples mentioned, Snowden made data open that had not previously been available to the public, his revelations also differ fundamentally from the others. The Snowden leaks in many ways, rather than promoting a vision of a culture of open data, instead actually have been taken by many as evidence that data should be subject to *tighter* constraints and controls — at least to the extent necessary to keep private personal information out of the hands of Five Eyes and from being shared within it.

3. Which data *ought* to be open?

Even the most ardent open data advocates, including those who feel that the social importance of open data justifies civil disobedience, do not believe that *all* data should be open. None believe that their own password or financial or health data should be freely available, nor does anyone adopt Immanuel Kant's position that the murderer at the door is entitled to an honest answer about where to find his intended victim. Indeed, although Mark Zuckerberg stated that privacy is no longer a "social norm",[27] the Facebook founder has as of yet failed to publish his own passwords; in fact, to protect his privacy, he purchased the four houses neighbouring his own Silicon Valley mansion.[28]

The question of *which* data should be open is not one to which early open data advocates devoted much thought. One of the reasons is that privacy interests were scarcely ever at stake in connection with the data they were fighting to open, such as software code or published creative works. This is still evident today, for example, in the so-called "8 Principles of Open Government Data":

> The Open Government Data principles do not address what data should be public and open. Privacy, security, and other concerns may legally (and rightly) prevent data sets from being shared with the public. Rather, these principles specify the conditions public data should meet to be considered 'open'.[29]

---

[26] J Assange, "The Non-Linear Effects of Leaks on Unjust Systems of Governance" (2006) available at http://cryptome.org/0002/ja-conspiracies.pdf (accessed 14 Dec 15).

[27] B Johnson, "Privacy No Longer a Social Norm, Says Facebook Founder" (2010) available at http://www.theguardian.com/technology/2010/jan/11/facebook-privacy (accessed 14 Dec 15).

[28] A Shontell, "Mark Zuckerberg Just Spent More Than $30 Million Buying 4 Neighboring Houses for Privacy" (2013) available at http://www.businessinsider.com/mark-zuckerberg-buys-4-homes-for-privacy-2013-10 (accessed 14 Dec 15).

[29] The Annotated 8 Principles of Open Government Data (2007) available at http://opengovdata.org/ (accessed 14 Dec 15).

In this section, I briefly analyse some of the relevant considerations in this previously neglected area, focusing on privacy and data protection.

Because it has traditionally focused on data with no associated privacy concerns (i.e. the low-hanging fruit), the Open Data Movement aimed to uniformly impose its strict set of conditions. The uniform approach certainly has strengths: the demand is clear and difficult to water down or whittle away. But it also has weaknesses: where certain data or services are not compatible for full openness by their very nature, the Movement has sometimes failed to imagine alternatives.

3.1 Privacy and data protection

To be fair, it is not always obvious how the twin goals of opening data and protecting privacy can be mutually reconciled. Most would agree that some data are too sensitive to be made open as per open data principles (e.g. HIV status of individuals), and other data, while not individual-level, may nevertheless be collected and analysed on an aggregate level but repurposed for inappropriate or socially disadvantage reasons (e.g. discriminatory credit and insurance risk profiling based on neighbourhood location). Some line must be drawn in the proverbial sand between open data and restricted data, and data that fall on the restricted access side should not be properly labelled "open data". Laws and government policies can and should rightfully restrict opening data full tilt to protect privacy as an enduring social norm. The difficulty lies in crafting security architectures and laws, be it through legislation or soft law approaches, to achieve principled approaches that strike an appropriate balance between openness and privacy.

Indeed, the leading techniques employed by experts to protect privacy are striking in that they represent the exact reverse of each of the demands in the open data definition necessary for data to be considered open: privacy is protected either by limiting access to the sensitive data, by limiting use of the data, by limiting transfer of the data, by limiting the purposes for which the data can be used, or by releasing only partial (or partially garbled) "de-identified" (i.e. anonymised) versions of the data. If the impossibility of fully reconciling openness and privacy is not already unsettling enough to open data and privacy advocates alike, even worse is that it is widely recognised that none of the privacy strategies—with the possible exception of outright deletion of the sensitive data — comes with a guarantee that privacy breaches will not nonetheless occur. No matter what compromise is reached, there will always be the potential to argue that open data, in its various modalities, should be curtailed further in the interests of an elusive claim to privacy, as well as the reverse under the banner of "public interest" or other claim.

But the newer, compromise approach to privacy and openness, although apparently necessary, poses risks to both. Adding flexibility to what is accepted as "open" or "privacy-protected" leaves each vulnerable to being undermined by institutional opportunism cast as an economic or research imperative.

The years-long legislative process of drafting the European Union's General Data Protection Regulation (GDPR) has been particularly revealing in this respect. Former EU Commissioner Viviane Reding observed of the process that lobbying "has been

fierce – absolutely fierce – I have not seen such a heavy lobbying operation."[30] This included lobbying from technology sector players seeking to limit privacy protections for reasons that had little or nothing to do with principles of open data as understood by most. For example, Facebook urged that a user right to opt-out of targeted advertising "impairs companies' ability to innovate and negatively impacts the users[']experience."[31]

But more sophisticated lobbying efforts have harnessed open data principles as a means of driving home their point. Yahoo, for example, has relied on anonymisation — the popular technique developed to allow increased openness of data while mitigating privacy concerns — to oppose requiring explicit consent from users before allowing collection of their personal information, on the basis that its stored data are anonymous.[32] Similarly, the Wellcome Trust, the UK's largest provider of non-governmental funding for scientific research, has lobbied to ensure that the GDPR allows "pseudonymised" data to be used in research without specific consent from the person to whom the data relates.[33]

The problem with anonymisation as a privacy-enhancing technique to allow openness is that it is ill-suited to being enshrined in law, especially as the sole protector of privacy. Debate has continued to rage about the continued relevance of anonymisation, but according to Princeton University's Ed Felton:

> The trend is toward treating [anonymisation] like cryptography, where 'I scrambled up the data a bunch' is not a valid argument and 'I can't think of an attack' is not a valid argument – you have to have a technically rigorous argument that no attack is possible.[34]

When a law requires only that data be "anonymised" without clearly specifying what this means, the mentality that "I can't think of an attack" is sure to predominate. On the other hand, when laws instead adopt a detailed rule-based approach to anonymisation, the result is even worse since anonymisation (and correspondingly, re-identification) is a technical, context-specific and rapidly evolving mechanism.[35] In short, data anonymisation is rather impervious to legal codification. The "Safe Harbor" provision in the Privacy Rule of the United States' Health Insurance

---

[30] M Warman, "EU Privacy Regulations Subject to 'Unprecedented Lobbying'" (2012) available at http://www.telegraph.co.uk/technology/news/9070019/EU-Privacy-regulations-subject-to-unprecedented-lobbying.html (accessed 14 Dec 15).

[31] D Liebelson, "First SOPA, Now Your Privacy: Facebook, Google Flex Lobbying Muscle in Europe" (2013) available at http://www.motherjones.com/politics/2013/03/google-facebook-sopa-privacy (accessed 14 Dec 15).

[32] *Ibid*.

[33] C O'Donoghue, "EU Research Group Condemns EU Regulation for Restricting Growth in Life Sciences Sector" (2014) available at http://www.globalregulatoryenforcementlawblog.com/2014/02/articles/data-security/eu-research-group-condemns-eu-regulation-for-restricting-growth-in-life-sciences-sector/ (accessed 14 Dec 15).

[34] C Doctorow, "Data Protection in the EU: The Certainty of Uncertainty" (2013) available at http://www.theguardian.com/technology/blog/2013/jun/05/data-protection-eu-anonymous (accessed 14 Dec 15).

[35] N Sethi and GT Laurie, "Delivering Proportionate Governance in the Era of eHealth: Making Linkage and Privacy Work Together" (2013) 13 *Medical Law International* 168–204.

Portability and Accountability Act of 1996 (HIPAA)[36] is the prime example of the shortcomings of the rules-based approach. The regulation allows otherwise personally identifiable health data to be considered "de-identified" if seventeen specified fields (a person's name, their email address, etc.) and "any other unique identifying number, characteristic, or code" have been removed and if the person responsible for the data does not in fact have knowledge that there is a way to re-identify the data.[37] HIPAA's Safe Harbor has been criticised on the grounds that it is trivial to construct a dataset that meets these requirements yet allows for reasonably easy re-identification and an Institute of Medicine report has found that the regulation inhibits both privacy and openness more than necessary.[38]

Some organisations have made noble attempts to operate in the liminal space of privacy and openness, often by attempting to seamlessly elide them. The strategy has been to invoke transparency in approaches to address privacy and data accessibility. The UK branch of the international Personal Genome Project, for example, describes its ethos as follows:

> We believe sharing is good for science and society. Our project is dedicated to creating public resources that everyone can access. Privacy, confidentiality and anonymity are impossible to guarantee in a context like this research study where public sharing of genetic data is an explicit goal. Therefore, our project collaborates with participants who are fully aware of the implications and privacy concerns of making their data public. Volunteering is not for everyone, but the participants who join make a valuable and lasting contribution to science.[39]

Yet, because privacy remains an important and enduring social value, it is problematic to extrapolate the Personal Genome Project's ethos outside this specific research context. The solution of wholesale abandonment of privacy, almost by definition, cannot be wholly desirable nor feasible as an overarching, long-term solution to legitimate and pressing privacy concerns. Open data and privacy remain in tension and elision is no win-win.

4. Conclusions

Open data offers much promise, yet it also raises complex questions. In this brief analysis piece, I have suggested that open data as a concept is subject to multiple interpretations, but in none does it equate to free data. Collection, curation and distribution of data, as well analysis and repurposing, require requisite technical

---

[36] Health Insurance Portability and Accountability Act of 1996 (HIPAA), Privacy Rule, 45 CFR 160, 164.

[37] United States, *Code of Federal Regulations, Title 45: Public Welfare, Part 164: Security and Privacy*, available at http://www.ecfr.gov/cgi-bin/text-idx?tpl=/ecfrbrowse/Title45/45cfr164_main_02.tpl (accessed 14 Dec 15).

[38] Institute of Medicine, *Beyond the HIPAA Privacy Rule: Enhancing Privacy, Improving Health Through Research* (Washington, D.C.: National Academies Press, 2009), at 2.

[39] Personal Genome Project: UK, "About the PGP" (2015) available at http://www.personalgenomes.org/uk/about-pgp (accessed 14 Dec 15).

know-how, time, cost and infrastructure. Moreover, the various consequences of the Open Data Movement, particularly in its relation to data privacy concerns, have yet to be fully explored. Where data do not transgress established privacy interests or other important social values such as justice, the mission of the Open Data Movement remains unwavering: much data should be treated as collective wealth and a public good, and made open according to the most robust standards. Where privacy interests (both individual and social) are at stake, and justice is in peril, on the other hand, the situation is drastically different. Here, value is to be gained in advocating for positive obligations and data controls, and controlled to some degree by those whose morally relevant interests are affected.

Whatever interpretative approach is adopted toward drawing the bounds of open data, *qualified and reflexive openness* must be the mantra of the Open Data Movement if it wishes sustain itself and the public trust. Recognising that powerful interests are now adopting the language of "open data" to serve purposes that do not necessarily promote either privacy or openness, much less the commonweal,[40] we must all remain attuned to and vigilant about the sociopolitical ramifications of this phenomenon's epistemic and ontological movements.

Acknowledgement. The author thanks Mark Phillips (Centre of Genomics and Policy, McGill University) for his valuable contributions to this article.

[40] JA Johnson, "From Open Data to Information Justice" (2014) 16 *Ethics and Information Technology* 263–274.