HENRIKSEN-BULMER, J., FAILY, S. and KATOS, V. 2018. *Translating contextual integrity into practice using CLIFOD*. Presented at the 2018 Networked privacy workshop: privacy in context: critically engaging with theory to guide privacy research and design, part of the 21st ACM conference on computer-supported cooperative work and social computing (CSCW 2018), 3 November 2018, Jersey City, USA.

Translating contextual integrity into practice using CLIFOD.

HENRIKSEN-BULMER, J., FAILY, S. and KATOS, V.

2018

Translating Contextual Integrity into practice using CLIFOD

Jane Henriksen-Bulmer
Bournemouth University
Fern Barrow, Poole, United Kingdom
jhenriksenbulmer@bmth.ac.uk

Shamal Faily
Bournemouth University
Fern Barrow, Poole, United Kingdom
sfaily@bournemouth.ac.uk

Vasilios Katos
Bournemouth University
Fern Barrow, Poole, United Kingdom
vkatos@bournemouth.ac.uk

Abstract
Public open data increases transparency, but raises questions about the privacy implications of affected individuals. We present a case for using CLIFOD (ContextuaL Integrity for Open Data), a step-by-step privacy decision framework derived from contextual integrity, to assess the hidden risks of making data obtained from Internet of Things (IoT) and Smart City devices before any data is released and made openly available. We believe CLIFOD helps reduce the risk of any personal or sensitive data being inadvertently published or made available by guiding decision makers into thinking about privacy in context and what privacy risks might be associated with making the data available and how this might impact prosumers.

Author Keywords
Privacy; Contextual Integrity; privacy by design; Privacy Risk; Decision Making

ACM Classification Keywords
K.4.1 [Computers and Society: Public policy issues]: Privacy; Use/abuse of power

Introduction
Most businesses rely on data to exist, this may be data relating to orders placed, the financial health of the company, staff records and details of customers and their preferences

and, as a result, we are now in data driven economy with data having become, arguably, the “new oil” fuelling businesses [14]. Internet of Things (IoT) devices in smart cities are leading to the production of even more data. As a result, organisations involved in this sector make decisions based on available data and intelligence derived from such data every day [7].

Significant amounts of the data produced is derived from public data collections, generated as part of open government initiatives; these seek to make public data more accessible by promoting three core values; participation, transparency and collaboration between government and its citizens [10]. Open data is data that may be downloaded, shared and used freely for any purpose by anyone[11], while public open data is open data that originates from a government controlled entity - a public body [12].

Further, as cities grow and costs rise, government and public bodies are keen to find new and innovative ways of improving quality of life for citizens within urban areas. To this end a lot of research and interest has been invested in Smart Cities and IoT devices, aimed at finding sustainable solutions to address some of the challenges this growth brings. These devices generate large volumes of data (*big data*) which, it is envisaged will be made available as open data [1].

However, although publishing open data serves to increase transparency and facilitates growth [4], it equally raises questions about the privacy implications of affected individuals. To what extent are the rights granted under the universal declaration of human rights in 1948 [18] respected (respect for privacy of home, correspondence and family life), when machines make decisions that may directly affect individuals? Moreover, do such decisions contravene privacy regulations like the General Data Protection Regulation (GDPR) in Europe? [3].

A popular method for assessing privacy risk is to conduct a Privacy impact assessment (PIA) [5]. However, while PIA frameworks help practitioners in assessing privacy risks, these frameworks tend to be directed at projects and ensuring the privacy risks or implications of implementing a new process or practice are accounted for, rather than looking at existing practices or processes [8].

We sought to address this gap by creating CLIFOD (ContextuaL Integrity For Open Data); a practical application of Nissenbaum’s theoretical privacy framework, Contextual Integrity (CI) [9]. CLIFOD was devised to facilitate informed decision making about the privacy implications of publishing data as open data by guiding decision makers through a step-by-step privacy risk assessment that considers all the elements of CI [6].

We contend that CLIFOD can be equally applied to consider the privacy implication of any existing processes and practices, such as making data from smart city sensor and IoT devices freely available to prosumers. Thus CLIFOD presents an exemplar model for how CI can be applied in practice to strategically support decision makers in making privacy decisions about the suitability of any data for open publishing.

Contextual Integrity

In explaining CI, Nissenbaum breaks the framework down into considering three key elements; (i) *explanation*, referring to existing data flows and practices; (ii) *evaluation*, assessing how the proposed changes to existing practices and data flows might affect privacy; and (iii) *prescription*, determining what the likely results and impacts of the proposed change in data flow will be ([9], p. 190). Further, in

considering these key elements, CI asks practitioners to consider the data flows by looking at this from three perspectives: actors, attributes and transmission principles. Actors, are the data *sender, receivers or subjects*, referring to those who handle the data (the *sender* and *receiver*) and the individuals whose data is being transmitted (the *data subject(s)*). Attributes are the individual data elements, and the transmission principles refer to how the data flows between actors, i.e. the data flow.

To account for context, Nissenbaum asks that each actor's different roles, and the capacity within which they handle the data is considered. This involves looking at perspectives, inter-actor relations, social and cultural norms and values and any power dynamics which could, or might have, a direct or indirect affect or impact on the data, the data subject or the data flow. Within this context, an actor will have multiple roles, one might depict their social standing and status, another, their work role. The context within which each of these roles is played is important because how the actor handles the data in one context will be different to how they might handle this in another.

For instance, an actor who is a prosumer may be (i) a citizen who contributes data through their smart meter; (ii) an employee of the electricity company tasked with processing smart meter data; and (iii) a friend or neighbour to another electricity customer who also contributes smart meter data. Thus, as a contributor, (*data sender*) the citizen and their neighbour will share data about their energy consumption including location data, time stamps etc. with the electricity company. As an employee, the actor might be processing the data, and therefore, privy to a lot of personal information about their own consumption and the consumption of others. In this scenario, while it would be appropriate for the actor to share their neighbour's usage information with the electricity grid for planning or predicting future capacity requirements, it would not be appropriate for the same actor to share this information with other friends in a social setting, that would constitute a breach of confidentiality and the privacy of the neighbout (*data subject*).

CI and CLIFOD

The CLIFOD framework takes Nissenbaum's three key elements and translates these into 3 phases; explanation, risk assessment and decision. Beneath each phase, the nine decision heuristics (DH) ([9], pp. 181-182) were then used to create a more detailed series of questions about the data, the actors, the transmission principles and the context. For example, DH2 asks that the *prevailing context* be described, e.g. the electricity company employee might work in the office of the local electricity company or work in a wider energy setting such as the national grid. A copy of these questions can be found at [1].

CLIFOD asks decision makers to consider the data transactions (i.e. the data flows) and how any changes, or suggested changes, might affect how the data flows and therefore, the context of the data transaction. CLIFOD was designed to facilitate informed decision making about whether or not a particular data set is suitable for publishing in open format. As part of this, in the decision phase (phase 3), where a data set is considered inappropriate for open data publication, decision makers are asked to consider and identify any mitigating steps that could, or need to, be taken in order to make the data suitable for publication.

We evaluated this framework in a case study setting working in collaboration with a local authority in the UK, where our findings showed that both personal and sensitive data

[1] Available to download at https://github.com/JaneHB/CIOpenData

had already been published in open format, which could have been prevented had they used CLIFOD prior to publication [6].

CLIFOD and Smart City Data

Smart Cities and IoT devices is a current focus for governments and public bodies, with a lot of work being conducted into how these technologies can help improve the lives of citizens [13]. To this end, big data gathered from sensors and other devices has been used to improve the quality of public service delivery [2], and geolocation data has been utilised to assist citizens with details of the safest route to their desired destination [17]. Moreover, citizens also contribute to this body of data by producing data from their IoT devices (e.g. smart meters), making them *prosumers*, as they both provide and supply data to the technical solutions being developed for smart cities.

Smart city and IoT data has been used to inform, for example, urban environment development and improve sustainability [15]. Further, because the data generated from these devises is also potentially made available in open format [1], it also provides prosumers with opportunities to capitalise on using the data, thereby facilitating sustainability and growth opportunities as well [4],[13].

While the intention of collecting and processing big data is invariably good, the data produced from these smart city and IoT devices may also contain personal or sensitive data, e.g. location data from the smart meter which may be used to determine supply levels. From the perspective of the individuals whose data is being analysed, the likelihood is they have had no input into these decisions and how they were arrived at, yet the decisions made may directly affect their lives. Thus, this may result in unintended consequences, and a potential threat to the privacy of the individual whose data may be processed.

For example, big data and automated decisions have been used to gauge individuals electricity usage. It has also been used to generate *heat maps* to predict which geographical areas are more likely to have higher crime rates [16]. The intention behind creating these maps is to reduce crime and deploy police resources where they are needed the most. One resulting unintended consequence of this however, might be a potential increase in crime in *heat map* high crime areas, because these have higher police presence. This effect, in turn, may negatively impact the heat map and distort the risks for already 'marked' high risk areas. Another unintended consequence would be a negative impact on house prices in that area because of the high risk rating. This would affect the citizens who live in the area by, for example, making their properties less desirable and reducing the property value. Most of these outcomes would arguably have a direct negative impact on the individual's quality of life. Therefore, some method of sorting this data and assessing the privacy risks and unintended consequences needs to be found.

We argue that, if decision makers had used CLIFOD to assess the risks associated with using smart city or IoT device data, some of these unintended consequences might have been avoided. For instance, while CLIFOD would not necessarily ask specific questions pertaining to this scenario, the process of answering the questions within CLIFOD would have asked first, what information would be gathered by whom and for what purpose. Next, the assessment would have asked that the decision makers (i.e. the assessors) consider and establish the context surrounding collection of this data and how it was used. Thus, taking the heat map instance, completing CLIFOD would show that an authoritative body (the police in this instance), was us-

ing machine learning to predict which parts of an area were more likely to be subjected to crime. This should, for the assessors, start alarm bells ringing and lead them to ask more details questions about the situation. These questions should include enquiry into, for example, what underlying data was used to make these calculations? and what level of human input or intervention went into the machine learning? In addition, questions need to be asked about where the data originated from?, is it based on historical crime data or collected from IoT or smart city devices who monitor different areas? From this, the assessors might then ask more probing questions around who has access to the data and whether this is shared with external bodies, and to what extend do citizens contribute data to this body of data? are they indeed prosumers as well as consumers?

Conclusion

If data producers (whether public or private entities) were to consider the privacy risks and potential impact of making the data available prior to releasing the data, we believe the likelihood and potential impact of any unforeseen consequences can be minimised. To achieve this, we contend that CI and CLIFOD may provide a first step towards a practical solution. Using CLIFOD before data is released will help organisations and/or public bodies to assess the potential *hidden* consequences of making data available before it is too late [6]. Further, by facilitating informed decisions being made up front, the likelihood of unforeseen consequences can be pre-empted and mitigation strategies can be put in place to avoid these pitfalls. We hope that this workshop will provide an interesting forum to discuss how CLIFOD can be improved to fully support these aims.

Acknowledgements

This work has received funding from the EU's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 778229 (IdealCities).

REFERENCES

1. Alliance for Telecommunications Industry Solutions. 2018. *Smart Cities Data Sharing Framework*. Technical Report ATIS-I-0000063. Alliance for Telecommunications Industry Solutions, Washington DC. http://www.atis.org/smart-cities-data-sharing/smart-cities-data-sharing.pdf

2. L. M. Amugongo, S. N. Nggada, and J. Sieck. 2016. Leveraging on open data to solve city challenges: A case study of Windhoek municipality. In *2016 3rd MEC International Conference on Big Data and Smart City (ICBDSC)*. IEEE, 1–6. DOI: http://dx.doi.org/10.1109/ICBDSC.2016.7460355

3. Maja Brkan. 2017. AI-supported Decision-making Under the General Data Protection Regulation. In *Proceedings of the 16th Edition of the International Conference on Articial Intelligence and Law (ICAIL '17)*. ACM, New York, NY, USA, 3–8. DOI: http://dx.doi.org/10.1145/3086512.3086513

4. Michael Chui, Diana Farrell, and Kate Jackson. 2014. *How Government can promote open data*. Technical Report. McKinsey & Company. http://www.mckinsey.com/industries/public-sector/our-insights/how-government-can-promote-open-data

5. Wright David, Finn Rachel, and Rodrigues Rowena. 2013. A Comparative Analysis of Privacy Impact Assessment in Six Countries. *Journal of Contemporary European Research, Vol 9, Iss 1, Pp 160-180 (2013)* 9, 1 (2013), 160 – 180.

6. Jane Henriksen-Bulmer and Shamal Faily. 2017. Applying Contextual Integrity to Open Data Publishing. In *Proceedings of the 31st British HCI Group Annual Conference on People and Computers: Digital Make Believe*. British Computer Society.

7. Dennis D. Hirsch. 2013. Glass House Effect: Big Data, the New Oil, and the Power of Analogy, The [article]. *Maine Law Review* 66, 2 (2013), 373.

8. Information Commissioners Office. 2014. *Conducting privacy impact assessments: code of practice*. Technical Report 20140225. Information Commissioners Office (ICO).

9. Helen Fay Nissenbaum. 2010. *Privacy in Context: Technology, Policy, and the Integrity of Social Life.* Stanford Law Books, Stanford: California.

10. Barack Obama. 2009. Transparency and Open Government: Memorandum for the Heads of Executive Departments and Agencies. (March 2009). https://www.whitehouse.gov/the_press_office/TransparencyandOpenGovernment

11. Open Data Institute. 2016. What is open data? (2016). https://theodi.org/what-is-open-data

12. Open Government Data Working Group. 2016. Open Government Data. (2016). http://opengovernmentdata.org/

13. Ahmed M. Shahat Osman, Ahmed Elragal, and Birgitta Bergvall-Kåreborn. 2017. Big Data Analytics and Smart Cities : A Loose or Tight Couple?. In *Proceedings of the International Conference on ICT, Society and Human Beings 2017 : Part of the Multi Conference on Computer Science and Information Systems 2017*. IADIS, 157–168.

14. Michael Palmer. 2006. Data is the new oil. [online]. (November 2006). http://ana.blogs.com/maestros/2006/11/data_is_the_new.html

15. S. Patel, Uday Kumar R.Y., and Prasanna Kumar B. 2016. Role of smart meters in smart city development in India. In *2016 IEEE 1st International Conference on Power Electronics, Intelligent Control and Energy Systems (ICPEICES)*. IEEE, 1–5. DOI: http://dx.doi.org/10.1109/ICPEICES.2016.7853363

16. David Robinson, Harlan Yu, and Aaron Rleke. 2014. *Civil Rights, Big Data, and Our Algorithmic Future*. Technical Report. Upturn. http://centerformediajustice.org/wp-content/up-loads/2014/10/Civil-Rights_Big-Data_Our-Future.pdf

17. G. B. Rocca, M. Castillo-Cara, R. A. Levano, J. V. Herrera, and L. Orozco-Barbosa. 2016. Citizen security using machine learning algorithms through open data. In *2016 8th IEEE Latin-American Conference on Communications (LATINCOM)*. IEEE, 1–6. DOI: http://dx.doi.org/10.1109/LATINCOM.2016.7811562

18. The General Assembly of the United Nations. 1948. The Universal Declaration of Human Rights. [online]. (December 1948).