Data management and use: case studies of technologies and governance

Produced for the British Academy and the Royal Society

Table of Contents

1. Acknowledgments

These case studies were produced for the British Academy and Royal Society working group on *Data Management and Use: Governance in the 21st Century* by Michael Veale, Doctoral Researcher: Responsible Public Sector Machine Learning at University College London. The detailed case study on Smart Meters includes input on the current regulatory landscape from Dr Fernanda Ribas.

The British Academy and the Royal Society would also like to acknowledge the following individuals for providing comment and input to the documents:

- Professor Ian Brown — University of Oxford
- Professor Geoff Gilbert — University of Essex
- Dr Yves-Alexandre de Montjoye — Imperial College London
- Dr Dirk Schaefer — University of Bath
- Professor Tim Baines — Ashton University
- Dr Ali Z Bigdel, — Ashton University
- Ian McKechnie — Ashton University
- Eleanor Musson — Ashton University

2. Introduction

These case studies were prepared in support of the *Data Management and Use: Governance in the 21st Century* project, carried out by the British Academy and the Royal Society. The purpose of these case studies was to stimulate thinking and discussion by the working group, in their deliberations on governance needs for data management and use across all sectors. This publication is an edited version of the case studies presented to the working group.

The case studies aim to give concrete examples of the kinds of social and ethical tensions that arise in contemporary data use and management – and they draw on the sets of social and ethical pairings, detailed in the main report of this project, that were identified at a cross-disciplinary expert workshop held in July 2016. They give current and forward-looking examples of the benefits and challenges of data collection, management and use across a range of sectors and the governance needs in different contexts.

Intended as they are solely to inform the working group, these case studies are not presented as the views of either Academy, and are not intended as appraisals or evaluations of any of the governance approaches identified in them. However they illustrate some of the issues that prompted the work behind the *Data Management and Use: Governance in the 21st Century* report.

The case studies were developed using desk research and informal interviews with researchers. Each case study has been reviewed by a relevant expert, as listed in the acknowledgements above.

3. Detailed Case Study: Smart Metering

3.1 Introduction

Smart meters are devices capable of real-time measurement and transmission of household electricity consumption. The UK Government, similar to many others around the world, aims to 'ensure that every home and business in the country is offered a smart meter by 2020, delivered as cost effectively as possible'. Rollout is non-compulsory, led by BEIS[1], regulated by Ofgem, and primarily funded and fulfilled by the energy suppliers. They promise to provide new opportunities for innovative markets, efficiencies in transmission, maintenance and billing, and spillover effects concerning 'smart homes' more generally. At the same time, they raise concerns around privacy, proportionality and security of energy systems.

This extended case study sets out the social and ethical tensions that are relevant to the management and use of (primarily domestic) energy data, from a selection discussed in greater length in the main report, *Data Management and Use: Governance in the 21st Century*. It considers the opportunities and challenges in using smart meter data, and looks at the ways that data management and use can be governed through technology and institutions. Finally it gives an overview of the governance arrangements in place in the UK and Europe. The case study also illustrates the challenge for governments in deploying cutting edge, technical data governance solutions where there is a one-off national roll-out of an infrastructure system.

3.2 Social and ethical issues in smart meter data management and use

Protecting personal information while safely using and linking open and non-sensitive data

The individual accounting of data at household level is the source of many of the attributed benefits of smart metering, yet also serves as the source of many of the identified risks. This individual accounting is essential for accurate billing, for example. However, live electricity consumption data can enable the inference of private data such as when you are at home to what you do when you are there.

Proportionality in the use of data while using data to protect public safety and wellbeing

The detailed inferences possible with smart metering data might provide opportunities for social uses, such as targeting services to vulnerable individuals, or building more detailed

[1] The rollout was formerly led by the Department for Energy and Climate Change (DECC) before departmental reorganisation in 2016 shifted the portfolio to the new Department of Business, Energy and Industrial Strategy.

maps of deprivation. However, consumers may be wary of such data collection, and maintaining public trust could prove difficult where there are demands for data repurposing.

Autonomy for individuals and communities while using data to achieve commercial benefit and efficiency in public services

In the UK, a single, centralised model for smart meter data governance was promoted by the regulatory approach, to create a specific new licensed body (a Data and Communications Company) for data processing in domestic energy regulation. To attempt to manage the privacy issues within this model, consumers were provided with rights over the granularity of data they send. However, this creates a challenge as European cost-benefit analyses demonstrate that the systems are only broadly cost effective to install if operational benefits are largely realised, based on detailed data being available. As discussed later in this example, the German system illustrates a way of addressing this tension.

3.3 Opportunities for use of smart meter data

Smart meters present opportunities for both energy efficiency, through incentivising the consumer and better grid management, and for future business models, such as the charging of electric vehicles away from home. Previous studies have distinguished between benefits emerging from better management of aggregated operations and benefits which emerge from more flexible billing[2]. A third category can be added: spillover benefits that might emerge from the repurposing of the metering infrastructure. Considering these opportunities together has been an important part of the calculus for smart meter installation. At a European level, the cost of smart meter installation is not fully offset by operational savings, requiring dynamic pricing provisions to make the present value of benefits overtake the costs[3].

Operational opportunities can result in efficiency savings for energy networks and providers. Data on electrical usage has previously only been routinely possible to obtain at a substation level at useable temporal resolutions. Smart meters providing data at higher resolution are expected to provide a range of benefits to suppliers and other energy decision-makers. These benefits include the better *projection of future network capacity requirements*; enabling *preventative maintenance*; and aiding in *understanding of faults, outages and quality issues*[4].

Billing opportunities allow for the creation of new markets, and are promoted as enhancing choice and efficiency. Capture of electricity usage information at a higher temporal resolution creates opportunities for *new tariff models*. Demand responsive pricing might serve to lower consumption at peak times, as well as incentivise action on energy wastage. Smart meters might also *incentivise consumer energy generation* through responsive feed-in tariffs[5], and

[2] Jawurek M, Kerschbaum F, George D. 2012 SoK: Privacy technologies for smart grids – A survey of options. *Microsoft Research*. See https://www.microsoft.com/en-us/research/wp-content/uploads/2012/11/paper.pdf (accessed 26 June 2017); Finster S, Baumgart I. 2015 Privacy-aware smart metering: A survey. *IEEE Communications Surveys Tutorials*. **17**, 1088-1101. (doi:http://dx.doi.org/10.1109/COMST.2015.2425958)

[3] Faruqui A, Harris D, Hledik R. 2010 Unlocking the €53 billion savings from smart meters in the EU: How increasing the adoption of dynamic tariffs could make or break the EU's smart grid investment. *Energy Policy*. **38**, 6222-6231. (doi: https://doi.org/10.1016/j.enpol.2010.06.010)

[4] Depuru SSSR, Wang L, Devabhaktuni V, Gudi N. 2011 Smart meters for power grid: Challenges, issues, advantages and status. *Renewable and Sustainable Energy Reviews*. **15**. (doi:http://dx.doi.org/10.1109/PSCE.2011.5772451)

[5] Römer B, Reichhart P, Kranz J, Picot A. 2012 The role of smart metering and decentralized electricity storage for smart grids: The importance of positive externalities. *Energy Policy*. **50**, 486-495. (doi:https://doi.org/10.1016/j.enpol.2012.07.047)

provide *cost savings using flexible energy storage devices*[6]. The ability to issue bills in a more timely fashion can strengthen this incentive link, in addition to reducing labour costs in call-centres and meter reading.

Smart meters also increase the capabilities of suppliers to *detect and act on fraud and non-payment*. Fraud detection can cost suppliers greatly, and increase electricity costs in general. British Gas reports that electricity theft costs the industry £400–500m a year, while worldwide energy theft equates to the total installed generation capacity of the UK, Germany and France together[7]. Even partial prevention might significantly reduce carbon emissions[8].

Some smart meters have the capability to be disabled remotely, which might be used to deter fraud and non-payment. Remote management of meters can aid in the management of pre-pay meters online and allow *accurate billing of costs* while away. It also allows for new business models, particularly around flexible charging of electric cars while away from home[9].

Infrastructural benefits have been highlighted in relation to the installation of these technologies. Smart meters allow the better *visualisation of energy consumption*, which might serve to reduce consumption[10]— although this can be accomplished 'offline' without transmission beyond the home[11]. Smart meters have also been proposed as *hubs for 'Internet of Things' technologies*, due in part to their wide public rollout. These hubs present both opportunities and risks, as discussed below.

3.4 Challenges in the management and use of smart meter data

In addition to these opportunities, smart meters also present certain privacy and security hurdles.

Privacy concerns stem from the fact that *real-time data from smart meters can leak information a user might want to remain private*. There are several categories of information possible to infer from smart meter readings:[12]

- Which appliances are being used at particular times[13]

[6] Malhotra A, Battke B, Beuse M, Stephan A, Schmidt T. 2016 Use cases for stationary battery technologies: A review of the literature and exisiting projects. *Renewable and Sustainable Energy Reviews*. **56**, 705-721. (doi:https://doi.org/10.1016/j.rser.2015.11.085); Stephan A, Battke B, Beuse MD, Clausdeinken JH, Schmidt TS. 2015 Limiting the public cost of stationary battery deployment by combining applications. *Nature Energy*. **1**. (doi: http://dx.doi.org/10.1038/nenergy.2016.79)

[7] Depuru SSSR, Wang L, Devabhaktuni V. 2011 Electricity theft: Overview, issues, prevention and a smart meter based approach to control theft. *Energy Policy*. **39**, 1007-1015. (doi:https://doi.org/10.1016/j.enpol.2010.11.037)

[8] Pyasi A, Verma V. 2008 Improvement in electricity distribution efficiency to mitigate pollution IEEE ISEE. *IEEE International Symposium on Electronics and the Environment.* (doi:dx.doi.org/10.1109/ISEE.2008.4562863)

[9] Clement-Nyns K, Haesen E, Driesen J. 2010 The impact of charging plug-in hybrid electric vehicles on a residential distribution grid. *IEEE Transactions on Power Systems*. **25**, 371-380. (doi:http://dx.doi.org/10.1109/TPWRS.2009.2036481)

[10] Fischer C. 2008 Feedback on household electricity consumption: A tool for saving energy? *Energy Efficiency*. **1**, 79-104. (doi:http://dx.doi.org/10.1007/s12053-008-9009-7)

[11] To note, some studies cast doubt on the ability of current technologies to incentivise lower energy consumption through visualisation alone. Buchanan K, Russo R, Anderson B. 2015 The question of energy reduction: The problem(s) with feedback. *Energy policy.* **77**, 89-96. (doi:https://doi.org/10.1016/j.enpol.2014.12.008)

[12] Jawurek M, Kerschbaum F, George D. 2012 SoK: Privacy technologies for smart grids – A survey of options. *Microsoft Research*. See https://www.microsoft.com/en-us/research/wp-content/uploads/2012/11/paper.pdf (accessed 26 June 2017).

[13] Molina-Markham A, Shenoy P, Fu K, Cecchet E, Irwin D. 2010 Private memoirs of a smart meter. *Proceedings of the 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building*. See http://dl.acm.org/citation.cfm?doid=1878431.1878446 (accessed 2 November 2016).

- How these appliances are used. Attacks by researchers have been able to detect particular TV channels or websites visited from electricity data[14]
- Behavioural patterns deduced from patterns of appliance use[15]
- Inferred information about other utilities or features of the building[16]
- Information that can be used to identify other pseudonymised records[17].

This information can lead to the inference of private knowledge. This knowledge might include times you tend to be home; length and frequency of your showers; whether you are protected by an alarm; how often you are out late; whether you tend to leave appliances on[18].

Security concerns can stem from remote shutdown, which can present a risk to national infrastructure if meters are insecure – allowing attacks to be targeted at times of peak demand, and potentially leading to grid damage[19],[20]. Understanding security threats to smart grids requires a cyber-physical framing, where both cyber attacks and physical attacks (and combinations of both) have both cyber and physical consequences[21]. There is also a risk of unauthorised third parties obtaining data flows. These concerns could be compounded by the potential for smart meters serve as 'hubs' for connected domestic 'IoT' devices, making them attractive targets for cybercriminals. Significant concerns have already been raised around the security protocols in meter-reading devices in the US[22]. Risks are exacerbated where homogenous hardware could create systemic vulnerabilities.

Proportionality and consent are issues in accessing smart meter data. There is a trade-off between giving detailed information that might be important to some users, and ensuring that consumers can understand the information presented to them to give truly informed consent (known as the 'transparency paradox)[23]. Secondly, tariffs using more granular data are likely

[14] Enev M, Upgta S, Kohno T, Patenl SN. 2011 Televisions, video privacy, and powerline electromagnetic interference. *Proceedings of the 18th ACM Conference on Computer and Communications Security.* (doi:http://doi.acm.org/10.1145/2046707.2046770); Greveler U, Glösekötter P, Justus B, Loehr D. 2012 Multimedia content identification through smart meter power usage profiles. *Proceedings of the International Conference on Information and Knowledge Engineering (IKE)* (The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp) 2012). See http://search.proquest.com/openview/059b0c797d4580cd7419dc90a2462602/1?pq-origsite=gscholar (accessed 26 June 2017); Clark SS, Ransford B, Sorber J, Xu W, Learned-Miller E, Fu K. 2013 Current events: Identifying webpages by tapping the electrical outlet. In Crampton J, Jajodia S, Mayes K (eds), *Computer Security – ESORICS 2013* (Springer Berlin Heidelberg 2013). See http://link.springer.com/chapter/10.1007/978-3-642-40203-6_39 (accessed 2 November 2016).

[15] Lisovich MA, Mulligan DK, Wicker SB. 2010 Inferring personal information from demand-response systems. *IEEE Security Privacy.* **8**, 11. (doi:http://dx.doi.org/10.1109/MSP.2010.40);
Beckel C, Sadamori L, Thorsten S, Silvia S. 2014 Revealing household characteristics from smart meter data. *Energy*. **78**, 397-410. (doi:http://dx.doi.org/10.1016/j.energy.2014.10.025)

[16] Beckel C, Sadamori L, Thorsten S, Silvia S. 2014 Revealing household characteristics from smart meter data. *Energy*. **78**, 397-410. (doi:http://dx.doi.org/10.1016/j.energy.2014.10.025)

[17] Jawurek M, Johns M, Rieck K. 2011 Smart metering de-pseudonymization. *Proceedings of the 27th Annual Computer Security Applications Conference*. (doi:http://doi.acm.org/10.1145/2076732.2076764); Tudor V, Almgren M, Papatriantafilou M. 2015 A study on data de-pseudonymization in the smart grid. *Proceedings of the Eighth European Workshop on System Security.* (doi:http://dl.acm.org/citation.cfm?doid=2751323.2751325)

[18] Quinn EL. 2009 Privacy and the New Energy Infrastructure. (doi:http://dx.doi.org/10.2139/ssrn.1370731)

[19] Anderson R, Fuloria S. 2010 Who Controls the off Switch? *First IEEE International Conference on Smart Grid Communications*. (doi:http://dx.doi.org/10.1109/SMARTGRID.2010.5622026)

[20] At a major 2014 security conference a flaw in Spanish smart meters was demonstrated which could theoretically result in an attacker being able to turn off the electricity in a wide region, potentially also causing significant grid damage. Illera AG, Vidal JV. 2014 Lights off! The darkness of the smart meters. *BlackHat Europe*. See https://www.youtube.com/watch?v=Z_y_vjYtAWM (accessed 26 June 2017).

[21] Mo Y *et al.* 2012 Cyber–physical security of a smart grid infrastructure. *Proceedings of the IEEE.* **100**, 195-209. (doi:http://dx.doi.org/10.1109/JPROC.2011.2161428).

[22] Rouf I, Mustafa H, Xu M, Xu W, Miller R, Gruteser M. 2012 Neighborhood watch: Security and privacy analysis of automatic meter reading systems. *Proceedings of the 2012 ACM conference on Computer and communications security*. (doi:http://dx.doi.org/10.1145/2382196.2382246).

[23] Nissenbaum H. 2011 A contextual approach to privacy online.*Daedalus*. **140**, 32-48. (doi:http://dx.doi.org/10.1162/DAED_a_00113)

to be cheaper, meaning that lower income households might be pushed to consent to greater data sharing for financial reasons.

Several technical solutions for these issues have been proposed, many using cutting-edge research in privacy-enhancing technologies to maximise the aggregate benefits of smart meters while not revealing individuals' individual consumption patterns. Yet these technologies have not been utilised in emerging governance systems, which differ between countries. The UK has chosen to centralise all data processing and access through a single new company while Germany has created a market for a new certified category of device, the smart meter 'gateway', which allows personalised end-to-end encryption between the home and various data users.

3.5 Governance of smart meter data management and use

Governance of smart meters to date can be seen through two interlinked lenses of **technical governance** and **institutional governance**.

Technical governance: Privacy-enhancing technologies

Privacy issues relating to smart meters have been the subject of many proposed **privacy-enhancing technologies (PETs),** which can be seen as a form of technical governance. PETs are

> *a system of ICT measures protecting informational privacy by eliminating or minimising personal data thereby preventing unnecessary or unwanted processing of personal data, without the loss of the functionality of the information system.*[24]

A range of PETs are applicable to both the operational and the billing functions of smart meters[25].

Operationally-relevant PETs attempt to maintain user informational privacy while allowing suppliers and other organisations to benefit from the rich information and functionality that these systems provide – potentially addressing the tension between protecting personal information and making use of non-sensitive data. A core privacy aim is to allow aggregate summary statistics of multiple readings without revealing information from individual meters. *Anonymisation* tools seek to remove the links between data creators and data users in ways that still allow desired calculations to be carried out. However, anonymisation tools can suffer from re-identification attacks and there is a trade-offs between more thorough anonymisation techniques and the accuracy of statistics generated from this data[26]. *Perturbation* seeks to

[24] Several wide-ranging surveys of these systems exist, see: Van Blarkom GE, Borking JJ, Olk J. 2003 Handbook of privacy and privacy-enhancing technologies. *Privacy Incorporated Software Agent (PISA) Consortium, The Hague*. See http://www.andrewpatrick.ca/pisa/handbook/Handbook_Privacy_and_PET_final.pdf (accessed 26 June 2017).

[25] Jawurek M, Kerschbaum F, George D. 2012 SoK: Privacy technologies for smart grids – A survey of options. *Microsoft Research*. See https://www.microsoft.com/en-us/research/wp-content/uploads/2012/11/paper.pdf (accessed 26 June 2017); Finster S, Baumgart I. 2015 Privacy-aware smart metering: A survey. *IEEE Communications Surveys Tutorials*. **17**, 1088-1101. (doi:http://dx.doi.org/10.1109/COMST.2015.2425958)

[26] Daries J *et al*. 2014 Privacy, anonymity, and big data in the social sciences.*Communications of the ACM*. **56**. See http://dl.acm.org/citation.cfm?doid=2663191.2643132 (accessed 3 November 2016); Angiuli O, Waldo J. 2016 Statistical tradeoffs between generalization and suppression in the de-identification of large-scale data set. *IEEE 40th Annual Computer Software and Applications Conference (COMPSAC)*. (doi:http://dx.doi.org/10.1109/COMPSAC.2016.198)

introduce noise into readings that preserves some privacy while maintaining the utility of the resultant calculations. *Cryptographic computation* encrypts data using a homomorphic protocol[27] which allows mathematical operations to be carried out while data is still encrypted, while allowing the decryption of the final result but not any of the individual parts.

PETs designed for billing require the waiving of anonymity, and the guarantee of accuracy, in order to correctly charge customers, and to enable dynamic pricing and tariffs. Billing however does not have to be reported in real time, but can be sent in batches at a coarse enough temporal resolution as to suppress most privacy concerns. This, however, can cause issues with the desire of suppliers to use dynamic pricing and tariffs. One technical solution to this is *secure multi-party computation*, which assumes the existence of one or more third parties external to the data user. The data user receives billing information resulting from other parties jointly computing a function over split inputs while keeping those inputs private to each other. A different solution, and one which avoids the use of third parties, is *verifiable computing*. Here the data creator themselves provides a record alongside the aggregate consumed which mathematically proves a computation was carried out in a particular way, while providing zero knowledge beyond the veracity of the calculation.

Institutional governance

There are a range of institutional governance needs in relation to data collection; transmission; processing, storage and access; and interoperability. This section sets out some of the responses to those needs.

Data collection

The Netherlands was an early actor in smart meter governance[28]. Initial legislative plans developed in 2006 focussed on energy efficiency, and mandated citizens to install a smart meter that would transfer data every fifteen minutes and was enabled with remote shutdown functionality. Failure to install would be met with a 17,000 EUR fine or even maximum of six months imprisonment. A report commissioned by the Dutch Consumer organisation, Consumentenbond, argued that the proposed rollout would violate Article 8 of the European Convention on Human Rights, the respect for private and family life[29]. This led to new legislation being proposed giving Dutch consumers both options to refuse smart meters with administrative shutdown capabilities or the ability to read a continuous stream of data, and to reject the installation of a smart meter entirely.

The Dutch case highlighted the phenomenon of 'function creep' with regards to data governance. The initial European-level proposals which led to the introduction of the smart meter legislation specified only the use of smart meters for providing information to end-users to help them save energy. The Dutch implementation of such legislation added abilities to combat electricity fraud and remote activation and shutdown, which were two of the features that caused public outrage.

---

[27] Gentry C. 2009 A fully homomorphic encryption scheme. *Stanford University*. See http://cs.au.dk/~stm/local-cache/gentry-thesis.pdf (accessed 26 June 2017).
[28] Cuijpers C, Koops B. 2013 Smart metering and privacy in Europe: Lessons from the Dutch case. In Gutwirth S *et al* (eds), *European Data Protection: Coming of Age* (Springer Netherlands 2013). See http://link.springer.com/chapter/10.1007/978-94-007-5170-5_12 (accessed 3 November 2016).
[29] Cuijpers C, Koops B. 2008 Het Wetsvoorstel "slimme Meters": Een Privacytoets Op Basis van Art. Universiteit van Tilburg. See https://www.vrijbit.nl/images/stories/files/pdf/onderzoek_uvt_slimme_energi1.pdf (accessed 26 June 2017).

In the UK, the then Department of Energy and Climate Change (DECC) began in 2010 to discuss privacy in relation to their own plans for smart meter rollout[30]. The UK consumer group, Consumer Focus, using their statutory powers found that trials of smart readers were already collecting half-hourly data readings.

While DECC was drawing up final plans, with reference to a call for evidence and a consultation, the Article 29 Working Party offered an opinion[31] suggesting that data from smart meter operations was personal data, and therefore data protection legislation applied to it. In the same opinion, they warned that the creation of detailed profiles might not be in line with data subjects' legitimate interests, and was not needed to achieve the basic purposes of smart metering.

Technical governance mechanisms, such as those described above, have not found great traction in the law, although there is general support for them in both the Data Protection Directive and the General Data Protection Regulation which takes force in Europe in 2018. For measurement relating to billing, metrology law[32] covers device accuracy, an important component of the governance of data collection and sensor validity.

Data transmission

The transmissions techniques used by smart meters vary widely[33]. The cheapest is power-line communication (PLC), which uses existing electrical cables to simultaneously carry data. PLC requires less fixed investment than other options, yet can provide lower signal qualities and data speeds[34], which might serve to limit the type of technical governance possible, particularly for types of trusted computing that require significant network interaction. Other options proposed have included SMS over GSM, local wireless and bluetooth networks, mobile 3G and 4G telecoms, and standard telephone lines. Where the meters use broader infrastructure, part of the governance of data during its transmission, including issues such as its security, falls under existing governance systems.

In the UK case, a mix of existing infrastructure and new infrastructure will be used to connect smart meters to a central data services provider (DSP) that will distribute data to suppliers, distribution network operators (DNOs) and authorised third parties. In the south and central parts of the UK, this infrastructure will be provided by Telefónica using existing cellular networks combined with a wireless mesh network in low-coverage areas, while Arqiva will build new infrastructure in the north.

Before the rollout of smart meters with the national data management system in the UK, many suppliers already installed 'remote access meters' with similar functionality. In 2015, Ofgem decided to extend the privacy requirements for smart meters to remote access meters, following a consultation[35], yet these plans faced hardware challenges, as many of the smart meters already installed were unable to be reconfigured to the consumer's wishes. Ofgem,

[30] Brown I. 2014 Britain's smart meter programme: A case study in privacy by design. *International Review of Law, Computers & Technology.* **28**, 172-184. (doi:http://dx.doi.org/10.1080/13600869.2013.801580)
[31] 00671/11/EN WP 183 Article 29 Data Protection Working Party: Opinion 12/2011 on smart metering.
[32] Of particular relevance is the Measuring Instruments Directive (2004/22/EC).
[33] Khalifa T, Naik K, Nayak A. 2010 A survey of communication protocols for automatic meter reading applications. *IEEE Communications Surveys Tutorials*. **13**, 168-182. (doi:http://dx.doi.org/10.1109/SURV.2011.041110.00058)
[34] Gungor VC *et al*. 2012 Smart grid and smart homes: Key players and pilot projects. *IEEE Industrial Electronics Magazine*. **6**, 18-34. (doi:http://dx.doi.org/10.1109/MIE.2012.2207489)
[35] Ofgem. 2015 Decision on extending the smart meter framework for data access and privacy to remote access meters. See https://www.ofgem.gov.uk/ofgem-publications/93187/dataprivacyextension-decision1.pdf (accessed 26 June 2017).

with the support of the Information Commissioner's Office (ICO), proposed that for those systems, the companies transmitting from the remote meters to the suppliers (so-called 'head end operators') are compelled by the regulator to filter user data in alignment with the given consent.

Data processing, storage and access

Different models of data processing and access exist in relation to smart meters. This section summarises three different implementations, from the UK, the Netherlands and Germany.

In **the UK**, data from smart meters is fed to a centralised authority, known as the Data and Communications Company (DCC)[36], envisioned as a central point for data to be provided to suppliers, operators of distribution systems, as well as authorised third parties[37]. The processes behind authorisation to use this data are split between government and regulators, which will oversee the processes underlying access to these data streams; and consumers themselves, required to consent to the making available of any data beyond that used for billing – though customers will not have to consent where data are 'required to fulfil regulated duties' [38]. In addition, while meters will be installed with some simple variable tariffs, such as algorithms based on common times of day, consent to transmission of more granular data will be necessary for consumers to access more advanced tariffs that are likely to save them money. No technical privacy-enhancing technologies are being publicly deployed within the data accessible to the DCC[39], and it was reported that DECC saw these technologies as yet immature for use, and industry argued that designers of privacy-enhancing technologies 'didn't understand the industry'[40].

In **the Netherlands**, the data from smart meters is fed into a central system, which often operates at a substation level. From this level, it flows to the operator of the distribution system and to energy suppliers. In theory, pre-processing *could* happen at this neighbourhood level, such as aggregation, where the supplier and other parties need only know that the substation was a trusted processor, rather than access consumers' detailed data. No technical governance of this sort has yet been published alongside the broader system design, however.

Germany presents a contrasting case to both the UK and the Dutch approaches. Firstly Germany affords the consumer the right to switch to a third party organisation to operate their smart meter and to collect their data. Secondly, as opposed to a system with data 'funnels' such as the DCC, data is centralised in consumers' homes through the use of a 'gateway' device. This is a device certified by the regulator, and provided by a consumer-chosen 'measuring point operator' which links to one or more smart meters.[41] The underlying data transfer principle is one of 'star-shaped communication', where this gateway can communicate with different protocols with the different users of smart meter data and third parties on the

[36] This organisational role was put out to tender and granted to Smart DCC Limited, a subsidiary of Capita plc.
[37] For example, the suggestion has been made that Citizens Advice could use such data to provide targeted services to vulnerable individuals: House of Commons Science and Technology Committee, *Evidence Check: Smart Metering of Electricity and Gas* (House of Commons 2016).
[38] They can however access at least the previous 13 months of data locally on their own meter. See Connor PM *et al*. 2014 Policy and regulation for smart grids in the United Kingdom. *Renewable Sustainable Energy Rev.* **40**, 269-286. (doi:http://dx.doi.org/10.1016/j.rser.2014.07.065)
[39] Ibid.
[40] Brown I. 2014 Britain's smart meter programme: A case study in privacy by design. *International Review of Law, Computers and Tech.* **28**, 172-184. (doi:http://dx.doi.org/10.1080/13600869.2013.801580)
[41] See the Gesetz zur Digitalisierung der Energiewende [trans: Law for the Digitisation of the Energy Transition] (adopted August 29 2016).

basis of consumer authorisation. This implies end-to-end encrypted communication with no middle data processing body[42].

It is important to note that products that somewhat resemble smart meters, and interact with home appliances, such as smart thermostats, may interact with smart technologies. By the same logic of the Article 29 Working Party in relation to government smart meter roll out, this data is also personal data, and therefore is similarly covered by data protection legislation. It does not, however, benefit from any of the extra rules that national governments have placed on data use and consent in specific relation to smart meter data, and consumers may not be aware that two or more transmission and processing activities are happening, often using two distinct transmission modalities.

Data interoperability

Meters adopted in the UK prior to mass roll-out pose significant barriers to interoperability. Data collection systems and transmission systems would not necessarily be able to work in 'smart' mode when supplier was switched. DECC noted that interoperability of these systems and the transfer of data-driven services was 'subject to agreement between energy suppliers', noting that the industry trade body Energy UK was 'working with energy suppliers on interim commercial and technical solutions for increasing the likelihood of consumers keeping a smart service when they switch' in a self-governance mode.[43]

3.6 The existing regulatory structure in the UK and Europe

Core regulatory bodies

The **Department for Business, Energy and Industrial Strategy (BEIS)**[44] holds the ministerial smart metering portfolio, previously held in the Department of Energy and Climate Change before a merger of departments in July 2016. On 30 March 2011, DECC (with Ofgem) published the Government's Response to the Smart Meter Prospectus. This set out a number of key dates relating to the rollout of smart meters in Great Britain. Since then, BEIS has continued to focus on ensuring all parties are making the necessary preparations for the main installation stage, so that energy suppliers are able to complete the rollout by the end of 2020.

A Smart Energy Code was put in place, necessary to enable the national data and communications system to go live[45]. The Code details the rights and obligations of industry parties who use smart metering systems and the information they provide. According to BEIS, customers will be able to choose:

[42] Pallas F. 2013 Beyond gut level - Some critical remarks on the german privacy approach to smart Mmtering.*European Data Protection: Coming of Age*(Springer 2013). See http://link.springer.com/10.1007/978-94-007-5170-5_14 (accessed 26 June 2017).
[43] Written evidence submitted by Department of Energy and Climate Change (SME0024) to the House of Commons Science and Technology Committee. See http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/science-and-technology-committee/smart-meters/written/34284.html (accessed 26 June 2017).
[44] About Us - Department for Business, Energy & Industrial Strategy - GOV.UK. See https://www.gov.uk/government/organisations/department-for-business-energy-and-industrial-strategy/about (accessed 19 April 2017).
[45] SEC and Guidance Documents. See https://www.smartenergycodecompany.co.uk/sec/sec-and-guidance-documents (accessed 19 April 2017).

- how much data their energy supplier collects from their smart meter, e.g. monthly, daily or half-hourly meter readings;
- whether their supplier shares details about their energy consumption with other organisations;
- whether their supplier can use their meter reads for sales and marketing purposes;
- how they can access information about their energy use to get the most benefit from it.

The **Office of Gas and Electricity Markets (Ofgem)**[46] is a non-ministerial government department and an independent National Regulatory Authority. Its principal objective is to protect the interests of existing and future electricity and gas consumers. The **Gas and Electricity Markets Authority (GEMA)** serves as Ofgem's governing body overseeing Ofgem's work and providing strategic direction[47]. The Authority's members are appointed by the Secretary of State at the Department for Business, Energy and Industrial Strategy.

Ofgem is providing independent regulatory expertise and advice to the government's central programme responsible for delivering the rollout of smart metering, taking steps to put in place appropriate consumer protections – including for customers who had smart meters installed before the completion of the government's regulatory framework for the smart meter roll-out. Ofgem has also taken on additional regulatory functions to support smart metering, including regulation of the new Data and Communications Company (DCC).

The Department of Business, Energy and Industrial Strategy (BEIS) held a competition to allocate a contract to a **Data and Communications Company (DCC)**[48] to establish and manage the data and communications network to connect smart meters to the business systems of energy suppliers, network operators and other authorised service users of the network. Smart DCC Ltd was granted a licence in September 2013, and is regulated by Ofgem. The data and communications infrastructure will:

1. operate consistently for all consumers regardless of their energy supplier;
2. provide smart metering data to network operators in support of smart grids;
3. allow authorised third parties to provide services to consumers who have granted them permission to use their data. Consumers can benefit by receiving energy services and advice on how to reduce their energy usage.

[46] About Us.30 April 2013. See https://www.ofgem.gov.uk/about-us (accessed 19 April 2017).
[47] Transition to Smart Meters. 17 June 2013. See https://www.ofgem.gov.uk/gas/retail-market/metering/transition-smart-meters (accessed 19 April 2017).
[48] Data Communications Company. See https://www.smartdcc.co.uk/about-dcc/ (accessed 19 April 2017).

European framework legislation[49]:

Energy Efficiency Directive

The 2012 Energy Efficiency Directive[50] establishes a set of binding measures aimed at helping the EU to reach its 20% energy efficiency target by 2020. Under the Directive, all EU countries are required to use **energy** more efficiently at all stages of the energy chain, from production to final consumption.

On 30 November 2016, the Commission proposed an update to the Energy Efficiency Directive, including a new 30% energy efficiency target for 2030, and measures to update the Directive[51] to make sure the new target is met. New national measures must seek to ensure major energy savings for consumers and industry alike. This includes ensuring that energy consumers should be empowered to better manage consumption. This includes easy and free access to data on consumption through individual metering.

Data Protection Act and GDPR

The Data Protection Act 1998 (and in future the GDPR) provides broad obligations that suppliers and other data controllers would need to meet. Following the developments outlined in the previous section, the Information Commissioner's Office (ICO) has supported the approach that the Government's policy on data access and privacy should address specific questions about the choices consumers should have, and the levels of energy consumption data that it is appropriate for suppliers and others to access to carry out essential functions connected to the provision of energy. The ICO also suggests that sector-specific provisions, that complement the Data Protection Act and the GDPR, might be appropriate in the case of smart metering.

Energy suppliers will need to comply with GDPR from May 2018. During the current transitional period, energy companies will need to work towards implementing GDPR provisions, including those relating to data portability, into their businesses. The supervisory authority under GDPR will be the Information Commissioner's Office, which is committed to working closely with other regulators, such as Ofgem.

It is the legal responsibility of all industry participants to ensure that they comply with the Data Protection Act and the GDPR to the extent that it applies to them. Generally speaking, suppliers, network operators and third parties accessing energy consumption data are likely to be data controllers, with the Data and Communications Company (DCC) potentially acting

[49] Relevant legislation include: Environmental Information Regulations 2004; The Gas Act 1986, the Electricity Act 1989; the Utilities Act 2000; the Competition Act 1998; the Enterprise Act 2002; the Energy Acts of 2004, 2008, 2010 and 2011; the Regulatory Enforcement and Sanctions Act 2008; Data Access and Privacy Framework (DAPF) 2012; Freedom of access to information Directive 2003; INSPIRE Regulations 2007; the Data Protection Act 1998; the General Data Protection Regulation, GDPR; Common Rules for the Internal Market in Electricity Directive (2009/72/EC); Common Rules for the Internal Market in Natural Gas Directive (2009/73/EC); Directive on the Processing of Personal Data (1995/46/EC); Energy Efficiency Directive (2012/27/EC); Regulation on guidelines for trans-European energy infrastructure ((EU) No 347/2013); Commission Recommendation of 9 March 2012 on preparations for the roll-out of smart metering systems (2012/148/EU); Commission Recommendation of 10 October 2014 on the Data Protection Impact Assessment Template for Smart Grid and Smart Metering; Directive 2014/94/EU on the deployment of alternative fuels infrastructure; Proposal for a revised electricity Directive.

[50] Energy Efficiency Directive - Energy - European Commission (*Energy*). See https://ec.europa.eu/energy/en/topics/energy-efficiency/energy-efficiency-directive (accessed 19 April 2017).

[51] Commission Proposes New Rules for Consumer Centred Clean Energy Transition - Energy - European Commission (*Energy*). See https://ec.europa.eu/energy/en/news/commission-proposes-new-rules-consumer-centred-clean-energy-transition (accessed 19 April 2017).

as a data processor on their behalf, although this will depend on the exact nature of the activity being undertaken and the contractual basis for it[52].

In addition to complying with the existing rules, the smart metering implementation programme considered privacy needs by adopting privacy by design principles throughout the smart metering regulatory regime. Under the GDPR, organisations have a general obligation to implement technical and organisational measures to show that they have considered and integrated data protection into their processing activities. Under the Data Protection Act, privacy by design has always been an implicit requirement of the principles - e.g. relevance and non-excessiveness. The programme recognised the confidentiality issues that may arise with respect to energy consumption data for non-domestic consumers and proposed that privacy by design principles should apply to both domestic and non-domestic consumers. According to Ofgem, in order to deliver "privacy by design" they will build on their initial work to carry out a detailed privacy impact assessment (PIA) and encouraging industry to carry out their own PIAs for their own internal systems[53].

Domestic energy regulation

The energy sector in the UK is mainly regulated through the Electricity Act 1989 and Gas Act 1986[54], both of which have been amended on numerous occasions to reflect developments in government policy. These Acts prohibit a number of activities, such as the supply of electricity, except under licence. Licence holders are required to comply with the relevant conditions contained within their licence. Compliance with these conditions is enforceable by Ofgem. Below these licence conditions sit a number of industry codes which contain the technical and commercial obligations that govern participation in licensed activities[55].

Significant changes to this regulatory framework have been made to achieve the government's vision for every home to have smart metering equipment. These changes include:

1. amendments to existing energy licences and industry codes, for example to require suppliers to roll out smart meters, and consequential changes to legislation, licences and codes;
2. the introduction of a new licensable activity relating to communications between suppliers and other parties and smart meters in consumer premises and the appointment of a Data and Communications Company to carry out this licensed activity;
3. the introduction of a new Smart Energy Code. This sets out the rules, rights and obligations for all parties for the new enduring metering arrangements in Great Britain.

Government has been making the regulatory changes to implement these arrangements, principally using powers conferred on the Secretary of State by the Energy Act 2008 and

[52] Smart Meter Data Access and Privacy - GOV.UK. See https://www.gov.uk/government/consultations/smart-meter-data-access-and-privacy (accessed 19 April 2017).
[53] Smart Meter Data Access and Privacy - GOV.UK. See https://www.gov.uk/government/consultations/smart-meter-data-access-and-privacy (accessed 19 April 2017).
[54] The Electricity Act 1989 provided for the privatisation of the electricity supply industry in Great Britain. The Act also established a licensing regime and a regulator for the industry called the Office of Electricity Regulation (OFFER), which has since become the Office of Gas and Electricity Markets (OFGEM). The Gas Act 1986 provided for the appointment and functions of a Director General of Gas Supplies and established the Gas Consumers Council. The Act also made new provisions for the supply of gas through pipelines.
[55] Smart Meters: Information for Industry and Other Stakeholders - GOV.UK. See https://www.gov.uk/guidance/smart-meters-information-for-industry-and-other-stakeholders (accessed 19 April 2017).

extended by the Energy Act 2011[56]. Any changes made under these powers, must first be the subject of consultation, including with Ofgem, before being presented to Parliament[57].The changes to these licences and codes have been phased in tranches of regulation to give businesses and other stakeholders the time to give input on the detail of the regulatory framework through individual and detailed consultations.

The Orders

As part of the tranches of regulation changes BEIS used their powers under the Energy Act 2008 to make amendments to the Electricity Act 1989 and Gas Act 1986 (the 'Acts')[58], which enabled them to make two new statutory instruments that supported the establishment of the DCC. These are The Electricity and Gas (Smart Meters Licensable Activity) Order 2012 (SI 2012/2400)[59] and The Electricity and Gas (Competitive Tenders for Smart Meter Communication Licences) Regulations 2012 (SI 2012/2414) (DCC Licence Application Regulations)[60]. The government also introduced a new statutory instrument to support the implementation of the Smart Energy Code (SEC). This is known as: The Electricity and Gas Appeals (Designation and Exclusion) Order 2013 (SI 2013/2429)[61].

The Smart Energy Code (SEC)

When the Department of Energy and Climate Change (DECC) granted the Smart DCC Ltd licence, stage 1 of the Smart Energy Code (SEC) also came into force. The SEC is a new industry code and sets out the terms for the provision of the DCC's services and specifies other provisions to govern the end-to-end management of smart metering. Ofgem regulates DCC[62] and like with other industry codes, they are responsible for approving any modifications to ensure consumers' interests remain protected.

The Smart Energy Code (SEC)[63] is a multi-party agreement, which requires energy suppliers, network operators and other relevant stakeholders to become a party to the SEC and to comply with its provisions in order to use DCC services. The SEC also sets out the rights and obligations of each Party. It is overseen by the SEC Panel and administered by the Smart Energy Code Administrator and Secretariat (SECAS). DCC is responsible for the development of a number of appendices known as subsidiary documents.

The Smart Energy Code Section 'Data Privacy' sets out the obligations regarding privacy, and are applicable to those DCC Users fulfilling the "Other User" role, as set out in the SEC. The Code acknowledges that, in providing the Services to a user, the DCC may act in the capacity of 'data processor' (as defined in the Data Protection Act 1998) on behalf of that user in respect of the Personal Data for which that user is the 'data controller' (as defined in the Data Protection Act), conferring on it duties set out in the DPA.

---

[56] Energy Act 2011.See http://www.legislation.gov.uk/ukpga/2011/16/contents (accessed 19 April 2017).
[57] Smart Meters: Information for Industry and Other Stakeholders - GOV.UK. See https://paperpile.com/c/LDyx2u/TGVY (accessed 19 April 2017).
[58] Ibid.
[59] The Electricity and Gas (Smart Meters Licensable Activity) Order 2012. See http://www.legislation.gov.uk/ukdsi/2012/9780111526545/contentshttp:/www.legislation.gov.uk/ukdsi/2012/9780111526545/contents (accessed 19 April 2017).
[60] The Electricity and Gas (Competitive Tenders for Smart Meter Communication Licences) Regulations 2012. See http://www.legislation.gov.uk/uksi/2012/2414/introduction/made (accessed 19 April 2017).
[61] The Electricity and Gas Appeals (Designation and Exclusion) Order 2013. See http://www.legislation.gov.uk/uksi/2013/2429/contents/made (accessed 19 April 2017).
[62] Transition to Smart Meters. 17 June 2013. See https://www.ofgem.gov.uk/gas/retail-market/metering/transition-smart-meters (accessed 19 April 2017).
[63] Smart Energy Code Home. See https://www.smartenergycodecompany.co.uk/ (accessed 19 April 2017).

Privacy Assessments are undertaken to assess Parties' compliance with the obligations defined in the Code. The assessments are undertaken by an Independent Privacy Auditor, who has been appointed by the SEC Panel to provide the audit services. The Assessments must be carried out in accordance with a Privacy Controls Framework, which provides the basis for enabling a consistent level of review across all Users. The Framework includes:

- Arrangements designed to ensure that Privacy Assessments provide reasonable assurance that Other Users are complying with their obligations under Sections 1.2 to 1.5;
- The Principles and criteria to be applied in the carrying out of any Privacy Assessment, including principles designed to ensure that Privacy Assessments take place on a consistent basis across all Other Users; and
- The Provisions for determining the timing, frequency and selection of Other Users for the purposes of Random Sample Privacy Assessments.

3.7 Concluding remarks

The case of smart meters is multi-faceted and touches upon individual and systemic aspects of data governance. The UK experience bases its governance around a regulated trusted body, the DCC. The centrality of this body is only possible due to the central organisation and rollout of smart meter technology with the government lead, which sees the system architecture as purposefully constructed rather than emergent as a result of market forces. This creates a centralised morphology for data transmission and storage. An alternative structure, one based on cryptographic proofs providing privacy-respecting aggregation for network operators, as well as personalised billing was not taken up, with a number of possible reasons — lack of cryptographic expertise within government and business; the novelty of the application domain of smart meters; or concerns that privacy-enhancing measures might detract from the profitability of proposed systems.

Germany, another country with a smart metering target and a similar timeframe for its rollout, also opted against installing a data governance infrastructure based on novel privacy-enhancing technologies. However, the regulatory approach they took to metering infrastructure allowed the flexibility, through the specification of a new piece of intermediate hardware, to create new markets for privacy down the line.

Data governance is important in the context of increasingly, or even ubiquitously, smart infrastructure. Given the one-off nature of some systems there are no out-of-the-box solutions but will need researchers, businesses and government to work hand in hand to create regulation.

4. Data and new markets for services[64]

4.1 Introduction

The line between manufacturer and service provider is becoming increasingly blurred as manufacturers increasingly offer services[65]. Servitization describes the trend for businesses to move from selling goods to 'bundles' of goods, services, support, self-service and knowledge, with services taking the lead role[66]. Such bundles, which can be thought of as hybrid product-services include installation and training, product repair and maintenance, customer support, recycling, inspection and insurance or finance[67]. These move all the spotlight away from the manufactured product towards whatever goal it aims to fulfil. Core aspects of this model are summed up by the new suffix *as a service* – as in 'software-as-a-service' (SaaS), 'energy storage–as-a-service' (ESaaS) or even 'anything-as-a-service' (XaaS). These resemble contractual rental models specified in terms of outcomes, rather than hardware.

Servitization is not a new phenomenon, originating in initiatives including Xerox's 'metered' copying, or Rolls-Royce's 50-year-old 'power by the hour' airplane rental. However, it is changing and today the logics for servitization are more diverse and varied than ever. New, more efficient logics of servitization have emerged, relying heavily on usage data. Computer chips are now inexpensive enough to justify inclusion in objects where they provide only a slim marginal benefit. The low-power, consumer-facing Raspberry Pi Zero costs only £4 and measures half the size of a playing card, yet has comparable specifications to the best-buy laptops developed a decade earlier in 2005[68]. The ubiquity of network connections, particularly over 3G, 4G and WiFi, has made communication cheap and easy, albeit often insecure[69]. Both these factors have spurred the use of electronics within product-offerings across sectors that never before had sufficient incentives or means to do so, with a variety of purposes.

4.2 Social and ethical issues in data-enabled services

We are looking at this use of data as it illustrates examples of the following social and ethical issues highlighted in the main report, which would need to be balanced through governance

[64] Dr Ali Z Bigdeli, Professor Tim Baines, Ian McKechnie and Eleanor Musson kindly donated time to review this work.
[65] Bustinza OF, Bigdeli AZ, Baines T, Elliot C. 2015 Servitization and competitive advantage: The importance of organizational structure and value chain position. *Research-Technology Management*. **58**, 53-60. (doi:http://dx.doi.org/10.5437/08956308X5805354)
[66] Vandermerwe S, Rada J. 2002 Servitization of business: Adding value by adding services. *Euro Manage J*. **6**, 314-324. (doi: https://dx.doi.org/10.1016/0263-2373(88)90033-3)
[67] Santamaria L, Nieto MJ, Miles I. 2011 Service innovation in manufacturing firms: Evidence from Spain. *Technovation*. **32**, 144-155. (doi:https://dx.doi.org/10.1016/j.technovation.2011.08.006)
[68] Titcomb J. 2015 Raspberry Pi's £4 computer is as powerful as the laptops of a decade ago. *The Daily Telegraph.* 26 November 2015. See http://www.telegraph.co.uk/technology/news/12017999/Raspberry-Pis-4-computer-is-as-powerful-as-the-laptops-of-a-decade-ago.html (accessed 10 February 2017).
[69] Walport M. 2014 The Internet of Things: Making the most of the second digital revolution. London: UK Government Office for Science.

of the sector. First is *incentivising innovative uses of data while ensuring that such data can be traded and transferred in mutually beneficial ways*, which is essential to getting the most value from services across all sizes of organisation in the sector.

Second, *using data relating to individuals and communities to provide more effective public and commercial services, while not limiting the information and choices available*. When services are hyper-personalised they can be very valuable to individuals, but they can potentially limit access to a wider choice of services or suppliers.

Third, and in line with the example of smart meters discussed above, *making use of the data gathered through daily interaction to provide more efficient services and security, while respecting the presence of spheres of privacy*. Many services might be domestic – including energy as a service – and will use and generate information about home life.

4.3 Opportunities for data-enabled services

From the standpoint of service providers', the aims of servitized products broadly fall into the categories of 'reducing costs' and 'increasing revenue'. More broadly however, one direction is to implement 'circular economy' approaches for 'predictive maintenance'. This direction aims to see both economic as well as social and environmental costs reduced by the more efficient use and maintenance of infrastructure. Modular, monitored, upgradeable systems, such as expensive MRI scanners, can be constantly monitored and repaired, with refurbished parts forming new scanners, or older models being sold to health systems with smaller budgets. Sensors allow manufacturers and engineers to manage buildings assets such as elevators, whole buildings through building information modelling (BIM), and energy infrastructure such as solar panels and batteries[70].

Servitization also offers several new potential revenue streams. Services and goods can be targeted in new ways that result from new revenue streams in advertising, complementary services, and sales of data. The Amazon Kindle e-reader, for example, has a lower 'special offers' price point for those opting to lock their screensavers to personalised adverts. Advertising efforts to cross-subsidise hardware businesses do not have to be consumer facing, as 'smart' television providers have been noted to be selling 'second-by-second' viewing data to 'authorized data partners including analytics companies, media companies and advertisers'[71]. Where goods can also be located remotely, such as servers, servitized 'cloud' approaches allow for the reduction of redundant and unused hardware and energy through sharing and dynamic allocation. In fact, flexible cloud rental systems such as Google Cloud, Microsoft Azure and Amazon AWS, are both highly servitized models of ICT provision, and power many other firms' servitized offerings. Servitized products also act as a vector to provide new services themselves. Smart home technologies, such as Amazon Alexa and Google Home, both play a practical role in linking pieces of hardware within and building and act as a channel to sell further products and services.

[70] Ellen MacArthur Foundation. 2016 Intelligent assets: Unlocking the circular economy potential. See https://www.ellenmacarthurfoundation.org/ (accessed 26 June 2017).
[71] Maheshwari S. 2017 Is your Vizio television spying on you? What to know. *The New York Times.* 7 February 2017. See https://www.nytimes.com/2017/02/07/business/vizio-television-vizio-collected-viewers-habits-consent.html (accessed 10 February 2017).

Example: Connected lighting

Lighting companies, such as Phillips, now provide lighting 'as-a-service' to a variety of sectors. In office environments, remote monitoring of energy usage and efficiency enables clients to receive money back from Philips were they to exceed spend beyond their expected energy usage. In turn, the company keeps the clients' lighting upgraded to the most efficient product in their lines over the 10–15 year contract, refurbishing the removed fittings and recycling them to clients with different needs and price demands.

Street lighting is also increasingly provided in this manner. Connected street lights are able to be flexibly, centrally controlled and programmed, as well as able to adapt more readily to environmental changes such as light and weather conditions than non-connected counterparts. Their position as nodal, internet-enabled infrastructure also places them in a ready position for alternative uses. Some of these, such as tracking or measuring movement of people and vehicles through low-energy Bluetooth, visual sensors, or radio-frequency identification (RFID) are controversial in some areas given the lack of explicit consent for such pervasive surveillance. Yet the data provided by these sensors might also be useful for planning purposes, creating an information feedback loop in the installation of new infrastructure that was not present before.

4.4 Challenges in data-enabled services

The collection, storage and management of new data streams that are essential to realising these opportunities can be personally or commercially sensitive. For example, data from factory machines, bought as a service, might leak information about new products being developed. As described in more detail below, data from sensors on machinery can reveal other aspects of performance beyond the technical aspects of the physical performance of machines, and this can have both value and sensitivity. Data from lights and home appliances are likely to reveal data about people's private lives, in terms of when they use appliances and in what way – much in the same way as discussed for the smart meters case study. Samsung's smart television series, equipped with voice and gesture recognition, achieved media attention for a statement in their privacy policy that requested customers not to speak about sensitive topics in front of their television, as information was sent to Samsung servers for processing[72]. These issues and others give rise to a range of governance needs surrounding the use, and generation, of data in the expanding service sector.

4.5 Governance needs for data-enabled services

Data-enabled services that change in real time

Servitization means products are no longer static, but change after they have been procured, sometimes with new features being trialled and tested live on users. This shift has not come

[72] Harris S. 2015 Your Samsung SmartTV is spying on you, basically. *The Daily Beast.* 6 February 2015. See http://www.thedailybeast.com/articles/2015/02/05/your-samsung-smarttv-is-spying-on-you-basically.html (accessed 1 December 2016).

without ethical consequences or concerns[73], creating challenging around privacy, research ethics, and scrutability.

Beyond privacy of data, the experimentation on users characteristic of modern software development introduces questions to servitized product-services that have previously been the domain of research ethicists[74]. Websites test features on individuals with 'A/B' experiments, gathering what essentially amounts to research data often without participants' awareness. Significant questions around the governance of both the research-like activities of new servitized products, and the use and reuse of this personal experimental data, are a challenge to existing regulatory frameworks.

The ways that software rapidly changes also makes audit and research difficult, by making it hard to explore or ascertain ethical or compliance issues within particular applications of data. To improve products, individuals are sorted into treatment groups, potentially based on how they interact with a service, and delivered different experiences as a result. Yet this makes it difficult for researchers to know the system they are studying, compare it over time, or even make basic assumptions about short term functionality[75].

Is data from servitization a threat to market competition?

Servitization and the bundling of product-services have historically been a concern to competition authorities, from older cases such as Xerox's bundling of 'black gold' toner with its leased copiers[76] to more recent, higher profile cases around Microsoft and the market advantage afforded to Internet Explorer. More recently both competition authorities[77] and intergovernmental organisations such as the OECD[78] have focused on the data accumulated in the operation of these product-services. They have expressed concern that accumulation of large, sector-specific datasets.

These perceived competition issues data can be considered from two angles, 'lock-out' and 'lock-in'[79]. Locking out implies that a product-service combination is not able to be effectively delivered in a competitive environment without a significant corpus of data. Servitized supply chains often need to be more reactive, real-time and responsive than production-based counterparts[80]. New firms may struggle to challenge incumbents because of their existing data strength, which will lead to the incumbents holding more market share for longer and increasing the data they hold on service design and delivery. On the other hand, it has been claimed that the digital economy has in fact seen many new entrants to markets despite

[73] Gürses S, van Hoboken J. 2017 Privacy after the Agile Turn. In Selinger E (ed), *The Cambridge Handbook of Consumer Privacy* (2017). See https://osf.io/27x3q/ (accessed 26 June 2017).
[74] Bird S, Barocas S, Crawford K, Diaz F, Wallach H. 2016 Exploring or exploiting? Social and ethical implications of autonomous experimentation in AI. See http://ssrn.com/abstract=2846909 (accessed 11 November 2016).
[75] Seaver N. 2013 Knowing algorithms. *Media in Transition*. **8**.
[76] Scherer FM. 2007 Technological innovation and monopolization. See http://papers.ssrn.com/Sol3/papers.cfm?abstract_id=1019023 (accessed 8 December 2016).
[77] Auchard E.2016 EU competition chief to eye "Big Data" concerns in merger probes. *Reuters*. 17 January 2016. See http://www.reuters.com/article/us-europe-data-competition-idUSKCN0UV0ZG (accessed 8 December 2016); Authorité de la concurrence and Bundeskartellamt. 2016 Competition Law and Data. See http://www.autoritedelaconcurrence.fr/doc/reportcompetitionlawanddatafinal.pdf (accessed 26 June 2017).
[78] OECD. 2016 Big Data: Bringing Competition Policy to the Digital Era. *Organisation for Economic Co-operation and Development*. See http://www.oecd.org/competition/big-data-bringing-competition-policy-to-the-digital-era.htm (accessed 26 June 2017).
[79] Hojnik J. 2016 The servitization of industry: EU law implications and challenges. *Common Market Law Review*. **53**, 1575-1624. See http://www.kluwerlawonline.com/document.php?id=COLA2016143 (accessed 8 December 2016).
[80] Johnson M, Mena C. 2008 Supply chain management for servitized products: A multi-industry case study. **114**, 27-39. *International Journal of Production Economics.* (doi:http://dx.doi.org/10.1016/j.ijpe.2007.09.011)

starting from a data-poor position[81]. Yet it is worth noting that the examples they use consider network effects of software and digital platforms, which arguably exhibit qualitatively different types of lock-in, fixed costs and path dependencies than physical infrastructures tend to.

Locking in implies that a consumer will struggle to move away from a particular service or provider for fear of losing the history that enables them to receive a hyper-personalised service. While the General Data Protection Regulation provides new rights to data portability designed to forestall the worst effects of these dynamics, these apply only to personal data, not to data that contains features of organisations, workplaces, systems or so on. Some voluntary programmes, such as Midata, which allows users to download de-identified copies of their banking records, have led to services that allow you to upload data to a price comparison website to understand which account would suit you best, although the benefits and uptake of this has not yet received rigorous evaluation[82].

Even if data are portable, they are not necessarily compatible across systems, making data from another provider potentially unhelpful in supporting maintenance or repair. This is particularly the case when the physical good differs significantly in characteristics, capabilities or underlying technology, as an office cleaning robot might.

If individuals and firms had the right, statutorily or contractually, to access their own data as collected by product-services, this might be valuable in secondary markets that provide data to new SMEs for the development and refinement of innovative ideas. What incentives are there for firms to sell such data?

Valuable secondary insights from data: access and protection

Due in part to servitization but also more broadly, more firms and NGOs are in possession of far-reaching datasets with latent insights. Many of these insights might not be of utility to the organisation holding the data, but might be of broader societal or economic benefit. The increased need for cloud processing in order to make and improve cutting-edge predictive models for features such as voice recognition also creates liabilities for a data holder — will such data be subject to legal pressure for release in particular circumstances? Amazon has recently come under media scrutiny when their Echo device was present at a murder scene in Arkansas, and prosecutors in August 2016 obtained a signed warrant for “audio recordings, transcribed records, text records and other data”. Amazon claim their policy is to object to “overbroad or otherwise inappropriate demands as a matter of course”[83].

There are also issues relating to data on more macro-level societal phenomena. A Japanese manufacturer of heavy machinery has been monitoring the use of over 400,000 devices, such as diggers, for nearly 20 years. This data is in high demand from investors and public authorities, as it is thought to give indication of the dynamics of slowdown and recovery across

[81] Balto DA, Lane MC. 2016 Monopolizing water in a tsunami: Finding sensible antitrust rules for big data. See https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2753249 (accessed 10 February 2017); Sokol DD, Comerford RE. 2016 Does antitrust have a role to play in regulating big data? See https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2723693 (accessed 10 February 2017).
[82] Hartfree Y, Evans J, Kempson E, Finney A. 2016 Personal Current Account Switching. *Personal Finance Centre, University of Bristol*. See http://www.bristol.ac.uk/media-library/sites/geography/pfrc/pfrc1604-personal-current-account-switching-report.pdf (accessed 26 June 2017).
[83] Alexa a witness to murder? Prosecutors seek Amazon Echo data. *Bloomberg Technology*. 28 December 2016. See https://www.bloomberg.com/news/articles/2016-12-28/alexa-a-witness-to-murder-prosecutors-seek-amazon-echo-data (accessed 6 January 2017).

the globe in particular sectors[84]. Yet releasing this data might have not only legal implications — the potential to de-anonymise and reveal commercial secrets, for example — but also PR questions, such as whether companies would want to deal with a firm that sells or releases data in this way.

Lack of clarity on who has claims on valuable secondary insights generates uncertainty among data controllers, who might be tempted to destroy data irrelevant to them which they do not wish to trade to avoid legal and PR issues. While in some cases, this may be a desirable course of action, and in line with reduced data retention and data minimisation practices, there may be a case for debate around determining legitimate external uses for secondary information.

4.6 Concluding remarks

Servitization entails increased data collection and analysis, which has large potential for efficiencies benefitting the economy and the environment. At the same time, these new devices can entail concerns and new challenges, leaving gaps for informal or formal governance systems. Where data or activities are less sensitive, these issues are less present or challenging, but discussion of these issues can benefit individuals, society and businesses.

[84] Lucas L, Lewis, L. 2016 Wanted: Japan digger group's secret trove of global economic data. *Financial Times*. 28 November 2016. See https://www.ft.com/content/6c43a27e-b46a-11e6-961e-a1acd97f622d (accessed 29 November 2016).

5. '-omics' data

5.1 Introduction

'-omics' data are concerned with the biological molecules that contribute and shape the structure and function of the organism. Since the sequencing of the human genome and the plummeting price of genome sequencing and analysis technologies, much hope has been placed onto the '-omics' sector to provide breakthrough treatments and diagnosis tools. The UK genomics market is currently estimated to stand at around £0.8bn, representing 10% of the global genomics market, with this share projected to outpace global market growth in the years to come due to significant investments and infrastructure in the scientific community[85]. The scope of the '-omics' field raises a broad range of data governance questions. Some of these surround the foundational research in the field, and how it is undertaken. Some relate to more applied research, and seeking that the research is not applied unfairly. Others concern the use of data-driven decision support tools in clinical decision-making, and the difficulties that come with reporting findings that relate to these model outputs.

5.2 Social and ethical tensions in the management and use of '-omics' data

This case study considers three areas at the confluence of data governance issues and '-omics' research and practice. The first surrounds how compatible these forms of data are and have been with open scientific inquiry. The second considers how findings developed from this research can remain fair, even where the risk of skewed research data exists. The final area concerns what happens when '-omics' data is used operationally — what are the rights and responsibilities of practitioners where incidental, uncertain but potentially high-impact findings about an individual's genome are discovered in the course of a scan or analysis that might not have been looking for them. The case study thus illustrates aspects of the following tension highlighted in the main report: *promoting and encouraging innovation, while ensuring that it addresses societal needs and reflects public interest*; and as with many issues around using data for medical research, it relates to *promoting and distributing the benefits of data use fairly across society while ensuring acceptable levels of risk for individuals and communities* and also to *providing ways to exercise reasonable control over data relating to individuals while encouraging data sharing for private and public benefit* – especially where individuals with rare diseases may be especially exposed.

[85] Deloitte M. 2015 Genomics in the UK: An Industry Study for the Office of Life Sciences. *Office for Life Sciences, HM Government.* See https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/464088/BIS-15-543-genomics-in-the-UK.pdf (accessed 26 June 2017).

5.3 Governance needs for '-omics data'

Open research and anonymity of individuals

Large scale research on the human genome is difficult. The huge sensitivity of the data required, in combination with the number of examples ideal for rigorous investigation, poses challenges for open scientific methods. To gather and share a large dataset that could benefit the international research community, the Personal Genome Project (PGP) was launched, first in the US (based at Harvard University), and then more broadly across the world. The UK's branch is based at University College London.

Yet genomic data is by definition both identifiable and highly sensitive, with the possibility to reveal many factors about an individual – such as a propensity for particular diseases. Consequently, the project does not operate within the context of assured privacy, but of 'open consent'. This involves participants undertaking a comprehensive examination in addition to eligibility and screening questionnaires. A recent study found that the main motivations for participants included the advancement of science, and the desire to be part of a pioneering community, although it cast doubt on how far this model of consent would work with individuals less familiar with academic culture[86].

Another element of the project governance was to create an open and participatory atmosphere among the participants. Users are free to post what they like on their deidentified websites, such as genomic results from other websites such as 23AndMe. Communities such as forums and LinkedIn groups have emerged to allow participants to share experiences[87]. This has opened an issue in the governance of the posted data, as the ability to use website profile pages to match the genomic data successfully[88], or by querying the genomic data to identify surnames using commercial databases[89], has been demonstrated. However, and as the PGP noted at the time, no participants are known to have withdrawn from the project on the basis of these events[90]. Important future research, such as that in the field of environmental genomics, requires explicit linkages between medical data and lifestyle data, thus heightening the risk of reidentification considerably. Robust governance mechanisms are likely to be needed for such research to be successfully and routinely undertaken.

Bias and data quality

Personalised or precision medicine is the tailoring of medical treatment to the individual characteristics, needs and preferences of a patient during all stages of care, including prevention, diagnosis, treatment and follow-up[91]. Personalised manufacturing of pharmaceuticals that target specific patient populations or even individuals is becoming

[86] Zarante OA, Brody JG, Brown P, Ramirez-Andreotta MD, Perovich L, Matz J. 2016 Balancing benefits and risks of immortal data: Participants' views of open consent in the Personal Genome Project. *Hastings Cent Rep*. **46**, 36-45. (doi:http://dx.doi.org/10.1002/hast.523)
[87] Ball MP *et al*. 2014 Harvard Personal Genome Project: lessons from participatory public research. *Genome Medicine*. **6**, 10. (doi:http://dx.doi.org/10.1186/gm527)
[88] Sweeney L, Abu A, Winn J. 2013 Identifying participants in the Personal Genome Project by name. See http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2257732 (accessed 29 November 2016).
[89] Gymrek M, McGuire AL, Golan D, Halperin E, Erlich Y. 2013 Identifying personal genomes by surname inference. *Science*. **339**, 321-324. (doi:http://dx.doi.org/10.1126/science.1229566)
[90] Ball MP *et al*. 2014 Harvard Personal Genome Project: lessons from participatory public research. *Genome Medicine*. **6**, 10. (doi:http://dx.doi.org/10.1186/gm527)
[91] US Food and Drug Administration. 2013 Paving the way for personalized medicine: FDA's role in a new era of medical product development. *Silver Spring*.

increasingly plausible due to significant advances in genomics and economic, flexible forms of drug manufacture such as 3D and inkjet printing[92].

Precision medicine creates challenges for evaluation of products, which in part relate to the availability of data. It is somewhat harder to test the efficacy and safety of highly personalised medicines, given the difficulties in finding and incentivising a test group, and the limited returns on the cost of building an evidence base. In particular, it can be difficult to make sure that the data being used to develop these technologies is representative of the genetic diversity in the general population. This can be difficult due to 'platform bias'[93], where the users of different genomic testing platforms are not representative of the genetic diversity in the general population.

Partly to pre-empt and tackle this, the US Food and Drug Administration (FDA) have developed an open online platform, *precisionFDA*. This is a public, cloud-based platform that "hosts shared tools, crowdsourced testing, and community challenges, to improve and share knowledge and methods for evaluating [next generation sequencing] bioinformatics pipelines"[94]. Among other areas, the FDA data governance approach seeks to incentivise the creation and use of 'gold-standard' datasets that adequately include "diverse, underrepresented populations" that may not otherwise be included in the benchmark datasets of different providers[95]. Reactions from incumbents have ranged from welcoming the initiative, to warning that over-regulation may deter smaller companies, or arguing that in-house benchmark datasets of large firms are still likely to outstrip the utility of those provided on crowd platforms such as the proposed[96].

Incidental findings

Once a whole human genome has been sequenced, the marginal cost of testing for further features beyond the purpose of the original test is greatly reduced. This has led to questions regarding what duties or obligations exist to both test for and report further features of a subject's genome that may be of medical interest or importance.

The concerns around incidental findings centre around two areas: the balance of rights and duties concerning patients, doctors and researchers regarding information generated as a result of genomic testing, and the relative costs and benefits of action or inaction in individual and public health terms[97].

---

[92] Voura C *et al*. 2011 Printable medicines: A microdosing device for producing personalised medicines. *Pharm Tech Eur*. **23**. https://www.researchgate.net/publication/223829350_Printable_medicines_A_microdosing_device_for_producing_personalised_medicines; Daly R, Harrington TS, Martin GD, Hutchings IM. 2015 Inkjet printing for pharmaceutics - A review of research and manufacturing. *Int J Pharm*. **494**, 554-567. (doi:http://dx.doi.org/10.1016/j.ijpharm.2015.03.017); Prasad LK, Smyth H. 2016 3D printing technologies for drug delivery: A review. *Drug Dev Ind Pharm*. **42**, 1019-1031. (doi:http://dx.doi.org/10.3109/03639045.2015.1120743)

[93] Altman RB *et al*. 2016 A research roadmap for next-generation sequencing informatics. *Science Trans Med*. **8**, 335-339. (doi:http://dx.doi.org/10.1126/scitranslmed.aaf7314)

[94] Altman RB *et al*. 2016 A research roadmap for next-generation sequencing informatics. *Science Trans Med*. **8**, 335-339. (doi:http://dx.doi.org/10.1126/scitranslmed.aaf7314)

[95] Altman RB *et al*. 2016 A research roadmap for next-generation sequencing informatics. *Science Trans Med*. **8**, 335-339. (doi:http://dx.doi.org/10.1126/scitranslmed.aaf7314)

[96] Petrone J. 2016 FDA wades into sequencing-based diagnostics regulation. *Nature Biotechnology*. **34**, 681-682. (doi:http://dx.doi.org/10.1038/nbt0716-681a)

[97] Damjanovicova M. 2016 Incidental findings. In Boniolo G, Sanchini V (eds), *Ethical Counselling and Medical Decision-Making in the Era of Personalised Medicine* (Springer International Publishing 2016). pg 89-95. (doi:http://dx.doi.org/10.1007/978-3-319-27690-8_9)

Many discussions of governance around the use of this data have centred primarily on the American College for Medical Genetics and Genomics (ACMG). A 2013 document[98] argued for an obligation on the part of physicians not just to report incidental findings, but actively to seek them out. In particular, 57 (now 59) "medically actionable genes recommended for return in clinical genomic sequencing". ACMG advocate use of the term "secondary finding" in place of "incidental finding", as they are advocating for intentional rather than incidental analysis[99]. To manage and curate this list, the Secondary Findings Maintenance Working Group (SFWG) was established. This working group takes proposals from ACMG members as input, considering factors such as severity and likelihood of materialisation, and more recently, the acceptability of any envisaged intervention in terms of risks and benefits. While ACMG previously argued for mandatory return and reporting, they have updated their policy to include an opt-out for patients, after 80% of respondents to a website survey of members supported an opt-out of processing[100].

The list can provide standards without which laboratories may report secondary findings of uncertain clinical utility[101]. The ethical rationale is that a list of abnormalities can be drawn up which would be highly likely to be beneficial to patient health to disclose, with the absence of certain limitations representing a balance-of-interests calculation carried out a priori. The list also allows firms to undertake such analysis with clear guidelines on their minimal reporting obligations, which provides both economic and accountability benefits. While these recommendations do not carry legal force, it has been suggested that they could be introduced as evidence of the standard of care[102].

Several tensions exist in practice around these policies both in clinical and in research settings. Patients may underestimate the significance of consent due to their low probability of arising in the population. Conceptualising the utility of particular disclosures can be difficult. Family members might see clinical benefit from the findings related to a particular individual, even if the individual themselves receives no clinical benefit. One case concerned the identification of a particular gene in a boy who presented with autism, which was deemed to have little immediate clinical use for the management of the boy but potential clinical use for the management of the family[103]. There is also concern that genomic testing may create or exacerbate psychological issues in certain individuals, particularly in relation to shorter-term anxiety around the future[104].

---

[98] Green RC *et al*. 2013 ACMG Recommendations for reporting of incidental findings in clinical exome and genome sequencing. GenetMedicine. **15**, 565-574. (doi:http://dx.doi.org/10.1038/gim.2013.73)
[99] Kalia SS *et al*. 2017 Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update (ACMG SF v2.0): a policy statement of the American College of Medical Genetics and Genomics. *Genet Med*. **19**, 249-255. (doi:http://dx.doi.org/10.1038/gim.2016.190)
[100] Scheuner MT *et al*. 2015 Reporting genomic secondary findings: ACMG members weigh in. *Genet Med*. **17**, 27-35. (doi:http://dx.doi.org/10.1038/gim.2014.165)
[101] McGuire AL *et al*. 2013 Point-counterpoint. Ethics and genomic incidental findings. *Science*. **340**, 1047-1048. (doi:http://dx.doi.org/10.1126/science.1240156)
[102] McGuire AL *et al*. 2013 Point-counterpoint. Ethics and genomic incidental findings. *Science*. **340**, 1047-1048. (doi:http://dx.doi.org/10.1126/science.1240156)
[103] Lewis A, James P. 2012 An incidental finding of a large genomic deletion of BRCA1 on a molecular karyotype for a 5 year old child. *Hered Cancer Clin Pract*. **10**. (doi:http://dx.doi.org/10.1186/1897-4287-10-S2-A73); Shkedi-Rafid S, Dheensa S, Crawford G, Fenwick A, Lucassen A. 2014 Defining and managing incidental findings in genetic and genomic practice. *J Med Genet*. **51**, 715-723. (doi:http://dx.doi.org/10.1136/jmedgenet-2014-102435)
[104] Meiser B, Dunn S. 2001 Psychological effect of genetic testing for Huntington's disease: An update of the literature. *West J Med*, **174**, 336-340. (doi:http://dx.doi.org/10.1136/jnnp.69.5.574); Lerman C, Croyle RT, Tercyak KP, Hamann H. 2002 Genetic testing: Psychological aspects and implications. *J Consult Clin Psychol*. **70**, 784-797. See https://www.ncbi.nlm.nih.gov/pubmed/12090383 (accessed 26 June 2017).

5.4 Concluding remarks

‘-omics’ data is highly impactful both at the level of individuals and the level of humanity. Governance issues range from how to responsibly carry out fundamental research, bearing mind issues of privacy and fairness, to how to utilise diagnostic and other tools that research provides. As ‘-omics’ data may have many suses, a forum for discussing emerging challenges and approaches across a variety of subdomains might prove important for robust development of the field as a whole.

6. Personal location data

6.1 Introduction

Location-based self-tracking technologies are gaining prominence and popularity. These technologies can be classified into five 'modes', clarifying the intentions involved, and levels of autonomy, in self-tracking[105]:

- **Private self-tracking** is undertaken by an individual to track and reflect on their own behaviour, experiences and biological activity, eg through wearable fitness trackers.
- **Pushed self-tracking** is incentivised by another actor, usually in order to promote behaviour change, or for financial incentives, such as demonstrating evidence of low risk to an insurer.
- **Imposed self-tracking** is a form of coercively enforced pushed self-tracking. This is often used in order to access a particular product or service that a consumer needs, such as the use of tracking boxes within cars to obtain certain insurance products.
- **Communal self-tracking** emerges from a culture of the pooling of private self-tracking data. An example of this is CleanSpace, a tag for monitoring air quality and providing a crowdsourced location-based dataset[106].
- **Exploited self-tracking** sells or utilises insights from what is ostensibly self-tracking data for other purposes. For example, AR games such as Pokémon Go gather data which can be of significant commercial value.

6.2 Social and ethical tensions in the management and use of location data

The key tension in using personal location data is between *making use of the data gathered through daily interaction to provide more efficient services and security, while respecting the presence of spheres of privacy*. Location data can underpin valuable services, but potentially presents an infringement of personal, physical privacy.

6.3 Opportunities for using location data

Evidence-based planning

A variety of firms whose services generate large, location-based datasets have seen value in trading this data with public authorities to improve evidence in the planning process. Strava is a commonly used application which allows runners and cyclists to record and share data about activities. While Strava do not share statistics on their user base, they report that the number of activities they have logged, such a runs or cycle routes, is in the order of hundreds of millions, increasing by about six million a week, with the GPS points collected in the hundreds

[105] Lupton D. 2014 Self-tracking cultures: Towards a sociology of personal informatics. *Proceedings of the 26th Australian Computer-Human Interaction Conference on Designing Futures: The Future of Design.* (doi:http://doi.acm.org/10.1145/2686612.2686623) (This list is edited, not a direct quote).
[106] Hopkinson N. 2016 Air quality—what's the point of warnings? *BMJ Blogs*. 8 December 2016. See http://blogs.bmj.com/bmj/2016/12/08/nick-hopkinson-air-quality-whats-the-point-of-warnings/ (accessed 26 June 2017).

of billions.[107] In London, around seven million cycle journeys are reported to have been uploaded to Strava in the 2015–16 period[108].

Such data are valuable for decision-makers, and Strava already sell their data to local authorities to aid planning decisions. Similarly, Uber are providing limited summary data on their journeys for planners for free through their new 'Uber Movement' service[109]. The navigation tool Waze has created collaborations with geographic information providers and with cities to enable better decision-support around local planning and public sector operations[110].

For many cities, this is a vast improvement in data quantity and quality, as well as a reduction in costs, compared to surveys or volunteer counting. Methodologies and tools that can be transferred across regions can also be built on these datasets, allowing insight even for parts of government that have little data analysis capacity, or little money to invest in this.

Whether this data is representative enough for decision-making is a core question for policy-makers. Around 90% of Strava's users are reported to be male[111]. This is compared to an approximate 75% male 25% female split of total cyclists in London, but near gender parity in Cambridge, according to data from the 2011 Census. To address both this bias and variability, Strava provide demographic metadata that helps contextualise the data that it provides[112]. Other biases, such as the disparate use of Strava in richer or safer parts of the city, such as gated communities[113], or the often subtle causes of digital divide, are harder to address. Perhaps more importantly is that Strava is used disproportionately by sports cyclists rather than casual commuters. While Strava argue that "when cyclists are in the urban core they optimise for the same kind of things as everyone else"[114], they do not publish technical justification for this. Use of Strava data therefore needs to take account of underlying assumptions.

[107] See http://metro.strava.com/faq, https://www.theguardian.com/lifeandstyle/2016/may/09/city-planners-cycling-data-strava-tracking-app and http://insights.strava.com (accessed 26 June 2017).
[108] Walker P. 2016 City planners tap into wealth of cycling data from Strava tracking app. *The Guardian.* 9 May 2016. See https://www.theguardian.com/lifeandstyle/2016/may/09/city-planners-cycling-data-strava-tracking-app (accessed 26 June 2017).
[109] Davies A. 2017 Uber's mildly helpful data tool could help fix streets. *Wired.* 8 January 2017. See https://www.wired.com/2017/01/uber-movement-traffic-data-tool/ (accessed 17 January 2017).
[110] Wze now shares its data with cities to improve roads and speed up journeys. *co.design.* 22 October 2016. See https://www.nexis.com/docview/getDocForCuiReq?lni=5M0N-PGM1-F11P-X4TH&csi=8399&oc=00240&perma=true (accessed 26 June 2017).
[111] Usborne S. 2013 Can cycling app Strava change the way we ride? *The Independent*. 4 July 2013. http://www.independent.co.uk/life-style/gadgets-and-tech/features/can-cycling-app-strava-change-the-way-we-ride-8685996.html (accessed 26 June 2017); Vanderbildt T. 2013 How Strava is changing the way we ride. *Outside Magazine*. 8 Jan 2013. See http://www.outsideonline.com/fitness/biking/How-Strava-Is-Changing-the- Way-We-Ride.html (accessed 26 June 2017).
[112] Strava Metro, *Comprehensive User Guide (v 2.0)* (Strava 2015).
[113] Selala MK, Musakwa W. 2016 The potential of Strava data to contribute in non-motorised transport (Nmt) planning in Johannesburg. *Int Arch Photogramm., Remote Sens. Spat. Inf. Sci. XLI-B2,* 587-594. (doi:http://dx.doi.org/10.5194/isprsarchives-XLI-B2-587-2016)
[114] Walker P. 2016 City planners tap into wealth of cycling data from Strava tracking app. *The Guardian.* 9 May 2016. See https://www.theguardian.com/lifeandstyle/2016/may/09/city-planners-cycling-data-strava-tracking-app (accessed 26 June 2017).

6.4 Challenges in the management and use of location data

Privacy challenges: re-identification and inference

A BCG survey seems to suggest that people consider location to be private[115], and when powerful actors have access not to just one location but large datasets, augmented with additional metadata such as time, user, and further characteristics, there is the potential for personal intrusion. It has been demonstrated that 87% of Americans are uniquely identifiable by postal code, date of birth and gender[116] and it is easy to see how these characteristics could be inferred from a stream of location data. In a well-known study, four spatiotemporal points are enough to uniquely identify 95% of individuals, and it is shown that even making the datasets significantly coarser does not have an overwhelming effect on reducing reidentifiability[117]. Indeed, when data are made significantly lower resolution it may take only a handful of extra points to reidentify. This effect may even be stronger for particular subsets, such as home and work areas pairs[118]. Location data do not only reveal identity, but also high-level behaviour — spending a lot of time somewhere at night; work, leisure and education patterns; presence at gender-specific shops or lavatories. Even with relatively low quality location data, it is possible to infer modes of transport[119], important locations for people[120], or income[121].

Machine learning can be used to infer future locations too. Theoretical bounds on the accuracy on the prediction of future location appear to be as high as 88%, with many predictive methods already approaching these levels[122], and could be used to infer potentially sensitive attributes like personality.

Traditional data linking usually requires auxiliary information which might be difficult or expensive to obtain, restricting the number of actors capable of such attacks. However, location data is particularly rich, even with the augmentation of only open data. A huge array of freely available online data, such as social media posts, infrastructure and landmarks, online ratings and reviews, news articles, and more already contain coordinate data, which creates many more opportunities for linking, insight and de-identification.

---

[115] bcg.perspectives. 2014 Data privacy by numbers. 19 Feb 2014. See https://www.bcgperspectives.com/content/Slideshow/information_technology_strategy_digital_economy_data_privacy_by_the_numbers/#ad-image-3 (accessed 27 June 2017).
[116] Sweeney L. 2000 Simple demographics often identify people uniquely. **671**, 1-34. See http://gqs685.pbworks.com/w/file/fetch/94376315/Latanya.pdf (accessed 26 June 2017).
[117] de Montjoye Y, Hidalgo CA, Verleysen M, Blondell VD. 2013 Unique in the Crowd: The privacy bounds of human mobility. *Scientific reports*. **3**. (doi:http://dx.doi.org/10.1038/srep01376)
[118] Gambs S, Killijian M, del Prado Cortez MN. 2011 Show me how you move and I will tell you who you are. *Transactions on data privacy*. **4**, 103-126. (doi:http://doi.acm.org/10.1145/1868470.1868479)
[119] Paterson DJ, Liao L, Fox D, Kautz H. 2003 Inferring high-level behavior from low-level sensors. pg 73-89. (doi:http://dx.doi.org/10.1007/978-3-540-39653-6_6/)
[120] Isaacman S *et al*. 2011 Identifying important places in people's lives from cellular network data. *Pervasive computing*. **6696**, 133-151. (doi:http://dx.doi.org/ 10.1007/978-3-642-21726-5_9/)
[121] Smith-Clarke C, Mashhadi A, Capra L. 2014 Poverty on the cheap: Estimating poverty maps using aggregated mobile communication networks. *Conference on Human Factors in Computing Systems – Proceedings*. pg 511-520. (doi:http://dx.doi.org/10.1145/2556288.2557358/)
[122] Lu X, Wetter E, Bharti N, Tatem AJ, Bengtsson L. 2013 Approaching the limit of predictability in human mobility. *Scientific Reports*. **3**. (doi:http://dx.doi.org/10.1038/srep02923/)

Potential biases in location data

Tracking data can provide valuable insights which might help tackle persistent societal challenges, but it is rarely the case in big data that 'n = all'[123]. The Strava case above illustrates the potential biases in location data and, more generally, the costs of self-tracking devices and the capacity to make use of their insights restricts them to certain groups. A platform for flagging potholes to a local government through the use of accelerometers in phones while driving was demonstrated to focus only on particular parts of the city due to the fact that it excluded groups in the population unlikely to have smartphones[124]. It is important to distinguish between datasets that could capture a reasonable part of the population and those that relate to a more limited group – eg users of particular apps.

With holistic analysis that critically considers assumptions and quirks in the data, it is possible to compensate for some of these biases and issues. However, the use of these insights uncritically is likely to result in perverse outcomes, particularly in areas of low analytical capacity.

Consent

Giving consent for use of location-based data is a persistent challenge for a number of reasons. It is not always clear to users when location-data is being captured. Even if the user is aware of the data being sent to or from GPS monitors on their device, other modalities also effectively capture location data. The wireless signals that are present can, with the help of calibrated database, be used to locate individuals, even without connecting to them. Microphones on smart phones, for example, can be used to pick up unique ultrasound signals inaudible to humans hidden within songs or adverts broadcast in public places, which can be used for tracking purposes[125]. Media stories have focused on large companies' location collection and storage practices[126,127].

Pushed self-tracking systems can provide incentivisation at a level that can be considered coercive to some. For example, if the cost of a generic, non-discriminatory car insurance plan is prohibitively expensive, then the user may have to opt for a tracking device in order to get a lower premium. Providing access to data is not only sharing something with current economic value, it is also conferring the ability to infer further information – and a consumer might not know the extent of such inferences. While this is governed by 'purpose limitation' in data protection, this can be a vague concept. Furthermore, the level of inference that is possible can be a function of other people's consent, such as people travelling with you who might be releasing additional information, or consenting to their data being used in different ways.

Many of the websites for communal self-tracking have data that is publicly available or published on semi-public social media sites. Recent controversies, such as a Master's student

[123] For an analysis of this, see Mayer-Schönberger V, Cukier K. 2013 *Big data: A revolution that will transform how we live, work, and think*. UK: John Murray.
[124] Crawford K. 2013 The Hidden Biases in Big Data. *Harvard Business Review.* See https://hbr.org/2013/04/the-hidden-biases-in-big-data (accessed 26 June 2017).
[125] Mavroudis V, Hao S, Fratantonio Y, Maggi F, Kruegel C, Vigna G. 2017 On the privacy and security of the ultrasound ecosystem. *Proceedings on Privacy Enhancing Technologies*. **2**, 95-112. (doi:http://dx.doi.org/10.1515/popets-2017-0018)
[126] Chen BX. 2011 Why and how Apple is collecting your iPhone location data. *Wired*. See https://www.wired.com/2011/04/apple-iphone-tracking/ (accessed 19 December 2016).
[127] Google faces Streetview wi-fi snooping action. *BBC News*. 11 September 2013. See http://www.bbc.co.uk/news/technology-24047235 (accessed 19 December 2016).

publishing a paper and open dataset based on publicly facing data scraped from OKCupid, a dating site[128], have highlighted that people see the reordering of publicly available data into a form more amenable to search and analysis as a breach of privacy, and go beyond the use of data that they have consented to. Current legal battles are ongoing to clarify the limits of researchers' use of bots and web-scraping in the US[129].

Health applications which aim to monitor elderly individuals, usually with the aim of prolonging independent living, involve great deals of self-tracking. Many of these individuals are unable to give genuinely informed consent.

Public safety and feedback effects

Waze is a navigation app that provides base maps and route recommendations on the basis of data shared by users. In return for sending Waze anonymous location data and for flagging incidents such as traffic accidents, commuters benefit from more optimised routing. Waze 'gamifies' its interface, showing users avatars of nearby other users, and encouraging them to thank each other as both a community-building mechanisms and a vector for reputation data[130].

The case of Waze has revealed tensions between open data and public safety. It was reported in the media that the Los Angeles Police Department wrote a letter to Alphabet, Google's parent company that acquired Waze in 2013, noting their concerns that the police tracking capabilities of Waze could be used by those who wish to either avoid the police or harm them. They noted that Ismaaiyl Brinsley, who killed two New York Police Department officers at the end of 2013, had used the data provided by the platform to track police in the days leading up to the shooting[131]. Waze has also been blamed by cities for stigmatising particular 'dangerous' areas[132], and for at least one further death due to misdirection into a "gang-controlled favela"[133].

These crowdsourced datasets suffer from selection bias due to feedback effects. If users are predominantly routed in a particular direction, then the knowledge about the situation on a different road is limited. There are also ethical issues relating to guiding users down a certain route with the intention of gathering more data to benefit the aggregate user base, which it has been argued should be seen through the lens of research ethics[134].

---

[128] Resnick B. 2016 Researchers just released profile data on 70,000 OkCupid users without permission. *Vox.* 5 December 2016. See http://www.vox.com/2016/5/12/11666116/70000-okcupid-users-data-release (accessed 26 June 2017).
[129] Farivar, C. 2016 To study possibly racist algorithms, professors have to sue the US. *Ars Technica UK.* 29 June 2016. See http://arstechnica.co.uk/tech-policy/2016/06/do-housing-jobs-sites-have-racist-algorithms-academics-sue-to-find-out/ (accessed 26 June 2017).
[130] Wang G, Wang B, Wang T, Nika A, Zheg H, Zhao BY. 2016 Defending against Sybil devices in crowdsourced mapping services. *Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services*. Pg 179-191. (doi:http://dx.doi.org/10.1145/2906388.2906420/)
[131] Hamilton M, Winton R. 2015 LAPD chief Beck concerned traffic app Waze puts police in harm's way. *Los Angeles Times.* 26 January 2015. See http://www.latimes.com/local/lanow/la-me-ln-lapd-chief-concerned-waze-20150126-story.html (accessed 26 June 2017).
[132] Qalandiya Drama Not the First Waze Controversy. *Arutz Sheva.* 1 March 2016. See http://www.israelnationalnews.com/News/News.aspx/208745 (accessed 26 June 2017).
[133] Darlington S. 2015 Waze app directions take woman to wrong brazil address, where she is killed. *CNN*. 8 October 2015. See http://edition.cnn.com/2015/10/05/americas/brazil-wrong-directions-death (accessed 26 June 2017).
[134] Rosenblat A, Hwang T. 2016 The wisdom of the captured. *Intelligence and Autonomy*. See http://datasociety.net/pubs/ia/Wisdom_of_Captured_09-16.pdf (accessed 26 June 2017).

Gaming attacks on systems using location data have been distinguished in academic work on the topic, both of which create fake data on the service for malicious ends[135]. The first consists of the reporting of fake events, such as a traffic accident, with the intention of rerouting users and reducing congestion for the attacker[136]. The second type of attack is the generation of fake vehicle data, using simulated location sensors, which can either cause the rerouting of traffic directly or can amplify the first kind of attack by lending it credibility.

6.5 Governance of the management and use of location data

Given the high re-identifiability of location data, it is listed as personal data inside the General Data Protection Regulation (GDPR)), though it is never explicitly defined in the text. Given the broad definition of personal data, location data seems likely to encompass variables including address, coordinates, addresses on a network (e.g. IP addresses) or addresses of hardware (MAC addresses, IMEI numbers) – though many of these virtual addresses can be spoofed. The GDPR, for example, applies to 'natural persons, whatever their nationality or place of residence', who fall within European jurisdiction. Yet for a business to be sure they are in compliance with the law, it would seem likely that *de facto* it would have to apply to anyone who appears, within cost-effective means of determination, to meet this category. The possibility of somebody located outside of a jurisdiction with GDPR or GDPR-equivalent law, but using technology such as a Virtual Private Network (VPN) poses interesting challenges for the future practical governance of location data.

Another issue is that the uses of location data are broad enough to allow gaming within the existing regulations, particularly regarding grounds for collection and processing. It is difficult to collect data about health, but the GDPR is more lenient towards the processing of data for health purposes. Consequently, applications could seek to collect location data for 'lifestyle' purposes, while processing it for so-called health purposes, to take advantage of this leniency.

A variety of technical approaches to better obscure and obfuscate location data have been written about, with some that have been trialled or made into products – though they do not always work and may give a false sense of security[137]. Many of these provide a layer between the hardware and software of smartphones and other devices that alters or replaces the signal from a sensor such as a GPS in a way that attempts to maximise functionality for the user (i.e. by 'fooling' apps). Other solutions are more physical, such as the growth in the market of RFID-proof bags and clothing[138], or RFID 'blocker tags' that jam readers by simulating all or a range of RFID codes simultaneously[139]. Initiatives which give individuals more control over their data can provide a useful governance function.

---

[135] Wang G, Wang B, Wang T, Nika A, Zheg H, Zhao BY. 2016 Defending against Sybil devices in crowdsourced mapping services. *Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services*. Pg 179-191. (doi:http://dx.doi.org/10.1145/2906388.2906420/)

[136] Hendrix S. 2016 Traffic-weary homeowners and Waze are at war, again. Guess who's winning? *The Washington Post.* 5 June 2016. See https://www.washingtonpost.com/local/traffic-weary-homeowners-and-waze-are-at-war-again-guess-whos-winning/2016/06/05/c466df46-299d-11e6-b989-4e5479715b54_story.html (accessed 26 June 2017).

[137] TrackMeNot. See https://www.schneier.com/blog/archives/2006/08/trackmenot_1.html (accessed 26 June 2017).

[138] Petsinger JA. 2000 Electromagnetic shield to prevent surreptitious access to contactless smartcards. See https://www.google.com/patents/US6121544 (accessed 26 June 2017).

[139] Juels A, Rivest RL, Szydlo M. 2003 The blocker tag: Selective blocking of RFID tags for consumer privacy. *Proceedings of the 10th ACM conference on Computer and communications security*. pg 103-111. (doi:http://dx.doi.org/10.1145/948109.948126)

6.6 Concluding remarks

Location data, when aggregated, has great potential in improving cities and planning, making evidence cheaper and more accessible for urban decision-makers in particular. Yet it can be highly revealing about individual characteristics and behaviour, some of which might be sensitive and unknowingly revealed. Governance needs in relation to use of location data concern understanding what is meaningful consent to use of this data, and finding means of protecting private information that can be inferred from it where possible. There are also research governance concerns about the reliability, accuracy and representativeness of this data, which can be biased or easily spoofed.

7. Data and humanitarian crises[140]

7.1 Introduction

Disaster situations – from natural disasters to outbreaks of disease – require rapid but targeted humanitarian response. This response is enabled by the availability of accurate data about areas and communities affected. However humanitarian crises can also be situations where traditional data infrastructures might have failed, necessary data sharing might not be in place, expertise might be external or acting at a distance, and normal procedures for checking and verifying data might be unreliable or unworkable. Secondary sources of data, such as from social media, can be untrustworthy, but might be some of the only sources of information that are accessible to feed into particular decisions.

While better decision-making in humanitarian crises is an obvious benefit, making data work in real-world circumstances to aid decisions can clearly be challenging. The coalescing field of 'disaster informatics' and data science for crisis must overcome significant technical and institutional hurdles before it is able to prove its worth in the field. Several data governance efforts have emerged in this space to tackle the many challenges that exist. This case study will outline some of the different challenges faced by this field, and how they have (or have not) been dealt with effectively. It illustrates the particular challenges of governance of data use when decisions have to be made rapidly, drawing on imperfect data.

7.2 Social and ethical tensions

This case study highlights the challenge of managing the common social and ethical tensions in data management and use in the context of a humanitarian crisis. It picks up on the tension of *incentivising innovative uses of data while ensuring that such data can be traded and transferred in mutually beneficial ways*, where the benefit of sharing data and making innovative use of it is to society and in particular to communities affected by a crisis; and it raises questions about how the usual arrangements for transfer of data, or data sharing, might be different in these circumstances. It also highlights an issues around the risks and benefits of acting on data such as social media posts which can be very valuable, but where accuracy and verifiability are a challenge.

7.3 Opportunities for using data for humanitarian response

Both pre-existing and real-time data can be used to enable and improve response in humanitarian crises.

[140] This case study benefitted from conversations with Professor Bartel van de Walle, Dr. Catherine Eaton, Dame Professor Wendy Hall and Dr. Pablo Suarez. Professor Geoff Gilbert kindly donated time to review this document.

Pre-existing data:

- Public sector data: Data useful to organisations on the ground is often already known in some form to public sector agencies. Some governments are making strong progress towards open data, though for others important data is still stored in fragmented and inconsistent ways. These data can also be used to feed into early warning systems, creating opportunities for schemes such as forecast-based financing (FbF)[141], where menus of action are triggered following the breach of certain risk thresholds in predictive models of natural hazards.
- Private sector data: Companies also hold information that might be useful in crisis situations. Logistical information from companies operating in the region can supplement mapping and access data. The World Food Programme working with UN Global Pulse has demonstrated that data from mobile phone signal can be used to estimate multidimensional poverty indices and food demand[142]. Information such as aerial imaging, which might be held primarily by private sector organisations, can be invaluable in a crisis[143].
- Volunteer-driven data: Often supported by a range of public and private data, initiatives such as Missing Maps/OpenStreetMap seek to create strong datasets, organised through a mix of physical meetups and virtual collaborations.

Real-time data:

- Sensor data: Data can be collected from volunteers on the ground, for example SMS reporting of natural or social events. New technology can also be deployed to gather data on the ground. Wireless health sensors known as STAMP2[144] were deployed during the Ebola crisis to aid with the remote monitoring of patients, particularly due to the time and effort needed for health workers to put on and remove protective equipment[145].
- Social data: Data from social media has can help identify trends and incidents, as well as providing useful contextual information. Users of social media outside of affected regions are also able to provide data processing services, such as labelling images in ways useful for response teams.

7.4 Challenges in using data for humanitarian response

Checking: reliability and verifiability of data

The high-stakes nature of humanitarian response makes validity of data a crucial issue. Many humanitarian funders and researchers are exploring the promise of social media to better

[141] Coughlan de Perez E, van den Hurk B, van Aalst MK, Jongman B, Klose T, Suarez P. 2015 Forecast-based financing: an approach for catalysing humanitarian action based on extreme weather and climate forecasts. *Nat Hazards Earth Syst.* **15**, 895-904. (doi:http://dx.doi.org/10.5194/nhess-15-895-2015)
[142] Decuyper A *et al*. 2014 Estimating food consumption and poverty indices with mobile phone data. *CoRR*. See https://arxiv.org/abs/1412.2595 (accessed 26 June 2017).
[143] Crowley J, Chan J. 2011 Disaster Relief 2.0: The Future of Information Sharing in Humanitarian Emergencies. Washington, D.C. and Berkshire, UK: UN Foundation & Vodafone Foundation Technology Partnership.
[144] Sensor Technology and Analytics to Monitor, Predict, and Protect Ebola Patients.
[145] Steinhubl SR, Marriott MP, Wegerich SW. 2015 Remote sensing of vital signs: A wearable, wireless “Band-Aid” sensor with personalized analytics for improved Ebola patient care and worker safety.*Global Health: Sciencea and Practice*. **3**, 516-519. (doi:http://dx.doi.org/10.9745/GHSP-D-15-00189)

inform crisis situations. Social media companies are also becoming important implicit and explicit partners in humanitarian endeavours.

Systems utilising social media data currently often scrape and transform microblogging data, such as those from Twitter, extracting aspects such as volume, geolocation, or semantic meaning[146]. Some of these systems work with generic and relatively automated processing systems, which can be applied to different sets of data, while others are built as a sociotechnical process. For example, systems that pipe out to a crowdsourced labelling system, which trains a supervised learning algorithm to predict labels for new tweets[147]. A main part of the challenge in this sector is separating the 'trustworthy tweet' from untrustworthy ones. At the moment, this kind of response is usually only deemed suitable when the risks of ignoring an accurate tweet outweigh the risks of acting on an incorrect one, often at the early, low-information parts of a response where uncertainty is high, but less for resource-scarce high stakes activities such as search and rescue[148]. Several systems have been proposed to measure the credibility of the tweet, most of which are crowdsourcing platforms combined with machine learning systems.

Establishing trustworthiness as a third party can be difficult to do without good and open access to social media dataflows. Out of context of a conversational flow, posts can be hard to parse and interpret, but the way that the social media providers allows access to their dataflows limits the ability to undertake this kind of analysis quickly and at scale[149]. Even where data is complete, it is unclear how different dynamics or patterns provide information about trustworthiness or not – eg whether multiple posts or tweets about the same thing lead to higher level of trustworthiness, or what a great volume of posts indicate about the size of an incident.

Sharing: interoperability, accessibility and sustainability of data

Technical barriers exist to the sharing of data — the creation of common formats, the transmission in areas with patchy or damaged infrastructure, the use of different language encodings or preferred regional software packages, for example. Archiving of data — and accessing these archives in crisis scenarios, also poses significant challenges and difficulties.

Social media use is also varied across the globe[150], but these different platforms give different access rights to humanitarian organisations and researchers. This stems not only from internal policies, but the cultural use of these platforms: many fewer people use Twitter on a private mode than those who apply some privacy settings to their Facebook accounts, for example. Social media platform use also varies with demographic within a particular geographic region, and the interaction of the openness of the platform with the response might have a distributive effect on response, rescue and relief.

---

[146] Imran M, Castillo C, Diaz F, Vieweg S. 2015 Processing social media messages in mass emergency: A survey *ACM Comput*. **47**, 67. (doi:http://dx.doi.org/10.1145/2771588)

[147] Imran M, Castillo, C, Lucas J, Meier P, Vieweg S. 2014 AIDR: Artificial intelligence for disaster response. *Proceedings of the 23rd International Conference on World Wide Web. 195-162.* (doi:http://doi.acm.org/10.1145/2567948.2577034)

[148] Tapia AH, Moore KA, Johnson NJ. 2013 Beyond the trustworthy tweet: A deeper understanding of microblogged data use by disaster response and humanitarian relief organizations. *ISCRAM.* See http://help.iscram.org/legacy/ISCRAM2013/files/121.pdf (accessed 26 June 2017).

[149] Palen L, Anderson KM. 2016 Crisis informatics—New Data for Extraordinary Times. *Science*. **353**, 224-225. (doi:http://dx.doi.org/10.1126/science.aag2579)

[150] World Map of Social Networks. See: http://vincos.it/world-map-of-social-networks/ (accessed 8 December 2016).

7.5 Governance of management and use of data for humanitarian response

OCHA, the UN Office for the Coordination of Humanitarian Affairs, has been responsible for fostering much of the high level data governance in the humanitarian domain. The 2002 Symposium on Best Practices in Humanitarian Information Management and Exchange, and the 2007 Global Symposium +5, were two of the most influential fora for formulating the initial shape of a more universal governance system[151]. The principles that were developed can be grouped into four broad categories, covering the checking, sharing, using of data, and broad humanitarian values that underpin the sector's ethos[152].

The Humanitarian Data Exchange[153] is a platform set up in July 2014 by OCHA which has over 4,400 datasets being shared for analytic purposes. This is accompanied by an emerging quality assurance framework for this kind of data[154] as well as the Humanitarian Exchange Language (HXL), which is a system designed to tag existing data formats, such as spreadsheet columns, in flexible and extendible ways to allow for fast comparability in crisis situations. HXL is developed by a multi-stakeholder working group with membership including the ICT4Peace Foundation, the Humanitarian Innovation Fund, the UN Office for the Coordination of Humanitarian Affairs, Save the Children, the UN High Commissioner for Refugees, UNICEF, USAID, the World Bank, and the World Food Programme. It was developed as a cooperative, not a competitive standard — one built for adoption in mostly existing systems. The new OCHA humanitarian data centre, currently being established in the Netherlands, also has responsibilities to promote standardisation and data literacy across many humanitarian actors.

OCHA is just one of several actors in this space that have developed relevant governance regimes. A range of organisations have diverse approaches and codes touching on data responsibility, including Oxfam, UN Global Pulse, UNICEF, USAID, ICRC, MSF and UNHRC[155].

Legally-grounded national and transnational data governance systems, such as data protection, also interact in important ways with the humanitarian system. In particular, where collaboration around personal or corporate data is required with little notice on the ground, organisations with important data (such as logistics data from a home delivery firm) face challenges in knowing whether the crisis allows them to legally waive the usual governance arrangements they have in place. Even where the law allows for such processing and exchanges (e.g. as part of a proportionality test), crises are often so rare that expertise on these workarounds is not known by individuals on the ground — and in a crisis situation, it is difficult to convene bodies that might usually discuss such proportionality and authorise or refuse data sharing requests. Unclear liability regimes could lead to inertia and stasis, as it

[151] Van de Walle B, Van Den Eede G, Muhren W. 2009 Humanitarian Information Management and Systems. In: Löffler J, Klann M (eds), *Mobile Response*. **5424**, 12-21. (doi:http://dx.doi.org/10.1007/978-3-642-00440-7_2)
[152] Van de Walle B, Comes T. 2015 On the nature of information management in complex and natural disasters.*Procedia Engineering*. **107**, 403-411. (doi: https://doi.org/10.1016/j.proeng.2015.06.098)
[153] The Humanitarian Data Exchange. See https://data.humdata.org (accessed 26 June 2017).
[154] United Nations Office for the Coordination of Humanitarian Affairs (OCHA). Humanitarian Data Exchange Quality Assurance Framework. See https://docs.humdata.org/wp-content/uploads/HDX_Quality_Assurance_Framework_Draft.pdf (accessed 26 June 2017).
[155] Berens J, Mans U, Verhulst S. 2016 Mapping and Comparing Responsible Data Approaches. *GovLab and Centre for Innovation*, Leiden University. See http://www.thegovlab.org/static/files/publications/ocha.pdf (accessed 26 June 2017).

can be unclear whether individuals are personally liable for release of data outside of established protocols. While certain humanitarian organisations have the right of non-disclosure[156], the way that data moves in fast-developing situations on the ground with a variety of actors is difficult to control, particularly in sensitive situations like border zones, or when working with or beside military organisations.

Some of these issues might be clarified by the use of anticipatory tools. In particular, a range of humanitarian organisations are turning to Privacy Impact Assessments or Data Protection Impact Assessments, grounded either in their own international practices and codes of conduct or on national/European data protection laws, to better anticipate risks from data use[157].

7.6 Concluding remarks

Humanitarian data governance produces challenges on a number of fronts. Firstly, the process of data location, compilation, cleaning and analysis for decision support is rapidly sped up as a result of the temporal urgency of disaster situations. Even where legal frameworks are ready and in place, the people who understand and operationalise them might not be, particularly where the events are extremely infrequent. Many tensions arise between the legal frameworks, institutional norms, high stakes nature of the decisions that need to be made, and the rapidity at which strategies have to be decided.

Disaster situations can also be situations where traditional data infrastructures might have failed, expertise might be external or acting at a distance, and normal procedures for checking and verifying data might be unreliable or unworkable. Secondary sources of data, such as from social media, can be untrustworthy, but might be some of the only sources of information that are accessible to feed into particular decisions. Consent, awareness and security are all hard to establish and maintain in these environments.

Many of the challenges of humanitarian data emphasise and echo broader data governance issues around documentation, better metadata, availability and more streamlined systems of need and access. Yet crisis situations also change the calculus around data sharing for many organisations — firm data on areas such as logistics, once secret for competitive advantage reasons, might suddenly have new users and new incentives for sharing. It could be said that data governance for humanitarian crises is largely about high quality general data governance which does not break apart in extreme situations.

[156] Council of the European Union. 2012 Applicability of the General Data Protection Regulation to the activities of the International Committee of the Red Cross (ICRC). Interinstitutional File: 2012/0011 (COD). See http://data.consilium.europa.eu/doc/document/ST-7355-2015-INIT/en/pdf (accessed 26 June 2017).
[157] Berens J, Mans U, Verhulst S. 2016 Mapping and Comparing Responsible Data Approaches. *GovLab and Centre for Innovation*, Leiden University. See http://www.thegovlab.org/static/files/publications/ocha.pdf (accessed 26 June 2017).