EU21 ***Recordkeeping, Open government Data and Privacy***

A literature review of the implications for information governance and recordkeeping of developments in data protection and privacy regulations in the UK and EU in the context of moves to extend open government data and access to public sector information in the context of access to medical records and patient data

Authors: Anna Sexton, Elizabeth Shepherd (UCL, DIS)

Introduction and Scope of Review

Medical records generated in the course of administering care to patients are fundamental to the functioning of healthcare systems around the globe. In an English context, since the introduction of the 'Lloyd George envelope' in 1911, information has been routinely recorded about patients in a relatively standardised form. However, it is true to say that 'in less than a generation' the recording of medical information on patients has gone from 'doctor's notes' to computer-based records that can capture data in increasingly standardised forms (Nuffield, 2014, p.6). Technological advances have underpinned this shift, and have also enabled new ways and means of sharing patient records and data. This increased ability to share is recognised as a strength, as it can provide an opportunity to better support the direct care of the patient, for example, through better information flows between care providers. In addition to supporting the clinical care of the individual patient, the data held in digital patient records can also be extracted and used for a wide range of additional purposes linked to improving health care services. These purposes include auditing of services, monitoring performance, resource planning, and evaluations on the cost effectiveness of care. This data can also be opened further for a variety of other extended re-uses including research. Such research, which often seeks to identify broad trends in health and disease, can have a collective benefit when it leads to improvements in public health.

At the same time there is an ever expanding number of opportunities for people to participate directly in health and medical research by providing data about their lifestyle, health, diseases, genetics and biology to projects that are supported by increasingly sophisticated information systems. Data can be generated from clinical trials and observations, or it can be generated directly by the patient (e.g. from life-logging) or it can be generated in the laboratory (e.g. from imaging or genome sequencing, and other 'omics'). Advances in both technology (faster information storage, retrieval, and processing) and data science (more powerful statistical techniques and algorithms) (Nuffield, 2014, p.4) have led to increasingly more sophisticated and powerful means available to researchers to 'collect, manage, combine, analyse and derive insight' from these different types of data (Nuffield, 2014, p.xvii).

The increased opportunities that shifts in technology and data science has wrought is indicated through the increase in scale in observational studies that have developed from the 1940s when 5,000 participants was seen as a large-scale project, to today, when the Million Women Study, has recruited over a million women to follow in order to learn about a range of health conditions including cancers,

osteoporosis, and cardiovascular disease. The speed of change can also be captured by the observation that 'within the span of an academic career' many researchers in the biomedical field 'have gone from using an edge-notched punch card system to digital data mining using cloud-based services several orders of magnitude more powerful' (Nuffield, 2014, p.6). These rapid advances have led to the development of the term 'big data' which is used as a catch-all by biomedical researchers to describe the opportunities and challenges in 'accessing, managing, analysing, and integrating datasets of diverse data types (e.g. imaging, phenotypic, molecular, health, behavioural) that are increasingly larger, more diverse, and more complex' (Nuffield: 2014, p.15).[1] Following Boyd and Crawford (2012, p.663), then, big data is a term that 'is less about data that is big than it is about increasing capacities to search, aggregate and cross-reference large data sets' (Moat et al, 2014, p.95).

There is clearly a public good in the use of medical records and patient data to support advances in medical, health and scientific research. These include advancements in the understandings of health and disease, improvements in health services, better detection and prevention of illness, and the development of innovative treatments. However, equal to this, there is also a public good in respecting and protecting privacy, maintaining confidentiality, and limiting the use of medical records and patient data (particularly when it is still identifiable). This becomes a particular issue in relation to the re-use of data generated routinely as part of clinical care, which is then opened up for a range of secondary purposes that may not have been anticipated or expected by the patient. Here, protecting patient data is not just in the individual's interest, but has a collective benefit as it enables the public to have the confidence necessary to participate in health services. Balancing these public goods (the public good in enabling research and the public good in protecting data) is increasingly challenging given the rapidly evolving mechanisms open to researchers and others to extract added value from data though linking and re-use. In these increasingly sophisticated technological environments these two public goods are entangled in complex ways.

Arguably, advocating for, and developing the means to exploit, the public good in sharing data lies at the heart of the open government data agenda. The broader concept of 'open government' is most usually positioned as intrinsically positive and is often linked to the related concepts of democracy, transparency and accountability (Davies and Bawa, 2012; Harrison et al, 2015). It is generally understood as a reaction to long-standing cultures of government secrecy, and, more recently, it has developed a rhetoric that places it as an antidote to the limited scope for citizen participation in policy making (Davies and Bawa, 2012, Yu and Robinson, 2012). The 'open government data' (OGD) agenda has often been positioned within this broader 'open government' narrative. As a form of proactive disclosure (Thurston, 2012), OGD is indicative of an open attitude, in which raw data collected by Government is made as transparent and as readily available as possible. However, other narratives on OGD focus on it as a technology, highlighting 'its role in facilitating new modes of production, the delivery of services and in supporting the role of competitive market forces in government services'

[1] See: US National Institutes for Health http://bd2k.nih.gov/about_bd2k.html#bigdata

(Davies and Bawa, 2012). This technological narrative fuses with one on economics, where OGD is positioned as a means of enabling data to be fully realised as an income generating 'asset'. In the UK Government's white paper on open data, aspects of all these narratives become fused: OGD is positioned as a means of enabling 'transparency to drive prosperity'. (2012, p.5) Given some of the threads within this narrative, a simplistic positioning of OGD as a 'public good' can unravel, particularly when OGD's role in supporting commercial interests emerges as a dominant underpinning force.

Whilst *sharing* data may lie at the heart of the OGD agenda, from a privacy perspective, the *limitation* of sharing personal information is not only seen as a positive, it is cast as a fundamental human right. However, its reach has to be viewed as 'non-absolute' and its application must be weighed against its functioning in society. A number of overlapping legal measures exist to protect privacy including privacy rights, which guarantee freedom from interference; rules of data protection, which controls the 'processing' of 'personal data'; and duties of confidentiality, which protect against unauthorised or unreasonable breaches of confidence (Nuffield, 2014, p.xix). Across these legal measures there is recognition that privacy is not always in the public interest if it is allowed to impede on other fundamental human rights and interests.

In the context of health, the tension between reaping the benefits of being open with records and data on the one hand, and maintaining data confidentiality on the other hand, is a tension that is sharply felt. In the context of NHS England, the most recent Information Governance review, by Dame Fiona Caldicott, summed this tension up perfectly in its title *To Share, Or Not to Share?* (2013) Indeed, that is one of the central dilemmas that this literature review seeks to explore.

The remit of this literature review has been to explore how the OGD and re-use of public sector information agendas balance against developments in data protection in the UK and the EU, with a focus on access to medical and health records, with an intention to draw out implications for record-keeping and information governance. As a way in and as a means of narrowing the scope, the situation of a medical and health researcher seeking to re-use routinely collected clinical care data (e.g. data taken from primary care and hospital records) has been used as the primary lens through which to to look at this landscape and open up the key issues.

Methodology

This literature review is based on an analysis of the applicable legislation and the relevant policy documents in the field of open government data, re-use of public sector information, data protection, and broader legislation affecting use of medical records and patient data. The analysis of the legislation is accompanied by a study of the literature around these themes. The focus has been predominately on England, but also seeks to look outward to legislation and policy in the devolved

nations of the United Kingdom (e.g. in the final section on Scottish approaches) as well as legislative developments in and across Europe.

Acknowledgements

With grateful thanks to Isabel Taylor for her thorough work in proofreading, editing and commenting on this text. Isabel has also shared numerous relevant references as well as some of her own unpublished notes which has helped to shape aspects of this text.

Open government data in the UK: What this means in relation to medical records and patient data

According to Bates (2014) the Open Government Data (OGD) agenda in the UK can be defined as:

> [A]n information policy which provides a particular framework for governing the *re-use* by third parties of datasets that are produced by public institutions. […] The proposal for Open Government Data argues that non-personal data that is produced by public bodies should be opened for all to re-use, free of charge, and without discrimination (p.390)

Proponents of the Open Government Data therefore argue that public services, including health services, generate large volumes of data that is simply not exploited to its fullest extent. When data is cast in these terms, as an asset ripe for exploitation, then the argument naturally follows that one way of increasing the possibility to exploit data to its fullest potential is through the proactive disclosure of open datasets that can then be used by any third party for any purpose. It has been suggested in the US, for example, that 'the benefits of opening up healthcare datasets 'could run into the hundreds of billions of dollars annually, realised through a combination of re-engineering health services and commercial exploitation' (Manyika et al. (2011) in Keen et al, 2013, p.229).

In reality there are a series of stumbling blocks to any realisation of exploiting patient data held in publicly maintained health systems through 'open data' initiatives. The first is related to the quality of the underlying records from which the data to be opened is drawn. As Anne Thurston (2012) points out, the quality of open data is only ever as reliable as the underlying records from which the data is extracted. Secondly, even if the underlying records prove to be authentic and reliable sources of data, there is an issue as to whether the data can maintain its authenticity and integrity during any process of extraction. Keen et al (2013) contend that datasets generated from public sector clinical care records are generally 'chronically incomplete and unquantifiably inaccurate'. While this might be 'good enough' for a number of purposes, it is certainly not 'good enough' when the data is intended to be opened to enable high-quality research (p.229). Thirdly, there is a fundamental problem with any assumption that the State has the right to decide the 'ifs, how, and when' in relation to opening up patient data held in public health systems. The assumption that they *do* have this right runs counter to strong trends in health care that stress the importance of patient's being able to control access to their data (Keen et al, 2013, p.229).

Clearly, to be published as wholly 'open data' without *any* restrictions, datasets must be anonymised (de-identified). The Article 29 Working Party (the European Advisory Body on Data Protection) describe anonymisation as a process that '"prevents all parties from singling out an individual in a dataset, from linking two records within a dataset (or between two separate datasets) and from inferring

any information in such a dataset."[2] It might be assumed on the surface that this then solves the problem of patient control: if individuals are no longer identifiable, then the data is surely no longer 'theirs'. This is an assumption that is, in fact, reinforced in data protection law, as data protection is only applicable to *identifiable* data, there are no data subject rights under data protection for data that has been de-identified. Yet the simple acceptance that anonymisation (de-identification) solves any issues that the patient might have with the creation of open datasets that are derived from patient records is problematic on a number of levels. Firstly, there is an inherent problem with assuming that data can ever be truly de-identified. It is simply not possible to make anonymity an intrinsic property within an individual level dataset. This is because identifiability is context dependent, in the sense that it depends on what other data is available to the individual seeking to make an identification. The outcome of this context-dependency is that the risks of being able to single an individual out from a dataset can never be categorically overcome. This fact will continue to ensure that some individuals will never be comfortable with their patient data being used in an open dataset even after robust de-identification processes and disclosure risk checks have been applied. The problem of anonymity is further compounded by the fact that controlling the levels of detail (the granularity) in the data is one relatively effective way to manage the risk of identification, but the irony here is that the less details are in the data, the less commercially valuable, or otherwise useful (i.e. to researchers), the dataset becomes. Keen et al (2013) suggest that it is 'not clear' if or how 'the circle of data protection and commercially valuable publication can be squared' (p.238). Secondly, following on from the points already made on anonymisation as a non-absolute, it is possible to suggest that anonymised data derived from patient records is always inextricably connected to the patient. As participants in the data, they still have a vested interest in its ongoing use. Even in the case of aggregated datasets where the focus is on group counts rather than person level data it is still plausible to argue that the patient's rights should extend to at least being notified on when data that has been derived from their records is published, and who has used it, for what purposes. This becomes a compelling argument when it is remembered that openness, transparency and accountability are supposedly at the *heart* of OGD initiatives. Proponents of the OGD agenda often frame open data as a direct benefit for the citizen. As taxpayers, citizens should have access to data relating to the services that they help to pay for, to re-use with as few restrictions as possible. Yet clearly, as the discussion above begins to tease out, positing open data as a 'citizen's right' can be in direct tension to the protection of the citizen's more fundamental rights and freedoms when the data being made open derives from personal records. For these reasons alone, applying the aspirations inherent in the 'open government agenda' becomes rather fraught.

Given the restrictions imposed by data protection, where data must become less nuanced in order to achieve anonymity, the basic utility of the OGD agenda becomes questionable. When time and money is spent pushing out vast quantities of supposedly re-useable information, the question of who is

[2] Article 29 Data Protection Working Party (2014) Opinion 05/2014 on anonymisation techniques, available at: http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf.

actually using this data, for what purposes, becomes an important one. There are examples of research from the Open Data Institute using open data released by the Health and Social Care Information Centre (HSCIC) that has been able to reveal disparities in GP prescribing patterns where GPs in some areas are prescribing high rates of patent statins instead of generic equivalents. By using this data, the research team were able to say that in 2011-2012, had every doctor prescribed the cheaper equivalents the cost would have been £200 million lower[3]. However, such examples of useful research outcomes from the use of open data appear to be relatively few and far between. As a general rule, quality and nuance in research are interlinked, and nuance requires the use of granular data, something that it is difficult to provide in an open data set without introducing an unacceptable risk of disclosure. Health researchers are therefore much more interested in negotiating with data providers for privileged and controlled access to either granular and/or linked data where the risk of disclosure are higher but this is counteracted by restricted access in a controlled environment. Researchers would therefore like to see a more proactive approach to the release of datasets for research use in safe haven environments and under other controlled mechanisms. Along these lines, in the UK the ESRC has funded the Administrative Data Research Network that aims to broker these access arrangements between data providers and researchers[4].

Data collection in NHS England

In the context of NHS England the extraction of data from patient records into aggregated data sets is not a new phenomenon, although the suggestion that these datasets should then be made publicly available in an open form is. The Department of Health (and its predecessors) have long had a role in dictating what data should be compiled in relation to the provision of health care services, to be reported up the chain. Keen et al (2013) provide a brief synopsis of the history of data collection in the NHS, which will be summarised here: In the early days of the NHS, most of this data was demanded from hospitals, with relatively little from primary care settings. In the 1980s, the data collection became much more systematic and covered all parts of the NHS. The 1990s brought fundamental changes through the introduction of New Public Management Policies. These were connected to creating an 'internal market' within the NHS by the separation of purchasers and providers of services as a means of decentralising decision making. These new arrangements required the generation of an increasing volume of data. It is important to understand that one of the NHS' longest running data sets, the Hospital Episodes Statistics (HES dataset), was initially created to aid the management of these new contractual arrangements across the NHS. Thus, it was created to aid decisions on payment, and therefore its coverage was directly related to fulfilling this aim. In the 2000s, there was another marked increase in the datasets NHS organisations were required to keep. Their nature and purpose varied greatly but included data on waiting times for hospital treatment, data from General Practitioners (GPs) to manage their performance in relation to national employment contracts, disease registers for diabetes

[3] See: Beggar thy neighbour; Open data and health care, *The Economist,* December 2012
[4] For details on ADRN see: https://adrn.ac.uk/

and stroke, and a system for reporting adverse events to the National Patient Safety Agency (pp.230-231).

Technology and Governance in NHS England

As summarised by Keen et al (2013), from the 1960s when mainframe computers were first introduced in the NHS, right up until the early 2000s, the IT infrastructure across the NHS was piecemeal with little linkage between separately developed systems. Often, data required centrally by the Department of Health had to be collated manually, taken from these separately developed systems. The National Programme for IT (NPfIT), launched by Labour in 2002, sought to change this fragmentation by creating a centralised IT system. The argument for doing so was strongly made; it would enable better sharing of information across the NHS, creating a more seamless service delivery and ultimately better care. The Programme, run by Connecting for Health (the predecessor body to the HSCIC), was ambitious in its intention to computerise and centralise every aspect of services and management across the NHS. Many aspects of the Programme, including the delivery of electronic health record systems, fell into disarray due to tangled disagreements with suppliers (pp.231-232). It was officially dismantled in 2013, with major objectives unachieved and amid public criticism of management failure and cost overrun. It has been described as 'the largest ever civilian IT project failure in human history'[5]. However, some parts of the envisaged centralised IT structure did come to fruition. Significantly, the programme was sufficiently developed to enable summary data from individual systems to be passed onto "Spine," now managed by the Health and Care Information Centre (HSCIC). This has allowed the creation of 'summary care records' which are held centrally and contain limited amounts of data on every individual living in England (e.g. demographic data, details about allergies, prescriptions etc) (Keen et al, 2013, p.231-232).

The push towards Open Data

As drawn out by Keen et al (2013), following the election of the coalition Government in 2010, in part as a reaction to the global financial downturn and the introduction of austerity measures, there has been a marked political focus on the potential to generate income from opening up public sector information. This included, in 2011, a £10 million pledge by the Government, to be delivered over five years, for the establishment of the Open Data Initiative. The Chancellor of the Exchequer gave this pledge in his Autumn 2011 statement to parliament in which he also further elaborated on the Government's underlying justifications for pushing forward with an OGD agenda:

[5] See: The NHS's national programme for information technology: a dossiers of concerns (2010), available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.304.9601&rep=rep1&type=pdf; House of Commons Public Accounts Committee, The national programme for IT in the NHS: progress since 2006, 2nd report of session 2008-09, HC 153, available at: http://www.publications.parliament.uk/pa/cm200809/cmselect/cmpubacc/153/153.pdf; Computer Weekly (22 September 2011) The world's biggest civilian IT project finally looks to have failed but is the NHS IT failure a surprise?, available at: http://www.computerweekly.com/blogs/outsourcing/2011/09/the-worlds-biggest-civilian-it-project-finally-looks-tohave-failed-but-it-is-no-surprise.html.

> Making more public sector information available will help catalyse new markets and innovative products and services as well as improving standards and transparency in public services. The Government will open up access to core public data sets on transport, weather and health, including giving individuals access to their online GP records by the end of this Parliament[6]

In a speech in December 2011, David Cameron stated:

> Now there's something else that we are doing…and that is opening up the vast amounts of data generated in our health service. From this month huge amounts of new data are going to be released online. This is the real world evidence scientists have been crying out for and we are determined to deliver it…We're going to consult on actually changing the NHS constitution so that the default setting is for patients' data to be used for research unless of course they want to opt out. Now let me be clear, this does not threaten privacy, it doesn't mean anyone can look at your health records but it does mean using anonymous data to make new medical breakthroughs and that is something that we should want to see happen right here in our country. Now the end result will be that every willing patient is a research patient; that every time you use the NHS you're playing a part in the fight against disease at home and around the world[7].

What is of interest here is how the market-orientated language employed by George Osborne in his November 2011 speech as a justification for pursuing an OGD agenda is replaced in David Cameron's December 2011 speech with a rhetoric in which the justification is based on increasing the opportunity for citizen participation. The invocation of the research-patient 'playing a part in the fight against disease' seeks to speak to notions of collective action and solidarity, with OGD positioned as a means of investing in the general good. Despite this glaze applied by Cameron, it is clear from George Osborne's earlier positioning that the UK's OGD agenda is as much connected to what Keen et al (2013) describe as "practical neo-liberalism" than it is to participatory citizenship. In "practical neo-liberalism" it is the relationship between the State and the private sector that takes precedence, and is ultimately reinforced. (p.241).

The Health and Social Care Information Centre (HSCIC)

As part of a major overhaul of the NHS, swept in through the Health and Social Care Act 2012, the Health and Social Care Information Centre was established as an arm's-length body to the Department of Health. As the national provider of information, data and IT systems for health and social care, its role is multifaceted. In relation to data, it has statutory powers to mandate disclosures from NHS service providers. The Act also requires the HSCIC to maintain and publish a register (catalogue) of

---

[6] George Osbourne's Autumn Statement to Parliament, Autumn 2011 (Paragraph 1.125) in Keen et al, 2013, p.232.

[7] Cameron, D. 2011. Speech on Life Sciences and Opening Up the NHS. December 6, 2011. http://bit.ly/s4hXEG in Keen et al (2013), p.232-233.

the data it has collected. HSCIC open data is available through a range of channels including the HSCIC's own website and through data.gov.uk, the UK government's portal for open data which contains nearly 25,000 data sets from all central government departments and a number of other public sector bodies and local authorities. In addition some of HSCIC's larger open data sets can be accessed through the UK Data Archive. The HSCIC currently also produces over 250 statistical publications each year where the data is provided in formatted Excel spreadsheets as well as machine-readable comma separated values (CSV) text files. Public-facing data on NHS organisations is also made available through NHS Choices, who provide downloads as well as a syndication service. In line with published commitments set out by the Government in 2011, the following datasets are also made available by the HSCIC under the terms of the Open Government Licence[8]:

- Prescribing by GP practice
- Complaints by NHS hospital
- Data from the National Lung Cancer Audit
- Data from the National Bowel Cancer Audit
- Data from the Oesophago-gastric Cancer Audit
- Data from the Diabetes Inpatient Audit
- Data from the Head and Neck Cancer Audit

In addition, it is possible to request access to a wider range of products and services through the HSCIC's 'Data Access Service,' for which charges are applied. For the purposes of this review, it is useful to distinguish this type of data access from fully open data initiatives, and to do this such data will be described as semi-open data here. This data is usually provided in pseudonymised[9] form. Each request for data, other than for anonymous data, is evaluated by the HSCIC's Data Access Advisory Group (DAAG)[10] in the first instance, and is subject to a data sharing contract which is made up of an agreement and a license. The most commonly accessed product is the Hospital Episodes Statistics (HES) described as 'over 1 billion records of patients attending Accident and Emergency units, admitted for treatment or attending outpatient clinics at NHS hospitals'[11]. The datasets can be provided in a variety of different ways including via a standard extracts service which provides cumulative data on a monthly basis (on an annual subscription) from any of the main datasets. Tailored record-level extracts from the main datasets is also possible on a one-off or regular basis. Datasets can also be linked, which can include linking two or more datasets held by HSCIC as well as linking data held by

[8] For a summary of HSCIC's approach to open data see: http://www.hscic.gov.uk/transparency

[9] Pseudonymous data is a category of data in-between identifiable and anonymous data. Pseudonymous data does not directly disclose a data subject's identity, but it may still identify an individual by way of association with additional information. Pseudonymisation usually takes the form of replacing identifiers with a code, which is then kept separate from the data. If two datasets are pseudonymised using the same codes by a trusted third party they can then be linked, making pseudonymisation a useful tool for researchers. Under these arrangements the researcher will not have access to the key to enable re-identification.

[10] For information on HSCIC's Data Access Request Service see: http://www.hscic.gov.uk/DARS

[11] For definitions of HSCIC's data products see: http://www.hscic.gov.uk/media/19239/DARS-products-and-services/pdf/DARS_Products_and_services.pdf

the customer (subject to relevant approvals). A particular secure service is provided for HES data known as the HES Interrogation System which 'allows registered users to securely access HES and use it in conjunction with a wide range of analytical functions, including the ability to access, manipulate, interrogate and then report on various full-year datasets including Admitted Patient Care, Outpatients and Accident and Emergency'[12]. The HES Interrogation system is provided through a secure portal and allows analysis to be performed on HSCIC servers without any local network load. There is also a patient tracking and status service where HSCIC provides the demographic status of a specific group of (usually de-identified) patients with the ability to track them over a period of time, providing regular updates including, where appropriate, cause of death.

Whilst this may sound as if data is easily and widely available through the HSCIC to enable its exploitation by a wide range of interested parties, a look at the literature emanating from the health research community highlights major problems with the control mechanisms that are in place. Far from realising a government vision of data as a raw material that with adequate safeguards can be made readily available to enable discovery, innovation and improvement, the research community point to how the information governance structure employed by HSCIC continually undermines and stymies the conduct of legitimate research. For example, Jonathan Filippon, a PhD researcher at Queen Mary University, describes in the *British Medical Journal* how time-consuming and expensive the HSCIC access system is:

> My objective was to compare treatment rates in Brazil with those in England and Scotland. Two nations with long established public healthcare systems. Despite the large economic and social differences between Latin America and the United Kingdom my goal was to improve the Brazilian public health system, with a particular focus on routine data collection systems and methods of disseminating information to inform policy makers. In May 2013 I applied to the Information Services Division of the NHS in Scotland for some anonymised data… the application procedure was straightforward, as was completing the mandatory online course, and I received the patient data extract within four months. Data from Brazil were even simpler to obtain: everything was available on the Ministry of Health's website. I analysed anonymised patient data with information on all hospital admissions to the public system since 2009… In the second year of my studies, in May 2014, we asked for some NHS England data, which, since the Health and Social Care Act 2012, are controlled by the Health and Social Care Information Centre (HSCIC)… The first surprise was that information about public healthcare in England had to be paid for, even though the NHS is publicly funded. My supervisor and I got an internal grant from the university, without which I could not have afforded the £1889…. The second surprise was how long it took. By August 2014 I had still not received any data. There was a long queue for applications to be processed: nothing happened for more than two months. Then the committee questioned the study's aims and

[12] Definition given in http://www.hscic.gov.uk/media/19239/DARS-products-and-services/pdf/DARS_Products_and_services.pdf

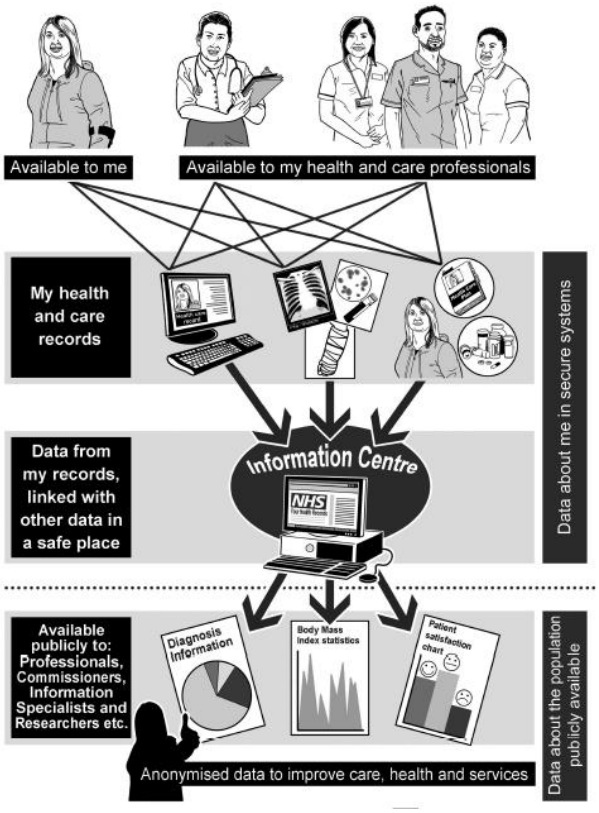
> objectives, on the grounds that a cross-country comparison would not benefit patients in England, and it refused to provide the data…We finally got approval after a long struggle that involved my supervisor having to write detailed notes. Then, however, the HSCIC would not supply the data extract we asked for because it was "under pressure not to release big extracts of data," even though the data were anonymised as they had been in Scotland and Brazil. We had to compromise and accept only tabulated data extracts instead of raw data, reducing the research possibilities; but the data asset owner—the person ethically responsible for keeping data under the HSCIC regulatory framework—still would not agree to release the data. Once again, the committee questioned the benefits of the study. Different HSCIC departments did not agree about the potential benefits of my research, and finally we were asked to liaise with all of the departments to ask them to come to an agreement internally. In January 2015 HSCIC finally approved the application with minor amendments, but our struggles were not over: HSCIC requires that all university computer servers comply with its "governance toolkit," a set of policy requirements. My university took two months to conform, but it still has no internal protocol for this kind of contract, and—14 months after the request and in the final year of my PhD—I am still waiting for the data to be released… So far this has taken more than 90 emails and more than 30 telephone calls, and I am still trying to resolve the issue despite my tight PhD schedule. I have had to ask for a one year extension from the Brazilian funders, at great cost to me and to them. I have heard similar stories throughout England, such as grants from the National Institute for Health Research being stopped because researchers cannot deliver the outcomes required (Filippon, 2015, p.351).

It appears from researcher accounts, such as the one above, that the information governance structure employed by the HSCIC in relation to data access requests for data that is intended to be semi-open is fundamentally flawed. Whilst questioning intentions against a 'public interest test', managing disclosure risk, and ensuring that the requester has adequate security systems in place to manage the data are entirely appropriate and necessary to ensure a balance between the public good of protecting data and the public good of making it available for legitimate purposes: an internal process that involves multiple committees, internal disagreements and long delays serves neither of these purposes well.

The NHS IT strategy of 2012 (*The Power of Information*) heralded the changes that have occurred with the establishment of HSCIC as a 'step-change in the health and care sector's approach to transparency, growth and open data' (Department of Health, 2012, p.13). The strategy went on to talk in ambitious terms about proposals for the 'release of big data,'the 'capture and release of My Data: provision of access to service users to their own identifiable data' and 'the creation of dynamic Information markets,' much of which is yet to be realised. It focused heavily on the positives of data integration in supporting direct care, and the importance of patients increasingly being able to access their own digital records as a means of ensuring that care operates on the basis of 'collaborative decision making'. Yet along with these aims the report also set out the goal of 'ensuring that data from personal

records is combined and linked together with other data in a secure environment, then made anonymous and open to be used', with the ultimate purpose that 'information extracted directly from combined records can, in the future, replace cumbersome data collections' (p.7). This is posited in the report as a positive step forward, since this data will be able to be used to better support 'quality audits, improve services and guide commissioning' as well as 'enabling the research and life sciences community to have access to a greater wealth of information to help drive improvements in health and care' (p.8). The model for the planned 'ideal' information flow given in the *Power of Information* is shown in Diagram One below. The mention of commercial access to this data, and the fact that requests can be made to access data that is not fully anonymised (if certain conditions are met) are conspicuous by their absence both in Diagram 1 and in the report more generally. In reality, despite ambitious plans to press ahead, the road to 'open data' across the health and care sector (undertaken largely under the auspices of the HSCIC) has been marred by controversy, as this review will unpack in due course.

Diagram One: Intended Information Flow for Health Data in England © Department of Health, 2012



The success of programmes to integrate data in the context of NHS England, and associated plans to open up this data, either under open government licence or through restricted access channels to third parties, ultimately turns on the question of public trust. The state, commercial firms, medical and health researchers, and health care professionals all have a vested interest in the project's progression, yet without gaining the social legitimacy that it needs from the citizen/patient it cannot move forward unhindered. Simply put, this public support has not been forthcoming, and planned data initiatives, such as the ill-fated care.data scheme, have had to be put on hold.

Re-use of Public Sector Information in the EU and the UK

Within the EU, the key piece of legislation on the re-use of public sector information is still the European Directive 2013/37/EU. This has been transposed, in the English context, into the Re-use of Public Sector Information Regulations 2015 (SI 2015 No. 1415). The National Archives have been responsible for this transposition and are the principal body offering guidance on its implementation. This guidance (which covers all aspects of implementation including best practice on standard licences, datasets and charging for re-use) is in line with the equivalent European Commission publication (European Commission, 2014). As drawn out by Janssen (2011), 'documents holding personal data are not as such excluded from re-use under the PSI directive, but such re-use has to be in line with the EU and national rules on the processing of personal data' (p. 446). This means that data may be derived from medical records or from patient interactions with services and made available for re-use as long as disclosure risk is effectively safeguarded through robust anonymisation.

The United Kingdom was one of only four countries in Europe that transposed the original 2003 European PSI Directive before the deadline of 1 July 2005, alongside Denmark, France and Northern Ireland. This was in stark contrast to Austria, Belgium, Luxembourg, and Spain, who were convicted by the European Court of Justice for non-transposition of the PSI directive, and Sweden, Poland and Italy, against whom the European Commission launched proceedings due to poor transposition of the Directive into national law (Janssen, 2011, p. 449).

The United Kingdom met the deadline due to its proactive approach, not only to the Directive's transposition, but also to its implementation. The government proactively established the Office of Public Sector Information (OPSI) in 2005 to lead on the transposition of the directive and advise on regulatory issues and operational procedures surrounding re-use. Following transposition, OPSI took a leading role in the provision of guidance; it managed the Information Fair Trader Scheme that endorses good practice in information trading by public sector data providers; and it investigated complaints under the PSI regulations 2005. It also provided an on-line licensing system and was responsible for the Government's own Information Asset Register. An Advisory Panel on Public Sector Information (APPSI) comprising representatives of public sector data providers as well as consumers of PSI from both academia and industry supported OPSI in the fulfilment of its role. OPSI came to an end in 2010 when its functions were absorbed and taken forward by the National Archives. Alongside the proactive role being taken by OPSI, in the summer of 2009 the British Prime Minister at the time, Gordon Brown, appointed Sir Tim Berners-Lee and Professor Nigel Shadbolt as advisors to the Cabinet Office for opening up public sector data, ensuring that open data remained firmly at the forefront of the agenda (Janssen, 2011, p.451). In January 2010, the web portal data.gov.uk was launched to provide a single access point to open data varying from information about house prices, school locations, and tax receipts, to commuting statistics and public transport routes and timetables (Janssen, 2011, p. 451).

Information from health and social care is included via the HSCIC and includes information on GP prescribing and statistics from cancer audits (see page 9 for the full description). Nevertheless, the question as to how demand-driven data.gov.uk really is, and whether it does in fact have a considerable contribution to make to innovation, are still open questions. After the elections in May 2010 the new coalition kept the momentum around open data and re-use going by establishing the Public Transparency Board to oversee the implementation of the new government's transparency commitments. These commitments include the publication of contracts and spending of central and local government, and other key government data such as crime data, as well as names and pay rates of senior civil servants (Janssen, 2011, p. 451). The UK push towards open data and re-use has finally been cemented by the introduction of a 'right to data' by the Protection of Freedoms Bill, amending the Freedom of Information Act, 2000, to include an obligation for public bodies to publish datasets available for re-use in a re-usable format either in response to a request or through their publication schemes (Janssen, 2011, p. 451).

As explained in the TNA guidance (2015a), the UK PSI Regulations 2005 are designed to:

- encourage the proactive publication of information that is easy to re-use
- enforce mandatory re-use permission for all information produced, held or disseminated within the course of a public task unless re-use is otherwise restricted or excluded (with some exceptions for the cultural sector)
- ensure the easy identification of public sector information that is available for re-use
- ensure the transparency of terms, conditions and licences
- make the availability of public sector information for re-use at marginal cost the default
- ensure clarity of any charges to be made for re-use (with explanation of the basis of the charge)
- ensure the use of standard licences that are as non-restrictive as possible, including through the Open Government Licence (OGL) for information for which no charge is made
- make the processing of requests for re-use timely, open and transparent
- ensure an accessible complaints process is in place with escalation to an authoritative body to make binding decisions
- ensure the retention of protection over personal data

The key changes from the earlier 2005 regulations around re-use are that:

- the cultural sector is now in scope (with the exclusion of performing arts): this affects libraries (including university libraries), museums and archives
- there is an obligation (rather than an encouragement) to allow re-use of most public sector

information. This is optional for cultural bodies unless the information is already available for re-use. When the cultural sector determines whether information is available for re-use and therefore carries this obligation, it must be remembered that non-published material can be considered available when it is used internally in a standardised form

- marginal cost pricing is the default (with certain exceptions)
- standard licensing is required; licences should be as non-restrictive as possible
- If a complaint cannot be resolved by a public sector body's internal system, it may be escalated to the Information Commissioner's Office —which can make binding decisions on most issues— and potentially to the First-Tier Tribunal for Information Rights

In relation to the forms of information that are in scope of the regulation, the guidance specifies:

> Any information (content) whatever its medium (form) – including print, digital or electronic, and sound recordings – produced, held or disseminated by a public sector body is considered public sector information. This includes an enormous range: corporate information such as reports and financial data, codes of practices, public records, statistics, still and moving images, press releases, publication schemes, and so on.
>
> If a public sector body holds the copyright for information it produces, holds or disseminates within its public task, then that information is in scope of the 2015 Regulations (TNA, 2015a, p.7).

The key activities recommended by the TNA (2015b) to ensure compliance are:

- Identify the benefits of re-use and transparency, the scope of the public task undertaken by the body, and the scope of public sector information available for re-use, and details of how the public body demonstrates compliance
- Publicise the definition of public task (i.e. whether it is statutory or by common administrative purpose)
- Create a statement of re-use to help users know what is available and what conditions apply (this should be aligned to the public task)
- Create (or update) a third-party intellectual property register including contact details of third-party rights holders where known
- Plan to make information and metadata open and machine readable wherever possible
- Offer open licences where possible, and make other licences as unrestrictive as possible. The Open Government License is preferred
- Offer information at marginal cost, and for free where possible
- Create or update internal complaints procedures

- Create a re-use team so that aspects of re-use can be managed effectively
- Ensure compliance with the Local Government Transparency Code (where applicable), the Data Protection Act and other related legislation
- Determine how compliance can be demonstrated

The intention behind the EU Directive, which is carried through into the Regulations, is to remove as may barriers to re-use as possible. As drawn out by Janssen (2011) in its 2010 Digital Agenda, the European Commission 'emphasised the importance of the availability of public sector data for stimulating markets for online content' (Janssen 2011, p.446). In its *Introductory Guide to the Amended PSI Directive*, the TNA also connects the ethos of the Directive to the delivery of economic benefits and employment opportunities across Europe through the development of a 'thriving information industry'. Janssen (2011) describes the PSI directive 'as a direct result of the European Commission's concern about the underdevelopment of the European information market and its inability to compete with the United States', a concern fuelled by the Commission's view that federal level data was widely available across the US at low cost (p.447). The purpose of the Directive was therefore two fold: 'on the one hand, enabling the availability of public sector data to third parties at low prices and unrestrictive conditions, and on the other hand, ensuring a level playing field between public bodies that operate in the information market in competition with the private information industry' (Janssen, 2011, p.447). In relation to exploitation of commercial value, the push towards openness at no (or marginal) cost effectively prevents the public sector *itself* from profiting form the information industry that it feeds. The promotion of openness necessarily entails a loss of control, and the negative implications are therefore felt by those who profited from the control mechanisms that were originally in place. The PSI Directive attempts to 'level the playing field' are actually designed to ensure that the private sector is not at a disadvantage to the public sector. As 'public sector information is often a monopoly asset for the public sector body that produces it, maintaining a level playing field among the different private sector users on the one hand, and between the public sector data provider and its customers on the other hand can be a challenge' (Janssen, 2011, p.448). Janssen (2011) summarises how the directive is designed to prevent public sector bodies from locking 'their data in exclusive deals with one private company or to maximise short term revenues by abusing their market power as monopolists' (p.448). The directive is also designed to mitigate any 'risk that public sector bodies fund (part of) their market activities with public tax money in order to keep their market prices low, and in this way use cross-subsidisation to distort the market (p. 448). Therefore, 'the directive seeks to mandate that any conditions for the re-use of public sector information have to be non-discriminatory for comparable categories of re-use' (p. 448). As drawn out by Janssen (2011), the PSI Directive also addresses the commercial use of public sector data by public sector bodies themselves:

> If these bodies 'provide added-value information products or services on the market based on their own data, they have to apply the same conditions and charges to their own re-use as to

> their private sector competitors. This may require the public sector body in question to have separate accounts or even separate its commercial activities from its public task activities in a different legal entity. Under the directive, public sector bodies are no longer allowed to conclude exclusive agreements. Re-use of public sector information has to be open to all potential actors in the market. An exception is made when exclusive rights are necessary for the provision of a service in the public interest (p. 448).

This reinforces the point that 'openness' is not wholly 'good' for all sections of society, all of the time. It is argued by those who see OGD as a form of neo-liberalism (Bates, 2012) that the private sector stands to benefit over and above both the citizen and the public sector. In regards to the citizen, commercial exploitation raises strongly felt privacy and security concerns. Can the private sector be trusted to act in the citizen's best interest? Rumours that the insurance industry may have used data released by the HSCIC's predecessor body to fix the costs of insurance provides plenty of fodder for the notion that the citizen and the private sector often have *competing* interests, with OGD placing the balance most firmly in the hands of the latter[13].

[13] For information on Parliamentary Health Committee enquiries into handling of NHS Patient Data including alleged misuse by the insurance industry see: http://www.parliament.uk/business/committees/committees-a-z/commons-select/health-committee/inquiries/parliament-2010/cdd-2014/

Data Protection in the UK and Europe

The development of data protection legislation across Europe is a complex picture involving the interplay of national approaches and legal systems, with tangential developments at a European level. Out of this complexity, several writers have sought to categorise the history of data protection legislation into distinct waves (Jori, 2007a; Mayor-Schönberger, 1997). The selective account provided here, which focuses on German, Swedish, French and UK contexts, will draw on the 'three-waves' framework provided by Jori (2007a). 'First-wave' data protection begins with the earliest instances in the 1970s, where the initial focus was on controlling and regulating the use of technology. In this phase it is possible to trace the emergence of the assertion of data subject rights, and thus the beginnings of a 'rights-based' focus for the concept of data protection. 'Second-wave' data protection law can be characterised by the further embedding and development of a 'rights-based' approach, commonly attributed to a landmark German Constitutional Court ruling in 1983 that cemented the concept of 'informational self-determination' as a doctrinal cornerstone in understandings of data protection. 'Third-wave' data protection law, in the wake of the 1995 European Directive, marks the continued adherence to a rights-based approach alongside an acceptance of the need for specific sectorial regulation and codes of conduct. A natural extension to Jori's framework would be to see the introduction of the General Data Protection Regulation (GDPR) at the European level as a 'fourth wave'. A close examination of this 'fourth-wave', and particularly its implications for medical and health data, is a major focus of this review. To contextualise this within the broader trajectory of developments across Europe, a brief account of the earlier legislative waves is provided here.

Historical background to data protection in Europe

First-wave data protection

In the context of the EU, the earliest developments in data protection law can be traced to the 1970s. These first generation laws are generally characterised in the literature as being a direct response to the technological advancements that had occurred throughout the 1970s, particularly in relation to the development of large-scale computerised systems that were enabling centralised and automated state control over information on the individual citizen (Jori, 2007, Mayer–Schönberger 1997). The very first data protection law, not just in a European context but anywhere in the world, was passed by the German Province of Hesse in 1970, with Sweden leading the way in passing the first *national* law in 1973.

In keeping with the general trend identified across first-generation data protection laws, the Hesse Act was directly connected to disputes around the procurement, distribution and use of large-scale computer systems in the performance of bureaucratic and administrative functions. It arose from the

context of a multi-faceted power conflict between the local community and the state administration; the legislature and the executive; and the citizen and the state. The thrust of the legislation was to regulate the use of the emerging large-scale computer systems to settle the issues around who could control the systems, and decide what programs to run on them. However, within the dispute, concerns around the confidentiality of the personal information being stored in the contested large-scale computerised record-keeping systems were actively voiced. Burkert (1999) argues that it was the addressing of the confidentiality of citizen data in the resulting Hesse Act that 'set a legislative program into motion' in which a right to confidentiality by default in relation to personal data was first enshrined (p.45).

The confidentiality clauses contained in the Hesse Act were not particularly substantial or robust. Confidentiality was established as a default, but the exemptions to the rule included 'if processing is necessary', rendering confidentiality by default difficult to establish in any circumstance. Leaving the limitations aside, the foundational status of the Act in relation to future data protection legislation in Europe is indicated by the fact that it was the first time that the term "Datenschutzgesetz" (literally translated as "Data Protection Act") had been used in law. Therefore, it was the instigator of the misnomer that has persisted today where 'data protection' is about the protection of the rights of the data subject rather than the protection of the data itself (Burkert, 1999, p.46). Arguably of greater significance are the Act's fundamental presumptions, which have become foundational for European data protection law: the negative default rule, where the processing of personal data is recognised as an interference with or a compromise of confidentiality that requires legitimisation; the acceptance that the data subject has rights and that access by the data subject needs no justification; and the requirement to establish a privacy protection institution and take an omnibus approach to regulating the application of data protection are all principles embodied in the Hesse Act which can be traced forward through subsequent European legislative developments (Burkert,1999, p.46).

The Swedish national data protection law of 1973 must also be considered in relation to its influence over future European legislation. Sweden was the first of the Scandinavian countries to introduce legislation. Denmark and Norway followed suit by introducing laws that came into force in 1978 and 1979 respectively, with Finland as a relative latecomer in 1988 (Burkert, 1999: 48). Swedish concerns over automated data and information management processes were somewhat unique in that they revolved around the Swedish national personal identification number system. This system had already been well established as an administrative tool, in operation since the 1940s. In the computerised landscape that was emerging in the 1960s, it was beginning to be realised that the personal identifier had the potential to have a vastly expanded role in pulling together administrative data in automated contexts and could serve 'as an integrator' for citizen information still held in 'largely decentralized files' (Burkert,1999, p.48). Since Sweden was also 'the most computerized country in the world (on a pro capita basis)' (Burkert,1999, p.48), the issue of how to introduce some controls around the potential integration of computerised citizen data gained a particular level of urgency. Data security issues connected to integration were actually at the forefront of Swedish data protection discussions, as it was felt that the centralisation of registers would leave Sweden open to data breaches by foreign powers,

with the potential to compromise Swedish neutrality. Against this backdrop of Swedish concerns, the resulting law introduced a concept that has been taken up subsequently in various national data protection regimes. Although not prevalent or favoured in a German context at a Federal level, the Swedish legislation enshrined the principle of a central register for personal data processing, established to serve the dual function of being open to public scrutiny, as well as operating as an enforcement tool for data protection agencies (Burkert,1999, p.48).

Data protection laws emerging in the latter half of the 1970s are characterised by a move away from the regulation and control of *technology* towards an increasing focus on establishing data subject *rights*. Of note here is the French context, where the establishment of data protection law was specifically fashioned against the backdrop of a public outcry over revelations of several secretive Government projects to pool citizen information in large-scale databanks. One of these databanks, GAMIN (Gestion automatiséee de la medicine infantile), was developed to process and track personal information on children, from birth through schooling, with the intention of identifying and following 'problematic' citizens (Burkert 1999:50). Post World War 2, the symbolic context of the databank was dramatically emphasised in the French press and, in the wake of this scandal, the French Data Protection Bill was developed in 1978, coming into force in 1980 (Burkert 1999: 50).. Drawing on the example set by Sweden, the French law also enshrined the role of registration into the administration of data protection to be overseen by the Commission Nationale de l'Information et des Libertés (CNIL). The CNIL was also given a unique advisory role in relation to the design of data systems. The French law is notable as a pro-active response to citizen concerns over state surveillance techniques, branded in the French media as mechanisms for 'chasing French citizens' (Burkert, 1999, p.55). The French bill is also notable for the impact it had on the development of data protection law in other French-speaking and Latin countries, notably Luxembourg, Belgium, Spain and Portugal (Burkert 1999, p.50).

Second-wave data protection

In a German context, the introduction of the Federal Data Protection Law in 1976, covering both the public and the private sector, was also undertaken against a backdrop of concerns over government plans to introduce a personal identifier as a coordinating mechanism for pooling citizen data. Earlier attempts to draft a Federal law that would regulate both the public and the private sector had stalled due to heavy criticism from the private sector. However, efforts were renewed in the light of moves by the Federal Government to introduce a personal identifier to co-ordinate the States' population registers. The Legal Committee of the Federal Parliament refused to pass the legislation enabling the introduction of the identification number without Federal data protection legislation in place. The resulting Federal Law introduced the concept of "informational self-determination" as a means of underpinning data subject rights (Burkert, 1999, p.49). In 1983, six years after the Federal Law was passed, a landmark Constitutional Court ruling relating to the Census Act passed earlier in the same year, cemented acceptance of the new Federal legislation. The Court's decision, which halted 300 million deutschmarks of administrative investment by ruling that that some provisions in the Census

Act were unconstitutional, had a profound impact (Burkert, 1999: 54). Jori (2007b) draws on Mayer-Schönberger to suggest that the decision's philosophy can be traced in the 1986 Amendment to the German Federal Data Protection Act, as well as in the Norwegian, Finnish and Dutch Acts, and the 1992 Hungarian Act (which was based on a decision of the Hungarian Constitutional Court, phrased in the wake of the German decision). Jori (2007b) also quotes Mayer-Schönberger to suggest that the shift between first and second-generation data protection can be summarised as an understanding that automation, with the introduction of the PC and networked computing, had permeated society to such an extent that regulating the spread of technological systems for data processing was no longer a feasible approach. Instead, principles for processing must be laid down, extending from the establishment of data subject rights. The thrust of the German approach to data protection therefore lay in developing a doctrine around the notion of 'informational self-determination': that as a social being an individual cannot avoid being the subject of data processing, but has the right to know and make decisions on that processing. Limitations on that right can only be accepted in cases of compelling public interest, and where clarity and proportionality is maintained. The German Constitutional Court ruling therefore established a series of data protection principles (purpose specification and limitation, correctness, timeliness); derived rights (access, rectification); derived controller obligations (fair and lawful obtaining of data, data minimisation, destruction when purpose fulfilled); and independent safeguards (overseeing institution) which can be seen echoed in the 1995 European Directive (Hornung and Schnabel 2009, p.87, Burkert, 1999, p.54).

Within this wave of second-generation data protection both the Netherlands and the United Kingdom are present as cautious adopters of data protection as law (Burkert 1999: 50). After a lengthy deliberation, the Netherlands national law finally came into force in 1990. The thrust of the approach was arguably more closely aligned to US traditions in its emphasis on the role of sector and self-regulation as a means of enforcement. Significantly, the Netherlands introduced "codes of conduct" as part of the regulatory regime. In the context of the UK, despite having passed a private members' bill on data surveillance as early as 1969, it wasn't until the 'Swedish incident', when Sweden threatened to restrict the export of personal data to the UK on the basis of inadequate data protection, that drafting legislation was seen as necessary. This necessity was compounded by the view taken at the time by the Council of Europe, who were moving towards legitimising the restriction of data in Europe on the basis of inadequate protection (Burkert 1999, p.51). The motivation behind the UK law, which was introduced in 1984, was therefore economic, and the thrust of the UK approach is frequently described as being 'trade-orientated' (Burkert 1999, p.51). Thatcherism dictated that the regulation of data protection should be a self-financing initiative, and accordingly in the UK context the need for registration, enshrined in the Swedish and French approaches to data protection, was developed into a mechanism for collecting fees that would support the role of the Information Commissioner in monitoring and enforcing good practice. The collection of fees on registration provided economic support for the scheme, and arguably also transformed the weight of the legislation into a more effective instrument (Burkert 1999: 55).

Third-wave data protection

With the introduction of piecemeal national legislation across Europe, it was inevitable that difficulties in the trans-border flow of information would lead to work on harmonisation. Third generation data protection in Europe was characterised by the introduction of the European Directive and national attempts to reconcile existing laws with the Europe-wide approach.

Early attempts to consolidate data protection through international and transnational guidelines came to fruition in the early 1980s through the respective work of the Council of Europe and the Organization for Economic Cooperation and Development (OECD). The latter of which drew an International membership including the USA. The international membership of the latter included the USA. Arguably, the European Council's text held greater influence, since, *if* ratified by members states, it would naturally lead to the adaptation of laws. The OECD instrument, on the other hand, was purely a statement of principles (Burkert, 1999, p.52), but was nevertheless important as the 'common denominator' between Europe and the USA (Jori, 2007a).

The Council of Europe passed two resolutions relating to privacy in 1973 and began working on a special convention in relation to automated processing. The text of the European Council Convention for the Protection of Individuals with regard to Automated Processing of Personal Data was finalised in 1981. It was initially hoped that the convention would be sufficient in solving the issues of trans-border information flow. A year after the publication of the convention, the European Commission issued a recommendation to member states to adopt the convention. However, increasing threats of interrupted data flows between Sweden and the UK, and in a later case, between France and Italy, highlighted the inadequacy of the Convention in solving the issue, and this prompted the Commission to act on European Parliament advice and begin the process of developing a Directive. Not all members were in support of this development. Great Britain, for example, was explicitly against union-level data protection legislation (Burkert, 1999, p.52).

In 1995, in recognition of both the differences in the levels of protection in each of its Member States, and the inadequacy of the law, the European Union passed a Europe-wide directive (95/46/E) to provide its citizens with a wider range of protections over abuses of their data. The directive on the *Protection of individuals with regard to the processing of personal data and on the free movement of such data* did not have regulatory power in itself, but was intended to set a clear benchmark for national law. Under the directive, each EU State was obligated to pass national legislation in order to implement the directive by October 1998. Member States such as Italy and Greece, who had taken a 'wait and see' approach to data protection, were in the somewhat fortunate position of being able to shape their national legislation directly around the Directive. For most other members, new national law was required, and in the UK context, a new Data Protection Act was introduced in 1998, in direct response to the EU Directive.

Jori (2007a) suggests that the European Directive's international impact was far-reaching, with influence over the development of legislation in South America as well as New Zealand, Hong Kong,

and Canada. Thus, a certain level of convergence in data protection approaches and legislation can be attributed to its arrival. However, a significant exception to this harmonisation remained in the form of the USA.

The UK Data Protection Act 1998: Implications for processing medical records and patient data

The UK Data Protection Act 1998 (DPA), was drafted in line with the European Data Protection Directive, 1995. It outlines the principles through which personal data can be collected, processed, maintained and disposed of in order for these actions to be considered lawful, fair and proportionate. In an English context, the DPA is just one piece of a complex legal and regulatory environment for the processing of medical records and patient data. The wider legislative and regulatory web is made up of national and transnational legislation, official policy, and professional guidance, supplemented by research governance frameworks under which a variety of research ethics committees operate at a variety of different levels. In relation to the legislative framework, the key reference points for those working with patient data in England other than the DPA are: the Human Rights Act 1998, the common law duty of confidence, the NHS Act 2006 as the parent act to the Health Service (Control of Patient Information) Regulations 2002, the Health and Social Care Act 2012 and the Care Act 2014. This wider legislative framework will be elucidated in the course of this review, but the focus in this section is on the DPA and its implications for the processing of medical records and patient data for secondary research purposes.

In line with the European Data Directive, the applicability of the DPA hinges on its definition of 'personal data'. This is defined as 'data which relates to a living individual who can be identified – a) from those data, or b) from those data and other information which is in the possession of, or is likely to come into the possession of, the data controller' (Section 1 (1)). The Act also defines 'sensitive personal data' as a specific kind of 'personal data' to which stricter controls need to be applied. Sensitive data includes, amongst other categories, data about a person's 'physical or mental health or condition' and 'sexual life' (Section 2 (e)). Therefore medical records and patient data are considered under the Act to be both 'personal' and 'sensitive'. The fair and lawful processing of sensitive personal data requires compliance with the eight general principles of the DPA. These are designed to ensure that personal data is:

(1) processed 'fairly and lawfully'
(2) obtained only for one or more specified and lawful purposes, and not further processed in any manner incompatible with that purpose or those purposes
(3) adequate, relevant and not excessive in relation to the purpose or purposes for which they are processed
(4) accurate and, where necessary, kept up to date
(5) not kept for longer than is necessary for that purpose or those purposes

(6) processed in accordance with the rights of data subjects provided for under the Act

(7) protected by appropriate technical and organisational measures against unauthorised or unlawful processing and against accidental loss, destruction, or damage

(8) not transferred to a country or territory outside the European Economic Area unless that country or territory ensures an adequate level of data protection

The first principle requires that processing of sensitive personal data must meet at least one of the conditions for 'fair and lawful processing' set out in Schedule 2 of the Act and at least one of the conditions in Schedule 3 of the Act. Schedule 2 of the Act permits processing of information under four
main conditions:

1. Where consent has been obtained
2. If processing is in the vital interests of the data subject
3. If processing is "necessary for the exercise of... functions of a public nature exercised in the public interest by any person"
4. If the processing is necessary for the purposes of legitimate interests pursued by the data controller or by the third party or parties to whom the data are disclosed.

Schedule 3 of the Act specifies the further conditions under which sensitive personal data can be processed. The three relevant in the context of justifying processing of sensitive personal data in relation to medial records and patient data can be summarised as:

1. Where explicit consent has been obtained
2. The processing is necessary to protect the vital interests of the individual (in a case where the individual's consent cannot be given or reasonably obtained), or another person (in a case where the individual's consent has been unreasonably withheld).
3. If processing is necessary for medical purposes and is undertaken by a health professional, or equivalent other professional who owes a duty of confidentiality with regard to patient information

In relation to processing medical records and patient data to support direct care, the 'vital interests' condition ensures that in emergency care situations, medical and health professionals can proceed to give treatment without the need to gain explicit consent.

Secondary processing for health research: with or without consent?

One of the key questions for researchers who seek to process medical records and patient data for

secondary research purposes has been whether, and on what grounds, the DPA makes it possible to conduct such research *without* explicit consent from the data subjects. There has been much discussion around this issue in the health and medical literature. The case for legitimising the possibility of conducting secondary research on sensitive personal data without explicit consent has been made on a variety of grounds, depending on the specificities of the context, but the argument can be grouped around three scenarios commonly presented in the literature: (1) contexts in which gaining explicit consent for secondary processing is either impossible or impractical; (2) contexts in which asking for consent would likely introduce an unacceptable bias into research findings or otherwise affect the quality of the research; and (3) contexts in which asking for consent may result in undue distress to the data subject (Academy of Medical Sciences, 2011, p.57; Forgo, 2015, p.58).

Determining whether or not such arguments are legitimate is not the focus here. Instead, the focus is on determining the extent to which the DPA enables this type of processing. Principle One of the DPA on 'fair and lawful processing' sets out the conditions for the processing of 'sensitive personal data' as outlined above, and is therefore a key part of the puzzle. However, all of the data protection principles need to be considered collectively rather than in isolation and, therefore, other DPA principles also have a role in determining the boundaries of what is and is not acceptable. Significantly, in relation to a consideration of processing for *secondary research purposes,* researchers must take heed of the second principle of the DPA which specifies that '[p]ersonal data shall be obtained only for one or more specified and lawful purposes, and shall not be further processed in any manner incompatible with that purpose or those purposes.' A question therefore opens up here as to whether further processing for research purposes can be viewed as compatible with the original purpose of the data collection. This is dealt with by Section 33 (2) of the Act which specifies that processing for 'research purposes' should not be interpreted as incompatible with the original purpose. The outcome of the application of this is that, as long as the 'relevant conditions'[14] also specified in Section 33 (2) are met, then conducting secondary research without consent is possible without breaching Principle Two. However, a potential difficulty lies in the fact that Section 33 (2) defines research purposes rather narrowly as research that is for 'statistical or historical purposes'. This leaves health related research in a potentially ambiguous position as "statistical or historical purposes" is not ideal terminology for covering all types of health research. Furthermore, while the so-called Section 33 'research exemption' provides a qualified exemption from the second (and fifth, while further modifying the sixth) data protection principles, the other principles still remain applicable and must be taken into account when considering whether it is possible to proceed without consent. Given these complexities, it is not surprising that health researchers have (particularly when the Act first came into force) bemoaned the difficulties of interpreting what they can and cannot legitimately do in relation to the processing of identifiable patient data under the DPA (Boyd, 2003). Several authors argue that the complexity of the DPA has directly contributed to the development of a dominant 'consent or anonymise' interpretation

[14] The relevant conditions are, '(a) that the data are not processed to support measures or decisions with respect to particular individuals, and (b) thatthe data are not processed in such a way that substantial damage or substantial distress is, or is likely to be, caused to any data subject' Data Protection Act 1998, Section 33(1)'

of the legal framework surrounding medical and health research (Al-Shahi and Warlow, 2000; Clarke and Weale, 2011, Academy of Medical Sciences, 2006; Haynes et al, 2007). Given that this is the case it will be useful here to look at the data protection principles, in particular Principle One, to consider the ways in which it *is* in fact possible to justify (under the DPA) the processing of sensitive personal data for secondary medical and health research purposes without explicit consent.

As stated above, under Principle One of the DPA, the processing of sensitive personal data must meet at least one of the conditions for 'fair and lawful processing' set out in Schedule 2 of the Act and at least one of the conditions in Schedule 3 of the Act. Guidance issued by the ICO for health researchers which appeared shortly after the Act came into force sought to make it clear that a Schedule 2 condition can be satisfied by demonstrating that the processing is necessary for the 'exercise of any function of a public nature exercised in the *public interest* by any person' (5(d), or if it is in the legitimate interest of the data controller (5(c)). In relation to fulfilling the additional requirements for sensitive personal data outlined in Schedule 3, obtaining 'explicit' consent is only *one* of the possible conditions under which processing can be justified. Schedule 3 can *also* be satisfied by demonstrating that the processing is necessary for medical purposes and is undertaken by a health care professional or person owing an equivalent duty of confidence (8 (1) and (2)). The position from the ICO articulated at the time when the Act came into force was relatively clear – that Schedule 2 and 3 combine in such a way that gaining explicit consent for the secondary processing of sensitive personal data for health and medical research is not the only means of satisfying the requirements of fair and lawful processing (see Boyd, 2003: 253).

Despite the reassuring position adopted by the ICO, the straightforwardness of relying on the 'medical purposes' condition for processing sensitive personal data under Schedule 3, as opposed to 'explicit consent,' is somewhat complicated by the fact that 'necessary for medical purposes' is open to interpretation. 'Medical purposes' are defined in the Act to include 'preventive medicine, medical diagnosis, medical research, the provision of care and treatment and the management of health care services' (Schedule 3, 8 (1) and 8 (2)). The explicit inclusion of 'medical research' under 'medical purposes' is therefore significant, although clarification is somewhat lost by the fact that 'medical research' is not in itself defined. Similarly, what the term 'necessary' means in the phrase 'processing is necessary for medical purposes' is also left unspecified. The Article 29 Working Party (an independent EU advisory body on data protection) have suggested that relying on the equivalent condition in Article 8 (3) of the European Directive as legitimation for the processing of personal sensitive data is only applicable in relation to activities that are required for the *direct provision of services*, and that this does not include medical research, even in the areas of public health and social protection. Furthermore, the Working Party has stated that 'the processing of personal data on grounds of medical purposes must be 'required' for the purposes of provision of services, and not merely 'useful'' (Article 29 Data Protection Working Party, 2007: pp10–12 in Clark & Weale, 2011:11)[15].

[15] According to the Article 29 Working Party: This derogation only covers processing of personal data for the specific purpose

Here we come to an important distinction between the wording of Article 8 (3) in the European Directive and its transposition into Schedule 3 of the DPA. Article 8 (3) does *not* include 'medical research' within its definition of medical purposes, whereas Schedule 3 of the DPA does[16]. It must be remembered here that the EU Data Protection Directive provides for derogation to the rules for the processing of sensitive data. This provision for derogation is expanded on in the EU Data Protection Directive Recitals, which state that 'Member States must also be authorized, when justified by grounds of important public interest, to derogate from the prohibition on processing sensitive categories of data where important reasons of public interest so justify in areas such as public health and social protection...scientific research and government statistics' (EU Data Protection Directive 95/46/EC, Recital 34). In the UK, derogations were introduced in the form of The Data Protection (Processing of Sensitive Personal Data) Order 2000. This provides for a number of scenarios in which the processing of sensitive data can proceed without specific consent, but these are focused on detecting crime and protecting the public from malpractice. Medical and health research is not covered in the order. Analysis of the debate in the House of Lords provides the key to understanding why this is the case. It was clearly felt that use of Schedule 3's 'medical purposes,' which includes 'research' within the condition, already provided sufficient coverage for enabling medical and health research in the public interest to proceed without the need for explicit consent. This is summed up by the Lord Chancellor, who stated in the House of Lords (in the year 2000) that:

> The 1998 Act allows medical data to be used for any medical research purpose without the need for consent of individuals. It is not necessary to define the term "medical research" nor to make specific provision for it to include the monitoring of public health, which for these purposes is regarded as medical research.[17]

As highlighted by Clark and Weale (2011), the other potentially problematic aspect of relying on the 'medical purposes' condition in Schedule 3 of the DPA is that it specifies that a 'health professional' or 'a person who owes an equivalent duty of confidentiality'[18] must undertake the processing in order to qualify. Relying on the 'medical purposes' condition is therefore relatively ambiguous in cases when

---

of providing health-related services of a preventive, diagnostic, therapeutic or after-care nature and for the purpose of the management of these healthcare services, e.g. invoicing, accounting or statistics. Not covered is further processing which is not required for the direct provision of such services, such as medical research, the subsequent reimbursement of costs by a sickness insurance scheme or the pursuit of pecuniary claims. Equally outside the scope of application
of Article 8 (3) are some other processing operations in areas such as public health and social
protection, especially in order to ensure the quality and cost-effectiveness of the procedures
used for settling claims for benefits and services in the health insurance system, as these are
mentioned in recital 34 of the Directive as examples for invoking Article 8 (4). Furthermore the processing of personal data on grounds of Article 8 (3) must be "required" for the specific purposes mentioned [above]. (p.10)

[16] Article 8 (3) of the European Directive allows for the processing of sensitive personal data under three cumulative conditions: the processing of sensitive personal data must be "required", and this processing takes place "for the purposes of preventive medicine, medical diagnosis, the provision of care or treatment or the management of health-care services" and the personal data in question
"are processed by a health professional subject under national law or rules established by
national competent bodies to the obligation of professional secrecy or by another person also
subject to an equivalent obligation of secrecy".

[17] Falconer C.*House of Lords Hansard* 2000 Nov 14. www.parliament.the-stationery-office.co.uk/pa/ld199900/ldhansrd/vo001114/text/01114-35.htm in Hotopf

[18] This is inline with the third condition under Article 8 (3) of the European Data Directive which specifies that the processing of sensitive personal data under the 'medical purposes' must be performed by medical or other staff subject to professional (medical) secrecy or an equivalent obligation to secrecy

researchers are not also medical professionals (p. 12). However, it has been argued by the Academy of Medical Sciences (2006, p.25) that almost all researchers employed by institutions are under obligations of confidentiality because of the legal principles within the common law of confidentiality. Research contracts typically include clauses that make misuse of confidential information a cause for disciplinary proceedings to reflect these obligations. As such, therefore all researchers are bound by the duty of confidentiality held by a health professional. Therefore, in their opinion, this renders the 'health professional' distinction an unhelpful misnomer (Academy of Medical Sciences, 2006: p. 25 in Clark and Weale, 2011, p.12).

What this amounts to is that the Data Protection Act *does* envisage and enable circumstances in which patient data may be accessed and used for medical research without explicit consent or full anonymisation through Schedule 3's 'medical purposes' condition. However, because 'necessary for medical purposes' is open to interpretation, it does not offer a straightforward blanket coverage, particularly when the broader European perspective is taken into account, in which the definition of 'necessary for medical purposes' is more tightly defined. In relation to cases where the researcher is not also a health professional, the application of the condition is further open to question. As pointed out by Iversen and Hotopf (2006), relevant case law to offer clarification on these issues is distinctly lacking, but in their opinion 'although the courts have not given an authoritative statement related to medical research, previous judgments suggest they would interpret current law as supporting large epidemiological studies that require record linkage, access to cancer registries, or data on names and addresses in order to identify potential participants' as contexts in which it is possible to proceed fairly and lawfully without consent (p.165). However, given the ambiguities, the emergence of a dominant 'consent or anonymise' culture, as a means of advocating for the safest option, becomes understandable.

The ongoing requirement to notify?

Confusion amongst medical and health researchers also arises in relation to processing for secondary purposes under the DPA in relation to the requirement to notify individuals, in order to provide them with information on any proposed processing. This is fundamentally different to the issue of consent. As already established in the above discussion, the second data protection principle states that 'personal data shall be obtained for one or more *specified* and lawful purpose, and shall not be further processed in a means incompatible with that purpose or those purposes. However, Section 33 (2) provides an exemption for 'research' that effectively enables the research to be viewed as compatible with the original purpose and therefore proceed even if it was not notified to the data subject. Despite this, in applying the Section 33 (2) exemption, the first data protection principle around fair and lawful processing must still be upheld. As drawn out by Taylor (2011) the DPA states that:

> Determining the purposes of the first principle whether personal data are processed fairly, regard is to be had to the method by which they are obtained, including in particular whether

> any person from who they are obtained is deceived or misled as to the purposes for which they are to be processed (Schedule 1, Part 2, s 1(1)).

This is not overridden by the 'research exemption' and, therefore, it has been argued that in order to avoid deception, notification of intention to process is still required (Taylor, 2011, p. 65). The European Data Directive deals with the obligation to notify by separating instances where data is obtained directly from a data subject (Article 10) from instances where data is obtained from a third party (Article 11). Through Article 11 (2), the Directive recognises that when data is being further processed by a third party there may be instances where the provision of such information proves impossible or would involve a disproportionate effort. Although it may seem useful to draw a distinction between first-hand and second-hand data collection on these lines, Taylor (2011, p.289) argues that there may be instances where a health researcher has collected data directly, and seeks secondary processing of that data after a significant time-lapse at a point where re-contacting might be extremely difficult if not impossible. According to Taylor (2011, p.290), the EU directive makes no apparent provision for non-notification in circumstances where the researcher was the original data controller. In the UK, the DPA is decidedly unclear on this point. It suggests that data collected directly can be exempt from notification in cases of secondary processing where notification is not practicable, but then goes on to suggest that *only* in the case of information received by a third party is a researcher excused from the responsibility to provide certain information if providing this information would involve *disproportionate effort.* Taylor (2011) suggests that:

> This is unsatisfactory on a number of counts: (i) it leaves researchers unclear exactly when they might (under English Law) be entitled not to provide the information, particularly if they have gathered data (historically) directly from a data subject (ii) it appears to leave English law out of step with the directive that it is supposed to implement (iii) it would seem to imply that researchers need to demonstrate a different level of disproportionate effort in cases where data has been gained directly than when it has been obtained from a third party...[Furthermore] it is unclear to what extent either the Directive or the 1998 Act understands the obligation to notify to be an on-going obligation. If research is already intended at the time that data are collected from a patient, then it would appear relatively unreasonable to not bring that fact to the patient's attention at the time. However, if research is not the intention at the time of collection, but is an intention subsequently formed, is there still an obligation to subsequently bring this to the attention of the data subject? (p.291-292)

Issues with anonymised and pseudonymised data

It is clearly stated in the EU Data Protection Directive that 'the principles of protection shall not apply to data rendered anonymous in such a way that the data subject is no longer identifiable' (Recital 26). However, a problem emerges here because the clear-cut distinction that is being made in the legislation between identifiable and unidentifiable data, which is in fact the fundamental definition on which the

applicability of the legislation hinges, simply does not reflect reality. In research contexts where identifiable data is de-identified (or anonymised) for research purposes it is vital to consider that 'anonymisation of data is never an absolute process, there are different degrees of anonymisation depending on the context' (Academy of Medical Sciences, 2011, p.49). As stated earlier in this review, anonymity can never be defined as an intrinsic property of the data, and this is because 'the level of anonymity of a given dataset depends on what other information is available to the person viewing the data' (Academy of Medical Sciences, 2011, p.49). Handling anonymity is therefore not about the provision of an *absolute* (as the crude position reflected in the EU directive and the DPA would appear to endorse), but instead is about *risk management*. One of the problems here is that managing these risks effectively involves the development of sophisticated technical controls drawing on complex statistical concepts such as k-anonymity, de facto anonymity or data cubes[19]. Needless to say, ensuring that data can legitimately pass under a definition of anonymity is such a complex issue that straightforward divisions between what is or is not 'anonymous' are actually impossible to make.

The fundamental distinction between identifiable and anonymised data made in the European Directive and the DPA, with no accompanying clarification around it, is different from the approach taken in equivalent legislation in other areas of the world. Forgo (2015) points to the USA, which has adopted a pragmatic and directly prescriptive approach: 'the Health Insurance Portability and Accountability Act of 1996 (HIPAA), for example, defines, on the one hand, individually identifiable health information and, on the other hand, provides a list of 18 precisely named identifiers that shall be removed in order to achieve de-identified data. There are no restrictions on the use of de-identified data. If the 18 identifiers are removed, this (purportedly) signifies a negligible risk that the information could be used, alone or in combination with other reasonably available information, by an anticipated recipient to identify an individual who is the subject of the information. This list includes values and entries such as names, dates, phone numbers, fax numbers , and e-mail addresses and also URLs, IP, addresses, and full-face photographs' (Forgo, 2015, p.57-58). Although this list may be exhaustive and easily understandable, the fact remains that even a prescriptive approach such as this cannot completely alleviate the potential risk of identification, however negligible that risk is perceived to be.

Further to this, Clark and Weale (2011) argue that neither the EU Directive nor the DPA are clear on whether the process of de-identification comes under the legislation, given that whoever performs the process will necessarily have access to identifying information in order to de-identify it. Clark and Weale (2011) point to the Information Commissioners' *Guidance on Anonymisation* (2002) which in fact indicates that, although anonymous data may fall outside the remit of the DPA, the act of anonymisation does not: "In anonymising personal data the data controller will be processing such data and, in respect of such processing, will still need to comply with the provisions of the Act" (p.13). This suggests that the act of anonymisation *is* covered by the Act, and as the processing will involve

[19] See for example Sweeney, k-anonymity: a model for protecting privacy. (2002) 10 Int J Uncertain Fuzz 557–570, http://arbor.ee.ntu.edu.tw/archive/ppdm/Anonymity/SweeneyKA02.pdf and E Kamateri et al, 'The Linked Data Access Control Framework' (2014) 50 J Biomed Inform 213–225. doi:10.1016/j.jbi.2014.03.002.

sensitive personal data, the conditions of both consent and notification for 'sensitive' data *would* apply. The ambiguity over when data can legitimately be classed as anonymised, and the confusion over whether the *act* of anonymisation is or isn't covered by the Act, are deeply problematic. This is summed up by Forgo (2015) in the following terms:

> Data protection law is (probably) the only legal field in which already the very first question of what it should and what it should not regulate is under constant debate. After 30 years of data protection law it is still hard to say when and under which conditions data may be seen as anonymous (p.57).

Is there a conflict between the practice enabled by the DPA and public expectations regarding notification and consent?

One of the key questions that can emerge from this is around whether public expectations over consent and notification for the use of patient data for secondary purposes (beyond the direct care of the patient) are out of step with the legal framework represented, in part, by the DPA. This does not just relate to when *identifiable* data is used for health related research, but also includes public perspectives on the use and linkage of pseudonymised data and the more open release of anonymised datasets. Research conducted for the Wellcome Trust and the Medical Research Council around public perspectives on the governance of biomedical research found that:

> The act of providing explicit consent, of being asked and agreeing to the request, is key in signalling the adequacy of research governance. This transaction, with which we are increasingly familiar in our every day lives, is not detached from everyday cultural concerns and obligations and is grounded in social conventions of courtesy: it is polite to ask. Not being asked is impolite and signals disrespect and being taken for granted. Implicit consent [in the mind of the public] is no consent at all (Medical Research Councul, 2007, p.87).

Wider research into public perceptions of the use of medical records and patient data for purposes beyond direct care provides a mixed picture. Reports written from a perspective that seeks to advocate for research, unsurprisingly, paint a picture of public opinion in which secondary uses for research purposes are viewed favourably (Academy of Medical Sciences, 2006). However, most of the research appears to indicate that patients want to at least be notified, and if possible, asked for consent *even* in cases where anonymised data is shared (Medical Research Council, 2007). Therefore, having procedures in place for providing notification *and* gaining consent is a considerably important component in building trust and support in ongoing uses of patient data.

The broad information governance strategy implemented across NHS England, which has been heavily influenced by a succession of Caldicott reviews (to be examined here in due course) has sought to develop a culture around the sharing of confidential patient information where there are 'no surprises'

for the patient (Caldicott, 2013, p.12). Yet despite this apparent tightening up across the NHS, stories of the mishandling of patient data regularly emerge in the press, and the concept of 'no surprises' is therefore continually undermined. A recent example of this hit the headlines through an article originally published in New Scientist entitled *Revealed: Google AI Has Huge Haul of NHS Patient Data* (Hodson, 2016). The article was based on the contents of a data sharing agreement between the Royal Free NHS Hospital Trust and a Google-owned artificial intelligence company called DeepMind which appears to show that 1.6 million identifiable patient records have been shared with Google by the Trust. The agreement sets DeepMind up as a data processor with the Royal Free as the data controller, and in press releases the Royal Free have been at pains to stress that the processing is for the purposes of *direct* clinical care. However, the data has been used to develop an application called Streams which helps staff monitor patients with kidney disease, but as there is no specific data set relating to kidney function, DeepMind has been given access to a wider pool of patient data. The data is encrypted and held by a trusted third party, and conditions have been placed on its destruction. Despite these safeguards, the public realisation that data has been released in this manner, without consent *or* notification, has caused a deep sense of misgiving, and has once again added to the weight of public mistrust as to how patient data is being shared. This mistrust is multi-layered: it is connected to concerns over secondary uses that involve commercial companies, but is also undoubtedly connected to a perceived lack of transparency (and therefore public accountability) over such data sharing agreements. When the public find out about data sharing of this kind through sensational media headlines rather than consultative mechanisms or notifications of intentions (which have a capacity to lead to robust public engagement and debate), a sense of betrayal inevitably takes hold.

Relatively recent research conducted by the Department of Health (2009) supports the emerging view that many patients have high expectations regarding notification and consent, *even* if their data have been effectively anonymised. What is significant about the Department of Health's piece of research (as described in Taylor (2011, p.276) is that it clearly indicates that there is a rift between 'public' and 'researcher' attitudes on these issues. This rift in perception and opinion was markedly drawn out in relation to a discussion on additional uses of particularly sensitive information (such as that placed within a 'sealed envelope' on an individual's health record):

> Almost half of all the respondents felt that data from sealed envelopes within a patient's record should be used for 'additional purposes' if anonymised without the need to get consent but the variation between stakeholder groups was particularly marked.
>
> (1) While three quarters of researchers (75%) thought this was acceptable, only a quarter of patients (25%) and three in ten of the general public (30%) thought it was.
>
> (2) Just over half of patients (51%) and about four in ten of the general public (43%) thought 'sealed envelope' data should only be used with the consent of the patient while two in ten researchers (20%) held this view.

(3) Around a quarter of patients (22%) and the public (26%) thought anonymised data from 'sealed envelopes' should *never* be used. Only a very small proportion of researchers (3%) held this view (in Taylor, 2011, p.276)

It is easy to frame the marked difference in attitudes between researchers and patients as indicative of a tension between the collective benefits of enabling research that leads to broad health improvements (more likely on balance to be advocated for by researchers), and individual privacy and data protection rights (more likely on balance to be advocated for by patients). Yet, presenting the tension between 'patient' and 'researcher' in terms of 'individual vs public interest' is misleading. As drawn out by Taylor (2011), protecting privacy is not just an 'individual interest'; there is also a collective 'public interest' in protecting patient confidentiality, because the maintenance of confidentiality is linked to public trust, and significant levels of public *mistrust* leads to significant withdrawal from the healthcare system. Therefore, the rift between 'researcher' and 'patient' attitudes is actually connected to the tensions between two *interconnected public interests* (that are in fact held to different degrees by *both* researchers and patients). The question is therefore how these interconnected public interests (of maintaining confidentiality and enabling research) should be traded off against one another.

Any attempt to resolve the tension between these public goods will have legal *and* ethical dimensions and therefore it would go beyond a mere discussion of what the law does or does not enable. However, what is clear from the discussion thus far is that the DPA does not offer an absolutely clear route to disentangling the current *legal* requirements surrounding how notification to, and consent from, the data subject should be handled in secondary uses of either identifiable *or* anonymised medical records and health data. The resulting confusion is highlighted continually across the literature on the DPA and it is broadly recognised that nowhere (except perhaps the direct marketing industry) has the purpose of the Data Protection Act and its practical impact been more widely misunderstood than in the area of medical research (Clark and Weale, 2011).

Fourth wave data protection: the European General Data Protection Regulation

The impact of the Lisbon Treaty

In December 2009, the EU's Lisbon Treaty came into force, introducing a new constitution for the European Union. Through an amendment to Article 16 of the Treaty on the Functioning of the European Union, a legal basis was provided for new Union legislation on data protection, effectively enabling the review of the existing Data Protection Directive.

As well as providing the legal basis for the review of the data protection directive, The Lisbon Treaty also incorporated the Charter of Fundamental Rights of the European Union (hereafter the EU Charter) into law. The EU Charter makes a significant distinction between privacy and data protection through Article 7, which recognises the right to privacy as a fundamental human right, and Article 8, which

specifically protects personal data as an autonomous human right, distinct from but closely connected to the right to privacy. Therefore, through the Lisbon Treaty, the right to the protection of personal data as described in Article 7 of the EU Charter became legally binding as a fundamental human right connected to, but *autonomous* from, the right for privacy. The incorporation into the Lisbon Treaty established that the EU Charter would have the same value as an EU treaty, and within this move personal data protection as an autonomous fundamental legal right became legally binding. The elevated positional and legal status of the Charter is therefore highly significant in seeking to understand the shifting landscape of privacy and data protection in the context of the EU (Gonzalez Fuster and Gellert, 2012).

Gonzalez Fuster and Gellert (2012) suggest that until the incorporation of the Charter, EU legislation on the protection of personal data was adopted on the basis of the fundamental rights and freedoms of individuals in general, but more specifically on the basis of the right to privacy laid down in Article 8 of the European Convention on Human Rights (hereafter ECHR). They argue that the elevation of Data Protection to the position of autonomous human right has had little effect on permeating what they refer to as the 'privacy thinking' enacted by the European Court of Justice in its on-going interpretation of data protection issues. They argue that the on-going merging between privacy and data protection, which sees data protection merely as a facet of privacy, is concerning because the nature of data protection as a human right is actually 'not homologous to the nature of the right to privacy' (p. 80). On the contrary, they suggest that both rights 'should rather be envisaged as of diverging essence' (p. 80). Gonzalez Fuster and Gellert (2012) argue that 'privacy' should be regarded as a 'positive freedom' in 'the sense of Isaiah Berlin: it consists of granting a prerogative to an individual. Yet, because this prerogative (or freedom) is not absolute, some criteria on how to lawfully limit it' must be laid down (p. 80). In contrast, data protection can be viewed as a 'negative freedom,' that is, 'protecting the freedom (and autonomy) of individuals not by empowering them with a determinate prerogative, but by channelling the behaviours of others, as they might infringe upon this very freedom' (p. 80). According to Gonzalez Fuster and Gellert (2012):

> The basic assumption underlying data protection law therefore should be that 'data processing is unavoidable in modern societies. It never disputes the fact that personal data can be processed, but merely determines how this processing should be undertaken. Yet, that is not to say that 'data subjects' (i.e. the individuals whose data are processed) are totally passive. On the contrary, they are endowed with a bundle of subjective rights (access or correct data) in order to exert some control over the data processor, thereby strengthening their (negative) freedom. Only by overtly placing the regulation of personal data protection under the scope of the new fundamental right (instead of under an imprecise reference to rights and freedoms of the individual, including the right to privacy, and the free flow of personal data) can the EU ensure progress towards the substantiation of personal data protection in the EU (p. 80).

The essence of Gonzalez Fuster and Gellert's (2012) perspective on the essential difference between privacy and data protection as fundamental human rights is given here, not as a means of endorsing their perspective, or of upholding it necessarily as the right interpretation, but to illustrate how the distinction between privacy and data protection is not at all clear or well established, so that the construction of data protection as a fundamental autonomous (non-absolute) human right remains somewhat ambiguous.

Justifications behind the General Data Protection Regulation (GDPR)

The justifications for the introduction of the General Data Protection Regulation were fundamentally connected to a rhetoric of modernisation. According to the European Commission, the 1995 Directive was introduced 'when many of today's online services and the challenges they bring for data protection did not yet exist'. Therefore, in an era of 'social networking sites, cloud computing, location-based services and smart cards' a new 'robust set of rules' is needed to ensure that 'people's right to personal data protection – recognised by Article 8 of the EU's Charter of Fundamental Rights – remains effective in the digital age' (European Commission, 2015). The rhetoric of modernisation has also been supplemented by calls for harmonisation, which highlight how 'differences in the way that each Member State implements the law have led to inconsistencies, which create complexity, legal uncertainty and administrative costs'—a lack of harmony which in turn affects the 'competitiveness of the EU economy.' Thus it has been argued that a single pan-European law for data protection, replacing the current inconsistent patchwork of national laws, is required so that companies deal 'with one law, not 28.' The benefits of the reform are estimated by the European Commission to be valued at €2.3 billion per year (European Commission, 2015). The aims of the GDPR are therefore multi-faceted, and focus on 'reinforcing individuals' rights, strengthening the EU internal market, ensuring stronger enforcement of the rules, streamlining international transfers of personal data and setting global data protection standards' (European Commission, 2015).

It has been argued from within the medical and health research communities that different transpositions of the 1995 EU Directive into national law have significantly hampered public health monitoring as well as research in and across some Member States (Di Lorio et al, 2014). Verschuuren et al (2008) show, through their research, that where Article 8 of the Directive on the processing of special categories of data (including health data) has not been fully or adequately transposed into national law, the legal constraints on processing have been overwhelming and prohibitive, thus producing what they refer to as 'a significant imbalance in favour of the right to data protection over and above the right to health' (p. 151). The authors point to the delicate balance that must be struck between data protection and enabling health monitoring and research, and highlight the need to balance the right to privacy and data protection against Articles 3(p) and 152 of the Lisbon Treaty that bind the commission to 'contribute to the attainment of a high level of health protection, to improve health, to prevent disease and to obviate sources of danger to health' (Verschuuren et al 2008, p. 151). Many in

the health research community have therefore actively called for data protection reform in the hope that it may provide a smoother path to accessing medical records and health data for research purposes.

The GDPR timeline

On 25 January 2012 the European Commission adopted a package for reforming the European data protection framework which included:

- a Communication entitled 'Safeguarding Privacy in a Connected World: A European Data Protection Framework for the 21st Century' ('the Communication') [COM(2012)9 final]
- a proposal for a Regulation on the protection of individuals with regard to the processing of personal data and on the free movement of such data ('the proposed Regulation') [COM(2012)11 final]
- a proposal for a Directive on the protection of individuals with regard to the processing of personal data by competent authorities for the purposes of prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and the free movement of such data ('the proposed Directive') [COM(2012)10 final]

The Commission's draft text for the Regulation marked the establishment of the main point of reference around which the data protection reform would proceed. The reception of the Commission's draft regulation was far from smooth, and its passage through the remaining legislative process has been marked by controversy and debate. However, its contents were broadly welcomed by the health and medical research communities, with Di Lorio (2014) calling it 'a major breakthrough for data protection in the field of public health and, in general, for health and scientific research' (p. 488).

On 16 January 2013, the rapporteur Jan Philipp Albrecht released suggested amendments to the Commission's proposed Regulation. These amendments were approved by the European Parliament's LIBE committee in October 2013. On 12 March 2014, the Parliament agreed on its consolidated text on the First Reading vote. This was a considerable political feat, considering that a total of 3133 proposals for amendments had been formulated by MEPS (Forgo, 2015). The LIBE committee amendments (which were themselves based on Albrecht's draft amendments) heavily influenced the Parliament's consolidated text of 12 March 2014. However, Parliament's consolidated text was broadly condemned by the academic research community, with a joint statement by academic, patient and non-commercial organisations labelling it as 'highly damaging' for research, and Di Lorio (2014) stating that 'the proposed amendments will make difficult or render impossible research and statistics involving the linkage and analysis of the wealth of data from clinical, administrative… and survey sources, which have contributed to improving health outcomes and health systems' performance and governance; and may illegitimise efforts that have been made in some European countries to enable

privacy-respectful data use for research and statistical purposes. If the amendments stand as written, the right to privacy is likely to override the right to health and healthcare in Europe' (p. 488).

On Monday 15 June 2015, Ministers representing the Member States at the EU Justice and Home Affairs Council finally agreed their approach to the proposed General Data Protection Regulation. This enabled the draft regulation to move into trialogue discussion between the Commission, Parliament and the Council of Ministers. The resultant text was welcomed by the medical and health research communities as a restoration of the 'positive' position for research (Wellcome, 2016).

On 15th December 2015, the EU Commission, Parliament and Council of Ministers reached agreement on the General Data Protection Regulation (GDPR), after months of trialogue negotiations.

The Council of the European Union confirmed agreement on the terms of the GDPR on 12 February 2016 by adopting a political agreement on its text, and formally adopted the GDPR at a Council meeting on 08 April 2016. The European Parliament's Civil Liberties Committee (LIBE) voted on the text on 12 April 2016 and the full Parliamentary Plenary voted on it at the early second reading on 14 April 2016.

On May 4 2016, the GDPR was published in the Official Journal of the European Union. The GDPR comes into effect on 25 May 2018.[20]

The i**mpact of various versions of the GDPR on medical and health research**

The passage of the GDPR, from the Commission's original draft through the amendments proposed by Parliament and the Council to the final Regulation, has been an anxious time for the medical and health research community. Particular alarm was caused by the publication of Parliament's consolidated text. Whilst the original Commission draft had been greeted with cautious optimism by the medical and health research communities (Wellcome Trust, 2015b), it was fiercely argued that Parliament's proposed amendments, if allowed to shape the final Regulation, would 'severely threaten' the possibility of carrying out vital medical and health research in the public interest (Wellcome Trust, 2015b). Therefore, in order to draw this out, the key aspects of three versions of the GDPR will be explored: the Commission's original draft, Parliament's consolidated text, and the final official published version of the GDPR.

The Commission's Draft Regulation (January 2012)

Like the Directive before it, the Commission's Draft Regulation (CDR) sets out the basis of its applicability upon its definition of 'personal data.' The implication was therefore intended to be clear-

[20] For a useful timeline of these key moments see: http://www.twobirds.com/en/practice-areas/privacy-and-data-protection/eu-framework-revision

cut: if the data falls within the regulation's definition of 'personal data' then the rules of the Regulation apply, if it does not then it is out of scope. In Article 4 (1) of the CDR, personal data is defined as data relating to a data subject who is defined as a person:

> [W]ho can be identified, directly or indirectly, by means reasonably likely to be used by the controller or by any other natural or legal person, in particular by reference to an identification number, location data, online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that person (Article 4 par. 1 CDR)

Recital 23 seeks to clarify that anonymous data where 'the data subject is no longer identifiable' is *out* of scope of the regulation. Recital 24, however, adds a layer of ambiguity to the CDR's definition of 'personal data' by stating that 'it follows that identification numbers, location data, online identifiers or other specific factors as such need not necessarily be considered as personal data in all circumstances.' When Article 2 and Recital 24 are taken together, the definition of 'personal data' and the characteristics which differentiate it from non-identifiable 'anonymous data' become decidedly loose and open to interpretation. This open approach was in fact very much welcomed by medical and health research lobbying groups such as those spearheaded by the Wellcome Trust in the UK, who have actively campaigned to keep the CDR's definitions in place. The key reason for this relates to research using pseudonymised data. The Wellcome Trust has been concerned to protect research using pseudonymised data from 'unnecessary regulatory burden,' and has argued that the CDR's open approach would enable a 'proportionate approach to determining the scope of the legislation.' The Wellcome Trust argued to keep what they referred to as the CDR's 'reasonably likely to be used' test as, in their opinion, this made it possible to view robustly pseudonymised research data (where the researcher uses the data in a 'safe haven' with no access to the key to enable re-identification) as out of the Regulation's scope (Wellcome Trust, 2015a). As this discussion will illustrate, that hope has not been realised in relation to the other versions of the draft under analysis here, where pseudonymised data is intentionally and unambiguously drawn *into* the scope of the Regulation.

In keeping with the approach in the EU Directive, the CDR recognises special categories of personal data (Article 9) for which processing is prohibited subject to limited exemptions. In the CDR, the special categories are defined as the 'processing of personal data, revealing race or ethnic origin, political opinions, religion or beliefs, trade union membership, and the processing of genetic data or data concerning health or sex life or criminal convictions or related security measures' (Par. 1). Article 9 Par. 2 specifies consent as the first exemption, but it also lifts the prohibition under a number of other exemptions, including when:

> (h) processing of data concerning health is necessary for health purposes and subject to the conditions and safeguards referred to in Article 81; or
> (i) processing is necessary for historical, statistical or scientific research purposes subject to the conditions and safeguards referred to in Article 83

Accordingly, then, Article 81 and Article 83 provide the possibility for health and medical research to be undertaken without data subject consent. However, determining which of the two Articles would apply in relation to a given health or medical research context is more complex. Article 81 applies (through Par. 1a) to 'purposes of preventative medicine, medical diagnosis, the provision of care or treatment or the management of health-care services…where those data are processed by a health professional subject to the obligation of professional secrecy or another person also subject to an equivalent obligation of secrecy under Member State law or rules established by national competent bodies.' This is, in the main, a transposition of the wording of Article 8 (3) in the former European Data Directive. However, Par. 1b extends the application of Article 81 to purposes undertaken for 'reasons of public interest in the area of public health, such as protecting against serious cross-border threats to health or ensuring high standards of quality and safety, inter alia for medicinal products or medical devices' and (through Par. 1c) to processing for 'other reasons of public interest in areas such as social protection, especially in order to ensure the quality and cost-effectiveness of the procedures used for settling claims for benefits and services in the health insurance system.' Par. 2 of Article 81 appears to clarify when Article 83 is more appropriate, by stating that 'processing of personal data concerning health which is necessary for historical, statistical or scientific research purposes, such as patient registries set up for improving diagnoses and differentiating between similar types of diseases and preparing studies for therapies, is subject to the conditions and safeguards referred to in Article 83.' This would seem to suggest that most research studies which rely on secondary access to patient data for identifying broad medical and health trends for collective health improvements are better classed under Article 83. Article 81 states that 'processing of personal data concerning health must be on the basis of Union law or Member State law which shall provide for suitable and specific measures to safeguard the data subject's legitimate interests.' In addition, Par, 3 of which Article 81 states that 'the Commission shall be empowered to adopt delegated acts in accordance with Article 86 for the purpose of further specifying other reasons of public interest in the area of public health as referred to in point (b) of paragraph 1, as well as criteria and requirements for the safeguards for the processing of personal data for the purposes referred to in paragraph 1.'

Article 83 (which deals with processing for historical, statistical, or scientific research purposes) suggests that personal data can be processed without consent if the research purpose *cannot* be achieved by using anonymous data. Where this is the case, robust pseudonymisation is advised, but this is qualified again by the inclusion of the phrase 'as long as the purpose of the processing can be fulfilled by doing so' (Par. 1b). This effectively opens up the possibility of carrying out health research using identifiable data *without* consent or pseudonymisation in cases where the purpose requires this.

Rather confusingly, within Article 83, a line is drawn between research processing (which is covered by Par. 1) and publication of research results (which is covered by Par. 2). In relation to the publication of results, Article 83 Par. 2 specifies that publication is possible if the data subject has consented (2a), or if the data subject has made the data public (2c). However, a third possibility for justifying publication is also introduced, which states that publication of results is possible (without

consent) if 'the publication of personal data is necessary to present research findings or to facilitate research insofar as the interests or the fundamental rights or freedoms of the data subject do not override these interests' (2b). The division between research and publication given in Article 83 does little in adding clarity to the balance that the CDR is advocating for between respecting privacy and enabling research, and in fact creates an unnecessary ambiguity. Under Article 83 (Par. 3), the CDR text also states that the Commission shall be empowered to adopt delegated acts for further specifying the requirements and criteria for processing. Therefore, in the case of Article 81 *and* Article 83, it appears that the Articles were not intended to represent a fully elaborated provision.

Article 5(b) of the CDR is highly relevant to a discussion of secondary processing of data for medical and research purposes because it specifies that 'personal data' must be 'collected for specified, explicit, and legitimate purposes and not further processed in a way incompatible with those purposes.' However, Recital 40 supports the notion of further processing of personal data 'where the processing is compatible with those purposes for which the data have been initially collected, in particular where the processing is necessary for historical, statistical or scientific research purposes.'

Article 5(e), which deals with data minimisation, also has a bearing on processing for research in that it specifies that personal data 'should be kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed; personal data may be stored for longer periods insofar as the data will be processed solely for historical, statistical or scientific research purposes in accordance with the rules and conditions of Article 83 and if a periodic review is carried out to assess the necessity to continue the storage'. The introduction of the concept of 'periodic review' is interesting here and suggests that ongoing and regular appraisal is needed to justify retention in the long term. Although this caused some concern within the archival community, this does not appear to have been commented on as problematic by the health and medical research community.

Article 6(e), which deals with fair and lawful processing, is also relevant to an understanding of how the CDR interprets processing for research, in that it emphasises that 'processing of personal data which is necessary for the purposes of historical, statistical or scientific research shall be lawful subject to the conditions and safeguards referred to in Article 83.'

The necessity of notifying data subjects of intended processing has been discussed earlier in this review in relation to the DPA, so it is useful here to consider how this is approached in the CDR. Here, the CDR appears to echo the earlier Directive in that Article 14 (5) effectively relinquishes the requirements for notification in cases 'when the data are not collected from the data subject and the provision of such information proves impossible or would involve disproportionate effort,' but there is nothing in the CDR to relinquish this responsibility in cases where the data has been collected first-hand. In relation to the data subject's right to be forgotten and to erasure contained in Article 17(3), this is relinquished for reasons of public interest in the area of public health in accordance with Article

8(b), and for historical, statistical and scientific purposes in accordance with Article 83 (c). Article 19 deals with the data subject's right to object, which is upheld 'unless the controller demonstrates compelling legitimate grounds for the processing which override the interests or fundamental rights and freedoms of the data subject.' International transfers, which can be significant in trans- and international medical and health research, are dealt with in Article 42 and are enabled by prior authorisation from the supervisory authority or through legally binding instrument.

The European Parliament's Consolidated Text (March 2014)

In his report for Parliament in January 2013, Jan-Philipp Albrecht 'proposed a very different equilibrium between privacy and research interests' (Forgo, 1215: 61). Despite significant attempts at lobbying MEPs by the health and medical research communities, the thrust of his amendments was carried through into the LIBE committee-approved text of October 2013 and also into Parliament's final consolidated text (hereafter PCT).

Firstly, it is important to draw out that the PCT contained a tighter definition of 'personal data' than that used in the CDR:

> 'personal data' means any information relating to an identified or identifiable natural person ('data subject'); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, unique identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social or gender identity of that person (Article 4, Par. 2)

Significantly, the PCT definition removes the phrasing in the CDR that linked the concept of being 'identifiable' to 'means reasonably likely to be used' by the data controller or others to identify the data subject. The implication of this tightening is that it becomes much more difficult to argue that robustly pseudonymised data may fall out of the scope of the Regulation. In fact, the PCT goes to the extent of introducing a definition of 'pseudonymous data' as a distinct category of personal data in Article 4. This change was highly criticised by the Wellcome Trust in its lobbying efforts (Wellcome Trust, 2015).

However, the area of greatest concern to the health research community was in relation to the PCT's amendments to Article 81 (Processing of Personal Data Concerning Health). Amendments to Article 81 Par. 2 of the CDR were introduced to ensure that, as an almost unbreakable rule, health research could *only* be undertaken with the explicit consent of the data subject. Accordingly, processing without consent could only be enabled in the PCT through national law in *exceptional* circumstances. This is laid out through amendments to Article 81, Par. 2, which states that 'processing of personal data concerning health which is necessary for historical, statistical or scientific research purposes shall be permitted only with the consent of the data subject and shall be subject to the conditions and safeguards

referred to in Article 83.' An entirely new sub-clause is then added (Article 81 Par. 2a) which states that:

> Member States law may provide for exceptions to the requirement of consent for research, as referred to in paragraph 2, with regard to research that serves a high public interest, if that research cannot possibly be carried out otherwise. The data in question shall be anonymised, or if that is not possible for the research purposes, pseudonymised under the highest technical standards, and all necessary measures shall be taken to prevent unwarranted re-identification of the data subjects. However, the data subject shall have the right to object at any time in accordance with Article 19.

As was to be expected, there was vehement opposition within the health research community to the idea that Member States could *only* provide an exemption from consent for research use of personal health data in cases where there is an 'exceptionally high public interest.' In May 2013, a joint statement (led by the Wellcome Trust) from non-commercial research institutions and academics drew attention to the likely impact that the proposed amendments would have on health and scientific research:

> Health and scientific research will be severely threatened if the amendments to Articles 81 and 83 of the Data Protection Regulation adopted by the European Parliament are taken forward. Scientific research generates important benefits by improving our understanding of society, health and disease. If implemented, the amendments would make much research involving personal data at worst illegal, and at best unworkable (Wellcome Trust, 2013).

The research community also objected strongly to the fact that Article 81 Par. 2a makes it explicitly clear that member states can *only* legislate for a research exemption to the requirement of consent for anonymised or pseudonymised data. If carried through into the final Regulation, this would effectively mean that health and medical research that made use of *identifiable* data would *never* under *any* circumstances be possible without consent, regardless of the safeguards in place. In response to this, the health research community has been extremely keen to stress that whilst anonymised and pseudonymised data can provide the basis for many research studies, equally there are cases when anonymised or pseudonymised data is simply not sufficient for the research purposes. In some of the cases where it *is* absolutely necessary to use identifiable data, there will be instances where gaining consent is either impossible, disproportionately impracticable or not in the best interests of the data subject (Di Lorio, 2014).

Other concerns from within the health research community focussed on the fact that pseudonymised data are often not sufficient to permit researchers to identify individuals in order to invite them to take part in an ethically approved study. In these instances researchers need access to the minimal amount of data required to identify eligible people. Obviously, people can only agree to join studies if they can be invited. It was therefore felt that the PCT proposals would outlaw this first step, which would render

even the *process* of gaining consent an impossibility.

Other issues with the PCT text, voiced by the health research community, related to Article 5(b) which specifies that personal data should be collected for 'specified, explicit and legitimate purposes and not further processed in a way incompatible with those purposes.' Whilst this may not be substantially different to the text in the CDR, there is an absence in the PCT of Recital 40, which in the CDR reinforces the concept of further processing in the context of historical, statistical or scientific research.

In relation to 'storage minimisation,' Article 5(e) also imposes tighter measures in the PCT then the CDR by stating that personal data should be:

> Kept in a form which permits direct or indirect identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed; personal data may be stored for longer periods insofar as the data will be processed solely for historical, statistical or scientific research or for archive purposes in accordance with the rules and conditions of Articles 83 and 83a and if a periodic review is carried out to assess the necessity to continue the storage, and if appropriate technical and organisational measures are put in place to limit access to the data only for these purposes (storage minimisation).

Standard protocols in research projects that access personal data for secondary health research already establish a need for periodic review of the necessity to hold the data and are also open to audit by the data controller in relation to the technical and organisational environment in which the data is stored and accessed. These protocols are usually laid out under data sharing agreements and licenses for use. Therefore, for the health and medical research communities, this did not necessarily raise additional concerns. However, the specific mention of 'archive purposes' here is potentially problematic for archival institutions. Parliament's text was the first to mention archives specifically, and whilst this could be interpreted as welcome recognition of the existence and role of the profession, its appearance in 5(e) is problematic. In tying archives to periodic reviews of holdings, and the necessity of ensuring that the appropriate 'technical and organizational infrastructure' is in place to limit access, this in effect put a tighter regulatory burden on archives than in the case of the CDR, where archives are not mentioned at all.

Parliament's approach to the Regulation has been described by health researchers as an 'ideologically-driven approach' that perhaps inadvertently created problems for research. Parliament's position was clearly dominated by the views of privacy advocates, and was undoubtedly fuelled by the Edward Snowden leaks regarding the use of personal data by national security agencies. The Policy Advisor at the Wellcome Trust, Beth Thompson, has commented that:

> Parliament had good intentions in taking a privacy-led approach, and MEPs did not intend to prevent research. However, many of their amendments to the Commission proposal were not

> based on a sound understanding of how research is conducted, or the potential impact these would have on research. In reality, the Parliament's amendments would have inadvertently tightly restricted the ways in which personal data could be used in research, with devastating impacts…For many working on the Regulation, inside Parliament and out, scientific research was a marginal issue. This may explain why the concerns of the research community were not heard when we first raised them in early 2013. The research community's input increased in volume as the situation grew more serious when the Parliament agreed that their amendments should form part of their position. Over the next two years a large coalition of research and patient organisations across Europe worked together to raise awareness of our concerns.

Overall the balance between the public good of data protection and the public good in enabling medical and health research is tipped vastly in favour of the former in the PCT. However, Data Protection is a delicate balancing act and it is acknowledged across the drafts of the GDPR that it requires a proportionate approach:

> The processing of personal data should be designed to serve mankind. The right to the protection of personal data is not an absolute right; it must be considered in relation to its function in society and be balanced against other fundamental rights, in accordance with the principle of proportionality (Recital 4, GDPR).

How well the final GDPR manages the delicate balance between the public goods of data protection and vital health research will be examined below.

Official General Data Protection Regulation (2016)

The final definition of personal data in Article 4 (1) of the GDPR is exactly equivalent to that in the PCT, apart from the change from 'unique identifier' to 'online identifier' shown below:

> 'personal data' means any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an *online identifier* or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person (emphasis added)

The PCT's definition of 'pseudonymous data' is replaced, in Article 4 (5), with a definition of 'pseudonymisation' as follows:

> ‘pseudonymisation’ means the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person.

The notion that pseudonymised data should be considered to be *in* the scope of the Act is made unequivocal through the first part of Recital 26, which states that:

> The principles of data protection should apply to any information concerning an identified or identifiable natural person. Personal data which have undergone pseudonymisation, which could be attributed to a natural person by the use of additional information should be considered to be information on an identifiable natural person.

This means that, contrary to the arguments put forward by the medical and health research communities, the processing of all pseudonymised data is in the scope of the Regulation regardless of whether robust safeguards are in place in ‘safe haven’ environments to ensure that use of the data takes place without the possibility of access to the key that could enable re-identification. However, the Wellcome Trust (2016, unpaginated) has since issued an analysis of the GDPR which argues that Recital 23 of the GDPR clouds the water on pseudonymisation, arguing that:

> Recital 23 can be read that all pseudonymised data should be considered personal data. The concept of identifiability also appears to have been expanded compared to the Directive by the reference to “singling out.” However, the scope of identifiability is qualified by the reference to “means reasonably likely to be used” as under the 1995 Directive. This suggests that there may be cases where pseudonymised data together with a combination of appropriate organisational, legal and technological measures can be considered anonymous data. A proportionate and context-dependent approach would take into account the range of measures used, including pseudonymisation, to determine whether the data is considered to be identifiable. In order to achieve this it is important to consider the text of Recital 23 in full to understand how the scope of the regulation relates to approaches commonly used in research.

Given that this is the position that the Wellcome Trust is taking, it is currently urging the European Data Protection Board (EDPB) to provide guidance on the interpretation of Recital 23 to clarify the scope of the Regulation with specific reference to a research context in which pseudonymisation is used in combination with strict organisational, legal and technological measures to reduce identification risks. They are also urging further clarity on the nature of the technical and organisational measures described under Article 4 (3b) to clarify whether additional information must always be kept separately by a different data controller, or whether this can be achieved *within* a data controller’s

infrastructure (Wellcome, 2016, unpaginated). Therefore, it appears that the issue of pseudonymisation and its applicability under the regulation is still far from clear.

The remainder of Recital 26 is concerned with clarifying that anonymised data is *out* of the Regulations's scope, and here an attempt is made to define the parameters for considering what constitutes 'anonymised data':

> To determine whether a natural person is identifiable, account should be taken of all the means reasonably likely to be used, such as singling out, either by the controller or by another person to identify the natural person directly or indirectly. To ascertain whether means are reasonably likely to be used to identify the natural person, account should be taken of all objective factors, such as the costs of and the amount of time required for identification, taking into consideration the available technology at the time of the processing and technological developments. The principles of data protection should therefore not apply to anonymous information, namely information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable. This Regulation does not therefore concern the processing of such anonymous information, including for statistical or research purposes.

These statements in Recital 26 amount to an acknowledgement that since anonymity is not an intrinsic attribute of the data, a risk based approach must be applied to determine the *likelihood* of identification given the costs, time and technology involved in managing to single an individual out. The Regulation has not found a way around the fact that 'anonymity' is context-dependent and therefore impossible to define as an absolute fixed category of data. The irony therefore remains that the applicability of the Regulation hinges on whether the data renders an individual identifiable or not, yet it is impossible to categorically state the boundaries of 'identifiability.' On the point of anonymisation, the Regulation (like the EU Directive and the UK DPA) also remains silent on the issue of whether the *process* of anonymisation is in or out of the scope of the Regulation. Given this silence, it must be presumed that since it involves processing identifiable data in order to *de-identify* then it must be *in* scope.

Article 9, which deals with special categories of data for which processing should be prohibited (unless one of the listed exemptions applies) defines the categories and the prohibition in the following terms:

> Processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation shall be prohibited.

Par. 2 and Par. 3 then deal with the exemptions. Alongside 2 (a), which relates to when explicit consent has been gained, the other exemptions relevant to health and medical research purposes are:

(h)processing is necessary for the purposes of preventive or occupational medicine, for the assessment of the working capacity of the employee, medical diagnosis, the provision of health or social care or treatment or the management of health or social care systems and services on the basis of Union or Member State law or pursuant to contract with a health professional and subject to the conditions and safeguards referred to in paragraph 3;
(i) processing is necessary for reasons of public interest in the area of public health, such as protecting against serious cross-border threats to health or ensuring high standards of quality and safety of health care and of medicinal products or medical devices, on the basis of Union or Member State law which provides for suitable and specific measures to safeguard the rights and freedoms of the data subject, in particular professional secrecy;

(j) processing is necessary for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Article 89(1) based on Union or Member State law which shall be proportionate to the aim pursued, respect the essence of the right to data protection and provide for suitable and specific measures to safeguard the fundamental rights and the interests of the data subject.

3. Personal data referred to in paragraph 1 may be processed for the purposes referred to in point (h) of paragraph 2 when those data are processed by or under the responsibility of a professional subject to the obligation of professional secrecy under Union or Member State law or rules established by national competent bodies or by another person also subject to an obligation of secrecy under Union or Member State law or rules established by national competent bodies

What is noticeable in the structure of the final GDPR is that there is no specific article for dealing with processing concerning health data. This is dealt with from *within* Article 9. This is, in fact, a transfer from the Council of Ministers' consolidated text (not under direct analysis here). For medical and health researchers, a decision is necessary to disentangle whether the processing purposes fall under (h), (i), or (j). As was the case with the CDR, where research is not explicitly linked to the direct care of individuals or the provision of services, but rather seeks to identify broader trends for the purpose of health improvement, it will fall best under (j) as a 'scientific research purpose' and will therefore need to take Article 89 into account. Before moving on to look more closely at Article 89, the significance of Par. 4 of Article 9 must be drawn out. It states that:

> Member States may maintain or introduce further conditions, including limitations, with regard to the processing of genetic data, biometric data or data concerning health.

Thus, contrary to one of its main stated intended purposes (which was to introduce a harmonisation of approach), in relation to health data the final GDPR *may* do little to counteract the piecemeal approach

to the processing of health data across the jurisdictions of the 28 Member States.[21] Arguably, however, a combined interpretation of the Articles and Recitals in the GDPR makes the position on processing for scientific research purposes much clearer than it was under the EU Directive. The elaborations that it provides and the degree of clarity in the way it is structured should diminish the possibility of misinterpretation of the rules, and should help to ensure that Member States adopt harmonised public health policies (Di Lorio, 2014).

Article 89 introduces the safeguards and derogations relating to 'processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes.' Par. 1 states that:

> Processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes, shall be subject to appropriate safeguards, in accordance with this Regulation, for the rights and freedoms of the data subject. Those safeguards shall ensure that technical and organisational measures are in place in particular in order to ensure respect for the principle of data minimisation. Those measures may include pseudonymisation provided that those purposes can be fulfilled in that manner. Where those purposes can be fulfilled by further processing which does not permit or no longer permits the identification of data subjects, those purposes shall be fulfilled in that manner.

In accordance with Par. 1, then, research using personal health data should fulfil its purpose using anonymised data if possible. If this is not possible, then measures for data minimisation should be employed, and pseudonymisation should be considered if the purpose can be enabled with its use. Clarification that research using personal health data can be undertaken without consent is given through Recital 54, which states that:

> The processing of special categories of personal data may be necessary for reasons of public interest in the areas of public health without consent of the data subject. Such processing should be subject to suitable and specific measures so as to protect the rights and freedoms of natural persons.

Returning to Article 89, Par. 2 goes on to state that:

> Where personal data are processed for scientific or historical research purposes or statistical purposes, Union or Member State law may provide for derogations from the rights referred to in Articles 15, 16, 18 and 21 subject to the conditions and safeguards referred to in paragraph 1 of this Article in so far as such rights are likely to render impossible or seriously impair the achievement of the specific purposes, and such derogations are necessary for the fulfilment of those purposes.

In effect, this means that a data subject's Rights of Access (Article 15), Right to Rectification (Article 16), Right to Restriction of Processing (Article 18) and Right to Object to the Processing (Article 21) can all be restricted through Member State law in cases where maintaining these rights would defeat the achievement of the processing (which could be the case in certain health and medical research

[21] Recital 53 elaborates further on this issue

contexts). The Right to Erasure ('right to be forgotten') in Article 17 is not included in the list, as a direct derogation is given for 'archiving purposes in the public interest, scientific or historical research purposes or statistical purposes' under Article 17 Par. 3 (d).

Article 14 deals with the information that should be provided where personal data have not been obtained from the data subject. Article 14 (5) (e) makes an exemption to the notification of intended processing detailed in Par. 1-4 in cases where:

> [T]he provision of such information proves impossible or would involve a disproportionate effort, in particular for processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes, subject to the conditions and safeguards referred to in Article 89 (1) or in so far as the obligation referred to in paragraph 1 of this Article is likely to render impossible or seriously impair the achievement of the objectives of that processing. In such cases the controller shall take appropriate measures to protect the data subject's rights and freedoms and legitimate interests, including making the information publicly available.

Thus, Article 14 Par. 5b provides a more nuanced understanding of the circumstances in which it may not be possible to notify a data subject. It includes alongside the notions of 'impossibility' and 'disproportionate effort' recognition that notification should not be necessary in cases where it would seriously impede the achievement of the objectives of that processing. The Health and Medical Research Communities are likely to welcome this, as health research is a context in which there are occasions when notification is feasible from a cost/time perspective but where it might be detrimental to the running of the research. However, previous points made in the discussion of the UK DPA also have resonance here. The GDPR makes no provision for recognising that there are some cases in which the health researcher may have been the original data controller, and is returning to the data after considerable time to further process the data. In such cases, the notions of impossibility or disproportionate effort do not apply under the GDPR.

Under the principles for processing, Article 5 (b) makes it explicit that the general principle that prevents further processing is not applicable to 'archiving purposes in the public interest, scientific or historical research purposes or statistical purposes' as these 'shall, in accordance with Article 89 (1), not be considered to be incompatible with the initial purposes.' In a similar way Article 5 (e) also makes it explicit that the rules on storage limitation do not apply to when data is processed for such purposes.

Of general relevance to health researchers is Article 25, which introduces the concept of data protection by design and default, Article 32, which deals with the security of processing, Articles 33- 34, which deal with data breaches, Article 35, which deals with data protection impact assessments (which will be applicable to large-scale data projects), and Articles 44-46, which deal with data transfers.

The response to the agreed GDPR from policy advisors, funders, and academics from within the health and research community can be summed up as relief in the face of the serious research restrictions that would have been imposed had Parliament's text been taken forward in the trialogue discussions.

Bernard Charpentier, President of the Federation of European Academies of Medicine, publicly commented following news of the agreements reached in December 2015, stating that:

> Personal data is used in health research to make important discoveries, which can save lives. Regulations must balance the use of personal data in research with the interests of individuals. We had been concerned that potential amendments to the EU General Data Protection Regulation could threaten health research studies. However, based on the outcome of the trialogue negotiations, we are optimistic that valuable health research studies will continue and we are pleased that the importance of this research has been recognised. We look forward to assessing the full implications for health research in due course (Wellcome Trust, 2015c).

Beth Thompson at the Wellcome Trust made remarks along similar lines, saying that the 'result is a significant achievement and a huge relief for the research community.' She perceptively adds 'however, this is not the end of the story. The devil will be in the detail of implementation. The medical and health research communities will need to continue to work to secure a strong outcome as the Data Protection Regulation enters this next phase, before the law becomes active in 2018' (London School of Eeconomics, 2016).

The broader regulatory and legal landscape for health and medical research in England

> The existing regulation and governance pathways [for health and medical research] have evolved in a piecemeal manner over several years. New regulatory bodies and checks have been introduced with good intentions, but the sum effect is a fragmented process characterised by multiple layers of bureaucracy, uncertainty in the interpretation of individual legislation and guidance, a lack of trust within the system, and duplication and overlap in responsibilities. Most importantly, there is no evidence that these measures have enhanced the safety and well-being of either patients or the public (Academy of Medical Sciences, 2006, p.37)

By delving into the regulatory and governance frameworks surrounding medical and health research, it is easy to understand why the Academy of Medical Sciences chose to make such a bleak statement.

The legal framework surrounding health and medical research is complex and involves UK statutory legislation, common law decisions, and various EU Directives. This is complicated further by the fact that there are numerous bodies responsible for issuing guidance on how the law should be interpreted in practice. These include (but are by no means limited to) the Information Commissioner's Office, the General Medical Council, the Medical Research Council, and the British Medical Council. Further complexities arise when the different Research Governance Frameworks (which outline the principles, standards and requirements for health and social care research) that are in place across the devolved nations are taken into consideration. The latest Research Governance Framework proposals are currently under consultation and will involve the withdrawal of the separate frameworks across the devolved nations in favour of a single UK policy framework. However, whilst the integration is useful, it must be remembered that the Research Governance Framework is a high-level policy document and will, as such, not be able to integrate the lower-level operational procedures for the management of health related research which will remain under the separate responsibility of the Health Research Authority and the equivalent bodies in the devolved administrations respectively. Additionally, ethical review for health and medical research in the UK is still a complex landscape despite consistent efforts to streamline the process. In an English context, the main ethics review committees (RECs) that can have a role in relation to health research are:

- The National Research Ethics Service (now under the auspices of the Health Research Authority) incorporates the NHS Research Ethics Committees and the Social Care Research Ethics Committee
- Independent Ethics Committees (IECs) which are involved in the review of clinical trials that take place outside of the NHS
- The Gene Therapy Advisory Committee (GTAC)
- Ministry of Defence (MoD) research ethics committees
- University ethics committees

(drawn from Academy of Medical Sciences, 2006)

The development of an Integrated Research Application System (IRAS) has done much to bring these different elements together. However, there is still residual complexity in navigating the landscape. One complication for research involving access to NHS patient data is that even when NRES approval for the research has been secured, as each NHS Trust is a legal entity in its own right, it must have its own procedures in place for reviewing its research and development activities. This leads to duplication in checks between the NRES and local NHS trusts, inconsistencies in approach between Trusts, and difficulties where patient data needs to be accessed across multiple Trusts. This brief glimpse of the research governance and ethical review frameworks that embrace health research suffices, for the purposes of this review, to draw attention to the complexities that exist in the broader regulatory landscape, beyond what is apparent when the legislation *alone* is in focus. This has led the medical research community to voice strong concerns.

The focus in this review, however, is on data protection legislation, and the broader legislation that also affects research use of medical records and patient data. Given this focus, the duties that are owed regarding health and medical data under the Common Law Duty of Confidence will be explored in relative depth here. In relation to the Common Law Duty of Confidence, Grace and Taylor (2014) assert that:

> The common law duty of confidence is one of the oldest equitable remedies recognised by English Law. In *Duke of Queensberry v Shebbeare*, it was established that a manuscript could not be published when publication was not one of the purposes for which it had been given. Since that time, and up to the modern day, English Law has continued to recognise that confidential information may not be disclosed for any purpose outside the 'reasonable expectations' of the party to whom it relates. The result is that, under English common law, confidential patient information may not be used for any purpose outside the reasonable expectation of a patient confiding personal information in a health care professional (p. 420).

The question mark over the interpretation of the common law duty of confidence in the context of medical and health research relates to what can be assumed to be within the 'reasonable expectations' of the patient regarding the sharing of their confidential information (Taylor, 2015, p.353). Taylor (2015) goes on to elucidate that 'reasonable expectations' is typically interpreted to mean that 'information that can identify individual patients, must not be used or disclosed for purposes other than healthcare without the individual's explicit consent, some other legal basis, or where there is a robust public interest or legal justification to do so' (p. 353). This duty to respect confidentiality through adherence to 'reasonable expectations' around disclosure is not only rooted in common law, but also in

the various professional codes of ethics and codes of conduct that are used across the health and social care sector[22].

Taylor (2015) suggests that while this may be a defensible position in both law and principle, it has not always been clear exactly what is required in practice. There is considerable ambiguity concerning the boundaries of the processes that can be construed as being part of a patient's direct healthcare, and therefore when explicit consent is needed to override the duty of confidentiality. For example, is explicit individual consent needed when using identifiable information to *audit* care and treatment? Similarly, what are the range of disclosures beyond direct healthcare that are permissible without explicit consent due to a 'robust public interest'? (p. 345) And how restrictive does an interpretation of a patient's 'reasonable expectation of privacy' need to be? (Grace and Taylor, 2013). This lack of clarity around permitted sharing and unlawful disclosure under the common law was increasingly recognised as problematic. To this end, the introduction of specific Health Service (Control of Patient Information) Regulations in 2002 was a direct response to a perception of ambiguity regarding when common law thresholds of confidentiality can be legitimately overridden and, therefore, when processing can be justified *without* the need for explicit data subject consent.[23] (Taylor, 2015, p.354).

An understanding of the broader socio-historical context in which the regulations came into being provides a useful backdrop for understanding the motivations and underlying ethos underpinning their introduction. Prior to the National Health Service's establishment, it was widely accepted as common practice across the UK that patient information could be used for secondary purposes beyond direct care, including research, at the discretion of the doctor responsible for treatment (Grace and Taylor, 2013, p.422). That ethos carried on within the NHS, with disclosure of confidential information for research purposes, without the need for explicit consent, being broadly accepted as being in the public's best interests, and not contravening an individual patient's 'reasonable expectations of privacy.' Acceptance of a 'doctor knows best' approach to data sharing and research continued to go unchallenged right up until the 1990s, when the beginnings of a shift in acceptance and practice of that approach can be detected (Higgins, 2003). The introduction of the DPA is significant, as it established in law the notion that data subjects have rights in relation to data processing, and data processing activities should therefore be carefully controlled. With the introduction of DPA it was no longer appropriate, or indeed lawful, to simply assume that 'doctor knows best' in relation to ongoing processing of patient data to support secondary activities such as research. However, the DPA was by no means the only instigator of attitudinal change operating in the late 1990s. Running parallel to the introduction of the DPA was a particularly influential review conducted by Dame Fiona Caldicott into uses of patient-identifiable information within the NHS across England and Wales. The review (Caldicott, 1997) was commissioned by the chief medical officer for England, in response to

[22] Such as the General Medical Council, Nursing and Midwifery Council and Health and Care Professions Council, which have the authority to strike people off the professional register for serious dereliction of duty

[23] Section 60 of the Health and Social Care Act 2001 (re-enacted by Section 251 of the NHS Act 2006) allows the Secretary of State for Health to make regulations to set aside the common law duty of confidentiality for defined medical purposes. The Regulations that enable this power are the Health Service (Control of Patient Information) Regulations 2002. Frequently, references are made within the research community to 'section 251 support or approval'. This refers to approval given under the authority of the Regulations. See: http://www.hra.nhs.uk/about-the-hra/our-committees/section-251/what-is-section-251/

'increasing concern about the ways in which patient information is used…due to the development of information technology in the service' (p. i). Recommendations made in the report were based on work conducted by the Caldicott committee into common information flows across the NHS. As a result of this work, the review formulated six principles for the control of patient-identifiable information (which have come to be known as the Caldicott principles):

Principle 1: Justify the purposes(s) - *Every proposed use or transfer of patient-identifiable information within or from an organisation should be clearly defined and scrutinised*
Principle 2: Don't use patient-identifiable information unless it is absolutely necessary
Principle 3: Use the minimum necessary patient-identifiable data
Principle 4: Access to patient-identifiable information should be on a strict need-to-know basis
Principle 5: Everyone with access to patient-identifiable information should be aware of their responsibilities
Principle 6: Understand and comply with the law

The Caldicott principles, which complement the DPA's principles (particularly around issues connected to 'data minimisation'), acted as a major signifier to NHS staff that a robust and careful approach to data sharing was what was expected and required. Significantly, the report also recommended that 'a senior person, preferably a health professional, should be nominated in each health organisation to act as a guardian, responsible for safeguarding the confidentiality of patient information.' The subsequent application of this recommendation has been extremely successful, and these Caldicott guardians have continued to be an enshrined element of the broad information governance structure adopted across the NHS.

Through the Caldicott principles, a broad attitudinal shift therefore occurred across the NHS, putting a greater emphasis on the need to justify and carefully control data sharing. The subsequent introduction of the 2002 regulations, which were more specifically connected to controlling the circumstances in which patient identifiable data can be further processed without patient consent, could be seen simply as part of this change. However, direct pressure on the British Government to further tighten controls in relation to the use of medical and health data without the explicit consent of patients came from the publication of two particularly damning reports that prompted the then Secretary of State, Alan Milburn, to acknowledge that patient expectations and clinical research practice had 'grown apart.' These reports, which focused on the Bristol Royal Infirmary and the Royal Liverpool Children's Hospital,[24]were related to the removal, retention and disposal of human tissue and organs from deceased children without next of kin consent. Whilst these inquiries specifically related to abuses of rights in relation to research involving human tissue, and therefore (in legal terms) concerned abuses of the Human Tissue Act, the emotional impact of the reports had broad ramifications in raising public awareness of the ethical and legal considerations surrounding *all* health and medical research practice

[24] For summary and recommendations from the Alder Hey report see: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/250914/0012_i.pdf

(Higgins, 2003). The Inquiry report into the Royal Liverpool Children's Hospital at Alder Hey labeled the *modus operandi* under which the events were allowed to unfold as paternalism in action (Redfern, 2001). This had the broader effect of calling into question whether paternalistic decision-making processes, which failed to adequately take into account the 'reasonable expectations' of the patient and/or their next of kin, was detrimentally endemic as an underpinning mechanism for *all* health and medical research practice within the NHS. In a key press release in 2001, Alan Milburn acknowledged that:

> The NHS can no longer assume that the benefits of science, medicine or research are somehow self-evident, regardless of the wishes of patients or their families. The relationship between patients and the service today has to be based on informed consent. That will require changes in practice and changes in policy (Milburn 2001 in Higgins, 2003, p.9).

Against this backdrop of concern, and with a focus on the importance of consent very much in the foreground of the political and public consciousness, the 2002 Regulations (which were themselves introduced through the Health and Social Care Act, 2001, reenacted by s. 251 of the NHS Act 2006) sought to iterate the exact circumstances in which the common law duty of confidence *could* be set aside to allow the use of confidential information for purposes beyond the direct care of the patient without explicit consent.[25] The regulations made it clear that despite any duty of confidence that may be owed in relation to patient data, it may be possible to establish a legal basis for processing confidential patient information for purposes beyond direct care without explicit consent when this is for public health purposes. Regulations 2, 3 and 5 define the exact purposes for which this is permissible. Regulation 2 specifies cancer registration, whilst Regulation 3 specifies public health purposes when processing is intended to 'diagnose, control, prevent, or recognise trends in communicable diseases and other risks to public health.'[26] Essentially, then, Regulation 3 relates to processing for the purposes of health *protection.* Regulation 5, on the other hand, specifies disclosure for purposes more broadly defined, namely purposes that fall under the remit of health *improvement*[27] (Taylor, 2015, p.352).

There is an independent advisory group with responsibility for advising on the application of the 2002 Regulations. Over the years, changes in the law that have affected the structure of health and social

---

[25] It is important to remember that the regulations only provide relief from the common law duty of confidence, which means that any activity taking place with the support of the regulation must still comply in full with the Data Protection Act. Grace & Taylor (2015) describe the relationship between the DPA and the regulations in the following terms: 'when offering advice on the use of the Regulations, the advisory group must take into account the restrictions and exclusions that are contained within the Regulations and also those that must be read down from the Parent Act…the Parent Act provides that the Regulations cannot make provision for processing 'in a manner inconsistent with any provision made by or under the Data Protection Act 1998'…The expectations of the Data Protection Act 1998 already discussed are thus explicitly imported into the context of secondary uses of confidential patient information.

[26] As noted by Taylor (2015) processing for health *protection* is also covered by further specific regulations, the most recent being The Health Protection (Notification) Regulations 2010 which 'set aside the common law duties by a statutory duty imposed on registered medical practitioners to notify certain specified infectious diseases and contaminations to the proper officer of the local authority via Public Health England (PHE). In addition, diagnostic laboratories are required to provide positive diagnostic test results on specified diseases to PHE' (p.351)

[27] The application of Regulation 5 is commonly referred to as an instances of 'class support' (Grace & Taylor, 2015: 424)

care management and provision have meant that the group has itself been reorganised a number of times during its history. Its first incarnation was in the form of the Patient Information Advisory Group (PIAG). The PIAG was established through section 61 of the Health and Social Care Act, 2001. Following the abolition of the PIAG, The National Information Governance Board for Health and Social Care (NIGB), established under s.157 of the Health and Social Care Act 2008, acquired the functions previously held by the PIAG, and its Ethics and Confidentiality Committee (ECC) had responsibility for the advisory role. Following the reorganisation of the National Health Service in England, outlined in the Health and Social Care Act 2012, these functions were transferred once more to the newly established Health Research Authority's (HRA) Confidentiality Advisory Group (CAG). The HRA was originally established as a Special Health Authority in December 2011, but became a non-departmental public body in January 2015 under the Care Act, 2014. The HRA also absorbed the functions that were previously held by the National Research Ethics Service to fulfil an ongoing remit 'to protect and promote the interests of patients and the public in health research, and to streamline the regulation of research.' Under the Care Act 2014, in response to concerns arising from the care data scheme, the CAG now has an additional statutory role in advising the Health and Social Care Information Centre (HSCIC) on aspects relating to their dissemination function (see more below).[28]

The terms of reference of the Advisory Group have subtly (but in some ways significantly) changed over the years. Originally, when the PIAG was first created, there was a statutory responsibility to consult the Advisory Group before new or amended Regulations were laid before Parliament, but when NIGB was abolished this responsibility transferred to the Care Quality Commission. Although the Advisory Group has always been invited to advise on decision-making under the 2002 Regulations, it is significant that the Care Act,2014, has made it a statutory requirement for the HRA to establish an independent committee for the purposes of giving advice in circumstances where a decision has to be made on whether to permit processing under the 2002 Regulations (Grace and Taylor, 2013).

In essence, the Health Service (Control of Patient Information) Regulations 2002 establish a legal basis for the common law duty of confidence to be set aside so that data can be disclosed for public health purposes without patient consent. However, under the Regulations a differentiation is made between types of processing purposes, with different safeguards and conditions attached. If processing without consent is for health *protection* purposes, under Regulation 3, then there is no need to go through the 'administrative inconvenience and delay' (Taylor, 2015, p.352) of applying for CAG support for that processing. However, when processing without consent is for health *improvement* purposes, under Regulation 5, an application to the CAG is required. Led by a diverse panel of experts and laypersons, CAG's role is to assess the necessity and proportionality of the interference with individual privacy. In doing so, the CAG must take into account the restrictions and exclusions laid out in the Regulations. In particular, where confidential patient information is processed under the Regulations, the researcher must not process that information any more than is necessary to achieve the permitted purposes. In

[28] See http://www.hra.nhs.uk/about-the-hra/our-committees/section-251/

addition, there is an emphasis on meeting the specified purpose in ways that, as much as possible, reduce identifiability within the data, restrict access to the data, and prevent unauthorised access through appropriate technical and organisational measures (which, in fact, neatly echoes the approach enshrined in the Caldicott principles). There is also an enforced requirement to review the necessity of the processing at twelve-month intervals, and to make available on request information on the steps taken to adhere to the regulations (Confidentiality Advisory Group, 2013).

According to the CAG's Principles of Advice, 'two of the most significant safeguards limiting the scope of any permissible interpretation and application of the regulations, concern the issues of 'public interest' and 'practicable alternative''(Health Research Authority, 2013, p.2). Under Section 251 of the NHS Act, the introduction of regulations is only possible if the regulations are considered necessary or expedient in the interests of improving patient care, and/or in the public interest. Accordingly, then, 'it follows that any advice given, in relation to the regulations, must be consistent with one or both of these purposes' (Health Research Authority, 2013, p.3). Therefore, support under the regulations (often referred to as s. 251 support) can only be given to research that is necessary or expedient in the interests of improving patient care, or in the public interest. Furthermore, the regulations may not be used to make provision for the processing of confidential identifiable patient information without consent "for any purpose if it would be reasonably practicable to achieve that purpose otherwise…having regard to the cost of and the technology available for achieving that purpose" (S. 251 (4) NHS Act 2006). What this means is that the regulations cannot offer support to research that uses patient-identifiable information without consent if there is the means to undertake that research in another way, either by gaining consent or by using de-identified data. In relation to assessing whether gaining consent is 'practicable,' due regard will be given to the costs involved and/or the availability of technology for doing so.

In evaluating whether processing is 'in the interests of improving patient care,' the CAG considers 'not only the intended consequences of the processing but also the potential damage to patient care that might follow from a loss of trust in the confidentiality of the information held by providers of healthcare services' (Health Research Authority, 2013, p.3). In this respect, the CAG is 'always mindful of the two public goods identified in previous sections in this review: the public good in a health care service which holds and processes patient information confidentially and the public good in improving that health care service through the use of confidential patient information for medical purposes without patient consent' (Health Research Authority, 2013, p.3). Having considered the potential impact of any particular proposal on each of these public goods, the group will only give support if it considers the balance to be 'in the public interest.' As a basic principle, the more readily acceptance for a proposal across a diversity of 'user groups' can be reasonably anticipated —not only from those who might benefit from the processing, but also from those whose data is to be processed without consent— the clearer the indication that the processing is in the public interest (Health Research Authority, 2013, p.3).

In assessing the availability of a 'practicable alternative' that would negate the necessity of using confidential information without consent, the CAG considers whether it is reasonable to expect the applicant to seek consent or adapt the research to draw only on data in de-identified form. Impossibility can sometimes be argued in retrospective studies including data of deceased patients, or when the patient may be incapacitated and therefore unable to give consent. A large cohort does not automatically mean that gaining consent is considered too difficult. The 'intolerable bias' that might occur if patients opt out when consent is sought is only seriously considered when the study involves small numbers. There are rare cases when the CAG considers that it may not be *reasonable* to seek consent even though it is perfectly *practicable* to do so. The example given in the CAG guidance is 'studies that have sought to investigate the records of the deceased and have had the opportunity to consent the subjects of those records while still living but where such opportunity could only have been realised by consenting a much larger cohort including those that had no reason to suspect that the treatment episode in question would result in their death' (Health Research Authority, 2013, p.12).

Alongside these considerations around 'public interest' and 'practicable alternative,' CAG's Principles of Advice make it clear that it is particularly important to demonstrate that it is possible to provide access to identifiable information without individual consent and still appropriately protect an individual's right to respect for their private life. Furthermore, 'a standard requirement imposed by the group is that *known* individual dissent to disclosure should be respected in all but the most exceptional circumstances' (Taylor, 2015, p.352). On this point, interpretation of the regulations on an individual's right to object actually goes beyond those upheld by the Data Protection Act 1998. Whereas the DPA specifies a right to object where processing will cause 'unwarranted substantial distress or damage,' the CAG maintains that the right does not need to be qualified in such specific terms and that all 'reasonable objections' will be upheld. It is important to note also that the advice given by the CAG (and the final decisions taken by the HRA and the Secretary of State based on that advice) is permissive but not absolute. Therefore an individual healthcare professional can decide not to provide the confidential information even when a legal basis has been established to permit them to do so (CAG, 2013, p.6). It is also vital to draw out from this that the emphasis within the CAG on assessing the 'public interest' and 'practicable alternatives' when considering whether the duty of confidence can be lifted is underpinned by the foundational notion that the obtaining of consent for breaking that duty of confidence should, in all but the most exceptional circumstances, be the default principle for enabling health and medical research that uses patient-identifiable information.

It is therefore a 'consent as the default' approach that has emerged as the accepted path to data sharing in NHS England, and this is echoed in amendments to the NHS constitution (2015). These amendments are indicative of what can be interpreted as an ongoing commitment to shift the ethos away from paternalism towards closer involvement for patients in decisions that affect them. The additions to the NHS constitution concerning 'consent, respect and confidentiality' read: 'You have the right to be informed about how your information is used' and 'You have a right to request that your confidential data is not used beyond your own care and treatment and to have your objections

considered, and where your wishes cannot be followed to be told the reasons including the legal basis' (NHS Constitution as quoted in Grace and Taylor, 2014, p.416). Taking all of this into account, Grace and Taylor (2014) sum up the English approach, and the legal premise on which it is based, in the following terms:

> The common law duty of confidence, the Human Rights Act 1998, the Data Protection Act 1998, and the Health Service (Control of Patient Information) Regulations 2002/1438, collectively establish a responsibility to consult a patient before confidential information is used for secondary purposes. There is a responsibility to provide information (where practicable) about the intended purposes of processing and also a responsibility to take conscientious account of any reasonable objection to disclosure. [Nevertheless]…there are at least two circumstances in which any responsibility to consult, at least inherent to the common law, might be limited. These relate to uses of de-identified data and uses of identifiable data 'in the public interest' (p. 426).

Having considered in the previous section the ambiguity in the DPA surrounding anonymous data, it is worth elucidating here on the inapplicability of the common law duty of confidence to anonymised data ,as explored by Grace and Taylor (2014), who analyse *R v Department of Health* ex parte *Source Informatics* as a means of establishing where the common law stands on this issue. The decision in this case (which has set an interpretive precedent) is described by Grace and Taylor (2014) in the following terms:

> The Court of Appeal…decided that a pharmacist could make use of de-identified [anonymised] patient information without breaching the duty of confidence that they held in relation to the data in identifiable form. Pharmacists were selling de-identified prescription information to companies interested in learning more about the prescribing habits of doctors. The Department of Health objected to this practice and claimed that disclosure amounted to a breach of confidence. The Court of Appeal decided that where data are de-identified by a health professional (who lawfully holds identifiable data), then disclosure of the de-identified data will not threaten the privacy of the individual. The process of de-identification needs to be sufficiently robust, and/or the subsequent information context sufficiently controlled, to prevent the re-identification of the patients, but—if anonymity is effectively assured—then the common law duty of confidence is lifted. The impact of *Source Informatics* has thus been to establish that, from a common law perspective, a patient does not need to be informed of any intention to de-identify data and to use it for further purposes. Nor would a patient appear to have any common law right to object to either the de-identification itself or the subsequent use of data in de-identified form (p. 426-427).

The connotations of this will be expanded on below in relation to an exploration of mandated data collections and further disseminations undertaken by the Health and Social Care Information Centre (HSCIC).

Mandated disclosures and the opt-out

Grace and Taylor (2013) suggest that the tenets of the position adopted in England, in which there is recognition of a general responsibility to consult patients on secondary uses of identifiable confidential data, has been put under considerable risk of being diluted by the exercise of powers introduced through the Health and Social Care Act 2012, which introduced a 'new legal framework for the flow of confidential patient information within the NHS' (Grace and Taylor, 2013, p.429). This new flow revolves around the newly created Health and Social Care Information Centre (HSCIC), which has been endowed through the Act with a number of specific powers. Grace and Taylor (2013) describe these powers as 'complex to the point of being Byzantine.' To summarise in greatly simplified form, the HSCIC was given power, under section 259, to acquire confidential patient data (and other information) from health and social care bodies across England. It has the power to require information where it is considered 'necessary or expedient' for the purposes of fulfilling any of its statutory functions, and in the case of confidential patient information, where the information is requested from either:

> (a) the Secretary of State or the Commissioning Board (section 254), or
> (b) a 'Principal Body'[29] which considers the information to be necessary or expedient to have in relation to discharge of their duties in relation to provision of health services or adult social care in England (sections 255 and 256) or
> (c) any person, who has requested that the Information Centre collect confidential information, if that person has (independent) authority to require disclosure of that information (section 256 (2) (b))
> (as laid out in Grace and Taylor, 2013, p. 430-431)

As elucidated by Grace and Taylor (2013), if a request made to the Information Centre to process confidential information falls outside of these conditions, then the Centre might still lawfully request disclosure of confidential information, but only if the request relates to 'information which may otherwise be lawfully disclosed to the Information Centre or to [the party requesting that the Centre collect the information] by the person holding it.' There is a general responsibility upon the Information Centre to publish all information collected through mandatory disclosure with the obvious exception of identifiable data. This means that 'whilst the Information Centre is empowered to collect, and indeed to require the disclosure of, confidential patient information it will not publish that

[29] 'Principal Body' means Monitor, the Care Quality Commission, the National Institute for Health and Care Excellence, and such other persons as may be prescribed in regulations (s 255(9))

information in identifiable form' (Grace and Taylor, 2013, p.431). In addition to the restrictions on the publication of identifiable data, 'there are further restrictions upon any kind of dissemination of identifiable data (other than by way of publication)' (Grace and Taylor, 2013, p.431). However, these restrictions on disclosure are 'lifted in a number of circumstances including where: the disclosure is made to any person in circumstances where it is necessary or expedient for the person to have the information for the purposes of exercising functions of that person conferred under or by virtue of any provision of the 2012 Act or any other Act' (Grace and Taylor, 2013, p. 431). This means, for example, that researchers with s. 252 support can request access to the identifiable information that is held by the HSCIC as a result of mandated disclosure.[30]

As the paragraphs above illustrate, the rules around the mandated disclosure of confidential patient data from health professionals to the HSCIC, and its further dissemination by the HSCIC, are indeed complex. In a detailed examination of the interaction between the Health and Social Care Act 2012, and the existing legal framework (provided by the 2002 Regulations, the Data Protection Act 1998 and the Human Rights Act 1998), Grace and Taylor (2013) have been concerned to establish the effect that mandated disclosure of this kind has on what patients can expect in terms of notification and consent in relation to the processing of their data. Through a detailed legal discussion, the authors argue that the common law duty of confidence does not apply to mandated disclosures made to the HSCIC, because the 2012 Act establishes a statutory gateway for the disclosure of confidential patient information. As such, the 2012 Act also removes any need to satisfy the requirements of the Health Service (Control of Patient Information) Regulations 2002. Therefore, the position established under the regulations that any reasonable patient objection should be respected is also no longer applicable. In relation to fair processing under the DPA, the Information Centre itself is released from any obligation to provide information about its processing —although the data controller from whom the data is obtained must make an effort to notify. The onus here is therefore not on the Information Centre, and this effectively reduces the opportunities that data subjects have to know about and challenge the processing. Therefore, in relation to mandated disclosures, even the more limited right given to data subjects to object to processing under the DPA is effectively curtailed (Grace and Taylor, 2013, p. 433-435).

The establishment of the HSCIC in 2013 can be connected to the increasing Government rhetoric, since 2011, around open data and the re-use of public sector information. Within NHS England itself, the pendulum swing between a cautious 'privacy-centred' approach to information sharing, epitomised in the Caldicott Principles and taken seriously following the Alder Hey scandal, was beginning to come under pressure, and the second Caldicott review of 2013 (Caldicott, 2013) sought to advocate for a balance between openness and data access restrictions which took account of the Government's increased rhetoric concerning the benefits of sharing. This balance was enshrined in the introduction of a new seventh Caldicott principle:

[30] For details on the process researchers go through to gain access to data held by HSCIC see: http://www.hscic.gov.uk/DARS

> Principle 7: The duty to share information can be as important as the duty to protect patient confidentiality

The second Caldicott report directly refers to the Government's White Paper on Open Data, in particular its 'twin track' approach in which 'data that can be published should be published' when coupled with a commitment to 'safeguard people's data from misuse and rigorously protect the public's right to privacy.' The implication in Caldicott II is that Information Governance in the NHS was failing to allow this balance to be effectively achieved:

> Over recent years, there has been a growing perception that information governance is often cited as a reason not to share information, even when this is in the best interests of the patient or service user…Paradoxically, criticism that the bureaucracy of information governance is standing in the way of sensible information sharing among professionals has gone hand in hand with equally vociferous criticism that the system is not doing enough to combat laxity in the protection of confidential data and information. There is a perception that too much information is being disclosed inadvertently as well as too little being shared deliberately. Furthermore there is uncertainty among many patients and users of services, who are unaware of how personal confidential data about them is collected and shared. If people do not know how their data will be used, it is wrong to assume they have given their implied consent to sharing (Caldicott 2013, p.25-26).

While the report sought to counteract a 'culture of anxiety' around information sharing that, it argued, had permeated the NHS, it equally sought to reinforce the need for rigour and proper handling of confidential information. Giving the public adequate information on the different ways in which their data may be used is highlighted as one of the fundamental cruxes of achieving the desired balance in information sharing, where the approach is both enabling and protective and in line with patient expectations.

However, the position on consultation and consent in relation to disclosure of confidential information under the statutory powers afforded to the HSCIC has continued to be an area of deep controversy. The publication of the HSCIC's Code of Practice on Confidential Information (mandated under the Health and Social Care Act 2012 and published in December 2014) has helped to clarify the ways in which the balance enshrined in Caldicott Principle 7, between sharing information and protecting patient confidentiality, is achieved. This code draws heavily on the full range of Caldicott Principles to ensure that collection, analysis and dissemination of confidential patient information for secondary purposes is legally sound, proportionate, and done in recognition of objections. In relation to the latter, the system that has emerged across England and Wales to manage patient consultation and consent in relation to mandated disclosures is an 'opt-out' model (Taylor and Taylor, 2014, p 3). This model has emerged as the means of bridging the gap in patient rights that develops in the context of the mandated disclosure of confidential patient information from health care professionals to the HSCIC as

introduced by the Health and Social Care Act, 2012. To this end, even though there is essentially no *legal requirement* for patient consent or opportunity for patient objection within mandated disclosure, the development of an 'opt-out' system goes some way to ensure that patient objection is respected in practice in the spirit of the 2002 regulations. The development of a national 'opt-out' model was, in fact, first introduced as part of the ill-fated care.data scheme, following effective lobbying by privacy campaigners. It currently consists of what has come to be known as Type-1 and Type-2 opt-outs, through which individuals can raise an objection to the use of their identifiable data for anything beyond their direct care. These are described on the HSCIC website in the following terms:

> **Type 1 opt-outs** If you do not want information that identifies you to be shared outside your GP practice, for purposes beyond your direct care you can register a type 1 opt-out with your GP practice. This prevents your personal confidential information from being used other than in particular circumstances required by law, such as a public health emergency like an outbreak of a pandemic disease.
>
> **Type 2 opt-outs** The HSCIC collects information from a range of places where people receive care, such as hospitals and community services. If you do not want your personal confidential information to be shared outside of the HSCIC, for purposes other than for your direct care you can register a type 2 opt-out with your GP practice (HSCIC, 2016).

Controversy surrounding these opt-outs has been fierce. It has been strongly argued that the wording of the opt-outs lack clarity, making it difficult for the public to know exactly what they are opting out of. There have also been substantial problems relating to the implementation of the type 2 opt-out. It came to light in 2015 that 700,000 'type 2' objections had, at that point, been registered under the opt-out scheme. These 'type 2' objections were registered by patients with GPs according to the protocol, but relate broadly to objections to data sharing undertaken by the HSCIC and the range of data sets that it collects under mandated disclosures. It emerged, amidst a public outcry, that the HSCIC could not initially extract the type 2 objections from the GP data without it also having adverse consequences for the direct care of the patient who, following the action on the objection, would not be approached for direct care services such as 'e-prescribing, bowel screening, e-referrals or e-pathology reporting.' Therefore, applying that objection was on hold as it was deemed unacceptable tht applying the objection would negatively impact on a patient's routine care. Consequently, confidential patient data was being shared by the HSCIC on an on-going basis against patient wishes due to a lack of means for respecting individual dissents *without* further consequences for direct care.[31] Following a direction from the Secretary of State, the HSCIC has now been mandated to not only extract but also apply type 2 opt-outs from 29 April 2016. Currently, then, the HSCIC collects information on type 2 opt-outs from GP practices on a monthly basis. These records are used to check against sets of data made available by HSCIC when it is intended that they will be released to another organisation. All personal confidential information is removed before that data set is made

[31] A description of this issue was given by the HSCIC as written evidence to the parliamentary health committee who were tasked with investigating the problem: http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/health-committee/handling-of-nhs-patient-data/written/17671.html

available. This brings the practice of the HSCIC broadly in line with the thrust of the 2002 regulations, where known dissents to the use of confidential identifiable patient data are respected. However, there are a range of extended scenarios laid out in the direction in which type 2 objections cannot currently be upheld, including where there are 'complex technical barriers' relating to particular datasets 'that make it very difficult to apply opt outs' (HSCIC, 2016b). The direction places an onus on the HSCIC to resolve these technical barriers, yet at the current time the system has yet to work in a way that can ensure the opt-outs are upheld in all reasonable circumstances. At the time of writing, 1 in 45 patients (2.2%) have registered type 2 objections (HSCIC, 2016c), which reportedly amounts to 1.2 million patients (Digital Health, 2016b), and there are currently 1, 461, 877 instances of type 1 opt-out codes occurring within GP records, preventing these records from being shared outside the practice for purposes other than direct care (HSCIC, 2016c). These type 1 objections will become applicable if and when the care.data scheme is reignited. In recognition of the importance of having a clear, understandable and easily applicable consent model, in September 2015 the Health Secretary Jeremy Hunt commissioned Dame Fiona Caldicott to conduct a review of the current 'opt-out' process. That review is complete, and it has been reported that Caldicott has devised 'a single, simple model for patients to use to determine the use of their data beyond their own personal care.' However, the publication of the review was delayed until after the UK's EU referendum on the 23rd of June 2016 (Digital Health, 2016a).

To be in any way effective, the anticipated Caldicott review of the 'opt-out' system will have to be accompanied by a consideration of the consent model upon which the 'opt-out' is based. Policy makers are aware that under any 'opt-out' scheme, most people accept the offered defaults, therefore the intended uses of the data are relatively unaffected by giving individuals the choice to dissent. Due to the limited number of choices, the 'opt-out' scheme currently adopted in England is also modelled on 'broad opt-out' as opposed to 'narrow opt-out,' as it is impossible under the current scheme to register a nuanced decision based on particular types of uses. An entirely different approach would be to enable some form of 'dynamic consent.' Dynamic models are being explored by international 'big data' initiatives such as those of the International Medical Informatics Association (IMIA) and the Global Alliance for Genomics and Health (Nuffield, 2014, p.75). However, consent models in a health context (dynamic or otherwise) have the further complication that, in the case of certain types of health data, while consent may be freely obtained from one individual, 'the data concerned may also significantly relate to the privacy interests of other individuals, including those not yet born' (Nuffield 2014, 76). Thus, for example, 'a DNA sequence may reveal probabilistic information or, in rare cases, disease traits or other characteristics in biological relatives, not merely about the person from whom it was obtained. In such cases, there is a divergence between the scope of autonomy (who gives or withholds permission for data access) and the scope of privacy (those whose privacy interests are affected by this permission) that current consent and data protection mechanisms find difficult to manage' (Nuffield 2014, p.76). The forthcoming Caldicott report has the difficult task of navigating these consent-related complexities.

Returning the focus to the HSCIC, it is fair to say that its remit and practices have come under heavy scrutiny. During an evidence session of the Parliamentary Health Select Committee in 2014, Labour

MP Valerie Vaz labelled the HSCIC as a 'shambles' in relation to its handling of patient data (O'Dowd, 2014). The issues surrounding the management of opt-outs is one small part of a wider debate on the ways and means in which the public is consulted, kept informed, and has the right to limit secondary uses of data beyond direct care. Questions from the Parliamentary Committee Chairman, Stephen Dorrell, focused on seeking to clarify in what ways the HSCIC was different from its predecessor, the NHS Information Centre, which had been rumoured to have sold de-identified data to the insurance industry, allowing companies to set prices over cover (O'Dowd, 2014, p.348; Pollock and Roderick, 2014, p.1524). Dorrell's questions highlight the fact that whilst the dissemination of *de-identified* health data to commercial companies may not contravene the common law duty of confidence or fall under the jurisdiction of the Data Protection Act, it may still be decidedly unethical. This is indicative of the complexity of the path which the HSCIC has to tread, and how its practices require a broader consideration of social legitimacy than a strict interpretation of the law alone allows for. Although the non-executive Chair of the HSCIC has argued publicly that the approach taken by the HSCIC is more robust than that taken by its predecessor, the complexity of disentangling legitimate commercial secondary uses of confidential data undertaken for health improvement from profit-making and market-orientated activities not in the public's best interest is an area of on-going public debate and review, particularly given that 'commercial contracts are now at the heart of NHS commissioning' (Pollock and Roderick, 2104, p.1524).

Much of the consternation surrounding the HSCIC has been connected to the mismanagement of the care.data scheme. This scheme, led by NHS England, sought to direct the HSCIC to collect identifiable GP data including NHS number, date of birth, postcode and gender alongside a long list of 'read codes' that record clinical data of various types (Taylor 2014, p.2). This collection was to be mandated as a means of expanding the HSCIC's existing Hospital Episodes Statistics (HES) dataset by linking it to the GP data and disease registry data. Such linkage is successfully in place in Scotland and Wales where the NHS is run separately. Researchers describe primary care data as the 'crown jewels,' as GP records have been 'electronic for decades and have been coded for decades and have wide population coverage of nearly 100%' (Callaway 2014, unpaginated). As NHS England is the world's largest public health system, caring for over 53 million people, the linked dataset would be far greater than any equivalent held in any other country, including Denmark and Sweden, which have had central health databases for years (Callaway, 2014). However, the scheme was put on hold following a heavily criticised public dissemination programme in which information sheets sent out to the public failed to mention the scheme by name, failed to provide adequate information on the foreseeable uses of the data and the potential risks, and failed to adequately explain the process of opt-out. This failure in oversight and transparency was confirmed by a BBC poll which found that only 29% of the English population could recall receiving information on the scheme (Taylor, 2014, p.2). This was mainly due to the fact that the information was sent out to households rather than individuals, and its appearance as junk mail meant that the messages it contained failed to reach a significant proportion of the population (Carter et al, 2016). In response to these concerns, four clinical commissioning groups were chosen to pilot the scheme, but this too was put on hold whilst the Health Secretary invited Dame Caldicott to review the opt-out process (Digital Health, 2016a, Carter et al, 2016). Carter et al (2016) argue that the

care.data scheme failed precisely because it did not attract an adequate 'social licence to operate.' According to their definitions a 'social licence' is given by society to an organisation, corporation or scheme when its conduct and activities meet the expectations of society 'beyond the requirements of formal regulation' (p. 405). Carter et al (2016) argue that 'warrants of trust' between the public and the scheme were never adequately established (p. 406-407). The rupture in traditional public expectations of the duty of confidence between GPs and their patients was not anticipated or addressed (p. 407), and the argument that care.data was in the public interest was inadequately expressed and unconvincingly made throughout the course of the planned roll-out (p. 407). The result was deep mistrust in the minds of patients and the wider public on the motivation behind the scheme and its potential uses, particularly in relation to commercial re-use. Led by the lobbying group MedConfidential, the rhetoric surrounding the risks that the scheme posed to privacy, and the untrustworthiness of the system to mitigate those risks, became powerful and ultimately convincing. On a legislative level, the Government was forced to enact a number of amendments through the Health Act 2014 to address the concerns raised by care.data. This included ensuring that the CAG's advisory role to the HSCIC was given a statutory footing. The Act also ensures that the HSCIC is placed under a general duty to "respect and promote the privacy of recipients of health services and of adult social care in England" (Taylor 2014, p.5). Further to this, additional restrictions on dissemination of information by the HSCIC, introduced by the Act, sought to make it clear that data collected by HSCIC should only be further disseminated in the provision of care, or the promotion of health (Taylor 2014, p.5).

One of the lessons drawn from the failures inherent in the care.data scheme is therefore that compliance with the law does *not* necessarily mean that any given instance of the collection and use of health data beyond the purposes of direct care is morally, ethically or socially acceptable. This is picked up on in a Nuffield report entitled *The collection, linking and use of data in biomedical research and health care: Ethical Issues* in which it is argued that strong information governance protocols and the law must be supplemented by the development of a 'morally reasonable set of expectations' for health data sharing which should be constructed through *participation* by those with morally relevant interests (p.84). The report argues that the involvement of interested parties (including patients and the wider public) would allow for an 'account to be given of how these interests are respected in decision making, thus fostering broader patient and public trust and cooperation' (Nuffield, 2014, p.84). Such a participative approach may therefore 'help to ensure that data initiatives remain more closely in touch with changing social norms' (Nuffield 2014, p.84). Arguably, a participative approach may also help to build more legitimacy for, and social acceptance of, the assertion that public-sector health data can be managed in a way that both protects confidentiality and privacy *and* releases the potential in the data to be used for health improvement.

Taylor (2014) puts forward a similar argument in his examination of the aftermath of the care.data scheme. He suggests that care.data fell into the trap of insisting that the data sharing it was promoting was in the 'public interest' without allowing that assertion to emerge from robust public engagement with the issues. Furthermore, he highlights the fact that the debate around care.data has 'at times

reflected the tendency, seen in (data protection) law, to consider the public interest as something opposed to the protection of an individual's fundamental rights and freedoms' (Taylor 2014, p.8). He argues that:

> There will undoubtedly be times when one must give way to the other but to present them as necessarily opposed can place too great an emphasis upon their differences. It invites judgment that when the public interest justifies it, then individual privacy may be sacrificed. This is an unhelpful way of framing important questions. It does not sufficiently recognise the public interest in privacy protection or the ways that using data for public interest purposes, where it is consistent with people's expectations and preferences, can respect privacy. In short, there are many ways and times that we can seek to improve protection of both privacy and the public interest in access *simultaneously* without having to sacrifice one for the sake of the other (Taylor 2014, p.8, emphasis added)

In keeping with what has been argued throughout this review, it does seems that the way forward through what is often framed as a dichotomy of opposing tensions is to ensure that privacy protection itself is framed, not just in terms of individual rights and freedoms, but as a collective public good alongside other public goods upheld in law, including the promotion of health. Only then will it be possible to engage the public in fruitful debate around the ways in which these equally reinforcing public goods can be upheld, protected and advanced. One of the main difficulties in beginning that debate, retrospectively, following care.data is that public confidence in the trustworthiness and legitimacy of such schemes has been deeply (and perhaps irrevocably) shaken. Gaining public trust and support for data linking initiatives involving personal health data will therefore be a much more difficult challenge going forward.

Achieving a balance between open data and data protection: what are the implications for information governance?[32]

The focus here will be on exploring the mechanisms of information governance that should frame secondary uses of medical records and patient data. These mechanisms must ultimately work to achieve a balance between the mutually reinforcing public goods of protecting privacy *and* enabling use that is in the public interest. While individuals have privacy interests in the use of data, they also share a collective interest in the wider use of data for health research. This 'broader public interest, which consists in securing objectives that are valued by society, may come into conflict with individual privacy' (Nuffield, 2014, p.158). However, the relationship between privacy and public interest in data is not simply one of opposition. The two are 'mutually implicated in each other': there are private interests in the achievement of common goals and a public interest in the protection of privacy, which encourage cooperation. Consequently, the issue is not finding a 'balance' between individual privacy interests and the collective public interest in a given data initiative, but resolving a 'double articulation' between the individual interest in protecting privacy and promoting the public good, and the public interest in protecting privacy and promoting the public good. 'We all have interests on both sides, and navigating among these different relationships with other individuals, professionals and institutions requires a subtle negotiation of many different norms of information access and disclosure, of when and how they may be modified, and where hard and fast limits on data sharing need to be applied'.(Nuffield, 2014, p.158). Adequate information governance in data initiatives involving medical records and patient data needs to comply with the law around privacy, data protection and confidentiality but must also go beyond the law. In other words, complying with the legal framework does not necessarily equate to complete ethical and moral freedom to act within it (Nuffield, 2014, p.112). This is perfectly exemplified by mandated disclosures by the HSCIC. Whilst it may be within the law for the HSCIC to collect patient data without offering patients notification or opportunity to dissent, it is not necessarily ethical or morally reasonable for them to do so. Mechanisms for information governancetherefore should work to bridge the gap between legal, ethical, and moral responsibilities. Adequate information governance in relation to data initiatives that draw on medical records and patient data therefore hinge on developing mechanisms that are based on:

- An agreed set of reasonable expectations about how data will be available for use in a data initiative. These 'reasonable expectations' must give proper attention to the different interests at stake and must be based on a robust reconciliation of the norms of privacy and disclosure among those who participate in the initiative (e.g. patients, the broader public, researchers and other interested third parties, health professionals, and data providers)
- Clarity and transparency around the ways in which individual freedoms are respected, for example, the freedom to modify these norms by consent

[32] The direction, content and thrust of the argument presented here has been derived from Nuffield (2014)

- Agreement on the form of governance that will give acceptable assurance that the expectations will be met
- Agreement on who is accountable for what in meeting the agreed 'reasonable expectations' of use with additional agreement around the mechanisms that need to be developed to hold the system and its participants to account

(Drawn from Nuffield, 2014)

This literature review has pulled together a variety of instances in which Information Governance procedures across NHS England are failing on all of the counts outlined above. In the care.data fiasco, NHS England pushed ahead with the planned mandated disclosures of GP data to the HSCIC in the absence of any robust debate with patients, the public, or data controllers within the NHS. An agreed set of 'reasonable expectations' was never established, and the result has been a devastating loss of confidence in the scheme, not just from patients, but also from GPs who were against mandated disclosure on the basis that it would damage the nature of the trusting relationships pre-established with patients. In relation to the release of identifiable patient records by NHS Royal Free Trust to Google, again there was no process of establishing what the public and patients' 'reasonable expectations' might have been be on this issue before the data was released. In relation to how the HSCIC provides researchers with access to its semi-open data via its data advisory service, there has been no engagement with the research community or patient representatives to establish common 'reasonable expectations' around the process. As a result, none of the interested stakeholders -- patient (and wider public), health professional, or third party requester (which includes researchers)-- are being served particularly well.

Secondary uses of medical records: Scottish approaches to Information Governance

In many ways the Scottish approach to the question of the reuse of data from healthcare has been the complete reverse of the English experience. Rather than forcing the public and other interested parties to engage *after* core decisions on infrastructure, mechanisms and programmes have been taken, the Scottish authorities *began* with public engagement to determine the public acceptability of the concept of re-use. The informatics structure was developed following engagement, hand-in-hand with internal governance mechanisms, which had the net effect of building public trust rather than undermining it. As a direct consequence of this, the information governance system which frames the extraction and use of patient data for secondary purposes has been relatively uncontroversial and appears to be working well on behalf of all the interested parties. It must be noted here that the Scottish health service is much smaller than the English equivalent, and is structurally and operationally different. Nevertheless, there are important lessons that can be drawn from the Scottish approach.

The Scottish Informatics Programme (SHIP – formerly the Scottish Health Informatics Programme) was initiated to develop a research platform to support more systematic collection, governance and research use of data derived from patient records, and to establish a research arm within the Information Services Division (ISD) of NHS National Services Scotland to support the efficient

functioning of public services more generally. The SHIP provides for linkage not only of primary and secondary care data (collected routinely as part of clinical care) but also data that have been gathered in cohort studies and other forms of publicly-held administrative data (Nuffield 2014, p.115).

Unlike in England, where the HSCIC (under statutory gateway) extracts data from local sources to a centralised data bank, in Scotland datasets are held in distributed collections, each of which is overseen by a data custodian who must agree to the release of data for the specific purposes of the research. Also, unlike in England, Scotland (like Wales) applies precautionary de-identification measures by splitting and encrypting demographic and clinical data prior to linking. Data from different sources (e.g. primary care, administration, etc.) are linked centrally and can be accessed in a safe haven on a temporary basis while a specific analysis is undertaken,and are destroyed as soon as practicable after the results have been obtained. It is intended that in the near future, one of the sources of data will be primary care records. In the Scottish system, GPs will exercise more control over data extractions than has been proposed under NHS England's care.data programme. The Scottish infrastructure and governance mechanisms are arguably more complex than the HSCIC approach, but, crucially, this complexity can be justified because it has come out of robust public engagement from which a case has been made to 'balance efficiency with public acceptability' (Nuffield2014, p. 116). The public have been supportive of the system *because* it has remained distributed and GPs have control over the data extractions. The system therefore relies on the trusting relationship already established between patients and their GPs. Significantly, then, GPs in Scotland have not had to fight, as have English GPs, to assert their powers and responsibilities as data controllers. The SHIP approach works on the broad presumption that both public and health professionalsin the main expect health data to be used for socially beneficial research, an assumption that can be backed up given the robust public engagement that has taken place (Nuffield 2014, p.116).

The information governance arrangements for the data services offered to researchers are built around the concept of 'proportionate governance,' which denotes an approach in which the balance of risks and benefits, and the appropriateness of means to ends, are central. At the heart of this approach is risk assessment. The SHIP continuously aims to improve existing data sharing approaches by 'including assessment of the relative merits of different governance mechanisms, the selection of appropriate governance pathways, and a choice of different governance tools appropriate to any given research application' (Nuffield 2014, p.117). It does not, for example, place a pre-reliance on any particular combination of anonymisation, consent or authorisation. Rather than applying a 'one-size-fits-all' solution, risk assessment enables flexibility in determining what tools and pathways are appropriate in a particular case. Risk assessment therefore has two functions in SHIP: 'one with respect to data protection and the risk of individual identification, and another with respect to authorisation of research in the public interest' (Nuffield2014, p. 117).

Under SHIP, a system of 'authorisation' for research operates to weigh up the balance between privacy and public interest in any given request for access and use. This moves away from the simple principle of minimising risk to the 'principle of optimising the balance and distribution of risks and potential

benefits, and therefore begins to be able to deal with the essentially political question of how this optimum is determined (and by whom, and in whose interest)' (Nuffield 2014, p.117-118). Answering this question requires those with interests at stake to reason together regarding their resolution. Therefore, the question of the 'balance of risks' is partially resolved through a commitment to public engagement as an input to governance. This commitment to public engagement has been taken on by the Farr Institute, Scotland, which raises awareness of medical research and its benefits and allows two-way communication between professional, academic and public/patient participants. As well as ad hoc activities, part of this commitment involves supporting a constituted Public Panel (currently with 20 members drawn from a cross-section of the Scottish public) that meets twice a year to review SHIP authorisation decisions and protocols (Nuffield 2014, p.118-119).

Information Governance as a means of moving beyond the law

The Scottish and English approaches to governance of the secondary reuse of patient data are vastly different in a number of ways, in relation to centralisation and distribution of resources: how data are disclosed and accessed; the range of acceptable users and purposes; and how control is exercised (Nuffield 2014, p.105). The HSCIC is centralised whereas the SHIP takes a distributed approach. TheHSCIC allows direct access to data by commercial companies whereas the SHIP is more guarded. The HSCIC and the SHIP also have different authorisation procedures in which individual preferences and values figure, in the case of the HSCIC through opt-out and in the SHIP through 'a more constitutive participation' (Nuffield 2014, p.105). Though there are lessons to be shared across these various experiences, perhaps the most significant relate to how information governance surrounding patient data must be infused with the principles of stakeholder participation and social engagement in order to be effective. Case studies drawn into this review from NHS England and the HSCIC strongly indicate that there are serious consequences for public trust and for the viability of data initiatives if steps are not first taken to identify the reasonable expectations of the interested parties involved.

It must be remembered that the ethical issues at stake are not merely to do with the privacy of individuals and the risk of disclosure. Significantly, and not touched on in this review,there are also critical social choices about the terms on which data are used that have moral consequences, both because they determine how specific individuals might be treated, and because they may have a broader social impact in informing political decisions. When data about groups is used, this can lead to discrimination and sigmatisation if the ethical consequences are not well thought through. The challenge around the 'balance of risks' must therefore also take these broader ethical issues into account so that data is used responsibly to promote the public interest, in a way that best reconciles the relevant interests of individuals and groups' concerns, in keeping with their fundamental rights (Nuffield 2014, p.153).

An ethical approach also 'recognises that that job of reasoning should not cease once the threshold of acceptability is passed but should continue throughout the process of establishing and governing a data initiative, and permeate it at every point' (Nuffield2014, p.154). Opportunities for ethical reflection

should therefore be built into data initiatives so that such reflection can assume a constructive role: ‘rather than that of the external conscience poised to say ‘no’ to certain practices that step over a notional line of acceptability, the recognition that there are ethical arguments on both sides of any question about data use allows them to be harnessed in the search of good and better solutions, not merely the delineation of acceptable ones’ (Nuffield 2014, p.154).

Following the principle of participation by those with relevant interests in a ‘deliberative procedure’ can optimise the relationship between public and private interests,. because ‘it allows values and interests to be transformed and reconciled through dynamic interaction’ (Nuffield 2014, p.155). This is in contrast to approaches that seek to demonstrate a pre-assumed public good in research, where the intention is simply to convince patient/public stakeholders to agree to a pre-determined goal and approach. Unlike top-down, one-way communication, participation is an embedded two-way process in which the ‘balance of risks’ can be co-constructed by those involved. In enabling stakeholders to engage in constructing the conditions of the future that affect them, it is therefore much more likely to produce outcomes that secure their commitment and build trust (Nuffield 2014, p.155).

The principle of accounting for decisions is a necessary complement to the principle of participation, ‘since not all interests can be represented through participation and not all interests may be satisfied with any outcome’ (Nuffield 2014, p.155). This ensures not only that a decision can be ‘accounted for,’ but also that there is an opportunity to challenge and even to re-evaluate the decisions, through formal structures, which may be regulation or appeal to a legitimate authority, and through continuing debate and re-evalaution. The principle ‘recognises the necessarily provisional nature of decisions about data management and governance, since the horizon of possibilities – and the values and interests invested in them – are constantly changing as the social, political, technological and information environments evolve’ (Nuffield 2014, p.155).

Taken together, the principles of participation and accounting offer the best chance of producing, for any particular data initiative, a morally reasonable set of expectations capable of being satisfied in practice. Establishing these expectations must be the first step in developingthe information governance that frames any data initiative. These expectations must be collectively established between stakeholders including how respect for diverse values and interests will be shown and how moral conduct of others will be assured, while at the same time resolving the ‘double articulation’ of public and private interests through a process of collective moral reasoning and establishment of the ‘balance of risks.’. In conclusion, then, in seeking to identify applicable norms, mere compliance with the law is inadequate to ensure that data use is ethical and morally reasonable. This is because ‘law both stands in a broadly derivative relationship with respect to morality and because it provides only a minimal framework for action rather than full determination for moral action’ (Nuffield 2014, p.155). It is therefore vital to build information governance frameworks that go beyond the law, based on an identification of reasonable expectations of privacy and data use held by all interested parties, which can then determine the moral thresholds for what is acceptable in a given use context.

Bibliography

Academy of Medical Sciences, 2011. *A new pathway for the regulation and governance of health research*, London.

Academy of Medical Sciences, 2006. *Personal data for public good: using health information in medical research*, London. Available at: http://www.acmedsci.ac.uk/viewFile/publicationDownloads/Personal.pdf [Accessed June 6, 2016].

Al-Shahi, R. & Warlow, C., 2000. Using Patient-Identifiable Data For Observational Research And Audit: Overprotection Could Damage The Public Interest. *BMJ: British Medical Journal*, 321(7268), pp.1031–1032. Available at: http://www.jstor.org/stable/25226020 [Accessed June 6, 2016].

Anderson, C.L. & Agarwal, R., 2011. The Digitization of Healthcare: Boundary Risks, Emotion, and Consumer Willingness to Disclose Personal Health Information. *Information Systems Research*, 22(3), pp.469–490. Available at: http://pubsonline.informs.org/doi/abs/10.1287/isre.1100.0335 [Accessed October 6, 2015].

Armstrong, C.L., 2011. Providing a clearer view: An examination of transparency on local government websites. *Government Information Quarterly*, 28(1), pp.11–16. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0740624X10001048 [Accessed October 6, 2015].

Article 29 Working Party, 2007. *Working Document on the processing of personal data relating to health in electronic health records (EHR)*.

Barkhuus, L., 2012. The mismeasurement of privacy: using contextual integrity to reconsider privacy in HCI. In ACM Press, p. 367. Available at: http://dl.acm.org/citation.cfm?doid=2207676.2207727 [Accessed October 6, 2015].

Bates, J., 2014. The strategic importance of information policy for the contemporary neoliberal state: The case of Open Government Data in the United Kingdom. *Government Information Quarterly*, 31(3), pp.388–395. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0740624X14000951 [Accessed October 6, 2015].

Batini, C. et al., 2009. Methodologies for data quality assessment and improvement. *ACM Computing Surveys*, 41(3), pp.1–52. Available at: http://portal.acm.org/citation.cfm?doid=1541880.1541883 [Accessed October 6, 2015].

Bennett, C., 1985. From the Dark to the Light: The Open Government Debate in Britain. *Journal of Public Policy*, 5(02), pp.187–213. Available at: http://journals.cambridge.org/article_S0143814X00003020 [Accessed June 5, 2016].

Birchall, C., 2015. "Data.gov-in-a-box": Delimiting transparency. *European Journal of Social Theory*, 18(2), pp.185–202. Available at: http://est.sagepub.com/cgi/doi/10.1177/1368431014555259 [Accessed October 6, 2015].

Blanchette, J., 2010. Total Recall: How the E-memory Revolution Will Change Everything; DELETE: The Virtue of Forgetting in the Digital Age.

Borglund, E. & Engvall, T., 2014. Open data?: Data, information, document or record? Anne Thurston, ed. *Records Management Journal*, 24(2), pp.163–180. Available at: http://www.emeraldinsight.com/doi/abs/10.1108/RMJ-01-2014-0012 [Accessed October 6, 2015].

Both, W., 2012. Open Data - what the citizens really want. *The Journal of Community Informatics*, 8(2). Available at: http://ci-journal.net/index.php/ciej/article/view/814 [Accessed October 6, 2015].

Bowcott, I.T.O., 2015. Facebook row: US data storage leaves users open to surveillance, court rules. *The Guardian*. Available at: http://www.theguardian.com/world/2015/oct/06/us-digital-data-storage-systems-enable-state-interference-eu-court-rules [Accessed October 13, 2015].

Boyd, P., 2003. The requirements of the Data Protection Act 1998 for the processing of medical data. *Journal of Medical Ethics*, 29(1), pp.34–35. Available at: http://jme.bmj.com/content/29/1/34 [Accessed June 6, 2016].

Böhm, C. et al., 2010. Linking open government data: what journalists wish they had known. In ACM Press, p. 1. Available at: http://portal.acm.org/citation.cfm?doid=1839707.1839751 [Accessed October 6, 2015].

Borglund, E. & Engvall, T., 2014. Open data?: Data, information, document or record? D. Anne Thurston, ed. *Records Management Journal*, 24(2), pp.163–180. Available at: http://www.emeraldinsight.com/doi/abs/10.1108/RMJ-01-2014-0012 [Accessed October 6, 2015].
Both, W., 2012. Open Data - what the citizens really want. *The Journal of Community Informatics*, 8(2). Available at: http://ci-journal.net/index.php/ciej/article/view/814 [Accessed October 6, 2015].
Boyd, P., 2003. The requirements of the Data Protection Act 1998 for the processing of medical data. *Journal of Medical Ethics*, 29(1), pp.34–35. Available at: http://jme.bmj.com/content/29/1/34 [Accessed June 6, 2016].
Boyd, D. & Crawford, K., 2012. Critical questions for big data. *Information, Communication & Society*, 15(5), pp.662–679. Available at: http://www-tandfonline-com.libproxy.ucl.ac.uk/doi/abs/10.1080/1369118X.2012.678878 [Accessed June 5, 2016].
Burdon, M., 2009. Commercializing public sector information privacy and security concerns. *IEEE Technology and Society Magazine*, 28(1), pp.34–40. Available at: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4799406 [Accessed October 6, 2015].
Burkert, H., 1999. Privacy, data protection: a German/European perspective. In second symposium of the German American Academic Council's Project 'Global Networks and Local Values. Woods Hole, Massachusetts. Available at: http://www.coll.mpg.de/sites/www.coll.mpg.de/files/text/burkert.pdf.
Burn-Murdoch, J., 2013. Europe deadlocked over data protection reform. *The Guardian*. Available at: http://www.theguardian.com/news/datablog/2013/aug/12/europe-data-protection-directive-eu [Accessed October 13, 2015].
Caldicott, F., 1997. *Report on the Review of Patient Identifiable Information*, London: Department of Health.
Caldicott, F., 2013. *To Share or not to Share? The Information Governance Review*, London. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/192572/2900774_InfoGovernance_accv2.pdf.
Callaway, E., 2013. UK push to open up patients' data. *Nature News*, 502(7471), p.283. Available at: http://www.nature.com/news/uk-push-to-open-up-patients-data-1.13958 [Accessed June 6, 2016].
Carter, P., Laurie, G.T. & Dixon-Woods, M., 2015. The social licence for research: why care.data ran into trouble. *Journal of Medical Ethics*, 41(5), pp.404–409. Available at: http://jme.bmj.com/content/41/5/404 [Accessed June 6, 2016].
Childs, S. et al., 2014. Opening research data: issues and opportunities. Anne Thurston, ed. *Records Management Journal*, 24(2), pp.142–162. Available at: http://www.emeraldinsight.com/doi/abs/10.1108/RMJ-01-2014-0005 [Accessed October 6, 2015].
Clark, S. & Weale, C., 2011. *Information Governance in Health: An analysis of the social values involved in data linkage studies*, London. Available at: http://www.nuffieldtrust.org.uk/sites/files/nuffield/information_governance_in_health_-_research_report-_aug11.pdf.
Cornford, J. et al., 2013. Local governance in the new information ecology: the challenge of building interpretative communities. *Public Money & Management*, 33(3), pp.201–208. Available at: http://www.tandfonline.com/doi/abs/10.1080/09540962.2013.785705 [Accessed October 6, 2015].
Correa, A.S. et al., 2014. Really Opened Government Data: A Collaborative Transparency at Sight. In IEEE, pp. 806–807. Available at: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6906875 [Accessed October 6, 2015].
Davies, T. & Frank, M., 2013. "There"s no such thing as raw data': exploring the socio-technical life of a government dataset. In ACM Press, pp. 75–78. Available at: http://dl.acm.org/citation.cfm?doid=2464464.2464472 [Accessed October 6, 2015].
Davies, T.G. & Bawa, Z.A., 2012a. The Promises and Perils of Open Government Data (OGD). *The Journal of Community Informatics*, 8(2). Available at: http://ci-journal.net/index.php/ciej/article/view/929 [Accessed June 5, 2016].

Dawes, S.S., 2010. Stewardship and usefulness: Policy principles for information-based transparency. *Government Information Quarterly*, 27(4), pp.377–383. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0740624X10000651 [Accessed October 6, 2015].
De Hert, P. & Papakonstantinou, V., 2012. The proposed data protection Regulation replacing Directive 95/46/EC: A sound system for the protection of individuals. *Computer Law and Security Review: The International Journal of Technology and Practice*, 28(2), pp.130–142.
Department of Health, 2012. *The Power of Information: Putting all of us in control of the health and care information we need*, London.
Department of Health, 2009. Summary of Responses to the Consultation on the Additional Uses of Patient Data. Available at: http://wwwww.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/documents/digitalasset/dh_110715.pdf.
Digital Health, 2016a. Caldicott and care.data stalled by EU referendum "purdah." Available at: http://www.digitalhealth.net/cybersecurity/47591/caldicott-and-care.data-stalled-by-eu-referendum-'purdah'.
Digital Health, 2016b. HSCIC actions 1.2m patient opt-outs. Available at: http://www.digitalhealth.net/analytics/47614/hscic-actions-1.2m-patient-opt-outs.
Dolgin, E., 2014. New data protection rules could harm research, science groups say. *Nature Medicine*, 20(3), pp.224–224. Available at: http://www.nature.com/nm/journal/v20/n3/full/nm0314-224b.html [Accessed October 12, 2015].
Esteve Casellas Serra, L., 2014. The mapping, selecting and opening of data: The records management contribution to the Open Data project in Girona City Council D. Anne Thurston, ed. *Records Management Journal*, 24(2), pp.87–98. Available at: http://www.emeraldinsight.com/doi/abs/10.1108/RMJ-01-2014-0008 [Accessed October 6, 2015].
European Commission, 2014. *Guidelines on recommended standard licences, datasets and charging for the reuse of documents (2014/C240/01)*. Available at: https://ec.europa.eu/digital-single-market/en/news/commission-notice-guidelines-recommended-standard-licences-datasets-and-charging-re-use [Accessed 18 July 2016].
European Commission., 2015. Press release - EU Data protection reform on track: Commission proposal on new data protection rules in law enforcement area backed by Justice Ministers. Available at: http://europa.eu/rapid/press-release_IP-15-5812_en.htm
Evans, A.M. & Campos, A., 2013. Open Government Initiatives: Challenges of Citizen Participation: Professional Practice. *Journal of Policy Analysis and Management*, 32(1), pp.172–185. Available at: http://doi.wiley.com/10.1002/pam.21651 [Accessed October 6, 2015].
Filippon, J., 2015a. Slow and costly access to anonymised patient data impedes academic research. *BMJ*, 351, p.h5087. Available at: http://static.www.bmj.com/content/351/bmj.h5087 [Accessed June 5, 2016].
Filippon, J., 2015b. Slow and costly access to anonymised patient data impedes academic research. *BMJ*, 351, p.h5087. Available at: http://www.bmj.com/content/351/bmj.h5087 [Accessed May 21, 2016].
Floridi, L., 2014. *Protection of Information and the Right to Privacy - A New Equilibrium?: Law, Governance and Technology Series 17*, Germany: Springer Verlag.
Forgó, N., 2015. My health data—your research: some preliminary thoughts on different values in the General Data Protection Regulation. *International Data Privacy Law*, 5(1), pp.54–63. Available at: http://idpl.oxfordjournals.org/content/5/1/54 [Accessed May 12, 2016].
Fuster, G.G. & Gellert, R., 2012. The fundamental right of data protection in the European Union: in search of an uncharted right. *International Review of Law, Computers & Technology*, 26(1), pp.73–82. Available at: http://dx.doi.org/10.1080/13600869.2012.646798 [Accessed June 6, 2016].
Gianluigi, V. et al., 2014. Compliance with open government data policies: An empirical assessment of Italian local public administrations. *Information Polity*, (3,4), pp.263–275. Available at: http://www.medra.org/servlet/aliasResolver?alias=iospress&genre=article&issn=1570-1255&volume=19&issue=3&spage=263&doi=10.3233/IP-140338 [Accessed October 6, 2015].
Grace, J. & Taylor, M.J., 2013. Disclosure of Confidential Patient Information and the Duty to Consult: The Role of the Health and Social Care Information Centre. *Medical Law Review*, 21(3), pp.415–447. Available at: http://medlaw.oxfordjournals.org/content/21/3/415 [Accessed May 20, 2016].

Gunasekara, G., 2014. Paddling in unison or just paddling? International trends in reforming information privacy law. *International Journal of Law and Information Technology*, 22(2), pp.141–177.
Gupta, S. & Kumaraguru, P., 2013. OCEAN: Open-source Collation of eGovernment data And Networks - Understanding Privacy Leaks in Open Government Data. Available at: http://arxiv.org/abs/1312.2784 [Accessed October 6, 2015].
Hamilton, I. et al., 2015. Co-benefits of Energy and Buildings Data: The Case For supporting Data Access to Achieve a Sustainable Built Environment. In *Procedia Engineering*. International Conference on Sustainable Design, Engineering and Construction. pp. 958–968. Available at: http://www.sciencedirect.com/science/article/pii/S187770581502192X [Accessed November 10, 2015].
Handley, O. & Giuliano, J., 2012. M03 Developing strategies for global compliance to data privacy and data protection in an observational non-interventional study of huntington's disease. *Journal of Neurology, Neurosurgery & Psychiatry*, 83(Suppl 1), p.A47.
Harrison, E., Shepherd, E. & Flinn, A., 2015. *InterPARES Report Team Europe EU19 project 2014-15: A Research Report into Open Government Data in NHS England*.
Hartzog, W., 9. Here's to the right to be partially forgotten.
Haynes, C.L., Cook, G.A. & Jones, M.A., 2007. Legal and ethical considerations in processing patient-identifiable data without patient consent: lessons learnt from developing a disease register. *Journal of Medical Ethics*, 33(5), pp.302–307. Available at: http://jme.bmj.com/content/33/5/302 [Accessed June 6, 2016].
Health Research Authority, 2013. *Principles of Advice: Exploring the concepts of "Public Interest" and "Reasonably Practicable,"* Available at: http://www.hra.nhs.uk/documents/2014/12/principles-advice-april-2013-v-2.pdf.
Hellberg, A.-S. & Hedström, K., 2015. The story of the sixth myth of open data and open government. *Transforming Government: People, Process and Policy*, 9(1), pp.35–51. Available at: http://www.emeraldinsight.com/doi/10.1108/TG-04-2014-0013 [Accessed October 6, 2015].
Henninger, M., 2013. The value and challenges of public sector information. *Cosmopolitan Civil Societies: An Interdisciplinary Journal*, 5(3), pp.75–95. Available at: https://epress.lib.uts.edu.au/journals/index.php/mcs/article/view/3429 [Accessed October 6, 2015].
Higgins, J., 2003. The Patient Information Advisory Group and the use of patient-identifiable data. *Journal of Health Services Research & Policy*, 8(suppl 1), pp.8–11. Available at: http://hsr.sagepub.com/content/8/suppl_1/8 [Accessed June 6, 2016].
Hodson, H., Revealed: Google AI has access to huge haul of NHS patient data | New Scientist. Available at: https://www-newscientist-com.libproxy.ucl.ac.uk/article/2086454-revealed-google-ai-has-access-to-huge-haul-of-nhs-patient-data/ [Accessed June 6, 2016].
Hornung, G. & Schnabel, C., 2009. Data protection in Germany I: The population census decision and the right to informational self-determination. *Computer Law & Security Review*, 25(1), pp.84–88. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0267364908001660 [Accessed October 13, 2015].
HSCIC, 2016a. Applying Type 2 Opt-Outs. Available at: http://www.hscic.gov.uk/article/7072/Applying-Type-2-Opt-Outs.
HSCIC, 2016b. New Figures Released on Data Sharing Opt-Outs. Available at: http://www.hscic.gov.uk/article/7104/New-figures-released-on-data-sharing-opt-outs.
HSCIC, 2016c. Your Personal Information Choices. Available at: http://www.hscic.gov.uk/yourinfo.
Höchtl, J. & Reichstädter, P., 2011. Linked Open Data - A Means for Public Sector Information Management. In K. N. Andersen et al., eds. *Electronic Government and the Information Systems Perspective*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 330–343. Available at: http://link.springer.com/10.1007/978-3-642-22961-9_26 [Accessed October 6, 2015].
Information Commissioner's Office, 2002. *Anonymisation:managing data protection risk code of practice*, Available at: https://ico.org.uk/media/1061/anonymisation-code.pdf.
Iorio, C.T.D. et al., 2013. Cross-border flow of health information: is "privacy by design" enough? Privacy performance assessment in EUBIROD. *The European Journal of Public Health*, 23(2), pp.247–253. Available at: http://eurpub.oxfordjournals.org/content/23/2/247 [Accessed May 12, 2016].

Iorio, C.T.D., Carinci, F. & Oderkirk, J., 2014. Health research and systems' governance are at risk: should the right to data protection override health? *Journal of Medical Ethics*, 40(7), pp.488–492. Available at: http://jme.bmj.com/content/40/7/488 [Accessed May 12, 2016].
Iversen, A. et al., 2006. Epidemiology: Consent, Confidentiality, And The Data Protection Act. *BMJ: British Medical Journal*, 332(7534), pp.165–169. Available at: http://www.jstor.org/stable/25455921 [Accessed June 6, 2016].
Janssen, K., 2012. Open Government Data and the Right to Information: Opportunities and Obstacles. *The Journal of Community Informatics*, 8(2). Available at: http://ci-journal.net/index.php/ciej/article/view/952 [Accessed October 6, 2015].
Janssen, K., 2011a. The influence of the PSI directive on open government data: An overview of recent developments. *Government Information Quarterly*, 28(4), pp.446–456. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0740624X11000517 [Accessed October 6, 2015].
Janssen, K., 2011b. The role of public sector information in the European market for online content: a never-ending story or a new beginning? L. Van Audenhove, ed. *info*, 13(6), pp.20–29. Available at: http://www.emeraldinsight.com/doi/abs/10.1108/14636691111174234 [Accessed October 6, 2015].
Janssen, M., Charalabidis, Y. & Zuiderwijk, A., 2012. Benefits, Adoption Barriers and Myths of Open Data and Open Government. *Information Systems Management*, 29(4), pp.258–268. Available at: http://www.tandfonline.com/doi/abs/10.1080/10580530.2012.716740 [Accessed October 6, 2015].
Jetzek, T., Avital, M. & Bjorn-Andersen, N., 2014. Data-Driven Innovation through Open Government Data. *Journal of theoretical and applied electronic commerce research*, 9(2), pp.15–16. Available at: http://www.scielo.cl/scielo.php?script=sci_arttext&pid=S0718-18762014000200008&lng=en&nrm=iso&tlng=en [Accessed October 6, 2015].
Jori, A., 2007a. The census decision and the second generation of data protection norms. Available at: http://web.archive.org/web/20071023201818/http://www.dataprotection.eu/pmwiki/pmwiki.php?n=Main.SecondGeneration.
Jori, A., 2007b. The development of data protection law and the outline if its history. Available at: http://web.archive.org/web/20071023201706/http://www.dataprotection.eu/pmwiki/pmwiki.php?n=Main.Development.
Joseph, R.C. & Johnson, N.A., 2013. Big Data and Transformational Government. *IT Professional*, 15(6), pp.43–48. Available at: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6560044 [Accessed October 6, 2015].
Kalampokis, E., Hausenblas, M. & Tarabanis, K., 2011. Combining Social and Government Open Data for Participatory Decision-Making. In E. Tambouris, A. Macintosh, & H. de Bruijn, eds. *Electronic Participation*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 36–47. Available at: http://link.springer.com/10.1007/978-3-642-23333-3_4 [Accessed October 6, 2015].
Kalampokis, E., Tambouris, E. & Tarabanis, K., 2011. Open Government Data: A Stage Model. In M. Janssen et al., eds. *Electronic Government*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 235–246. Available at: http://link.springer.com/10.1007/978-3-642-22878-0_20 [Accessed October 6, 2015].
Keen, J. et al., 2013. Big data + politics = open data: The case of health care data in England. *Policy & Internet*, 5(2), pp.228–243. Available at: http://onlinelibrary.wiley.com/doi/10.1002/1944-2866.POI330/abstract [Accessed June 5, 2016].
Kosinski, M., Stillwell, D. & Graepel, T., 2013. Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*, 110(15), pp.5802–5805. Available at: http://www.pnas.org/cgi/doi/10.1073/pnas.1218772110 [Accessed October 6, 2015].
Krotoski, A. & Krotoski, A., 2015. The power of privacy – video series. *The Guardian*. Available at: http://www.theguardian.com/media-network/ng-interactive/2015/sep/29/the-power-of-privacy-video-series [Accessed October 13, 2015].
Lachmann, P.J., 2003. Consent and confidentiality—where are the limits? An introduction. *Journal of Medical Ethics*, 29(1), pp.2–3. Available at: http://jme.bmj.com/content/29/1/2 [Accessed June 6, 2016].
Linders, D., 2012. From e-government to we-government: Defining a typology for citizen coproduction in the age of social media. *Government Information Quarterly*, 29(4), pp.446–454.

Available at: http://linkinghub.elsevier.com/retrieve/pii/S0740624X12000883 [Accessed October 6, 2015].
LSE Blog, 2016. Data Protection: How Medical Researchers Persuaded the European Parliament to Compromise. Available at: http://blogs.lse.ac.uk/brexitvote/2016/02/11/data-protection-how-medical-researchers-persuaded-the-european-parliament-to-compromise/.
Luna-Reyes, L.F., Bertot, J.C. & Mellouli, S., 2014. Open Government, Open Data and Digital Government. *Government Information Quarterly*, 31(1), pp.4–5. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0740624X13001147 [Accessed October 6, 2015].
Martin, C., 2014. Barriers to the Open Government Data Agenda: Taking a Multi-Level Perspective. *Policy & Internet*, 6(3), pp.217–240. Available at: http://onlinelibrary.wiley.com/doi/10.1002/1944-2866.POI367/abstract [Accessed October 6, 2015].
Mayer-Schonberger, V., 1997. Generational Development of Data Protection in Europe. In P. Agre & M. Rotenberg, eds. *Technology and Privacy: The New Landscape*. Cambridge, Massachusettes: MIT Press.
McCall, B., 2014. European Parliament supports data protection reforms. *The Lancet*, 383(9923), p.1115. Available at: http://www.sciencedirect.com/science/article/pii/S0140673614602312 [Accessed October 12, 2015].
McDermott, P., 2010. Building open government. *Government Information Quarterly*, 27(4), pp.401–413. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0740624X10000663 [Accessed October 6, 2015].
Mcgoldrick, D., 2013. Developments in the Right to be Forgotten. *Human Rights Law Review*, 13(4), pp.761–776.
McLeod, J., 2014. Editorial. *Records Management Journal*, 24(2). Available at: http://www.emeraldinsight.com/doi/abs/10.1108/RMJ-06-2014-0028 [Accessed October 6, 2015].
McLeod, J., 2012. Thoughts on the opportunities for records professionals of the open access, open data agenda. *Records Management Journal*, 22(2), pp.92–97.
Meijer, R., Conradie, P. & Choenni, S., 2014. Reconciling Contradictions of Open Data Regarding Transparency, Privacy, Security and Trust. *Journal of theoretical and applied electronic commerce research*, 9(3), pp.32–44. Available at: http://www.scielo.cl/scielo.php?script=sci_arttext&pid=S0718-18762014000300004&lng=en&nrm=iso&tlng=en [Accessed October 6, 2015].
Mergel, I., 2014. The Long Way From Government Open Data to Mobile Health Apps: Overcoming Institutional Barriers in the US Federal Government. *JMIR mHealth and uHealth*, 2(4), p.e58. Available at: http://mhealth.jmir.org/2014/4/e58/ [Accessed October 6, 2015].
Milić, P., Veljković, N. & Stoimenov, L., 2012. Framework for open data mining in e-government. In ACM Press, p. 255. Available at: http://dl.acm.org/citation.cfm?doid=2371316.2371369 [Accessed October 6, 2015].
Misuraca, G. & Viscusi, G., 2014. Is Open Data Enough?: E-Governance Challenges for Open Government. *International Journal of Electronic Government Research*, 10(1), pp.18–34. Available at: http://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/ijegr.2014010102 [Accessed October 6, 2015].
Moat, H.S. et al., 2014. Using big data to predict collective behavior in the real world. *Behavioral and Brain Sciences*, 37(01), pp.92–93. Available at: http://journals.cambridge.org/article_S0140525X13001817 [Accessed June 5, 2016].
Murillo, M.J., 2015. Evaluating the role of online data availability: The case of economic and institutional transparency in sixteen Latin American nations. *International Political Science Review*, 36(1), pp.42–59. Available at: http://ips.sagepub.com/cgi/doi/10.1177/0192512114541163 [Accessed October 6, 2015].
Napoli, P.M. & Karaganis, J., 2010. On making public policy with publicly available data: The case of U.S. communications policymaking. *Government Information Quarterly*, 27(4), pp.384–391. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0740624X10000547 [Accessed October 6, 2015].
Newman, A.L., 2015. What the "right to be forgotten" means for privacy in a digital age. *Science (New York, N.Y.)*, 347(6221), pp.507–8.
Nuffield Council on Bioethics, 2014. *The collection, linking and use of data in biomedical research and health care: ethical issues*, London.

O'Donnell, G.A., 1998. Horizontal Accountability in New Democracies. *Journal of Democracy*, 9(3), pp.112–126. Available at: http://muse.jhu.edu/content/crossref/journals/journal_of_democracy/v009/9.3odonnell.html [Accessed October 6, 2015].
O'Dowd, A., 2014. Handling of patient data by Health and Social Care Information Centre is a "shambles," say MPs. *BMJ*, 348, p.g2702. Available at: http://www.bmj.com/content/348/bmj.g2702 [Accessed May 21, 2016].
O'Hara, K., 2015. The Right to Be Forgotten: The Good, the Bad, and the Ugly. *Internet Computing, IEEE*, 19(4), pp.73–79.
O'Neill, O., 2003. Some limits of informed consent. *Journal of Medical Ethics*, 29(1), pp.4–7. Available at: http://jme.bmj.com/content/29/1/4 [Accessed June 6, 2016].
P. P. Craig, author, 2015. *EU law : text, cases and materials / Paul Craig and Gráinne De Búrca.* Sixth edition., Oxford: Oxford University Press.
Parkes, S.E., 2004. Legal aspects of records based medical research. *Archives of Disease in Childhood*, 89(10), pp.899–901. Available at: http://adc.bmj.com/content/89/10/899 [Accessed June 6, 2016].
Parycek, P., Höchtl, J. & Ginner, M., 2014. Open Government Data Implementation Evaluation. *Journal of theoretical and applied electronic commerce research*, 9(2), pp.13–14. Available at: http://www.scielo.cl/scielo.php?script=sci_arttext&pid=S0718-18762014000200007&lng=en&nrm=iso&tlng=en [Accessed October 6, 2015].
Paulan Korenhof et al., 2014. Timing the Right to Be Forgotten: A Study into "Time" as a Factor in Deciding About Retention or Erasure of Data.
Payne, D., 2015. Google, doctors, and the "right to be forgotten." *The BMJ*, 350, p.h27. Available at: http://www.bmj.com/content/350/bmj.h27 [Accessed October 13, 2015].
Pilkington, E., 2015a. Edward Snowden calls for global push to expand digital privacy laws. *The Guardian*. Available at: http://www.theguardian.com/us-news/2015/sep/24/edward-snowden-international-laws-digital-privacy-video [Accessed October 13, 2015].
Pilkington, E., 2015b. Edward Snowden to attend meeting via video on "treaty" to improve privacy laws. *The Guardian*. Available at: http://www.theguardian.com/us-news/2015/sep/24/edward-snowden-meeting-treaty-privacy-laws [Accessed October 13, 2015].
Pollock, A.M. & Roderick, P., 2014. Trust in the time of markets: protecting patient information. *The Lancet*, 383(9928), pp.1523–1524. Available at: http://www.sciencedirect.com/science/article/pii/S0140673614607273 [Accessed May 21, 2016].
Press Association, 2015. NHS details released against patients' wishes, admits data body. *the Guardian*. Available at: http://www.theguardian.com/society/2015/jun/06/nhs-data-released-against-patients-wishes [Accessed May 21, 2016].
Presser, L. et al., 2015. Care.data and access to UK health records: patient privacy and public trust. *Technology Science*. Available at: http://techscience.org/a/2015081103/ [Accessed October 6, 2015].
Radu, R., Chenou, J.-M. & Internet Policy Review, 2015. Data control and digital regulatory space(s): towards a new European approach.
Redfern, M. 2001. *The Royal Liverpool Children's Inquiry Report.* Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/250934/0012_ii.pdf [Accessed 1 July 2016].
Reding, V., 2012. The European data protection framework for the twenty-first century. *International Data Privacy Law*, p.ips015. Available at: http://idpl.oxfordjournals.org/content/early/2012/06/25/idpl.ips015.short [Accessed October 13, 2015].
Rees, C. & Heywood, D., 2014. The "right to be forgotten" or the "principle that has been remembered." *Computer Law & Security Review: The International Journal of Technology Law and Practice*, 30(5), pp.574–578.
Reuters, S., 2015. Edward Snowden: mass surveillance does not help combat terrorism – video. *The Guardian*. Available at: http://www.theguardian.com/us-news/video/2015/sep/25/edward-snowden-treaty-glenn-greenwald-mass-surveillance-terrorism-video [Accessed October 13, 2015].
Roeber, B. et al., 2015. Personal data: how context shapes consumers' data sharing with organizations from various sectors. *Electronic Markets*, 25(2), pp.95–108.

Sartor, G., 2013. Providers' liabilities in the new EU Data Protection Regulation: A threat to Internet freedoms? *International Data Privacy Law*, 3(1), pp.3–12.
Saxby, S., 2012. The 2012 CLSR-LSPI seminar on privacy, data protection & cyber-security – Presented at the 7th international conference on Legal, Security and Privacy Issues in IT law (LSPI) October 2–4, 2012, Athens. *Computer Law and Security Review: The International Journal of Technology and Practice*.
Scotl, S.C. & editor, 2015. Police spying report delayed by failure to appoint commissioner, IOCCO says. *The Guardian*. Available at: http://www.theguardian.com/uk-news/2015/oct/02/iocco-investigation-claims-police-spied-journalists-scotland [Accessed October 13, 2015].
Shadbolt, N. et al., 2012. Linked Open Government Data: Lessons from Data.gov.uk. *IEEE Intelligent Systems*, 27(3), pp.16–24. Available at: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6171150 [Accessed October 6, 2015].
Simeonovski, M. et al., 20150619. Oblivion: Mitigating Privacy Leaks by Controlling the Discoverability of Online Information.
Steinman, J.S. and J., 2015. Did You Really Agree to That? The Evolution of Facebook's Privacy Policy. *Technology Science*. Available at: http://techscience.org/a/2015081102/ [Accessed October 6, 2015].
Sweeney, L., 2015. Only You, Your Doctor, and Many Others May Know. *Technology Science*. Available at: http://techscience.org/a/2015092903/ [Accessed October 6, 2015].
Taylor, M.J., 2011. Health Research, Data Protection, and the Public Interest in Notification. *Medical Law Review*, 19(2), pp.267–303. Available at: http://medlaw.oxfordjournals.org/content/19/2/267 [Accessed June 6, 2016].
Taylor, M.J., 2015. Legal Bases for Disclosing Confidential Patient Information for Public Health: Distinguishing Between Health Protection and Health Improvement. *Medical Law Review*, 23(3), pp.348–374. Available at: http://medlaw.oxfordjournals.org/content/23/3/348 [Accessed May 20, 2016].
Taylor, M.J. & Taylor, N., 2014a. Health research access to personal confidential data in England and Wales: assessing any gap in public attitude between preferable and acceptable models of consent. *Life Sciences, Society and Policy*, 10(1), p.1. Available at: http://lsspjournal.springeropen.com/articles/10.1186/s40504-014-0015-6 [Accessed June 6, 2016].
Taylor, M.J. & Taylor, N., 2014b. Health research access to personal confidential data in England and Wales: assessing any gap in public attitude between preferable and acceptable models of consent. *Life Sciences, Society and Policy*, 10(1), p.1. Available at: http://lsspjournal.springeropen.com [Accessed May 21, 2016].
Thielman, S., 2015a. Experian hack exposes 15 million people's personal information. *The Guardian*. Available at: http://www.theguardian.com/business/2015/oct/01/experian-hack-t-mobile-credit-checks-personal-information [Accessed October 13, 2015].
Thielman, S., 2015b. Privacy groups hail "freedom from surveillance" in European court's Facebook ruling. *The Guardian*. Available at: http://www.theguardian.com/business/2015/oct/06/europe-court-right-to-privacy-max-schrems-us-tech-companies [Accessed October 13, 2015].
Thomas, H., 2013. *What challenges does the current legislative, regulatory and organisational context within the NHS pose for Records Management?* MA Dissertation. London: University College London.
Thurston, A., 2014. Guest Editorial. *Records Management Journal*, 24(2). Available at: http://www.emeraldinsight.com/doi/abs/10.1108/RMJ-06-2014-0027 [Accessed October 6, 2015].
Thurston, A., 2013. Transparency can "break cycle of poor governance" in developing world. *The Guardian*. Available at: http://www.theguardian.com/public-leaders-network/2013/mar/12/transparency-break-cycle-poor-governance [Accessed October 6, 2015].
Thurston, A.C., 2012. Trustworthy Records and Open Data. *The Journal of Community Informatics*, 8(2). Available at: http://ci-journal.net/index.php/ciej/article/view/951 [Accessed October 6, 2015].
TNA, 2015a. *Checklist for public sector bodies: Get ready for re-use of public sector information*, Available at: http://www.nationalarchives.gov.uk/documents/information-management/checklist-for-public-sector-bodies.pdf.

TNA, 2015b. *Introductory Guide to the Amended PSI Directive*, Available at: http://www.nationalarchives.gov.uk/documents/information-management/psi-directive-transposition-intro-guide.pdf.
UK Government, 2012. *Open Data White Paper: Unleasing the Potential.* Available at: https://data.gov.uk/sites/default/files/Open_data_White_Paper.pdf [Accessed 18 July 2016].
Veljković, N., Bogdanović-Dinić, S. & Stoimenov, L., 2014. Benchmarking open government: An open data perspective. *Government Information Quarterly*, 31(2), pp.278–290. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0740624X14000434 [Accessed October 6, 2015].
Verma, N. & Gupta, M.P., 2013. Open government data: beyond policy & portal, a study in Indian context. In ACM Press, pp. 338–341. Available at: http://dl.acm.org/citation.cfm?doid=2591888.2591949 [Accessed October 6, 2015].
Verschuuren, M. et al., 2008. The European data protection legislation and its consequences for public health monitoring: a plea for action. *The European Journal of Public Health*, 18(6), pp.550–551. Available at: http://eurpub.oxfordjournals.org/content/18/6/550 [Accessed May 12, 2016].
Walker-Osborn, C., Fitzsimons, L. & Ruane, J., 2013. Data Protection. *ITNOW*, 55(3), pp.38–39. Available at: http://itnow.oxfordjournals.org/cgi/doi/10.1093/itnow/bwt050 [Accessed October 12, 2015].
Warso, Z., 2013. There's more to it than data protection – Fundamental rights, privacy and the personal/household exemption in the digital age. *Computer Law & Security Review: The International Journal of Technology Law and Practice*, 29(5), pp.491–500.
Weber, R.H., 2015. The digital future – A challenge for privacy? *Computer Law & Security Review: The International Journal of Technology Law and Practice*, 31(2), pp.234–242.
Wellcome Trust, 2015a. *Academic research perspective on the European Commission, Parliament and Council texts of the proposal for a General Data Protection Regulation - 2012/0011(COD)*,
Wellcome Trust, 2016. *Analysis: Research and the General Data Protection Regulation*, Available at: https://wellcome.ac.uk/sites/default/files/new-data-protection-regulation-key-clauses-wellcome-may16.pdf.
Wellcome Trust, 2015b. *Ensuring a healthy future for scientific research through the Data Protection Regulation 2012/0011(COD): Position of academic, patient and non-commercial research organisations*, Available at: https://wellcome.ac.uk/sites/default/files/ensuring-healthy-future-for-scientific-research-data-protection-regulation-joint-statement-dec15.pdf.
Wellcome Trust, 2015c. Press Release - Vote by European Parliament on data protection welcomed by research community. Available at: https://wellcome.ac.uk/press-release/vote-european-parliament-data-protection-welcomed-research-community.
Wellcome Trust (joint statement), 2013. *Impact of the draft European Data Protection Regulation and proposed amendments from the rapporteur of the LIBE committee on scientific research*, Available at: https://wellcome.ac.uk/sites/default/files/wtvm054713.pdf.
Whitmore, A., 2014. Using open government data to predict war: A case study of data and systems challenges. *Government Information Quarterly*, 31(4), pp.622–630. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0740624X14001154 [Accessed October 6, 2015].
Wilson, S., 2015. Data protection: Big data held to privacy laws, too. *Nature*, 519(7544), pp.414–414. Available at: http://www.nature.com/doifinder/10.1038/519414a [Accessed October 12, 2015].
Woolf, N., 2015. Court hears first arguments in case challenging bulk data collection by NSA. *The Guardian*. Available at: http://www.theguardian.com/us-news/2015/sep/25/first-arguments-case-challenging-bulk-data-collection-nsa [Accessed October 13, 2015].
Yannoukakou, A. & Araka, I., 2014. Access to Government Information: Right to Information and Open Government Data Synergy. *Procedia - Social and Behavioral Sciences*, 147, pp.332–340. Available at: http://linkinghub.elsevier.com/retrieve/pii/S187704281404018X [Accessed October 6, 2015].
Yu, H. & David, R., 2012. The New Ambiguity of "Open Government. *UCLA Law Review Discourse*, 59, pp.178–208.
Yu, H. & Robinson, D.G., 2012. The New Ambiguity of "Open Government." *SSRN Electronic Journal*. Available at: http://www.ssrn.com/abstract=2012489 [Accessed October 6, 2015].
Zanfir, G., 2012. The right to Data portability in the context of the EU data protection reform. *International Data Privacy Law*, 2(3), pp.149–162.
Zuiderwijk, A. & Janssen, M., 2014a. Open data policies, their implementation and impact: A framework for comparison. *Government Information Quarterly*, 31(1), pp.17–29. Available at:

http://www.sciencedirect.com/science/article/pii/S0740624X13001202 [Accessed October 6, 2015].
Zuiderwijk, A. & Janssen, M., 2014b. The Negative Effects of Open Government Data - Investigating the Dark Side of Open Data. In *Proceedings of the 15th Annual International Conference on Digital Government Research*. dg.o '14. New York, NY, USA: ACM, pp. 147–152. Available at: http://doi.acm.org/10.1145/2612733.2612761 [Accessed October 6, 2015].
Wellcome Trust (joint statement), 2013. *Impact of the draft European Data Protection Regulation and proposed amendments from the rapporteur of the LIBE committee on scientific research*, Available at: https://wellcome.ac.uk/sites/default/files/wtvm054713.pdf.
Whitmore, A., 2014. Using open government data to predict war: A case study of data and systems challenges. *Government Information Quarterly*, 31(4), pp.622–630. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0740624X14001154 [Accessed October 6, 2015].
Wilson, S., 2015. Data protection: Big data held to privacy laws, too. *Nature*, 519(7544), pp.414–414. Available at: http://www.nature.com/doifinder/10.1038/519414a [Accessed October 12, 2015].
Woolf, N., 2015. Court hears first arguments in case challenging bulk data collection by NSA. *The Guardian*. Available at: http://www.theguardian.com/us-news/2015/sep/25/first-arguments-case-challenging-bulk-data-collection-nsa [Accessed October 13, 2015].
Yannoukakou, A. & Araka, I., 2014. Access to Government Information: Right to Information and Open Government Data Synergy. *Procedia - Social and Behavioral Sciences*, 147, pp.332–340. Available at: http://linkinghub.elsevier.com/retrieve/pii/S187704281404018X [Accessed October 6, 2015].
Yu, H. & David, R., 2012. The New Ambiguity of "Open Government. *UCLA Law Review Discourse*, 59, pp.178–208.
Yu, H. & Robinson, D.G., 2012. The New Ambiguity of "Open Government." *SSRN Electronic Journal*. Available at: http://www.ssrn.com/abstract=2012489 [Accessed October 6, 2015].
Zanfir, G., 2012. The right to Data portability in the context of the EU data protection reform. *International Data Privacy Law*, 2(3), pp.149–162.
Zuiderwijk, A. & Janssen, M., 2014a. Open data policies, their implementation and impact: A framework for comparison. *Government Information Quarterly*, 31(1), pp.17–29. Available at: http://www.sciencedirect.com/science/article/pii/S0740624X13001202 [Accessed October 6, 2015].
Zuiderwijk, A. & Janssen, M., 2014b. The Negative Effects of Open Government Data - Investigating the Dark Side of Open Data. In *Proceedings of the 15th Annual International Conference on Digital Government Research*. dg.o '14. New York, NY, USA: ACM, pp. 147–152. Available at: http://doi.acm.org/10.1145/2612733.2612761 [Accessed October 6, 2015].

UK Legislation under analysis

Data Protection Act 1998 (UK). Available at: http://www.legislation.gov.uk/ukpga/1998/29/contents

Human Rights Act 1998 (UK), Available at: http://www.legislation.gov.uk/ukpga/1998/42/contents.

The Health Service (Control of Patient Information) Regulations 2002 (UK), Available at: http://www.legislation.gov.uk/uksi/2002/1438/contents/made

NHS Act 2006 (UK), Available at: http://www.legislation.gov.uk/ukpga/2006/41/contents.

Health and Social Care Act 2012, Available at: http://www.legislation.gov.uk/ukpga/2012/7/contents/enacted.

Care Act 2014 (UK), Available at: http://www.legislation.gov.uk/ukpga/2014/23/contents/enacted.

Re-Use of Public Sector Information Regulations (UK), 2015. Available at: http://www.legislation.gov.uk/uksi/2015/1415/pdfs/uksi_20151415_en.pdf.

European Legislation under analysis

European Parliament and Council, *Protection of Individuals with regard to the processing of personal data and on the free movement of such data, European Data Protection Directive*, 95/46/EC, 1995. Available at: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:en:HTML.

European Commission, *Proposal for a regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation)* 2012/0011 (COD), Available at: http://ec.europa.eu/justice/data-protection/document/review2012/com_2012_11_en.pdf [referenced as CDR in text].

European Parliament *Position adopted at first reading with a view to the adoption of Regulation (EU) No .../2014 of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation)*, 12 March 2014. Available at: http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+TA+P7-TA-2014-0212+0+DOC+XML+V0//EN#BKMD-6 [Referenced as PCT in text].

Council of the European Union*, Proposal for a Regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation) - Preparation of a general approach*, 2015. Available at: http://data.consilium.europa.eu/doc/document/ST-9565-2015-INIT/en/pdf.

Regulation (EU) 2016/679 of the European Parliament and of the Council on *The protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)* 2016. Available at: http://ec.europa.eu/justice/data-protection/reform/files/regulation_oj_en.pdf [Referenced GDPR in text].

Re-Use of Public Sector Information Directive, 2013 Available at: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2013:175:0001:0008:EN:PDF.