What are data? A categorization of the data sensitivity spectrum

John MM Rumbold,* Barbara K Pierscionek

School of Science and Technology,
Nottingham Trent University
Clifton Campus
Clifton Lane
Nottingham NG11 8NS, UK

John.Rumbold@ntu.ac.uk (corresponding author)

Abstract

The definition of data might at first glance seem prosaic, but formulating a definitive and useful definition is surprisingly difficult. This question is important because of the protection given to data in law and ethics. Healthcare data are universally considered sensitive (and confidential), so it might seem that the categorisation of less sensitive data is relatively unimportant for medical data research. This paper will explore the arguments that this is not necessarily the case and the relevance of recognizing this.

The categorization of data and information requires re-evaluation in the age of Big Data in order to ensure that the appropriate protections are given to different types of data. The aggregation of large amounts of data requires an assessment of the harms and benefits that pertain to large datasets linked together, rather than simply assessing each datum or dataset in isolation. Big Data produce new data via inferences, and this must be recognized in ethical assessments. We propose a schema for a granular assessment of data categories. The use of schemata such as this will assist decision-making by providing research ethics committees and information governance bodies with guidance about the relative sensitivities of data. This will ensure that appropriate and proportionate safeguards are provided for data research subjects and reduce inconsistency in decision making.

Keywords: Big Data; information ethics; data science

Funding: This work was supported by a Horizon 2020 grant from the European Union ICT 2014/1

Introduction

The definition of data might at first glance seem prosaic, but formulating a definitive and useful definition is surprisingly difficult. This question is important because of the protection given to data in law and ethics. Healthcare data are universally considered sensitive (and confidential), so it might seem that the categorisation of less sensitive data is relatively unimportant for medical data research. This paper will explore the arguments that this is not necessarily the case.

The terms data and information are sometimes used as synonyms and sometimes distinguished. Data protection legislation often does not distinguish the two concepts, except for using data to denote digitally stored information (although data protection laws may also protect non-digital data). The definition of data is surprisingly difficult. Communication (the transfer of data) has over 100 different definitions [1,2]. There are many discipline-specific definitions of information. One generic definition states that:

> *information is produced by all processes and it is the value of characteristics in the processes' output that are information* [3].
>
> Information has been more narrowly (and usefully) defined as "data that has been processed into a meaningful form" [4]. Other definitions include
>
> *Information is data that has been processed into a form that is meaningful to the recipient* [5].
>
> *Data is the raw material that is processed and refined to generate information* [6].
>
> *Information equals data plus meaning* [7].

The definition of data acquires great importance in the area of data protection and privacy, as the issue of what are personal data determines which data are protected by law and are confidential. The Open Data movement makes the issue even more important [8]. Whilst the status of healthcare data as sensitive personal data is enshrined in law, there are many other types of data used in data linkage studies. Their 'capacity' to contribute to the reidentification of subjects increases their potential sensitivity. Although there has been a great deal of discussion over which data are in the sensitive category, there is little examination of the different levels of sensitivity within the personal data category [9]. The two main ethical and legal issues in data protection – autonomy and privacy are interrelated concepts as encapsulated in the concept of informational self-determination. These issues are common to all healthcare research, although the harms are lesser in data research. Although there are a number of narrower issues and rights, they can all be rooted in one of these two concepts. Autonomy is relatively easy to define – it is the ability of a person to make decisions and act upon these. Privacy is more complex and covers several distinct concepts. The protection of physical space or person is not relevant to data protection, except by analogy. There have been several attempts to provide comprehensive taxonomies [10,11], and several seminal works on privacy in the USA, starting with Warren and Brandeis in 1890, and developing via Prosser, Westin, and Altman [12-15]. The German concept of 'informational self-determination' (which is part of the right to development of the personality in Article 2 of the German constitution), that one has the right to decide what personal information should be communicated to others and under

what circumstances (Westin's definition of privacy) [14], arguably covers all the issues relating to privacy and data.

Neither consent nor anonymisation is necessary nor sufficient in law or ethics for the use of personal data for research. Requiring consent or anonymisation will not guarantee protection of data subjects in all circumstances [16]. Neither of these rights are absolute, and there are provisions to override them in particular circumstances (however, an express refusal to research use must be respected unless there are exceptional circumstances). The opportunity to opt out of even anonymised data processing indicates that respect for privacy alone may be insufficient (although the right being protected is unclear – it may simply reflect a desire to maintain the social licence). Where consent is impossible or impracticable, governance mechanisms will permit processing of personal data if necessary and proportionate.

Do the public understand what data are, how their data are used, who controls it, and who will have access to it? The answer according to several studies is "in the negative" [17-22]. In particular, it has been found that privacy controls can give false reassurance to users [23]. Further, it has been shown that intentions do not translate into action [24]. The term "personal data" covers a massive range of data from the totally trivial to the extremely intimate.

British law has lagged behind Continental jurisdictions and the USA in the protection of privacy (although there is a common law duty of confidentiality), but now a right to privacy has emerged in the UK through case law (*Douglas v Hello!*), or based on Article 8 rights to a private and family life [26]. There are definite and well-recognized interests in protecting and keeping private personal data. It is possible that there are some data about a person that relate to nothing sufficiently important or personal and therefore no strong justification for them to be protected or kept private. However, there is a large amount of data about which their value and sensitivity depends both on context and the motivations and trustworthiness of the person accessing the data [27]. This is the justification for a wide definition of data for the purposes of legislation and regulation, but the use of a more detailed classification of different types of data could potentially increase the utility and reduce the risks of Big Data if it allowed a more nuanced definition of personal information [28]. This paper will examine whether more finely defining the categories of data in the context of research could enable more flexible and responsive approaches to privacy and autonomy. This is important for the maintenance of the social licence whilst maximising the utility of data for research projects [25].

Etymology of data

The term 'datum' (plural data) comes from Latin, meaning "a thing given". This says something about the nature of data - that it has its value in transmission. This concept of value in transmission can also be related to the legal status of a database in property as a "thing in action" (as opposed to a "thing in possession" – see below).

Definition of Big Data

The term 'Big Data' has been appropriated to mean many different things [29,30]. Collection of vast quantities of data have become economically feasible due to the massive decrease in the cost of digital storage [31] and data collection (due to the proliferation of smartphones and other devices) [32]. Big Data could be characterized as the value of vast amounts of data, which

are of little if any value in small quantities. This can result in the tragedy of the anticommons, i.e. the inability to use resources because so many people have the right to exclude others which may occur - in particular, if the data subjects wish to monetise their data (see below) [33-35].

Data versus Personal Data versus Information versus Metadata

Returning to the opening question, what are data? What are the distinctions between data, personal data, and information? The terms 'data' and 'information' are often used interchangeably in legislation and regulations. In terms of public understanding, there is a useful differentiation to be made between statistics, data, and information. The term "statistics" better conveys the nature of aggregated anonymised data that cannot be traced to any individual.

Legally, personal data are defined in the EU by the Data Protection Directive (to be replaced in May 2018 by the General Data Protection Regulation). In the United Kingdom, the relevant transposition of the Data Protection Directive is the Data Protection Act 1998. Many European jurisdictions have a Personal Data Protection Act or equivalent, which is a more accurate title. There is no protection for data that are not linked to a person, directly or indirectly. Any data that can be linked to a person, directly or indirectly, are personal data.

Metadata is the term used by legislators for data about communications other than the actual text. For example, they might be the recipient, time, date and duration of a telephone call. The term is misleading – these "metadata" are still data. Their content might be less sensitive than the content of the communication, but they can permit more sensitive inferences.

Ownership of data

An issue related to privacy is the issue of data "ownership". However, control over data pertaining to oneself is about more than privacy – it is informational self-determination. One model that has been developed to acknowledge data subjects rights is the Nordic MyData model[36].Legal possession of a thing connotes the ability to exclude others from its possession or use. Legally, it is clear that data *per se* cannot be owned (*Oxford v Moss, Your Response Ltd v Datateam Business Media Ltd*). There have been calls for patients to have ownership of their data to facilitate data sharing [37]. This risks the tragedy of the anti-commons by making the aggregation of healthcare data for research more difficult [33,34]. This does not mean there are not rights in data [38]. The social license model requires recognition of data subjects' rights [25].

Databases are legal property as a "thing in action". This may be as intellectual property or as the *suis generis* database right such as provided by the Database Directive 96/9/EC. The act of possession of the paper or hard disk on which data are recorded does not equate to possession of the data or database in the legal sense. The issue of data "ownership" is particularly problematic with the Internet of Things, for example the output of medical monitors such as continuous glucose monitors or implantable cardioverter-defibrillators. Despite the data about the patient being stored on the device within the patient's body, the patient may have no rights to access the raw data.

Data in the public domain (including surveillance by CCTV)

Data that relate to what is publicly visible are not necessarily non-sensitive data. Hair, skin and eye colour can be observed by anyone (except communities where veiling of the face and/or eyes is commonplace). The counter-example is that a person may not wish to have certain characteristics disseminated to a wide audience. Race may be immediately apparent on the basis of skin colour, but again this would be widely seen as a protected characteristic. Certain characteristics are sensitive due to socio-cultural issues of perceived or actual discrimination, rather than privacy as such.

The protection of privacy encompasses more than the protection of confidential information. There are privacy issues even in public spaces. A Russian photographer demonstrated that he could identify via social media strangers he had photographed in the street using a Russian app called "FindFace" [39]. An art student developed make-up schemes and clothes to supposedly prevent surveillance by CCTV and drones and whilst this has not been seriously treated (the make-up schemes drew human attention to the person) [40], it highlights approaches taken to protect privacy in public places. There are commercial websites advertising clothing with electromagnetic field (EMF)shielding incorporated [41]. Solove discusses the privacy issues of identity – the fact that this enables the linkage of an individual with the aggregated facts about them makes it a form of disclosure [11]. Data about individuals can be produced from other data initially gathered that was not predicted by those individuals who consented only to the data that was collected. This can be a violation of privacy even if the individual gave consent, which may have seemed innocuous individually or even collectively. A celebrity who eats or sunbathes in public might not wish for photographs of these activities to be published in a magazine (*Hannover v Germany*). The CCTV footage of a man attempting suicide in public cannot be broadcast on TV (*Peck v UK*). In some countries and territories (including Brazil, France, Hungary, Macau, Spain, and Switzerland), there are restrictions on taking or publishing photographs of persons in public spaces, which may be greater or lesser for persons in the public eye.

Equally, the harm caused by surveillance is not limited to the retention of data. The simple fact of monitoring the actions of individuals by CCTV cameras or wire taps is an intrusion that requires justification, because it has a chilling effect on behaviour. As the case of *Nader v General Motors Corporation* demonstrates, even the surveillance of actions in public where sufficiently pervasive will invade privacy. The court held that

> A person does not automatically make public everything he does merely by being in a public place.

There are numerous examples of acts performed in public that would not be expected to be subject to close scrutiny. The issue with Google StreetView and the broadcast of CCTV footage is not just scrutiny, which may or may not be any closer or wider than that of the general public (there has been an issue with the height of the cameras, which provided a viewpoint higher than that of a pedestrian), but dissemination to a far wider audience [11 at p 491,41]A person vomiting in public or exiting adult shops is

highly unlikely to want this behaviour broadcast to all the world. This distinction was the basis for the decisions in *Hannover v Germany* and *Peck v UK*. An aerial vantage point for

observation may or may not amount to an invasion of privacy, depending on the modality (*Kyllo v. United States*; *Florida v Riley*; *Dow Chemical Co. v USA*).

The increasing availability of devices that can record personal data has the potential to evade data protection legislation. For example, a UK police force recently asked for dashcam footage of drivers on their mobile phones [43]. Therefore, even when data are gathered from a public place, it is erroneous to assume there are no privacy rights on that basis.

Context

The value of data often lies in linkage with other data, and thereby creating new data. The woman of fertile age purchasing a certain type of lotion and certain supplements could be surmised to be pregnant based on previous data analysis – so these simple purchases reveal a far more personal fact [44]. Aggregation of data produces new data, which the data subject may be uncomfortable with the data user knowing. This production of new knowledge is the aim of data linkage research and in this context it is seen as a positive thing; in many other contexts it may be seen as a negative consequence. It erodes the ability of the individual to control who knows what about themselves[45]. Aggregation of data also leads to the major issue with de-identified data – re-identification. It can also lead to automated profiling and behavioural tracking. A rule identified by data mining might be ethically sensitive, and there has been a tool developed to flag up potentially problematic findings [46].

Re-identification and temporal variations

Some data are proxies for the individual and readily recognized as potential re-identifiers – the registration mark of a car, for example. The tracking of the location of a vehicle has been deemed not an invasion of privacy that required a warrant in *United States v Knotts* on the grounds that this amounted to a following of the vehicle on the public highways. The combination of date of birth and postcode clearly allows easy identification of the individual. The combination of date of birth, sex and the five-digit zip code has been found to identify 87% of US residents [47]. Dynamic or static IP addresses have been variously held to be personal data or not depending on the jurisdiction. Static IP addresses are generally considered personal data, although there are circumstances of multiple users of a computer where that is not the case. Dynamic IP addresses often require additional information to become personal data. The CJEU examined the issue following a referral from the German Federal Court of Justice. The GDPR states

> *Natural persons may be associated with online identifiers provided by their devices, applications, tools and protocols, such as internet protocol addresses, cookie identifiers or other identifiers such as radio frequency identification tags. This may leave traces which, in particular, when combined with unique identifiers and other information received by the servers, may be used to create profiles of the natural persons and identify them (Recital 30).*

Breyer v Germany confirmed that dynamic IP addresses by themselves are not personal data *per se*; it depends on the associated information.

The Article 29 Working Party acknowledged the need to reassess the risk of re-identification over time:

> *data controllers should consider that an anonymised dataset can still present residual risks to data subjects. Indeed, on the one hand, anonymisation and re-identification are active*

> *fields of research and new discoveries are regularly published, and on the other hand even anonymised data, like statistics, may be used to enrich existing profiles of individuals, thus creating new data protection issues. Thus, anonymisation should not be regarded as a one-off exercise and the attending risks should be reassessed regularly by data controllers*[48].

Therefore, data are dynamically defined, and data might to be either re-categorized or subject to further anonymisation procedures for future uses in order to continue protecting the data subject. The Australian Privacy Commissioner counselled that *all* data, even pseudonymised or anonymised, should be treated as personal because of this possibility [49]. The frequency of reassessment is not addressed by the Article 29 Working Party (which will be replaced by the European Union Data Protection Board once the GDPR is in force). This adds a temporal dimension to the definition of data, whereby the nature of the data may change with time from being non-identifiable to re-identifiable. When further data are available in the public domain, re-identification may become possible. Open Data initiatives propose a massive increase in the amount of data available to the public. This has the potential for rendering large swathes of anonymised data effectively personal data again. The other temporal aspect relates to the so-called "right to be forgotten". This relates to another aspect of privacy, namely the right to be able to move on from past history and re-invent oneself.

The implications for this are that future risks must be considered when releasing any data to the public, and therefore the strictest possible anonymisation standards must be applied to minimise the risks of re-identification in the future. Data cannot be recalled, and so the greatest risk to breach from techniques, skills, additional information and motivations of any user should be assumed. It is notable that the Information Commissioner's Office recognizes the difficulties in predicting the future risk of re-identification, nonetheless it ruled that Queen Mary's University of London had to release anonymised patient data in response to a freedom of information request [50].

Data versus information

Data has several definitions, but the common theme is that data are more concrete and information is more abstract. Individually, data are rarely useful. A date alone may be–an appointment, a holiday or an anniversary. However, data are often used to mean specifically digitally stored quantified information, especially as the substrate for computerised processing. It usually refers to unprocessed data, although the term "raw data" reveals that this is not always the case.

Information is formed by an organization and/or by analysis of data. The process of producing information from data may increase *or* decrease the personal nature of the data. Using the bare term 'data' in all these circumstances is confusing for the public. The abstract nature of massed data used for statistics is vastly different from identifiable medical records, yet this is not always communicated effectively.

The Nuffield Council on Bioethics distinguished between data they call 'raw materials' used for analysis and the information such data provide, which is governed by their relationship to 'other facts or conclusions [51].' Data must have some context to have any information value.

The amount of water used by the public might seem inconsequential (unless there is a drought and a hosepipe ban), but these data have been used by Arkansas police to build a case against a murder suspect. It was suggested that the use of large amounts of water in the early hours of

the morning was to wash away blood residues [52]. Pacemaker data were recently used to charge a man with arson and insurance fraud [53]. At the other extreme, the mantra of the Personal Informatics AKA Quantified Self movement is "You are your data". It is currently only in the realms of science fiction that one's entire consciousness can be uploaded. There is no strict dichotomy between data and information; rather there is a spectrum, and a datum will move along the spectrum of informational value according to added information, deductions, and inferences.

If the case of a patient with an unusual first name and unusual medical condition is presented at a conference, the relatives might be able to recognize the patient involved. If no one else could identify the patient from the details provided, it is questionable whether there has been a breach of confidentiality? It would be likely to depend on what information was presented. If coronary angiograms were posted online without any identifiers, would they be personal data? Each coronary angiogram is unique, but only persons who already have access to patient records would be able to link that coronary angiogram with an individual – people who already had seen that medical image in another setting. The same applies to unprocessed retinal images (as opposed to the same images processed for biometric purposes). This difference is explicitly acknowledged in the GDPR.

Anonymised data are generally outside the remit of data protection laws, which are concerned with personal data. In terms of autonomy, it might be argued that the person has the right to refuse the use of their anonymised data for medical research they find objectionable. There is no issue of privacy in this case. The requirement in the DPD and GDPR to respect opt-outs *may* reflect a desire to respect autonomy (it might also simply reflect political pressures and a desire to secure social licence).

Another possibility is the use of any opt-out to secure money. There was a move to secure payment for data in exchange for not exercising an opt-out during the Iceland Genome Project. This prompted a backlash from the Icelandic public [35]. The public are used to the concept of a value exchange particularly on social media. The public expect their healthcare data to be shared appropriately to ensure safe healthcare. Data subjects in research projects ought to be seen as 'data donors' (or 'data philanthropists') with the value being created for the common good [54]. There are other ways of providing value, including the provision of information about the research being conducted

Motivation and trustworthiness of user

Data have value. Millennials understand this, and their digital life reflects this acceptance of an exchange of value. Users of social media understand that their data secures free use of a valuable platform. However, they may not realise exactly how much of their data is being gathered, and who is using it [55,56]. In particular, automated profiling may result in decisions being made about them with no input, awareness of the process, or ability to detect erroneous information.

Empirical studies have confirmed that the public have different expectations about the protection of their healthcare data depending on the data user [57,58]. One study of British attitudes to the commercial use of health data found that 17% of the public did not wish commercial companies to use their data for research under any circumstances. In particular,

insurance companies were distrusted by the public, who apparently saw no social good in providing their healthcare data to them.

The same study established that data subjects go through four questions consecutively when deciding on whether or not their data can be used:

- **WHY** Is the data being used for public benefit or commercial gain?
- **WHO** Can the data user be trusted to produce public benefit?
- **WHAT** What data are being used?
- **HOW** Are the data being stored and processed securely?

These questions all relate to the motivation and trustworthiness of the data user (apart from "What?"). Trust is a subjective feeling that the data user cannot control, but trustworthiness is an objective state of being deserving of trust that the data user *can* control. SHIP addressed the issues with the concept of safe havens, which addressed three aspects of information governance:

- Safe people
- Safe data
- Safe environments

Safe people can be researchers bound by a professional or contractual duty of confidentiality. This can be extended to a commitment not to attempt re-identification (it should be noted that once the data subject has been re-identified, then the data are re-classified as personal data). They require vetting, training, accreditation, institutional guarantors. The possibility of legal sanctions for breaches of data protection (except good faith mistakes) is an important ingredient for accountability. However, if sanctions apply to all breaches then guiding principles become *de facto* regulations [59]. The Wellcome Trust research suggests that some of the public do not see commercial concerns as safe people, although when in partnership with academia and/or the NHS they are more acceptable.

Safe data have been processed to remove identifiers (direct and indirect) except where this will reduce the utility of the data to an unacceptable level. In some case, synthetic data are an alternative. The reliability of analyses performed on synthetic data is a concern[60], but they currently a substrate for testing platforms where real data are not yet available. Further iteration, when real data have been introduced, may be needed and the extent of subsequent iterative effort will depend on the reliability of the synthetic data.

Safe environments require a number of measures. Safe havens or remote access which does not enable extraction of raw data will prevent the dissemination of personal data by accident or malice – if data are downloaded onto USB sticks they can and will be lost [61]. Removing this facility eliminates that possibility.

Some organizations add safe projects and safe outputs [62]. Safe projects are ensured by appropriate governance. They are projects that comply with data protection law and where the potential for harm has been balanced against the potential for benefit. Safe projects may include the consideration of public benefit, which does not exclude commercial concerns (but may exclude gross economic exploitation without benefits for the community). The assessment of a suitable project may include factors beyond safety to include an assessment of public benefit.

Safe outputs require appropriate anonymisation which takes into account all factors including context and population studied. Anonymisation metrics should be examined looking at k-anonymisation and differential privacy, for example. Data masking, grouping or deletion may achieve sufficient anonymisation, or synthetic data can be substituted.

The purposes for which data will be used is relevant. Registering a product enables consumers to be warned about dangerous defects and so take appropriate action and arrange repairs. The use of healthcare data to contact people for UK NHS screening may be viewed rather differently from their use by private companies to offer screening. The consent for communications is often a blanket consent, with no options to limit the number of communications in a particular time period. A more granular approach that allows the selection of different privacy levels is preferable.

It is difficult to categorize any data relating to opinions or preferences *a priori* as innocuous. A favourite colour may not seem personal, but pets' names may give vital clues to someone's passwords. Robert Bork's video rental history was leaked after his nomination for the US Supreme Court in 1987. There were no shocking revelations from this leak (intended to demonstrate the folly of Bork's stance on privacy), but nonetheless the Video Privacy Protection Act was passed in 1988 by the Reagan administration. For some (notably Senators Pat Leahy and Paul Simon), the fact that Robert Bork watched *A Day at the Races* was worthy of protection [63]. On the other hand, it is apparent that the revelation that the husband of Jacqui Smith (UK Home Secretary at the time) watched two adult films in the privacy of their home caused a scandal and might be considered sensitive information [64].

These concrete examples demonstrate how difficult it is to define data without knowledge of the motives and trustworthiness of the user. It also indicates that the sensitivity of a category ought to be determined by the users with the most sensitive data in that category. Heterosexuals may well consider sexual orientation less worthy of protection than homosexuals in certain societies. If sexual orientation is designated as non-sensitive when the person is heterosexual, then it enables identification of non-heterosexuals indirectly when the data is classified as sensitive and obscured.

It is clear that merely staying within the bounds of the law alone will not guarantee that data subjects will feel their rights are protected, as the troubled Care.data project demonstrated. The requirements for social licence include

1) Reciprocity
2) Non-exploitation
3) Service of the common good [25]

The mere provision of NHS services may not be seen as a fair value exchange for the selling of data. The Wellcome Trust study contrasted the activities of 'buying' and 'doing' (in commercial transactions such as using Facebook or shopping online) with 'being' and 'service using' (in relation to use of the NHS). Some individuals have refused to use NHS services because of the improper use of their data [50, box 2.3].

Given the conflicting demands of data for healthcare research and data for public release, there is a compelling argument for different standards of anonymisation for different uses. This will enable the best combination of data utility versus privacy protection. There is no need for the

general public to be privy to the level of detail necessary for quality research. Many operations can be performed on synthetic data instead of personal data.

Categories of data

Some data has no connection to any natural person. The only protection for such data would relate to the interests of the controller e.g. proprietary interests in the intellectual property of a novel invention. The sensitivity of data categories *may* be reflected in legislation, but neither financial details and criminal convictions are classified as sensitive under the Data Protection Directive.

The UK Anonymisation Network (UKAN) classifies data into four types (Table1):

Table 1 **About People**	**Non-Identifiable data**	**Identifiable data**
Yes	Anonymised Data	Primary Personal data
No	Apersonal Data	Secondary Personal Data

Table 2.1 from The Anonymisation Decision-Making Framework under Creative Commons Attribution-Noncommercial-No Derivatives 4.0 License [27]

This classification emphasizes the fact that data that are not *about* people can nonetheless be personal data. Registration marks and addresses are not data about people, but they often relate to identifiable people. Dynamic internet provider (IP) addresses are data that may or may not be identifiable depending on what other data are associated with them (*Breyer v Germany*). The GDPR recognizes this in Article 4 and Recital 30 [65,66].

We propose the following six categories of data, based on their connection with a natural person, their propensity for being kept private, whether they are ascriptive or not, whether they are legally protected from discrimination or not, and their connection with sensitive issues such as beliefs or health. An address, a vehicle registration or a dynamic IP address might be strongly associated with an individual, so data about them is indirectly data about that individual. The data may reveal where they go, what they spend, what websites they view and much more. The fact that certain facts are readily observed in public does not mean there are no privacy implications, and the case law on photography of celebrities demonstrates this (*Hannover v Germany*). Distinction can be made between ascriptive and non-ascriptive characteristics. Non-ascriptive characteristics are not assigned by society and can be changed, unlike ascriptive characteristics such as race or caste. Gender is non-ascriptive, but sex is ascriptive. Immutability is not a criterion for ascriptive characteristics, but it is persuasive. Human behaviour is subject to social control and our thoughts and opinions can be subject to social judgment, as these variables do reflect character. Our ability to choose to protect our thoughts and opinions from being known or broadcast is an important safeguard [44]. For these reasons, our proposed classification divides the UKAN's primary personal data category into four further categories, and merges the anonymised and apersonal data categories into one category of non-personal data creating a new system of six categories.

a) Non-personal data

This applies when there is no connection that can be made to an individual or group of individuals; examples of this are details of particular new materials, new technologies or computer models and corresponds to the apersonal data category in the UKAN classification.

Thoroughly anonymised data would come into this category, but the difficulty is in defining what anonymised data are. UKAN define different types of anonymisation: formal, guaranteed, statistical and functional [27]. This differentiation emphasizes that anonymisation is not a straightforward process. Guaranteed anonymisation reduces the utility of data dramatically. Functional anonymisation reduces the risk of re-identification to an acceptable level. Most useful data are potentially re-identifiable.

b) Human-machine interactions

Data related to a motor vehicle potentially has a connection to a person (unless it is a fully autonomous vehicle). Tracking that motor vehicle is tracking the occupants by proxy. Any human-machine interaction could log human behaviour – be it driving patterns or browsing history. Privacy protection on the Internet of Things (IoT) is a major issue [67]. This should be of concern since a large proportion of Big Data will come from the IoT in future.

This category is secondary personal data in the UKAN classification.

c) Human demographics, behaviour, thoughts and opinions

Human demographics are socio-economic variables such as age, income, education and employment. They may reflect status but not character per se. Although race is a demographic, for other reasons it is a sensitive characteristic (see below).

The football team an individual supports could be problematic if this allegation indicates other things such as religious or political sympathies – for example, in Northern Ireland or the West of Scotland, supporting Celtic or Rangers is strongly linked with being Republican or Loyalist respectively. Thought and opinions are considered part of a persona, hence the notion that "thought crime" is an unacceptable invasion of personal privacy. Privacy concerns are among the objections to the use of functional magnetic resonance imaging studies as brain-based lie detection for the courts. Political and religious affiliations are protected and considered sensitive data because of the potential for discrimination.

d) Readily apparent human characteristics (unprotected)

This means unprotected human characteristics readily apparent to the human senses (sight, hearing, touch, smell) without any aids, such as an ophthalmoscope. In this context 'protected' and 'unprotected' refer to whether or not the characteristics are covered by EU anti-discrimination law.

Certain aspects of physical appearance cross over with behaviour, as they have been assumed by the person as an expression of their personality – piercings and hair style or colour, for example. People with tattoos or "Goth" styles can be discriminated against, but these are not currently protected under discrimination law in the UK, nor elsewhere in Europe to our knowledge (although some police forces in the UK record these as "hate incidents", they are not hate crimes as such) [68].

e) Readily apparent human characteristics (protected)

Such characteristics include race, sex, ethnic group, pregnancy (in the later stages). These categories are essentially closed, although caste was added as a potential category to race by an amendment to s9 of the Equality Act 2010.

f) Medical or healthcare data

This category involves data from some of the above categories in certain contexts. The presence of a plaster cast (visible characteristic) suggests a fracture, but simply blogging about this may infringe data protection law on the grounds that this is health data (*Bodil Lindqvist v Åklagarkammaren i Jönköping*).

There is a cross-over between human behaviour and medical data in the form of "wellness data", often produced by wearables or apps. An example is the fitness wristband that tracks activity levels. It is indubitably personal data, but could also be classified as healthcare data depending on the details recorded, how they are processed, and their use.

The above four categories would all come under the UKAN primary personal data category.

The sensitivity of data in these categories varies considerably within the categories, so these broad categories require further subdivision. The range of potential sensitivity of some subcategories is depicted visually in the graph below. This graph illustrates the difficulty in definitively quantifying the sensitivity of a particular type of data, with different subcategories covering a large portion of the spectrum and overlapping considerably.

The perceived sensitivity of a particular type of data varies widely both between societies or ethnic groups and within those groups. There are several factors related to the data subject that affect the level of protection required, such as capacity and vulnerability. Also, a particular datum might only be sensitive where the individual belongs to a group that is discriminated against, e.g. the issue of gender at birth is unproblematic for those that are not transgender. The classification ought to reflect the sensitivity of the members of those vulnerable groups.

Socio-cultural factors can dramatically alter the sensitivity of certain data; sexual orientation is much more sensitive in countries where Shariah law is practised, for example. Some characteristics are sensitive only because of the potential for discrimination, for example the categorisation of race and ethnic groups as sensitive in data protection law is due to this. The potential for misuse after accidental or malicious release also affects sensitivity considerations. Religion is highly sensitive in areas where there is a high degree of sectarian conflict.

Sensitivity of data

Table 2 details the potential spectrum of sensitivity for particular subcategories of data, with explanations in the appendix. The numbers in the cells refer to the relative frequency with which data would fall into that part of the spectrum for that data category e.g. occupation would rarely fall into the most sensitive data category

Table 2: The data sensitivity spectrum

Sensitivity increases from 0-10 with colours from green to red used as a visual aid. Relative frequency with which data would occur in any part of the sensitivity spectrum is given by numbers 1-4.

	10	9	8	7	6	5	4	3	2	1	0
A: Data not related to any human being											
Relating to objects											4
Anonymised data related to persons			1	1	1	1	2	3	3	4	3
B: Human/machine interaction											
Recordings of human/machine interaction						2	3	3	3	3	
Location data that act as proxy for human location			3	3	3	3	3	2			
C: Human demographics, behaviour, thoughts & opinions											
Purchasing habits			1	2	3	3	3	3	3	3	
Income			2	3	3	3	3	3			
Occupation	1	2	3	3	3	3	3	3			
Social class			3	3	3						
Address			2	3	3	3	3				
Location		1	2	3	3	3	3	3			
Opinions				3	3	3	3	3	3	3	
Religious or political beliefs			3	3	3	3	3				
Lifestyle or wellness data		3	3	3	3	3	3	3			
Sexual orientation			3	3	3	3	3				
Transgender status		3	3	3	3	2					
D: Readily apparent non-protected											
Facial images (non-processed)				3	3	3	3	3	2		
Body images (non-nude)			3	3	3	3	3	3	2		
Body images (nude)		3	3	3							
Any traits processed for biometrics		3	3	3							
E: Readily apparent protected											
Sex							3	3	3		
Age				3	3	3	3	3			
Pregnancy		3	3	3	3	3	3				
Race			3	3	3	3	3				
Ethnic group			3	3	3	3	3				
F: Medical or health data											
Wellness data					3	3	3	3			
Lifestyle data		3	3	3	3	3	3	3			
Genetic data	3	3	4	3	3						
Diagnoses		3	3	3	3						
Highly sensitive diagnoses	4	3									

These assessments would be highly contextual and the following caveats should be added.

i) the risk with anonymised data relates to the risk of re-identification (plus the sensitivity of the underlying data)

ii) the sensitivity of data relating to opinions, beliefs or sexual orientation would be potentially shifted to the left in an authoritarian or theocratic society

The sensitivity of data is a one-dimensional measure, but informational privacy concerns map onto several dimensions. The Internet Users Information Privacy Concerns scale looks at three dimensions – collection, control and awareness [69]. The Concern For Information Privacy scale looks at four dimensions – collection, unauthorized secondary use, improper access, and errors [70].

Conclusion

The expectations of privacy differ radically from person to person. It is impossible for any definition of personal data to encompass the expectations of the entire population. The law is interpreted to reflect the reasonable expectations of the public. Recent research in the UK has enabled greater insight into public attitudes towards the use of their healthcare data in different contexts. There is a need for more empirical work with different populations.

The law is not prescriptive about the safeguards for healthcare data, as statutory bodies are in place to make the decisions about individual cases. Therefore, ethical assessments are the chief means of regulation of healthcare research on data in the UK and the rest of the European Union. The opinions of ethicists may not reflect those of the public; in particular, they may be too risk averse (a paternalistic approach which denies many participants true autonomy). Although these decisions are highly fact-specific, they can be guided by an appreciation of the spectrum of data and the contextual factors that dictate the sensitivity of any given situation. Any categorization of personal data can only be a rough guide to the issues involved in particular circumstances. A structured approach to understanding concerns can however lead to a more nuanced evaluation of personal data.

Appendix: Details on Sensitivity of Data Table

The rationales for the designation of the subcategories of data identified are laid out below.

Related to objects: these data are totally apersonal, being unrelated even indirectly to an identifiable person.

Anonymised data related to persons: The categorisation of these data relates to the risk of re-identification, which is nearly always greater than zero. If the anonymisation is sufficiently weak and the data once identified is sufficiently sensitive, then even the anonymised data may warrant being treated as sensitive.

Recordings of human/machine interactions: These are generally secondary personal data. These range from the innocuous to voice data that record conversations (examples include smart televisions [71] and the Amazon Echo device). These data might be highly sensitive for all sorts of reasons – they might incriminate a murder suspect, for example. In the same case, water usage from the water meter was also used to build a case [72].

Driving patterns from "black boxes" (often installed for insurance purposes) are arguably primary personal data – certainly where additional data ties that behaviour to an individual. Dashcams record personal data.

IP addresses may or may not be personal data, depending on what data are associated with them (see above).

Location data that act as proxies for human location: See below for the issues relating to location data. Vehicle data is not synonymous with tracking a named individual, but in practice it may amount to the same thing.

Purchasing habits: As the infamous Target anecdote illustrates, purchasing habits can potentially reveal highly sensitive data through Big Data-driven inferences. They may also reveal an unhealthy lifestyle, illegal activity, and provide location and time data. Purchasing of economy brands has been linked to financial difficulties, and so resulting in a lowering of credit rating.

Income: These data are not considered sensitive in some cultures e.g. Sweden. The example encountered at the first UK Biobank Annual Meeting demonstrates how this information is considered almost more sensitive than health data by some subjects.

Occupation: Some occupations will be sensitive either for national security reasons or because of social stigma. Some professionals on social media will wish to hide their professional status. It will be an indicator of social class.

Social Class: The sensitivity of this datum relates almost entirely to social factors. There may be fear of discrimination.

Address: Privacy of address details (including telephone numbers, email accounts etc) are a key ingredient of privacy. "Doxing" (the unauthorised publication online of personal details) is an emerging form of online harassment [73]. An address can also be a proxy for other personal characteristics such as race and religion.

There are also data *about* an address, which may be indirectly about an individual or family. Infra-red cameras can detect abnormal heat signatures and so find possible cannabis-growing locations (see *Kyllo v. United States*). This was deemed intrusive both because it revealed details of activity inside the house and because it was a modality not commonly available. By contrast, visual surveillance from a police helicopter was not deemed intrusive, even though the helicopter offered a special vantage point (*Florida v Riley*). Whether or not they are deemed intrusive, they are personal data by the broader definition that they relate to one or more of a particular household.

Location: Location data has the potential to be very sensitive if it implies certain traits about the person e.g. religious beliefs, sexual orientation, or reveals illegal or immoral behaviour. It has the potential to identify the person – the combination of work and home location would identify most of us [74]. It is also an intrusive form of surveillance, even though the person can already be observed in public places.

Opinions: This is highly variable, and often depends on factors such as occupation and/or employer. Public officials will be expected to not express certain views in public fora. Participations in certain events might be considered incompatible with certain jobs.

Insurance companies have considered using social media accounts to adjust premiums for their customers [75].

Religious or political beliefs: These data will be very sensitive in particular socio-political environments. In sectarian societies, religious affiliation (regardless of practice) will be very sensitive.

Lifestyle or wellness data: The distinctions between these data and healthcare data are not easy to delineate. The main criterion is their application, which is the approach taken by the European Union.

Sexual orientation: The sensitivity of these data is largely dependent on social mores (and in some cases national laws).

Transgender status: Transgender status is a sensitive datum. Data about previous sex will implicitly reveal transgender status, and is also highly sensitive to many transgender individuals on the basis that they reject the sex formerly attributed to them. Transgender individuals may wish to have official documents such as birth certificate of their children changed.

Facial images (non-processed): Although in most societies, it is the exception for people to cover their face in public, where face and/or head coverings are mandated in public, photos showing the face and/or hair might be considered sensitive. The publication of images of minors, even when in a public place, is considered sensitive in some societies.

Body images (non-nude): These are capable of being sensitive data in some circumstances. The publication of photographs of actor Gorden Kaye in hospital after a head injury by the Daily Sport newspaper in the UK attracted opprobrium in the 1990s [76]. Additionally, they can be used for identification purposes.

Body images (nude): Most people would consider nude body images sensitive. The criminalisation of so-called “revenge porn” in the UK, North America and elsewhere reflects

this sensitivity. The level of exposure that is classified as "nude" or sufficiently explicit will vary. Many people would consider that images of breastfeeding should be treated differently from erotic images of breasts.

Any traits processed for biometrics: There are several traits that can be processed for biometric applications, ranging from fingerprints to facial images to the vein pattern of the forearm or palm. There is a difference between biometrics being used for verification and identification [77]. For example, the use of thumbprints to provide security on Apple and other devices does not infringe privacy, as the user is the custodian of this data that is only used to confirm they are the authorised user (rather than to identify them as such). These data can be processed from ordinary photographs to "tag" people in photographs, which is problematic and has been the basis for regulatory action against Facebook [78].

Any traits processed for other reasons: Facial expressions and body language can be processed to identify emotions. Pupillary diameter may indicate the level of physiological arousal of an individual. CCTV systems may be used to identify suspicious behaviour in public places. At the current time, these abilities are unlikely to exceed the capacities of human observation, yet they introduce another element to electronic surveillance.

Sex: Sex can be a sensitive datum where there is potential for discrimination. Where the person is transgender, it can be extremely sensitive as it will indirectly reveal transgender status. The designation of fatherhood can be problematic for a transgender individual (see above).

Age: This can be sensitive, particularly where age will preclude an eligibility (or vice versa). For example, if reaching a certain age makes someone eligible for military service during an active conflict, this is a very sensitive datum.

Pregnancy: The potential for discrimination particularly in employment issues makes this a sensitive datum. This is also a sensitive datum where the pregnancy is due to a sexual relationship that is disapproved of.

Race: Although race is often a visible characteristic, recording of race is a sensitive issue because of the perceived potential for racial discrimination.

Ethnic Group: The membership of an ethnic group is not necessarily a visible characteristic. The perceived potential for discrimination makes this a sensitive datum.

Wellness and lifestyle data: This category covers a large range of data. Geolocation data may be used to calculate distance travelled. Jogging routes may indicate home or work addresses. Heart rates may signify medical issues. Variations on weight may have medical significance. The European Data Protection Supervisor (EDPS) issued Opinion 1/2015 on Mobile Health that states in Para 12 that:

> Lifestyle and well-being data will, in general, be considered health data, when they are processed in a medical context (e.g. the app is used upon advice of a patient's doctor) or where information regarding an individual's health may reasonably be inferred from the data (in itself, or combined with other information), especially when the purpose of the application is to monitor the health or well-being of the individual (whether in a medical context or otherwise) [79].

The Article 29 Working Party states in an annexe to a letter sent to the European Commission on Feb 5th 2015 that personal data on medical apps is health data where:

1. The data are inherently/clearly medical data

2. The data are raw sensor data that can be used in itself or in combination with other data to draw a conclusion about the actual health status or health risk of a person

3. Conclusions are drawn about a person's health status or health risk (irrespective of whether these conclusions are accurate or inaccurate, legitimate or illegitimate, or otherwise adequate or inadequate) [80].

There are no published definitions of lifestyle and wellness data. They can be distinguished on the basis that lifestyle data purely pertains to actions, whereas wellbeing data will also include some input from sensors that measure the effects of a particular lifestyle. Lifestyle and wellness apps can measure, log and/or aim to improve: distance walked/jogged and where; physical activity; calories consumed; sleep quantity and quality; levels of stress; and relaxation, amongst other things.

Genetic data: Currently there are several gene testing services which bypass medical regulation by simply looking at associations, rather than testing for specific disease-causing genes. Genetic data pose unique difficulties, as it is not only the individual that has an interest in genetic abnormalities but also other family members. Genetic diseases may reveal hitherto unknown parentage issues.

Diagnoses: Medical data are universally recognized as sensitive, even when the matter is relatively inconsequential (*Lindqvist*). The sensitivity will depend on a number of factors including severity, chronicity, terminal illness, data relating to children and minors [9].

Sensitive diagnoses: The definition of sensitive diagnoses is very dependent on socio-cultural factors. Often diagnoses are sensitive because there is a stigma associated with the disease or the cause e.g. mental health problems, reproductive health, addictions, alcohol or drug-related conditions such as cirrhosis, and sexually transmitted infections.

Any conditions or predispositions that might preclude getting certain treatments or insurance coverage could be classified as more sensitive.

References

[1] F. Dance, The concept of communication, Journal of Communication 20 (1970) 201-210.

[2] J. Wood, Communication Theories in Action: An Introduction, Wadsworth/Thomson Learning, Belmont, 2000.

[3] R.M. Losee, A Discipline Independent Definition of Information, Journal of the American Society for Information Science 48 (1997) 254.

[4] J. Feather, P.Sturges (eds), International Encyclopedia of Information and Library Science, Routledge, 2003.

[5] G.B. Davis, M.H. Olson, Management Information Systems: Conceptual Foundations, Structure, and Development, 2nd ed, McGraw-Hill, New York,1985.

[6] G.A. Silver, M.L. Silver, Systems Analysis and Design, Addison Wesley, Reading MSA, 1989.

[7] P.B. Checkland, J. Scholes, Soft Systems Methodology in Action, John Wiley & Sons, New York,1990.

[8] M. Janssen, Y Charalabidis, A. Zuiderwijk, Benefits, adoption barriers and myths of open data and open government, Information System Management, 29 (2012) 258-268.

[9] S.O. Dyke, E.S. Dove, B.M. Knoppers, Sharing health-related data: a privacy test? Nature Partner Journals Genomic Medicine 1(2016) 16024.

[10] B. Koops, B.C. Newell, T. Timan, I. Škorvánek, T. Chokrevski, M. Galič, A typology of privacy, University of Pennsylvania Journal of International Law, 38 (2016) 483.

[11] D.J. Solove, A taxonomy of privacy, University of Pennsylvania Law Review 154 (2006) 477-564.

[12] S.D. Warren, L.D. Brandeis, The right to privacy, Harvard Law Review Dec 15 (1890)193-220.

[13] W.L. Prosser, Privacy. Californian Law Review, 48 (1960) 383.

[14] A. Westin, Privacy and Freedom, Atheneum, New York,1970.

[15] I. Altman, The Environment and Social Behaviour: Privacy, Personal Space, Territory, and Crowding, Brooks/Cole Publishing, Monterey,1975.

[16] G. Laurie, J. Ainsworth, J. Cunningham, C. Dobbs, K.H. Jones, D. Kalra, N.C. Lea, N. Sethi, On moving targets and magic bullets: Can the UK lead the way with responsible data linkage for health research? International Journal of Medical Informatics, 84 (2015) 933-40

[17] S.B. Barnes, A privacy paradox: Social networking in the United States, First Monday 11(2006).

[18] B. Debatin, J.P. Lovejoy, A. Horn, B.N. Hughes, Facebook and online privacy: Attitudes, behaviors, and unintended consequences, Journal of Computer- Mediated Communication 15 (2009) 83-108.

[19] E. Zheleva, L. Getoor L, To join or not to join: the illusion of privacy in social networks with mixed public and private user profiles, Proceedings of the 18th International Conference on World Wide Web (2009) 531-540.

[20] P.G. Lange, Publicly private and privately public: Social networking on YouTube, Journal of Computer-mediated Communication 13(2007) 361-380.

[21] C.J Hoofnagle, J.M. Urban, Alan Westin's Privacy Homo Economicus, Wake Forest Law Review 49 (2014) 261.

[22] P.M. Schwartz, Privacy and democracy in cyberspace. Vanderbilt Law Review, 52 (1999)1607.

[23] L. Brandimarte, L. Acquisti, G. Loewenstein, Misplaced confidences: Privacy and the control paradox, Social Psychological and Personality Science 4(2013) 340-347.

[24] P.A. Norberg, D.R. Horne, D.A. Horne, The privacy paradox: Personal information disclosure intentions versus behaviors, J Consumer Affairs 41(2007)100-126.

[25] P. Carter, G.T. Laurie, M. Dixon-Woods, The social licence for research: why *care.data* ran into trouble, Journal of Medical Ethics 41 (2015) 404-409.

[26] G. Phillipson, Transforming breach of confidence? Towards a common law right of privacy under the Human Rights Act, The Modern Law Review 66 (2003) 726-58.

[27] M. Elliott, E. Mackey, K. O'Hara, C. Tudor C,The Anonymisation Decision-Making Framework, UK Anonymisation Network, 2016.

[28] O. Tene, J. Polonetsky, Big Data for All: Privacy and User Control in the Age of Analytics, Northwestern Journal of Technology and Intellectual Property 115 (2013) 239-73.

[29] T. Elliott, 7 definitions of Big Data you should know about. Available at: http://timoelliott.com/blog/2013/07/7-definitions-of-big-data-you-should-know-about.html, 2013 Accessed Jul 19th, 2017.

[30] D. Beer, How should we do the history of Big Data, Big Data & Society Jan-Jun (2016) 1-10.

[31]M. Komorowski, A history of storage cost (update). Available at: http://www.mkomo.com/cost-per-gigabyte-update, 2014. Accessed Jul 19th, 2017.

[32] A. McAfee, E. Brynjolfsson, T.H. Davenport. Big data: the management revolution. Harvard Business Review 90 (2012) 60-8.

[33] W.W. Lowrance, Privacy, Confidentiality, and Health Research, Cambridge University Press, Cambridge, 2012.

[34] M.A. Heller, R.S. Eisenberg, Can Patents Deter Innovation? The Anticommons in Biomedical Research, Science 280(1998)698.

[35] G. Palsson, P. Rabinow, The Icelandic genome debate, Trends in Biotechnology 19(2001)166

[36] Ministry of Transport and Communications (Finland), MyData – A Nordic Model for human-centered personal data management and processing. Available at https://julkaisut.valtioneuvosto.fi/handle/10024/78439, 2015. Accessed Oct 5th 2017.

[37] L.J. Kish, E.J. Topol, Unpatient - why patients should own their medical data, Nature Biotechnology 33(2015) 921-24.

[38] R. Kemp, P. Hinton, P. Garland, Legal Rights in Data, Computer Law & Security Review 27 (2011) 139-151

[39] C. Cuador, From Street Photography to Face Recognition: Distinguishing between the Right to Be Seen and the Right to Be Recognized, Nova Law Review 41 (2016) 237.

[40] J. Nash, Fashion statement: Designer Creates Line of Drone-proof Garments to Protect Privacy. Available at https://www.scientificamerican.com/article/drone-proof-anti-infrared-apparel/, 2013 Accessed Feb 21st, 2017.

[41] Less EMF, EMF Shielding Clothing. Available at: http://www.lessemf.com/personal.html . Accessed Feb 21st, 2017.

[42] Electronic Privacy Information Center, Investigations of Google Street View. Available at https://epic.org/privacy/streetview/, 2012. Accessed Apr 24th, 2017.

[43] Norfolk Constabulary, #OpRingtone - Police target drivers using phones. Available at: https://www.norfolk.police.uk/news/latest-news/opringtone-police-target-drivers-using-phones, 2017. Accessed Jul 25th, 2017.

[44] L. Golgowski L, How Target knows when its shoppers are pregnant - and figured out a teen was before her father did. Available at: http://www.dailymail.co.uk/news/article-2102859/How-Target-knows-shoppers-pregnant--figured-teen-father-did.html, 2012. Accessed Nov 26th, 2015.

[45] [42] H. Nissenbaum, Privacy as contextual integrity, Washington Law Review 79(2004) 119

[46] P. Fule, J. Roddick, Detecting Privacy and Ethical Sensitivity in Data Mining Results, Proceedings of the 27th Australasian Conference on Computer Science 26 (2004) 159-166

[47] L. Sweeney, Simple demographics often identify people uniquely, Health (San Francisco) 671 (2000) 1-34

[48] Article 29 Data Protection Working Party, Opinion 4/2007 on the concept of personal data. 01248/07/EN WP136. Brussels: Directorate C, European Commission (2007).

[49] P. Cowan, Treat anonymised data as personal information: Pilgrim. Available at http://www.itnews.com.au/news/treat-anonymised-data-as-personal-information-pilgrim-411949, 2015. Accessed May 17th 2016

[50] FS50565190 Queen Mary's University of London [2015] Information Commissioner's Office.

[51] Nuffield Council Working Party, The collection, linking and use of data in biomedical research and health care: ethical issues, Nuffield Council for Bioethics, London, 2015.

[52] J. Hart, Smart Meter Data at Crux of Arkansas Murder Case. Available at: http://stopsmartmeters.org/2016/08/26/smart-meter-data-at-crux-of-arkansas-murder-case/, 2016. Accessed Apr 24th, 2017.

[53] R. Abel, Privacy Issue? Pacemaker data used to charge suspected arsonist. Available at: https://www.scmagazine.com/police-use-pacemaker-date-to-charge-suspect/article/635665/, 2017. Accessed Aug 11th 2017.

[54] R.T. Titmuss, The Gift Relationship, George Allen & Unwin, London, 1970.

[55] T. Ridley-Siegert, Data privacy: What the consumer really thinks. Journal of Direct, Data and Digital Marketing Practice 17(2015) 30-5.

[56] J.L Spears, The effects of notice versus awareness: An empirical examination of an online consumer's privacy risk treatment, System Sciences (HICSS), 46th Hawaii International Conference (2013) 3229-3238).

[57] M. Aitken, SHIP Public Engagement: Summary of Focus Group Findings, Scottish Health Informatics Programme, Edinburgh, 2011.

[58] Ipsos MORI, The One-Way Mirror: Public attitudes to commercial access to health data, Wellcome Trust, London, 2015.

[59] S. Schwarcz, The "Principles" Paradox, European Business Organization Law Review 10 (2009)175-184

[60] J. Drechsler , S. Bender, s. Rässler. "Comparing Fully and Partially Synthetic Datasets for Statistical Disclosure Control in the German IAB Establishment Panel." Trans. Data Privacy 1 (2008) 105-130.

[61] BBC News, East Sussex NHS Trust apologies over data breach. Available at: http://www.bbc.co.uk/news/uk-england-sussex-33401806, 2015. Accessed Feb 16th, 2016.

[62] NHS Digital, Accredited Safe Haven Accreditation Process Stage 1. Available at: http://content.digital.nhs.uk/media/12203/Accredited-Safe-Haven-Accreditation-Process-Stage-1---June-2013/pdf/safe-haven-accred-proc-stage-1.pdf. 2016. Accessed Aug 11th 2017.

[63] P.M. Schwartz, Privacy and democracy in cyberspace, Vanderbilt Law Review 52 (1999)1607

[64] BBC News, Smith 'frozen rather than angry' about porn expenses. Available at http://www.bbc.co.uk/news/uk-12535038, 2011. Accessed on Aug 11th 2018

[65] Breyer v Germany C582/14, 2016.

[66] European Union, General Data Protection Regulation (GDPR) (Regulation (EU) 2016/679), Official Journal of the European Journal 2016 May 4th(L119).

[67] S. Higginbotham, Companies need to share how they use our data. Here are some ideas. Available at http://fortune.com/2015/07/06/consumer-data-privacy/, 2015. Accessed May 17th 2016.

[68] BBC News, Subculture abuse classed as hate crime. Available at http://www.bbc.co.uk/news/uk-england-leicestershire-34919722, 2015. Accessed on Aug 11th 2018

[69] N.K. Malhotra, S.S. Kim SS, J. Agarwal, Internet users' information privacy concerns (IUIPC): The construct, the scale, and a causal model, Information Systems Research 15(2004) 336-55.

[70] K.A. Stewart, A.H. Segars, An empirical examination of the concern for information privacy instrument, Information Systems Research, 13(2002) 36-49.

[71] K. Rushton, Samsung warns viewers: Our smart TVs could be snooping on your private conversations. Daily Mail 2015 Feb 9th.

[72] E. Ortiz, Prosecutors Get Warrant for Amazon Echo Data in Arkansas Murder Case. Available at: http://www.nbcnews.com/tech/internet/prosecutors-get-warrant-amazon-echo-data-arkansas-murder-case-n700776, 2016. Accessed Feb 21st, 2017.

[73] D.M. Douglas, Doxing: a conceptual analysis. Ethics and Information Technology 18(2016) 199-210.

[74] P. Golle, K. Partridge, On the anonymity of home/work location pairs, Pervasive Computing (2009) 390-7.

[75] Guardian, Admiral to price car insurance based on Facebook posts. Available at https://www.theguardian.com/technology/2016/nov/02/admiral-to-price-car-insurance-based-on-facebook-posts, 2016. Accessed Aug 11th 2017.

[76] Morning Star Online, Nasty side of the press shows its face in the wake of Grenfell tragedy. Available at https://www.morningstaronline.co.uk/a-e70d-The-nasty-side-of-the-press-shows-its-face-in-the-wake-of-the-Grenfell-tragedy#.WY2HtVF942w, 2017. Accessed Aug 11th 2017.

[77] E.J. Kindt, Privacy and data protection issues of biometric applications, Springer, 2016.

[78] Guardian, Facebook facial recognition software violates privacy laws, says Germany. Available at https://www.theguardian.com/technology/2011/aug/03/facebook-facial-recognition-privacy-germany, 2011. Accessed Aug 11th 2017.

[79] European Data Protection Supervisor, Opinion1/2015 on Mobile Health. Available at https://edps.europa.eu/sites/edp/files/publication/15-05-21_mhealth_en_0.pdf, 2015. Accessed Aug 11th 2017.

[80] Article 29 Working Party, Annex to letter to the Commission on health apps and devices. Available at http://ec.europa.eu/justice/data-protection/article-29/documentation/other-document/files/2015/20150205_letter_art29wp_ec_health_data_after_plenary_annex_en.pdf, 2015. Accessed Aug 11th 2017.

Cases

Breyer v Germany C-582/14 (2016)

Douglas v Hello! EWCA Civ 595 [2005]

Dow Chemical Co. v USA 476 U.S. 227, 238-39 (1986)

Florida v Riley 533 US 27, 40 (2001)

Hannover v Germany 40 EHRR 1 (2005)

Kyllo v. United States, 533 U.S. 27, 29 (2001)

Bodil Lindqvist v Åklagarkammaren i Jönköping ECR I 12971 (2003)

Nader v General Motors Corporation 25 N.Y.2d 560, 255 N.E.2d 765, 307 N.Y.S.2d 647 (1970)

Oxford v Moss 68 Cr App R 183 (1979)

Peck v UK 36 EHRR 41 (2003)

United States v Knott, 460 U.S. 276, 103 S. Ct. 1081, 75 L. Ed. 2d 55 (1983)

Your Response v Datateam Business Media Ltd EWCA Civ 281 [2014]