

Chapter 6

Data initiatives in health systems

Chapter 6 – Data initiatives in health systems

Chapter overview

This chapter discusses developments in the functions and purposes of health information systems and draws lessons from some specific initiatives.

Health-care IT systems were originally introduced to facilitate basic administrative tasks but have evolved to provide business intelligence for service improvement and to support observational research. These come together in the concept of a 'learning health system'.

Information systems in the English NHS have moved towards a centralised approach now overseen by the Health and Social Care Information Centre (HSCIC). Debate around the 'care.data' programme focused attention on assumptions about the relationship between privacy norms relevant to NHS patients and the legal norms under which HSCIC operates. It highlighted the absence of reflection on this difference and of a capacity to address it, and raised questions about how the rights of individuals were respected, resulting in a damaging loss of public and professional trust.

The Scottish Informatics Programme involved initial public consultation to identify relevant social norms. It developed a model of bespoke data linkage using a safe haven, subject to proportionate governance that takes into account both privacy risk and public benefit. Governance refers to an explicit, potentially revisable, statement of guiding principles and best practices that takes account of the findings of public engagement.

The 100,000 Genomes project involves linking data from genome sequencing with individuals' NHS records to investigate some cancers and other diseases. Authorised researchers from all sectors may access a firewall-protected, pseudonymised dataset with the broad consent of patients. The dataset is administered by a Government-owned company, Genomics England Ltd. The claimed public interest lies explicitly in securing economic as well as scientific and therapeutic benefits, by stimulating the commercial sector.

A number of recommendations are made in relation to defining reasonable expectations of data use, accounting for data use and delivering outcomes in the public interest.

Introduction

- 6.1 This chapter and the following one examine concrete contexts in which data initiatives have taken shape. In this chapter we will show how the introduction of information technology has wrought a transformation in our understanding of whole health care systems, particularly the NHS in the UK. From being a system focussed on the delivery of health care to those in immediate need, the introduction of IT in the health services has broadened and blurred the horizons of the system.
- 6.2 In the context of health care, perhaps more than elsewhere, the potential uses of data have influenced choices about information technology infrastructure, which is only partly about the delivery of health care. Other functions include the delivery of research insight, better resource allocation, support for innovation and evidence to inform policy. Data and IT initiatives also seek to extend the boundaries of health care beyond responding to illness, for example, to predicting who might get ill before they experience any symptoms and to understanding the rich combination of medical, behavioural and environmental factors relevant to health conditions.

6.3 The health services in the UK offer a number of examples of data initiatives at different scales, from the macro (e.g. the National Programme for IT – NpFIT), through the meso (e.g. the ‘100,000 Genomes’ project) to the micro (e.g. a clinical evaluation of a particular intervention). These have not always gone smoothly. While their possible failings as infrastructure projects are of moral relevance (when they miss opportunities to deliver benefits, or when they use public resources inefficiently, or undermine public trust) our interest will be primarily to identify good practice and areas for improvement in how they manage the relationship between underlying norms of access and disclosure, respect for individual values and interests, and governance in the public interest. Drawing on our discussion so far, the critical decisions for each initiative will be those that place them at different points on the following critical axes:

- The arrangements for storage (whether data are retained close to the point of collection or gathered together in safe havens or in a single, central repository)
- The arrangements for data disclosure/access (whether data are published, subject to controlled disclosure, controlled access or mediated access)
- The role given to individual patients (from explicit individual consent, through implicit consent with opt-out, to no individual authorisation)
- The range of users and purposes approved for access/disclosure (from restricting access to particular classes of users, such as academic researchers, or an expectation that the broader public interest will be served by any responsible use of data, including by commercial users).

6.4 The optimum relationship between norms, private freedoms and public objectives may be found through consideration of the interests and values at stake in the practical context. Health services in the UK have a long and complicated history, and giving attention to the way in which information use has developed will help us to understand the interests and drivers involved.

IT innovation and developing information requirements

6.5 As in other sectors, health care has a number of basic information needs that are increasingly met by computerised systems. The way in which these systems were introduced to health care paralleled other sectors in many respects, usually starting with routine administrative tasks.

Tracking patients

6.6 General Practitioners (GPs) were relatively quick to adopt computers to manage patient medical records, primarily to facilitate repeat prescriptions and later to manage complete patient records. Some of the earliest systems in the UK were developed on microcomputers in the late 1970s by technically adept GPs themselves.²⁹¹ The Department of Health (DH) introduced standards (‘requirements for accreditation’) in the late 1990s, along with government subsidies for compliant systems. These ensured that public administrative requirements would be incorporated alongside clinical requirements. In 2000 GPs were officially allowed to stop keeping paper records and

²⁹¹ One of the first GP computer systems, the Integrated General Practice system was developed by an Essex GP in collaboration with IT staff from a local shoe factory (an example of skills clustering leading to innovation) and successfully marketed to other practices via a venture capital funded company (Value Added Medical Products, or VAMP). For a brief history, see: Health Service Journal (12 August 1999) *VAMP comes alive*.

move to a paperless system although many had converted to a fully electronic system before then, albeit without necessarily having transferred legacy data from the retained paper files. In 2003 the new GP contract transferred effective ownership of GP systems away from GPs to Primary Care Organisations, which thus became the software vendors' customer.²⁹²

- 6.7 Hospitals adopted electronic patient administration systems from the late 1960s to manage patient appointments and track patients across the hospital, but these were not primarily for managing medical data, which often continued to be held separately in paper notes. Patients were usually identified by an assigned 'hospital ID' and, where available, their NHS Number.²⁹³ Hospitals might also have local clinical or laboratory systems to capture information directly from different types of machine, such as analytical instruments, or from clinicians. These vary from simple spreadsheets of results to complex scientific systems. However, they would often be only loosely integrated with the central patient administration system, perhaps having the patient's hospital ID as the only common feature. Regional health care systems often developed on the back of national requirements to facilitate collection of datasets in the days before universal use of the Internet (or a Virtual Private Network such as NHSNet). The data collected could then be used for regional planning and commissioning.
- 6.8 The English NHS Number used to be managed electronically through the National Strategic Tracing Service (NSTS), introduced in 1999 to allow NHS organisations to discover or check the NHS Number for individuals. This followed the recommendation of the first Caldicott Committee to increase the use of the NHS Number so that simple anonymisation could be used for administrative analyses of patient data, minimising the need to transmit data with full identifiers in order to link and de-duplicate datasets.²⁹⁴ There were a number of restrictions on access to NSTS to prevent it being used to re-identify anonymised records or to trace individuals. Unfortunately, these were so cumbersome or intrusive that many health regions set up their own 'master patient index' to provide a more rapid service locally, while using the NHS number only for those patients coming from outside their region.²⁹⁵ In Scotland a separate Community Health Index (CHI) Number is used to identify patients nationally, which must be used for each health care episode.²⁹⁶ Some countries, such as the USA, have no such universal health number, and it is often argued that this makes national collation of data difficult.²⁹⁷ This argument is somewhat suspect, however: universal naming systems are often sold as a means of improving data quality but in practice there are many difficulties and they rarely fulfil their potential.²⁹⁸ Similarly, while some people oppose universal numbering on privacy or even religious grounds, most patients are easily identified even by traditional methods such as name and date of birth.

²⁹² Primary Care Organisations (PCOs) are NHS England in England, Health and Social Services Boards in Northern Ireland, Local health boards in Wales, and primary care divisions within area health boards in Scotland. See: http://www.rcgp.org.uk/training-exams/gp-curriculum-overview/~media/Files/GP-training-and-exams/Curriculum%20previous%20versions%20as%20at%20July%202012/curr_archive_4_2_IMT_v1_0_mar06.ashx.

²⁹³ The old-style NHS Number (XXXXX NNN) originally appeared on a baby's birth certificate; the new NHS Number (NNN-NNN-NNNN) is allocated from birth (see: <http://systems.hscic.gov.uk/nhsnumber/patients/yournumber>).

²⁹⁴ The Caldicott Committee (1997) Report on the review of patient-identifiable information, available at: http://webarchive.nationalarchives.gov.uk/20130107105354/http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_4068403.

²⁹⁵ See, for example: <http://www.tamesidehospital.nhs.uk/documents/EnsuringAccuratePatientInformationPolicy.pdf>.

²⁹⁶ Those in the Border regions may have both, as their GP may be in England but their nearest hospital in Scotland. <http://www.ehealth.scot.nhs.uk/support-documentation/document-holder2/>.

²⁹⁷ See, for example, <http://www.cardiosource.org/en/Advocacy/Issues/Health-Information-Technology/ACC-Policies-and-Activities/Unique-Patient-Identifier-Principles.aspx>.

²⁹⁸ See "Naming" in Anderson R (2008) Security engineering – a guide to building dependable distributed systems (Wiley), available at: <http://www.cl.cam.ac.uk/~rja14/book.html>.

Observational research

- 6.9 Various GP system providers have set up research databases based on proprietary extracts from their systems. First was VAMP (now InPractice Systems) whose original business model was to provide free PCs to GPs and then sell access of the aggregated data to pharmaceutical companies. This facility was donated to the Department of Health in 1994, where it became the GP Research Database. Most GP systems support a facility, MIQUEST, introduced in 1996, that allows queries to be run by participating GP systems at the discretion of the GP practice and return aggregated data for research or other purposes. Exceptionally, individual-level data could be requested, but only age and postal area (not full postcode) would be returned to limit identifiability. In 2012, the Clinical Practice Research Datalink (CPRD) was established. This combines the activities of the MHRA's General Practice Research Database (GPRD) and the Department of Health's NIHR Research Capability Programme.²⁹⁹ The CPRD is accessible to researchers for a variety of research purposes, including observational research and planning interventional trials and can support clinical decision making by providing clinicians with relevant, real-world data to inform their consultations with patients.³⁰⁰
- 6.10 A number of condition-specific registries have been developed over the years to help to understand the effect of those conditions and provide integrated services to those affected. For example, cancer registries (of which there are 11 run by eight regional organisations, also confusingly called 'registries').³⁰¹ These cancer registries link with the Office of National Statistics death register to identify when patients on the various registries die (so mortality statistics can be generated – and compared with other countries' experience).³⁰² Recruitment to registries used to be seen by recruiters as part of clinical care with the possibility for patient opt-out rather than specific opt-in consent. It was a challenge to this presumption (from some wording in the GMC Confidentiality guidance in 2001) that led to the introduction of Section 60 of the Health & Social Care Act 2001 (now Section 251 of the NHS Act 2006) and the Health Service (Control of Patient Information) Regulations 2002 to enable the use of this information for research without specific consent under some conditions (see chapter 4). The path established by cancer registries subsequently provided a template for other specialties.

Performance evaluation and improvement

- 6.11 Organisations need feedback in order to improve quality and ensure safety. For this purpose, health systems generally, health organisations separately, clinical teams as well as individual clinicians compare their performance with the norm or with others' performance. If there is a public interest in health care, there is also a public interest in supporting good practice and identifying and eliminating poor provision. In the UK, health care intelligence providers such as Dr Foster have created a business in monitoring and analysing outcomes so that patients and funders can see which

²⁹⁹ See: <http://www.cprd.com/home/>.

³⁰⁰ Ibid.

³⁰¹ Some people reserve the term 'register' for the database and 'registry' for the organisation where registering takes place.

³⁰² One of our reviewers drew to our attention the fact that delays in death registration in England, Wales and Northern Ireland can have the effect of undermining the evidence-base for epidemic monitoring, record-linkage research and policy development. The Royal Statistical Society has recommended that registration of the fact of death (pending determination of cause of death) should take place in a timely fashion (<http://www.publications.parliament.uk/pa/cm201213/cmselect/cmpublicadm/406/406we08.htm>).

hospital departments have the lowest mortality rates.³⁰³ On the one hand, this can be useful in detecting failing institutions; on the other, if not interpreted with reference to the context, it can penalise hospitals that tackle more complex, high-risk cases turned down by others. Nevertheless, the trend for increased public monitoring of medical performance now appears to be well established. In the USA, websites like ZocDoc and Vitals are establishing patient feedback as a norm, as TripAdvisor has done for hotels.³⁰⁴

6.12 The distinction between striving for continual improvements in productivity through efficiency and innovation, and improving health through developing better patient information and treatment – that is, through research – begins to disappear in the concept of a so-called ‘learning health care system’. While these two senses of ‘service improvement’ (business productivity and improved care) have always been related, they have been the subject of separate information and governance systems. The integration of these systems represents a more recent innovation.

6.13 Three moral justifications have been suggested for this integration: the need for a just system, for high quality care, and for a system that supports economic well-being.³⁰⁵ These are said to entail moral requirements on both clinicians and patients to participate in research aimed at service improvement (‘learning’).³⁰⁶ The key claim is that there is no ‘do nothing’ option because of the growing pressure of circumstances: at an individual level, people are getting ill (healthy people becoming ill, ill people getting more ill, and new illnesses appearing); at a system level, resources to meet these demands are more or less tightly constrained (although the level of constraint is, of course, a result of political decisions). Proponents of learning health systems may accept that the risk of data abuses increases as a result of a learning activity, and that such activities may impose additional burdens on patients (such as extra visits to clinics). However, they believe that these can be minimised through appropriate controls and that the residual risk is justified.³⁰⁷

Public administration and service delivery

6.14 Over the last 25 years, there has been continuous pressure for more administrative access to records for purposes such as clinical audit, service planning and cost control.³⁰⁸ Indeed, these may often have been the real drivers behind centralisation efforts, with ‘research’ promoted as a desirable further purpose and often as the public rationale. The biggest single driver of centralisation in the UK was, however, the

³⁰³ See: www.drfooster.com.

³⁰⁴ The Economist (26 July 2014) *DocAdvisor*, available at: <http://www.economist.com/news/international/21608767-patients-around-world-are-starting-give-doctors-piece-their-mind-result/>.

³⁰⁵ See Faden RR, Kass NE, Goodman SN, et al. (2013) An ethics framework for a learning health care system: a departure from traditional research ethics and clinical ethics *Hastings Center Report* 43(s1): S16-S27, available at: <http://onlinelibrary.wiley.com/doi/10.1002/hast.134/full>.

³⁰⁶ Faden et al. (op.cit) offer a Rawlsian account of common good to justify a ‘norm of common purpose’. This is based on “the reciprocal obligation that arises among strangers who occupy the role of patient over time” (at page 23). In their system, the first 4 principles they advance are designed to protect individual rights and freedoms, and limit the claims that can be made of any individual participant.

³⁰⁷ Ibid.

³⁰⁸ The collection of UK national statistics was a recommendation of the Körner Committee, which carried out a major review of health service information between 1980 and 1984. In the 1980s the NHS Exeter system was developed to bring together data sets from across England with 21 regional centres collating data and transmitting them to the national centre at Exeter (building on an earlier initiative to develop a national PAS system there). The data covered a range of topics from GP registrations, organ donor registration, national screening programmes, to GP capitation payments. The Exeter system is now known as the National Health Applications and Infrastructure Services (NHAIS).

purchaser/provider split, introduced in 1991.³⁰⁹ The fact that medical procedures performed in hospitals had to be paid for meant that information about the procedure, the patient, and the cost had to flow on a large scale. This led to the establishment of the Hospital Episodes Statistics (HES) database, amongst others, to track activity across the NHS as a whole.³¹⁰ The HES database was largely populated using data from the NHS-wide Clearing Service which handled payments. Otherwise data was only available on a piecemeal basis, gleaned from individual audits or research studies.³¹¹

- 6.15 In 1992, the Department of Health published the Information Management and Technology (IM&T) Strategy, whose vision was a single electronic health record (EHR) for each patient, accessible to everyone working within the NHS.³¹² The British Medical Association objected that making patient records available beyond the teams responsible for a patient's direct care would compromise both safety and privacy.³¹³ This led to the first Caldicott review and the creation of 'Caldicott Guardians'.³¹⁴ In 1998, NHS England released a new IT strategy, *Information for Health*, promoting the adoption of both EPRs that would be held at a particular care provider and EHRs to be shared across all providers.
- 6.16 In 2002, with high-profile backing from the (then) Prime Minister, the Department of Health announced a major infrastructure initiative, the National Programme for IT (NpFIT). This was to be implemented by Connecting for Health, a new agency established to drive forward 'ruthless standardisation' across the NHS in England.³¹⁵ (Wales and Scotland had separate initiatives taking rather different approaches.) The aspiration was to deliver the EPR systems promised in *Information for Health*, in a time frame of about three years.³¹⁶
- 6.17 The NpFIT supported a number of features. Some, like the Picture Archiving and Communication System, which provided an accessible electronic archive for radiology images, proved to be both useful and successful.³¹⁷ Others, like the 'Choose and Book'

³⁰⁹ NHS and Community Care Act 1990, available at: <http://www.legislation.gov.uk/ukpga/1990/19/contents>. The internal market was brought to an end in 1997 but the split continued through the introduction of Primary Care Groups (then Primary Care Trusts).

³¹⁰ Initially data were collected annually, then quarterly. At the time of writing they are collected on a monthly basis (<http://www.hscic.gov.uk/hes>).

³¹¹ It is notable that it was only in 1999 that Professor Sir Brian Jarman at Imperial College used HES data to produce comparative mortality statistics, thus creating quite a storm about why these figures were unreliable, though few suggestions on how to improve on them and reduce adverse incidents. See: Jarman B, Gault S, Alves B, et al. (1999) Explaining differences in English hospital death rates using routinely collected data British Medical Journal 318: 1515-20, available at: <http://www.bmj.com/content/318/7197/1515.full>.

³¹² Department of Health (1992) Getting better with information: an IM&T Strategy for the NHS in England (London: HMSO). See: <http://www.intosaiitaudit.org/9899371.pdf>.

³¹³ R Anderson (1996) Security in clinical information systems, British Medical Association, available at: <http://www.cl.cam.ac.uk/~rja14/policy11/policy11.html>.

³¹⁴ See paragraph 2.43.

³¹⁵ Department of Health (2002) *Delivering 21st century IT support for the NHS*, national strategic programme, available at: http://webarchive.nationalarchives.gov.uk/20130107105354/http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/@dh/@en/documents/digitalasset/dh_4071684.pdf.

³¹⁶ Computer Weekly (18 Feb 2008) *Secret papers reveal Blair's rushed NpFIT plans*, Computer Weekly, available at: <http://www.computerweekly.com/blogs/public-sector/2008/02/secret-papers-reveal-blairs-ru.html>; *BBC News* (25 October 2007) *NHS IT time-frame 'ludicrously tight'*, available at: <http://news.bbc.co.uk/1/hi/health/7061590.stm>; Campion-Awwad O, Hayton A, Smith L and Vuaran M (2014, MPhil Public Policy 2014, University of Cambridge) The national programme for IT in the NHS – a case history, available at: <http://www.cl.cam.ac.uk/~rja14/Papers/npfit-mpp-2014-case-history.pdf>, at page 13.

³¹⁷ Picture Archiving and Communications System (PACS), though not part of the original specification for NpFIT, moved radiology images from film-based processing to electronic recording. The systems were generally set up on a regional basis as a shared service to hospitals (with feeds to GP practices as well) for economies of scale. In this case, the technology was well developed if not widely adopted: a few leading sites (mainly in the USA) had taken on such facilities and proved them to

electronic booking system to enable a GP to book a specialist appointment for a patient while the patient was still in the surgery, were much less so. One feature strongly promoted by Connecting for Health was the Summary Care Record (SCR), a nationwide system containing a GP record summary (initially, current prescriptions and allergies) that would facilitate out-of-hours care and could also enable patients to view their own records via a mechanism called HealthSpace.

Box 6.1: The Summary Care Record

The first national care record to be implemented was the Scottish Emergency Care Summary (ECS), starting in 2002 and achieving near complete coverage by 2007.³¹⁸ The role of the ECS was simply to allow the most immediate medical details to be available in an emergency, by out-of-hours GPs, clinicians at Accident and Emergency (A&E) departments in hospitals, or call-centre staff at NHS24.

It was limited to basic allergies, adverse drug reactions, and current (<12 months) medications data. Patients were able to opt out and would be asked for 'consent to view' the record by clinicians treating them.³¹⁹

The introduction of the Summary Care Record (SCR) in England was more ambitious, carried out on a reduced timescale (2½ years) as the first stage in a move to full electronic health records and involved a wholesale replacement of almost all NHS systems as part of the National Programme for IT. Initially it would contain a summary of information from the patient's GP system and referral and discharge correspondence (though not clinical details), and, later, Common Assessment Framework (CAF) care plan documents to permit 'joined up' care delivery between health and social care services.

The decision to provide access to the system through a browser-based stand-alone application raised security and confidentiality concerns as it separated the access to the SCR from any local governance process. It also created some confusion between the SCR as a record and the SCR 'application', and how each might be accessed.

The SCR has been bedevilled by technical and policy difficulties that derived largely from the scale of the ambition and the number of variables, including arguments about what options individuals might exercise (which information could be uploaded from detailed care records, including, for example, 'sealed'/'sealed and locked' envelopes), the legitimacy of default options (the RCGP and BMA argued for an 'opt in' approach, for example), and how individuals' choices could be given effect through the technical architecture of the system.

be effective. The technology also required relatively little change to clinical practice (calling up an image on screen was often easier than obtaining the film) but it enabled a number of further enhancements. It raised the possibility of rationalising radiology reporting services across hospitals, rather than requiring each hospital to have its own staff; teams could be merged, relocated, or even outsourced if necessary, though this led to important debates about patient understanding of how their images might be shared and whether reduction of clinician-radiologist interaction mattered.

³¹⁸ Although this was the first national record, the Hampshire and Isle of Wight Care Record was implemented even earlier.

While all 14 Health Boards participated, in 2007 some 7 GP practices had chosen not to participate, so their patients' records were not present. It is not clear if this was for reasons of principle or practicalities, such as incompatible systems.

³¹⁹ Nevertheless the system was the subject of a high-profile confidentiality failure when a doctor accessed the records of a number of public figures. See: "Medical records of Gordon Brown and Alex Salmond hacked" Daily Record, 1 March 2009, available at: <http://www.dailyrecord.co.uk/news/scottish-news/medical-records-of-gordon-brown-and-alex-1011941>.

6.18 However, a review of the SCR in 2010 found little evidence of benefit. It found that even staff in Accident and Emergency services did not often access it and that patients also made little use of HealthSpace.³²⁰ A significant number of people also opted out of the SCR owing to concerns about data would be handling.³²¹ (Criticism that the patient information circulated in pilot areas made opting out difficult led the campaign 'The Big Opt Out' to produce an opt-out letter, which was extensively downloaded.) The NPfIT was officially dismantled in 2013 with major objectives unachieved and amid public criticism of management failure and cost overrun.³²²

Box 6.2: Connecting for Health – an assessment

NPfIT has been described as 'the largest ever civilian IT project failure in human history'.³²³ It has been documented by many articles in the computing and health IT press and by successive parliamentary committee inquiries. A recent case history classifies the problems according to three main themes:³²⁴

- **Haste.** Politicians and programme managers rushed policy making, procurement and implementation processes, allowing insufficient time for consultation with key stakeholders and failed to deal with confidentiality concerns;
- **Design.** The government pursued an overambitious and unwieldy centralised model, without giving consideration to how this would impact user satisfaction and confidentiality issues; and
- **Culture and skills.** NPfIT lacked clear direction, project management and an exit strategy, meaning that the inevitable setbacks of pursuing such an ambitious programme quickly turned into system-wide failures. Furthermore, the culture within the Department of Health and government in general was not conducive to swift identification and rectification of strategic or technical errors.

One further problem with the centralised commissioning and ownership of health IT is that IT suppliers respond to the priorities of the customer rather than the clinical user of the system, which allowed the collection of data for central use to compete with the objective of improved care or efficient local service delivery.³²⁵

³²⁰ Greenhalgh T, Stramer K, Bratan T, *et al.* (2010) *The devil's in the detail: final report of the independent evaluation of the Summary Care Record and HealthSpace programmes*, available at: <http://www.ucl.ac.uk/news/scriefullreport.pdf>. This was equally true of patient-centred health record offerings from Google and Microsoft: the former has been closed and the latter developed into a more conventional 'white label' product. For a discussion of online personal health records, see Nuffield Council on Bioethics (2010) *Medical profiling and online medicine: the ethics of 'personalised healthcare' in a consumer age*, available at: <http://nuffieldbioethics.org/project/personalised-healthcare-0/>.

³²¹ As of 18 December 2014 the number of people who had opted out was 528,034 according to the HSCIC (see: <http://www.hscic.gov.uk/article/2220/FOI-disclosure-log>).

³²² House of Commons Public Accounts Select Committee (2013) *The dismantled National Programme for IT in the NHS* (Nineteenth Report of Session 2013–14) HC 294 (London: TSO), available at: <http://www.parliament.uk/business/committees/committees-a-z/commons-select/public-accounts-committee/news/npfit-report/>.

³²³ See: The NHS's national programme for information technology: a dossier 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/cmpublicacc/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-to-have-failed-but-it-is-no-surprise.html>.

³²⁴ A recent case history classifies the problems according to three main themes. See: Champion-Awwad O, Hayton A, Smith L and Vuaran M (2014, MPhil Public Policy 2014, University of Cambridge) *The national programme for IT in the NHS – a case history*, available at: <http://www.cl.cam.ac.uk/~rja14/Papers/npfit-mpp-2014-case-history.pdf>.

³²⁵ Foundation for Information Policy Research, Evidence submitted by the Foundation for Information Policy Research (EPR 61), Select Committee on Health, 15 March 2007; at <http://www.publications.parliament.uk/pa/cm200607/cmselect/cmhealth/422/422we22.htm>.

6.19 The experience of the NPfIT may, nevertheless, be salutary for health care data initiatives more generally because it highlights the risks of external drivers overtaking the establishment of data initiatives (see chapter 2) and of lack of involvement or imbalance of key interests, and the need adequately to address values and norms relating to confidentiality (chapter 5). These lessons are important, moreover, because the SCR service, along with many of the other projects previously overseen by Connecting for Health, have themselves continued, with responsibility for their delivery having been passed to the Health and Social Care Information Centre (HSCIC).

The Health and Social Care Information Centre

6.20 The Health and Social Care Information Centre (HSCIC) has replaced Connecting for Health as the Department of Health's agency for driving the implementation and use of health IT for business intelligence and to equip the NHS to be a learning health system.

Box 6.3: The Health and Social Care Information Centre

The Health and Social Care Information Centre (HSCIC) is an executive non-departmental public body that took its current form following the Health and Social Care Act 2012 (HSCA). The HSCIC was created with the intention of being a national focal point for information collection across health and social care that is responsible for collecting, transporting, storing, analysing and disseminating the nation's health and social care data.³²⁶ Continuing the work of its predecessor, the NHS Information Centre, and absorbing continuing elements of the Connecting for Health programme (e.g. the NHS Spine, the National Back Office and SUS), a key aim of the HSCIC is to provide a trusted 'safe haven' for health data.³²⁷ NHS health data comprises approximately a quarter of this data; the majority is social care data.

The stated aim of the HSCIC is to 'revolutionise' the ability to unlock NHS and care data. It has been given the legal and administrative power and responsibility to:

- Collect information from health and social care bodies
- Hold that information within a secure environment
- Make that information readily available for others to turn into "actionable business intelligence"³²⁸

At present, the HSCIC collects a range of information, including (monthly) Hospital Episodes Statistics (HES) which relate to in- and out-patient appointments, and accident and emergency admissions. This is intended to be supplemented by primary care data through NHS England's 'care.data' programme (see below).

6.21 The intention is that the HSCIC will hold comprehensive and integrated information about the care patients receive from all parts of the health service, including hospitals and GP practices. It is hoped that by collecting this information and analysing it,

³²⁶ The responsibility to provide information under the Health and Social Care Act 2012 was analysed in Grace J and Taylor MJ (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)**: 415-47.

³²⁷ Health and Social Care Information Centre (2014) A strategy for the Health and Social Care Information Centre 2013-2015, available at: <http://www.hscic.gov.uk/media/13557/A-strategy-for-the-Health-and-Social-Care-Information-Centre-2013-2015/pdf/hscic-strategy-2014.pdf>.

³²⁸ Health and Social Care Information Centre, 2012. Exploiting the potential of health and care data, available at: <http://www.hscic.gov.uk/ourrole>.

researchers can compare the safety of different NHS hospitals, monitor trends in different diseases and treatments and use the data to plan new health services.

6.22 There are two levels of information release relevant to the HSCIC: general publication (i.e. open data) and limited access.³²⁹ Data disclosed for limited access may include potentially identifying individual-level data or, with the subjects' consent or if it is in the public interest, identifying data. Disclosures are governed by agreements about purposes for which the data will be used, who will have access, and whether they may be sold on to third parties.³³⁰ (In theory, the HSCIC could audit compliance with these agreements but in practice it does not and this does not seem to be budgeted for in the costs recovered through the charges made for data extracts.) The HSCIC also offers to link its own data to data supplied by the customer (see Box 6.5 below).

Box 6.4: Data available from HSCIC

Product	Description
Tabulation	A statistical table of aggregate data.
Bespoke extract - pseudonymised	A one-off extract tailored to the customer's requirements of specified data fields containing patient identifiable data, sensitive data items or both.
Standard extract	Cumulative data for the financial year to date, delivered on a monthly basis via a subscription service. Users sign up to receive a year's worth of data, delivered in monthly increments
Bespoke data linkage	A bespoke service linking one or more datasets held by the HSCIC to data supplied by the customer.
Patient status and/or tracking	Products designed to enable customers to receive one-off or on-going notifications of mortality and morbidity events affecting a specified patient cohort.
List cleaning	Validating demographic data to ensure it is accurate and improve linkage outcomes.

Source: Data Linkage and Extract Service: Service Charges 2013/14³³¹

6.23 While HSCIC will not exclude particular organisations from using their data, access to datasets will only be granted if the data will be used broadly for purposes beneficial to health or health care.³³² In providing data, the HSCIC operates on a 'cost recovery' basis: it does not charge for data itself but applies charges to cover the costs of processing and delivering its service. The cost is determined by the amount and type of data required (between a few hundred and a few thousand pounds). The HSCIC takes

³²⁹ The Health and Social Care Act 2012 (s.260) imposes a general duty on the HSCIC to publish data that it collects, although identifying data are exempt from this requirement.

³³⁰ 'commercial reuse licenses' enable firms to resell personal health information to third parties. See also the HSCIC register of approved data releases: <http://www.hscic.gov.uk/dataregister>.

³³¹ Available at: www.hscic.gov.uk/media/12443/data-linkage-service-charges-2013-2014-updated/pdf/dles_service_charges_2013_14_V10_050913.pdf.

³³² This was clarified in the Care Act 2014, s.122; which amends s.261 of the Health and Social Care Act 2012 to specify that the HSCIC may only further disseminate information for the purposes of 'the provision of health care or adult social care' or 'the promotion of health'. The breadth allowed of these purposes was debated during the passage of the Act.

advice on data access from its Data Access Advisory Group (DAAG), which makes available details of all approved projects that utilised HSCIC's datasets containing 'sensitive' data. Sensitive data may include a patient's NHS number, postcode, date of birth and/or death, physical and mental health, etc. This arrangement was later fortified, but not before the HSCIC had been exposed to considerable political turmoil.

6.24 As we have argued in this report, questions about the terms under which information may be collected and disclosed by the HSCIC need to be answered by first establishing the norms of access and disclosure that govern the kinds of information transactions involved. These questions found a public focus in the debate that arose around the programme to extract GP data for inclusion in the HSCIC database. This brought many GPs and civil society groups into conflict with NHS England, resulting in delayed implementation and redesign of the initiative.

Box 6.5: The NHS England 'care.data' programme

Care.data is a programme commissioned by NHS England and promoted with the caption 'better information means better care'. Its purpose is to bring together routinely collected information from NHS organisations within the HSCIC. Care.data has been particularly focussed on the extraction of data from primary care records (GP data) and widening the range of data collected from secondary care (hospital data), to add to the hospital episode statistics (HES) data already routinely collected by HSCIC.³³³

The collection of GP data is through a new General Practice Extraction Service (GPES), which extracts information stored electronically within GP practices and sends it to the HSCIC.³³⁴ This information does not include patient names but does include each individual's NHS number, date of birth and full post code, as well as the patient's history of diagnoses, prescriptions, vaccinations, etc. At present, particularly sensitive information (e.g. pregnancy termination) and handwritten or 'free text' notes will not be extracted.

In the future, it is anticipated that care.data will join up with other NHS projects, for example, allowing phenotypic data to be linked to the genomic data produced by the 100k Genome Project being delivered by Genomics England Ltd (GeL).³³⁵

³³³ In Scotland, the mandated use of the CHI Number in relation to all health episodes, the comprehensive computerisation of routine clinical data and the centralisation of NHS data has allowed the Scottish Informatics Programme (SHIP) to develop a platform for EPR research involving the linkage of health and other records (now based at the Farr Institute @ Scotland (http://www.farrinstitute.org/centre/Scotland/3_About.html). Though different (and clearer) in its aims from care.data, SHIP has had a less troubled development than care.data, involving more academic reflection and a programme of public engagement in relation to its governance arrangements.

³³⁴ See: <http://www.hscic.gov.uk/gpes>. This was originally due to begin in Spring 2014 but was delayed owing to opposition from GPs and civil society groups, with a pilot finally announced in late 2014. A function of GPES is to feed the Quality and Outcomes Framework (QOF) used to calculate a significant part of GPs' remuneration. QOF payments are made for meeting multiple targets for activities from screening through immunisation, and were previously claimed using paper forms. The proposal that QOF payments be computed automatically on live patient data that would be subjected to bulk upload without the practical possibility of an opt-out for either GPs or patients led to widespread protest. GPES is owned and managed for the NHS by the HSCIC and has an Independent Advisory Group (IAG), which considers applications to extract data from GP clinical systems from other organisations (e.g. NHS England, Public Health England) according to a defined approvals process (see: <http://www.hscic.gov.uk/article/3472/Custom-requirements>).

³³⁵ Indeed, Sir John Chisholm, Executive Chair of GeL, has recently become a Non-Executive Director at HSCIC. GeL state that it is intended that the findings of the 100k Genome Project will be linked with identifiable data from primary care and hospital records and that this can be linked to relevant clinical data. See: 'On the progress and outcomes from the NHS England Genomics Strategy Board and the genetics lab reconfiguration', <http://www.england.nhs.uk/wp-content/uploads/2013/07/180713-item16.pdf>. See also: Tim Kelsey, *The Guardian* (5 November 2013) *Five ways to enable transparency in the NHS*, available at: <http://www.theguardian.com/healthcare-network/2013/nov/05/transparency-operating-principle-nhs>.

6.25 A number of arguments have been deployed to establish the moral reasonableness of the care.data programme. The example is highly instructive because the sites of these arguments have moved from the extension of implicit norms, through attempts to rebalance these in legislation, to a public account of political decision making.

The moral justification of the 'care.data' programme

6.26 When care.data was initially proposed it was framed as an extension of previous programmes, such as the Secondary Uses Service (SUS) initiated by Connecting for Health.³³⁶ Analogies can also be drawn to other programmes such as the Clinical Practice Research Datalink (see paragraph 6.9 above), which already links a number of the same data sources that will be linked by the HSCIC and which has been implemented without critical attention. A first argument, then, is that if there is no morally significant distinction between these initiatives, and others have been accepted, then to object to care.data would be inconsistent.

6.27 Any argument by analogy to previous initiatives is, however, only as robust as the justification for those previous initiatives. The identification of an inconsistency in attitudes between care.data and SUS or CPRD (if it is an inconsistency) does not indicate which (if either) of the inconsistent beliefs is correct. Care.data attracted a high level of media attention whereas SUS did not: however, it might be that people would have objected to SUS more strongly or in greater numbers if there had been a greater level of awareness of it.³³⁷ This raises important questions about how those norms were established (questions that we suggest may be addressed by our third and fourth principles of participation and accounting for decisions— see chapter 5). It has since become common ground that there was insufficient communication and consultation with key stakeholders prior to the planned introduction of the care.data programme.³³⁸

6.28 A second potential weakness of this argument is the robustness of the analogy. That is, whether the HSCIC's proposed activities are constrained within the scope of those previously accepted (in the case of SUS, for example) and established in data protection law (see below) or, on the other hand, whether they transgress previously accepted moral and, arguably, legal norms of privacy and disclosure. Crucial in this respect is the question of who might have access to data or might receive extracted data from HSCIC, and for what purposes (which has been taken up through legislative action to re-establish a legal norm and associated mechanisms).³³⁹

6.29 A second site of justification, then, concerns the legal norms applicable to care.data, specifically those of data protection and data sharing. The cornerstone of data protection law, that data processing should be fair, requires reasonable steps to be taken to give notice to the data subject of the purposes for which their data will be used.³⁴⁰ In August 2013 NHS England maintained that it was the responsibility of GPs,

³³⁶ See <http://www.connectingforhealth.nhs.uk/systemsandservices/sus>. It is interesting that care.data is widely seen as controversial where SUS was not. Indeed, it has been argued that care.data became controversial only because the Secretary of State, in undertaking to honour the choice to 'opt out', actually gave prominence to the question of choice, and that the media became widely engaged with the issue.

³³⁷ In this case, the argument turns in the opposite direction and becomes a criticism of previous failings to inform the public.

³³⁸ See: <http://www.england.nhs.uk/2014/02/19/response-info-share/>; <http://www.england.nhs.uk/wp-content/uploads/2014/04/cd-stakeholder-lett.pdf>.

³³⁹ See the Care Act 2014, discussed below.

³⁴⁰ DPA 1998, Sched.1, Part II. See also: Grace J and Taylor MJ (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): 415-47.

as controllers of primary care patient data, to inform their patients about the data extractions. Owing to the extraordinary demands this would place on GPs' resources, an attempt was made to inform patients through a NHS England leaflet, 'Better information means better care', which was sent to households in England.³⁴¹ This was, however, widely criticised for its uninformative design, its limited circulation and its resulting lack of success in generating awareness; it did little to persuade many GPs that patients had been adequately informed.³⁴² This information would be all the more important if patients were to be able meaningfully to exercise the opt-out they had been promised.³⁴³ This notwithstanding, however, the Health and Social Care Act 2012, in effect, mandated the submission of data to the HSCIC. The effect of the two Acts placed GPs at the intersection of potentially inconsistent requirements – not to submit data without giving adequate notice to the data subjects and to submit data to the HSCIC – that each invoked norms of privacy and public interest.³⁴⁴ On the eve of the first extractions, pressure from GP bodies and civil society groups caused a postponement of the data extraction to allow time to 'build understanding' of the aims and benefits of care.data.³⁴⁵

6.30 A third site of justification is the public account given of the conditions to which care.data was the proposed response, conditions that implicitly entailed the recognition of new norms in the light of which the care.data approach would be justified. Alongside considerations of individual privacy, there is a public interest in making use of data if doing so results in more efficient health service planning and delivery, better treatment and the development of scientific knowledge. Other things being equal, there is

³⁴¹ This was a *volte-face* from the position in August 2013, when the NHSE maintained that it was the responsibility of GPs to inform their patients about care.data. For the leaflet, see: <http://www.england.nhs.uk/wp-content/uploads/2014/01/cd-leaflet-01-14.pdf>. The leaflet did not go to households that have registered with the Royal Mail's 'door to door opt-out'. However, it was delivered to households using the Mail Preference Service (see: https://www.whatdotheyknow.com/cy/request/royal_mail_contract_for_caredata).

³⁴² In relation to its design, it was criticised for biased presentation of the benefits and risks of care.data, for failing to provide adequate information about how to opt out, and for not taking advice (Dame Fiona Caldicott, speaking to the BBC Radio 4 PM programme, said that NHS England had gone ahead without waiting for her IIGOP committee's advice on the leaflet (see: <http://www.bbc.co.uk/news/health-27069553>). It was also criticised in terms of its circulation (for example, an ICM survey for the BBC found that only 29% of 860 adults polled recalled receiving the leaflet and about 45% of people remain unaware of the plan to share some data from GP medical records; see: <http://www.bbc.co.uk/news/health-26187980>). Further criticisms included that the leaflet was not accessible for people with visual impairments and some others, and that it did not include the term 'care.data' anywhere.

³⁴³ Patients were promised the opportunity to opt out of the programme by the Secretary of State for Health in the wake of the publication of the 2012 Information Governance Review report. As the HSCIC privacy impact assessment acknowledges, "patients have a right to object to personal information about them being collected and used by the Health and Social Care Information Centre" (http://www.hscic.gov.uk/media/12931/Privacy-Impact-Assessment/pdf/privacy_impact_assessment_2013.pdf, at page 16). In law, this 'opt out' is constituted by a right to object to the use of data (with the ICO as an arbiter), which, pursuant to the Secretary of State's undertaking, would be automatically upheld. Even in June 2014, however, an Ipsos MORI poll of 1,958 UK adults commissioned by the Joseph Rowntree Reform Trust found that 51% of respondents said they had never heard of the scheme and 13% had heard of it but did not know what it was. On the substantive issues, a small majority were opposed to the 'opt out' basis on which care.data is proceeding, with 40% saying that their GP should only be allowed to share their data with explicit consent, and 13% saying that said data shouldn't be shared by GPs under any circumstances. (This compares with 27% who said GPs should be able to share patient records if they had been informed and given a chance to opt out and 10% who said that they should be shared even without informing them.) See <http://www.jrrt.org.uk/sites/jrrt.org.uk/files/documents/IpsosJRRTPrivacypollMay2014full.pdf>.

³⁴⁴ It is important to appreciate, as we have argued, that norms of privacy and public interest are implied in both the applicable Acts (i.e. it is not the case that one defends individual privacy and the other the claims of public good), although arguably to different effect. This is also the case with the Care Act 2014. This is not inconsistent with the proposition that legislative action in this case is attempting to affect moral norms rather than merely to reflect them.

³⁴⁵ In a letter to NHS England on the eve of the first intended extractions, the RCGP said: "While we recognise the substantial programme of activity and materials that has already been developed to communicate care.data, we believe that there is a deficit of awareness and understanding regarding the scheme amongst many members of the public and professionals." (Letter from RCGP Honorary Secretary, Professor Nigel Mathers, to NHS England (18 February 2014), quoted in RCGP news release; see: <http://www.rcgp.org.uk/news/2014/february/rcgp-calls-for-reassurances-before-controversial-data-scheme-goes-ahead.aspx>). See Pulse (17 February 2014) *GPC calls for urgent talks over public awareness of care.data scheme*, available at: <http://www.pulsetoday.co.uk/your-practice/practice-topics/it/gpc-calls-for-urgent-talks-over-public-awareness-of-caredata-scheme/1/20005884.article?&PageNo=1&SortOrder=dateadded&PageSize=20#.VLZvx3uj9Oc>.

likewise a public interest in using health data to generate economic growth by stimulating economic activity around it, and generating revenue or saving costs for the health service. The proponents of care.data have argued that there is such an overwhelming public interest in the programme that privacy interests should be qualified proportionately.³⁴⁶ This argument has been stated in terms no less than that 'the future survival of the NHS depends upon it.'³⁴⁷

- 6.31 The cogency of this argument depends on accepting its premises. These include (1) that changing conditions require such a measure to improve the efficiency of health care so that it can continue to be provided as a public good; (2) that the proposed approach is likely to contribute to this objective in the way intended³⁴⁸; and (3) that the proposed approach is preferable to all alternatives.³⁴⁹ These premises rest on a number of further factual, political and speculative claims. It is empirically true, for example, that, with ageing populations and increasing technological intensity, inflation in health care costs has outstripped the rise in other prices worldwide. Nevertheless, UK health care is largely funded through taxation and the adequacy of resources is substantially a matter of political decision rather than a hard constraint. It is likely, therefore, that the NHS will continue to be caught in major political arguments about funding and resource allocation, with or without the business intelligence and advances in treatment promised by the HSCIC.³⁵⁰

Moving the arguments forward

- 6.32 With moral and, arguably, also legal norms unresolved a second layer of debate concerned how individual freedoms should be respected. The default setting (choice of opt-out or opt-in) and the level of public information became a new focus of debate. The justification of default inclusion might be seen as extending the norm of social solidarity that is conventionally assumed to underwrite the provision of public health care. However, the content of those norms is not well developed or widely discussed (as the previous section shows). For example, the 'imagined community' to which these norms are referred is often taken to exclude commercial firms. However, disclosure of data to commercial firms is within the intentions of the HSCIC, so long as their objects are consistent with the purposes specified in law.³⁵¹

³⁴⁶ See ECHR, Art.8(2).

³⁴⁷ See evidence to House of Commons Health Select Committee, February 2014. Tim Kelsey: "My view is that it is one of the most important public debates we are having and it is about the future of the health service." (Oral Evidence: Care.data database, HC 1105, 25 February 2014, available at: <http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/health-committee/handling-of-nhs-patient-data/oral/6788.pdf>, at page 34).

³⁴⁸ We noted in chapter 2 that the policy discourse has been consistently ambitious. For example, a 2014 report from a big data solutions company claims that better use of data analytics could free £16.5 billion and £66 billion worth of NHS capacity.

³⁴⁹ Some have argued that if the same project were set on a voluntary basis people would decline to participate in such numbers that the data would be significantly less valuable. A major concern of those who opposed an opt-in approach is that it would disproportionately disadvantage those who are most vulnerable and in need of its benefits (i.e. those who have low social power and status, and are already disadvantaged, including the elderly and those with long-term illness and disability). The second premise has been challenged, however: it can be argued that a certain level of non-participation (whether through 'opting out' or choosing not to 'opt in') would be tolerable and still allow the aims of the programme to be met. (See: evidence to HC Health SC, February 2014, <http://www.parliament.uk/business/committees/committees-a-z/commons-select/health-committee/news/14-02-25-cdd-ev/> and <http://www.parliament.uk/briefing-papers/SN06781/caredata>. On mandatory inclusion and the (un)acceptability of 'free riding', see paragraph 5.12.

³⁵⁰ See Nuffield Council on Bioethics (2014) *Funding pressures in the NHS: an ethical response* (Forward Look background paper), available at: http://nuffieldbioethics.org/wp-content/uploads/NHS_resource_pressures_final_FL_paper1.pdf.

³⁵¹ On 'imagined communities', see Busby H and Martin P (2006) Biobanks, national identity and imagined genetic communities: the case of UK biobank *Science as Culture* 15(3): 237-251. As we noted, the research and innovation system is complex, so

- 6.33 The question of who might have access to the data and to whom it might be disclosed in turn raised the question of governance of data use by the HSCIC. This turned the spotlight on the organisation's decision making procedures and their (or their predecessor, the NHS information Centre's) management of data sharing agreements. HSCIC provided little clarity on how data would be used in the run up to the initial launch date of the care.data extraction programme and, indeed, had not at that time published a code of practice.³⁵² Civil society groups opposed to the proposed introduction of care.data and the media began to draw attention to cases in which individual-level health data had been disclosed widely (including by the HSCIC's predecessor body, the NHS Information Centre) in ways that presented re-identification risks.³⁵³
- 6.34 In these ways, the key moral elements of a data initiative that we have discussed in this report, namely, the relationship between the underlying moral norms (including their relationship to legal norms), the way of respecting individual moral interests and values, and the way in which decisions are made and accounted for, finally became explicit and the subject of public and political discourse.

Solutions for the HSCIC

- 6.35 Care.data has raised to public prominence a debate about the justification, beyond the purely legal basis, of the HSCIC's collection and release of data. According to our approach, outlined in chapter 5, the moral question that confronts the HSCIC is to define a public set of morally reasonable expectations about the use of data generated by health and social care services. This should be done within a framework of principles that takes the mutual implication of public and private interests that we discussed in chapter 3 into account.
- 6.36 The first task should be to identify the relevant norms. There may have been an assumption that there were no novel features and therefore that the norms of social and health care, of the 'presumed broad consent' of patients to the use of data within the NHS, would apply unaltered to the HSCIC initiative. However, the HSCIC initiative appeared to want to keep the scope of potential uses broad, in the spirit of treating data as a resource with multiple and undefined potential uses. It is not at all clear that all these possible uses would fall within the expectations of patients, and the reaction of GPs, civil society groups and the media demonstrated this. Indeed, findings from public opinion research suggest, as we noted in chapter 5, that while there is broad public support for some further uses of care data, such as biomedical research, many individuals still want to be asked about this, and there are other uses, such as commercial uses, that are viewed with suspicion.³⁵⁴ It may be that these views would

that even if medical research is carried out in English universities on an entirely non-profit basis, the commercialisation of any drugs discovered will almost inevitably involve the pharmaceutical industry.

³⁵² A code of practice has now been published by the HSCIC whilst this report was being finalised: HSCIC (2014) *Code of practice on confidential information*, available at: <http://systems.hscic.gov.uk/infogov/codes/cop/code.pdf>.

³⁵³ Complaints were made to the ICO regarding the publicised 'Staple Inn incident' where individual-level data were apparently released to insurance actuaries. See: Information rights and wrongs blog (24 February 2014) *Hospital records sold to insurance companies – in breach of the data protection act?*, available at: <http://informationrightsandwrongs.com/2014/02/24/hospital-records-sold-to-insurance-companies-in-breach-of-the-data-protection-act/>. A review of data releases made by the NHS IC was subsequently carried out by a non-executive director of the HSCIC, Sir Nick Partridge. The review identified a number of incidents relating to the previous body and also to HSCIC; some refer to ongoing practices at HSCIC and made recommendations, which have been accepted by the HSCIC. See: <http://www.hscic.gov.uk/datareview>.

³⁵⁴ Hill EM, Turner EL, Martin RM, and Donovan, JL (2013) "Let's get the best quality research we can": public awareness and acceptance of consent to use existing data in health research: a systematic review and qualitative study *BMC Medical Research Methodology* 13: 72, available at: <http://www.biomedcentral.com/1471-2288/13/72>.

not withstand further reasoning: we noted in chapter 2 that the involvement of commercial firms is a feature of the innovation system on which the medical developments that people want rely. It may be that increasingly broad use of data will play an important role in a sustainable future for the NHS. Such arguments gesture to the way in which private interests may be implicated in the broader public interest. But there was no open debate of these issues and arguments in public. One place in which these questions were opened up to deliberation, albeit belatedly, was through the ‘autonomous’ advisory group, which was set up when the planned GPES extractions were postponed and which includes a number of NGOs and other stakeholders.³⁵⁵ Broader consultation was also promised which may help at least to identify norms specifically relating to the proposed uses, as will the ‘pathfinder’ exercises in four areas, which will examine ways of communicating with patients to make the ability to opt out meaningfully exercisable.³⁵⁶

- 6.37 As we have noted in this report, the existence of a framework of legal norms offered by various legal instruments (governing data protection, the duties of public health care services, and human rights – discussed in chapter 4) does not entail complete moral freedom to act within it. While these offer a starting point, they are clearly not sufficient and may even come into conflict.³⁵⁷ (No more can moral norms be shifted simply by passing new legislation: the law and public morality must correspond with each other, procedurally as well as structurally.) Though the Care Act 2014 was intended to clarify the purposes for which HSCIC data may be used, ‘the provision of health care or adult social care, or the promotion of health’ merely sets apparently more limited but by no means more precise criteria for decision making.³⁵⁸ What the Care Act 2014 (and currently awaited secondary legislation) does recognise is the need for further moral guidance on data access for which the HSCIC, as an executive agency, was not itself resourced. The solution it provides is to take advice from the HRA Confidentiality Advisory Group (CAG), which, through its predecessors, the Patient Information Advisory Group (PIAG) and the Ethics and Confidentiality Committee of the National Information Governance Board (ECC), has experience with the questions of data access, moral norms and reasonable expectations, with which the HSCIC is faced.³⁵⁹ The elusive – and possibly illusory – significance of the difference between individual-level data and identifying data, as we have argued in this report (see chapter 4), makes the reliance on the Confidentiality Advisory Group (CAG) for advice on disclosures a natural link in the scheme of governance as it presently exists. This is because, in effect, the question that must be posed for all disclosures of individual-level data (and a lot of aggregate data too) is the one that CAG is implicitly established to address, namely, whether the proposed disclosures would fall within the scope of reasonable expectations of disclosure. This advisory system is also buttressed by the appointment

³⁵⁵ See: <http://www.england.nhs.uk/ourwork/tsd/ad-grp/>.

³⁵⁶ See: <http://www.england.nhs.uk/wp-content/uploads/2014/10/caredata-pathfinder-proposal.pdf>.

³⁵⁷ The position of GPs with regard to the DPA v. HSCIC, see paragraph 6.29.

³⁵⁸ Care Act 2014, s.122(3), which amends section 261 of the Health and Social Care Act 2012. The scope of this amendment was debated during the passage of the Act: “Here the current wording of the Commons Amendment, although well meant—we are pleased to have it—leaves open too many questions for interpretation. Their amendment suggests that use of patients’ data may be allowed for “the promotion of health”. This leaves us open to two types of possible interpretation that may be undesirable. For example, “promotion” could be taken to mean that food manufacturers could use data in their marketing campaigns for so-called healthy foods. That may or may not be desirable but it would put many off if it appeared that their data were being used for commercial gain in a competitive market.” per Lord Turnbull, HL Hansard, (7 May 2014) c1514.

³⁵⁹ New Economics Foundation (NEF) research found public opposition to the use of s.251: see recommendation 5 of http://b.3cdn.net/nefoundation/2cb17ab59382fe7c67_bfm6bdoas.pdf.

of Dame Fiona Caldicott as the first National Data Guardian for health and care, a role described as ‘the patients’ champion on security of personal medical information’.³⁶⁰

6.38 While this entails that particular decisions of the HSCIC will receive support, it does not make any more explicit the scope of what are reasonable expectations, nor does it require an engagement with interested participants (although CAG is very likely to take account of social research to inform its own reasoning). Rather it suggests that the norms will be established and elaborated through developing expertise and precedent. This has the advantage of flexibility to respond to the evolving debate. However, it fails to make any clearer to the public what expectations they may have about who will have access to individual-level NHS data or for what purposes, and therefore from what they might be opting out when they consider whether to do so.

Recommendation 6

We recommend that an independent group of participant representatives is convened to develop a public statement about how data held by the Health and Social Care Information Centre should be used, to complement the Code of Practice on confidential information. This should clearly set out and justify what can reasonably be expected by those from whom data originate and be able to demonstrate that these expectations have been developed with the participation of people who have morally legitimate interests in the data held by the HSCIC, including data subjects, clinical professionals and public servants.

Recommendation 7

We recommend that, in addition to implementing the recommendations of Sir Nick Partridge’s review, all Data Sharing Agreements held by the HSCIC should be published, along with the findings of a periodic independent audit of compliance with those agreements.³⁶¹

Recommendation 8

We recommend that HSCIC Data Sharing Agreements should include a requirement to maintain an auditable record of all individuals or other legal entities who have been given access to the data and of the purposes to which it has been put; this should be made available to all data subjects or relevant authorities in a timely fashion on request.

³⁶⁰ See: <https://www.gov.uk/government/news/national-data-guardian-appointed-to-safeguard-patients-healthcare-information>.

³⁶¹ Sir Nick Partridge (2014), Review of data releases by the NHS Information Centre, available at: <http://www.hscic.gov.uk/datareview>. The recommendations included that HSCIC develops one Data Sharing Agreement to be used for all releases of data, and which includes clear sanctions for any breaches.

The Scottish Informatics Programme and the Farr Institute

- 6.39 In many ways the Scottish approach to the question of the reuse of data from health care has been the inverse of the English experience. The Scottish authorities began with public engagement to determine acceptability (rather than being forced to engage after the fact) and developed an integral governance approach along with the informatics infrastructure (rather than having to invent one to fill a vacuum between the legal framework and operational decision making), building public trust rather than undermining it. The Scottish Health service is, of course, much smaller than the English NHS, and organised and run differently. Data linkage is facilitated by the almost universal use of the CHI in Scotland, and the Scottish experience of moving to electronic patient records (EPRs) has been both more measured in ambition and smoother in execution than in England.
- 6.40 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 linked EPRs, and to establish a research arm within the Information Services Division (ISD) of NHS National Services Scotland and to support the efficient functioning of public services more generally. It provides for linkage not only of care data but also data that have been gathered in cohort studies and other forms of publicly-held administrative data.³⁶²

Box 6.6: Health record linkage in Scotland

SHIP began in 2008 as a collaboration between the universities of Dundee, Edinburgh, Glasgow and St Andrews and the Information Services Division (ISD) of NHS Scotland. It was established with a £3.6 M grant from the Wellcome Trust, the Medical Research Council and the Economic and Social Research Council. The development of the infrastructure involved three workstreams:

- Public engagement to ascertain acceptability
- Legal review and development of proportionate governance approach
- Development of data linking methodology and associated IT infrastructure

In 2014 the programme was moved to the Farr Institute @ Scotland, one of four Farr centres across the UK (see paragraph 2.31 above).³⁶³

A central element of the infrastructure is a National Safe Haven.³⁶⁴ Unlike in the English HSCIC, data are not collected and held centrally; centralisation of data was found to be less acceptable during the prior public consultation. Instead, locally held datasets are linked within the safe haven for the purposes of specific analyses. The safe haven comprises three elements, provided by three separate organisations to increase privacy protection:

- An **indexing service** provided by National Records Scotland, which establishes links

³⁶² SHIP (2012) A blueprint for health records research in Scotland, available at: http://www.scotship.ac.uk/sites/default/files/Reports/SHIP_BLUEPRINT_DOCUMENT_final_100712.pdf. These requirements should apply more specifically in the areas of clinical trials, pharmacovigilance, diabetes epidemiology, and research resulting from the linkage of EPRs to socioeconomic and environmental data.

³⁶³ There is a Farr Institute – CIPHER – in Wales. For reasons of space we have not included a discussion of the situation in Wales, but the conditions in Wales are more akin to those in Scotland than those in England.

³⁶⁴ The SHIP Blueprint cites the 2008 Data sharing review report as an inspiration and SHIP's use of safe havens develops proposals within that report. See: The SHIP Blueprint 2012, available at: http://www.scotship.ac.uk/sites/default/files/Reports/SHIP_BLUEPRINT_DOCUMENT_final_100712.pdf.

- between cases in different datasets using a CHI look-up table or probabilistic linkage
- An **analytic platform** housed in the University of Edinburgh's Advanced Computing Facility, which holds the research dataset (without direct personal identifiers) and is accessed by researchers to conduct analyses
- A **researcher Advice and Disclosure Control Service (eDRIS)** provided by National Services Scotland NHS, which provides advice to researchers and statistical disclosure control.

Data linking and governance

- 6.41 Unlike in England where the HSCIC extracts data from local sources to a central data warehouse, in Scotland different 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 (as does Wales) applies precautionary de-identification measures by splitting and encrypting demographic and clinical data prior to linking. This is achieved in the following way. Individual-level data are sent by each contributing data source to an indexing service where identifiers are removed and a code assigned. A different code is assigned to each individual case for each data source, but an index of corresponding codes is produced by the indexing service, which links the different codes assigned to the same individual where they occur in different data sources. The coded data are then returned to the source and sent on to the national safe haven for analysis by authorised researchers. The researchers are sent the index of corresponding codes by the indexing service so that they can match up cases from each source. It is therefore only within the safe haven that the data from the multiple sources can be linked, as a result of knowing the corresponding codes assigned to cases in the different sources. Data from different sources (e.g. primary care, administrative, etc.) are linked centrally only for the purposes of a specific analysis; the linked data are held securely in the safe haven and destroyed as soon as practicable after the results have been obtained.³⁶⁵ This is apparently more complex than the HSCIC approach, but represents a balance between efficiency and acceptability (as determined with reference to prior public engagement).
- 6.42 One of the sources of data will be primary care records, although GPs in Scotland exercise more control over data extractions than is proposed under NHS England's care.data programme. The Scottish Primary Care Information Resource (SPIRE) will be built from Quality Outcomes Framework (QOF) data about GP activity, as in England (see note 336 to Box 6.6 above), a standardised set of data about patients using primary care services that GPs may choose to provide (via an opt-in mechanism), and other bespoke extracts obtained with the permission of GPs.³⁶⁶ As in England, GPs are in a crucial position and bear significant responsibility to protect their patients' interests alongside their own, although GPs in Scotland have not had to fight, as have English GPs, to assert their powers and responsibilities as data controllers.

Proportionate governance

- 6.43 The SHIP approach begins with an assumption that the public and medical doctors are in most cases expecting health data to be used for socially beneficial research, an

³⁶⁵ See also: http://www.isdscotland.org/Products-and-Services/Health-and-Social-Care-Integration/docs/IF_Framework_HSC_Integration_Route_Map_V1.7.pdf.

³⁶⁶ See <http://www.spire.scot.nhs.uk/>. SPIRE is expected to be operational from March 2015.

assumption supported by a dedicated research programme to review and build on existing evidence to this effect.³⁶⁷ The governance arrangements for data services are built around the concept of ‘proportionate governance’ and a stated set of guiding principles and related best practices.³⁶⁸ Among the guiding principles are protection of privacy and promotion of the public interest.³⁶⁹

- 6.44 ‘Proportionate governance’ denotes an approach to information governance in which the balance of risks and benefits and appropriateness of means to ends are central.³⁷⁰ It aims to improve on existing 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. It does not place reliance on any particular combination of anonymisation, consent or authorisation (see chapter 4). Furthermore, the approach is not focussed solely on the ‘kind’ of data in play but also, importantly, on the context in which it is in play. Rather than applying a ‘one-size-fits-all’ solution, risk assessment is used to determine what tools and pathways are appropriate in a particular case. Risk assessment 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.
- 6.45 The use of risk assessment explicitly orientates SHIP initiatives towards ‘proportionate’ governance as distinct from ‘precautionary’ governance.³⁷¹ This approach is possible where the context is one that is reliably controlled in such a way as to minimise risk (‘safe data, safe people, safe environment’).³⁷² A particular difficulty of assessing risk in this area is lack of evidence. It makes good sense to ground risk assessment in evidence because it offers an explicit, public reference point rather than remaining a matter of private opinion or judgement. However, there are two problems in this case, relating to uncertainties about the future conditions and about present facts (inductive and epistemological uncertainties). The first is a standard problem of induction, which points to the difficulty of using past evidence to make judgements about the future *when conditions or circumstances are changing significantly*.³⁷³ That is because it is the regularity of the circumstances that underwrites the inference about the future. We

³⁶⁷ See: SHIP public engagement work-stream, <http://www.scot-ship.ac.uk/c4.html> and SHIP (2011) SHIP public engagement: summary of focus group findings, available at: http://www.scot-ship.ac.uk/sites/default/files/Reports/Focus_Group_Findings_Briefing_Paper.pdf.

³⁶⁸ The ‘good governance framework’ has four key elements: (1) guiding principles and best practice, (2) proportionate governance, (3) roles and responsibilities of data controllers and (4) researcher training. See: Sethi N and Laurie G (2013) Delivering proportionate governance in the era of eHealth: making linkage and privacy work together *Medical Law International* **13(2-3)**: 168-204; Laurie G and Sethi N (2013) Towards principles-based approaches to governance of health-related research using personal data *European Journal of Risk Regulation* **1**: 43-57.

³⁶⁹ See the *SHIP blueprint*, Appendix 7, available at <http://www.scot-ship.ac.uk/publications.html>.

³⁷⁰ “Proportionality is the overarching principle that ties the varying components of good governance together and should be the ultimate benchmark against which to assess the appropriateness of conduct – both at the level of individual linkage decisions and the choice of what counts as appropriate governance over those linkages.” Laurie G and Sethi N (2012) Laurie G and Sethi N (2012) Information governance of use of health-related data in medical research in Scotland: towards a good governance framework, University of Edinburgh, School of Law Research Paper Series No 2012/13, available at: http://www.scot-ship.ac.uk/sites/default/files/Reports/Working_Paper_2.pdf, at page 12.

³⁷¹ A precautionary approach is often thought to be appropriate in conditions of significant uncertainty, where there is a potential for widespread and/or irreversible harm. A precautionary approach might be thought to entail that where there is any risk of re-identification, for example, then data should be treated as ‘personal data’ and, reliance on other grounds being insecure, the consent requirements of the Data Protection Act duly engaged.

³⁷² Laurie G and Sethi N (2012) Information governance of use of health-related data in medical research in Scotland: towards a good governance framework, University of Edinburgh, School of Law Research Paper Series No 2012/13, available at: http://www.scot-ship.ac.uk/sites/default/files/Reports/Working_Paper_2.pdf.

³⁷³ Hume D (175 [1748]) *Enquiries concerning the human understanding and concerning the principles of morals* (3rd Edition) (Oxford: Oxford University Press), section IV.

have already discussed the significant problem of rapid developments in data science and the accumulation of data; indeed it is this scale and pace of these developments that provoked this report. The SHIP approach of not retaining linked data and not disclosing raw data, limits the dimension of uncertainty that arises from holding linked data for indefinite time periods in changing circumstances. In such a context, a negligible risk of re-identification may obviate the need for additional consent from patients where it would otherwise be required (albeit that, in line with the findings of public consultation, the burden falls on the researchers to demonstrate why consent is either unnecessary or inappropriate). A second problem with an evidence-based approach to risk is the fact that, while some relevant evidence might exist, there are reasons to think that undesirable outcomes are significantly under-reported and intrinsically difficult to find, rather than simply absent. This is strongly suggested by our commissioned research into harms associated with data abuse (see chapter 2).³⁷⁴

- 6.46 Even if it were reliably determinable, however, risk alone cannot be a sufficient basis for governance. This is because the nature of the questions with which we are concerned, which are partly moral questions, are not tractable solely by the application of evidence, no matter how much evidence is available. Their resolution requires, additionally, a form of reasoning that reveals and resolves the values and tolerances associated with different possibilities and consequences.³⁷⁵ Whereas robust approaches to data handling may minimise the risk of re-identification of individual patients, they do not in themselves address questions of privacy and potential harms to privacy.

Authorisation, decision making and accountability

- 6.47 Under the SHIP model a system of ‘authorisation’ of research operates alongside and somewhat independently of any requirement for consent and measures to de-identify the data, addressing instead the duty of care owed by professionals to the public out of respect for them as moral agents. (This is consistent with our conclusion, at which the SHIP analysis also explicitly arrives, that consent is neither necessary nor sufficient to protect the interests of those involved.)³⁷⁶ Whether or not consent is judged to be required, the second function of risk assessment is to support the authorisation process in considering the ‘relative risk’ of a research initiative, taking account of the full range of interests at stake and the balance of likely hazards and benefits (not merely the risk of re-identification). This is to recognise that, from a public policy perspective, there are risks of both using and not using data, and that these may be distributed differently among the range of those with interests in the initiative. (Thus, the minimisation of risks for some, taken to its logical conclusion of simply locking down data and preventing all re-use, is not without consequences in terms of benefits foregone, both for those people and for others.)
- 6.48 When one moves from the simple principle of minimising risk to the principle of optimising the balance and distribution of risks and potential benefits, the essentially political question of how this optimum is determined (and by whom, and in whose interest) becomes salient. This is, in effect, the most fundamental question of privacy: how we construct the norms that enable cooperation in pursuit of common goals but

³⁷⁴ One response to this is to pursue further research as we have recommended (see recommendation 2).

³⁷⁵ See our third and fourth principles (participation and accounting for decisions).

³⁷⁶ Laurie G and Sethi N (2012) Information governance of use of health-related data in medical research in Scotland: towards a good governance framework, University of Edinburgh, School of Law Research Paper Series 2012/13, available at: http://www.scot-ship.ac.uk/sites/default/files/Reports/Working_Paper_2.pdf.

keep the unreasonable demands of others in check. Of course, this question is, in reality, always present and is only held in abeyance by a restricted framing of the question as if it were simply about *this* particular risk and how to manage it (the risk of re-identification of given individuals, for example); in reality, reducing the exposure to risks for one person almost always increases it for others. It is for this reason that, in chapter 5, we presented data initiatives as *social* practices, embedded in the wider life of the political society, which require those with interests at stake to reason together regarding their resolution.

- 6.49 As we have done in this report, SHIP sets out principles that provide a reference point for deliberation and decision making without prescribing what should be done in any particular set of circumstances.³⁷⁷ Though the principles are not the same, they demonstrate a clear and explicit engagement with the question of the relationship between public and private interests that led us to posit our principles of respect for persons and human rights (see chapter 5). They provide a common language and frame of reference to consider what morally relevant interests are at stake, to reason through the issues, and to ground justifications offered for particular outcomes. It is important, therefore, to ask who should participate in this process and to whom and by what means their decisions are accounted for.
- 6.50 In Scotland, the authorisation for release of data is acknowledged as being initially in the hands of GPs and other data custodians of the contributory datasets (e.g. NSS, NHS or within Health Boards). (The English system, by contrast, seems designed to shift these powers as far as possible to the centre. The fact that the GPES extraction of primary care data, rather than HES data, for example, became the main battleground owes much to the involvement of GPs as interested participants – as participants, of course, with their own interests in the possible data initiatives, alongside researchers, firms, patients and others.) Further authorisation for use of data is required (by law) from the data custodians and may be sought, additionally, from a body such as the Scottish Privacy Advisory Committee (PAC).³⁷⁸ In addition to the legal question of whether data *may* be released, the SHIP approach explicitly requires consideration of the moral question of whether data *should* be released. It is here that the question of ‘balance of risks’ is engaged.
- 6.51 The SHIP approach connects to the social basis of data sharing through a commitment to public engagement as an input to governance.³⁷⁹ The commitment to public engagement has been taken on by the Farr Institute @ Scotland, to raise awareness of

³⁷⁷ The SHIP principles contain references to human rights and respect for individuals as moral agents. They are nevertheless more operational (and more numerous) than the higher-level principles set out in this report, which are intended to serve a broader range of applications. For the SHIP principles, see: http://www.scot-ship.ac.uk/sites/default/files/Reports/Guiding_Principles_and_Best_Practices_221010.pdf.

³⁷⁸ PACs play a role somewhat analogous to the one now marked out for CAG in the English HSCIC system, although they have a formal approval function where CAG remains advisory. See: http://www.nhs.uk/pages/corporate/privacy_advisory_committee.php.

³⁷⁹ “13. Public and stakeholder engagement. **Principles** (1) Public and stakeholder engagement is an integral part of good governance. As far as possible, account should be taken of the full range of stakeholder positions in the development and implementation of governance arrangements. (2) The interests of one (or a few) stakeholder(s) should not dominate use/linkages or the conditions of the same, especially where this might be at the expense of other stakeholder interests. Robust justifications must be given for any departure from this principle. **Best Practice** (1) Stakeholder interests and expectations should be monitored over time by an appropriate body or individuals with appropriate expertise for the task. Where necessary, governance arrangements should be adapted to take account of shifting stakeholder needs and expectations. (2) Active engagement exercises should be developed and implemented over time to monitor and respond to stakeholder interests.” (http://www.scot-ship.ac.uk/sites/default/files/Reports/Guiding_Principles_and_Best_Practices_221010.pdf, at page 16).

medical research and its benefits and to foster trust and to allow two-way communication between professional and public 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.³⁸⁰ (This is consistent with recommendations we make in relation to research in the next chapter.)

- 6.52 One upshot of this has been the prior identification of expectations with regard to the involvement of the commercial sector. The stipulation that research should be in the public interest does not, in principle, bar private companies (e.g. pharmaceutical companies) using the resource. However, a key feature of the way the SHIP approach has been developed by the Farr Institute @ Scotland is that it takes seriously public guardedness about engaging with commercially motivated researchers by not allowing commercial users direct and un-chaperoned access to the data. At the same time, the approach acknowledges that in the science and innovation ecosystem as it currently exists, commercial actors can have an important role to play in promoting the public interest. In order to make use of the resource, therefore, private sector researchers must demonstrate that the research is in the public interest and form a partnership with NHS or academic researchers. It is these NHS or academic researchers who will undertake the analysis and have direct access to the raw data.³⁸¹ If a private company is involved, they will have access to the results of the research although results will only be released following approval from the PAC to ensure that no identifying data will be released. An explicit consideration is the reputational risk and the impact on public trust that commercial involvement may represent in any particular case.
- 6.53 The SHIP initiative demonstrates a number of elements of good practice according to our analysis and principles. It pays regard to context rather than simply the 'type' of data in use; it acknowledges the importance of responsible behaviour on the part of professionals over and above the duty to respect the consent of patients, even where data with a low risk of re-identification are used; it aims to resolve the 'double articulation' of public and private interests that we described in chapter 3, partly through a commitment to public engagement; and it takes seriously the need for trust and concerns about the involvement of commercial interests (which we consider further, in another context, in the next section).

The '100,000 Genomes' Project

- 6.54 The rich phenotypic and, increasingly, laboratory data held by the NHS and other health services, and the NHS's continuing relationship with patients, offer scientifically and politically attractive opportunities to carry out biomedical research alongside treatment. One such initiative is the UK 100,000 Genomes Project.³⁸² This project brings into conjunction more explicitly than any other current initiative, the research, policy, and national economic drivers that we discussed in Chapter 2, and embodies a commitment to the prospects of genomic medicine and to the idea that the NHS should be their proving ground. This has led to a notable two-stranded rhetoric around the

³⁸⁰ See www.farrinstitute.org/centre/Scotland/21_Public-Engagement.html.

³⁸¹ The electronic Data Research and Innovation Service (eDRIS) provides a single point of contact for researchers. See: www.isdscotland.org/Products-and-Services/eDRIS/.

³⁸² A member of our Working Party, Professor Michael Parker, chaired the initial CMO's ethics working group leading up to the launch of the 100k Genomes Project. He is currently a Non-executive Director of GeL and chair of its Ethics Advisory Committee.

project, which freely mixes the objectives of advancing science to improve human health with ambitious ‘techno-nationalism’³⁸³.

“...the race is on. The benefits to human health (better and earlier diagnoses as well as personalised care) are so enormous that everyone will want to be in the game. Even so, the insights we can unlock are so numerous, there’s enough potential reward for all players. But when it comes to building the critical mass of data needed to tackle some of our most serious healthcare challenges, there will be one winner, and that will be Britain.”³⁸⁴

Box 6.7: The UK 100K Genomes Project and Genomics England Ltd

The 100K Genome Project, announced in a speech by the Prime Minister in December 2012, and launched formally on 1 August 2014, is a project to generate 100,000 whole genome sequences from NHS patients in England.³⁸⁵ The project gives shape to the ambition to realise the benefits of genomic medicine in the NHS.³⁸⁶ Its focus is on generating sequences from cancer patients and their tumours, patients with rare diseases (those affecting <1:1500 people) and their parents, and those with infectious diseases (HIV, tuberculosis, and hepatitis C) and antibiotic resistance. The project design was initially informed by reports from three working groups (on strategic priorities, ethics and data) established by the Chief Medical Officer for England.³⁸⁷ Following these, the decision was taken to establish a private limited company, Genomics England Limited (GeL), wholly owned by the UK Government, to deliver the project and to manage the extraction of value from it.³⁸⁸ The project has an embedded ethics team and an ethics advisory group. The aims of the project are stated as follows:

- to bring benefit to patients
- to create an ethical and transparent programme based on consent
- to enable new scientific discovery and medical insights
- to kickstart the development of a UK genomics industry³⁸⁹

Suitable NHS patients will be invited to participate by their health professionals and complete consent forms outlining the aims of the initiative, the possible uses of their data and the mechanisms for governing this. No immediate therapeutic benefits are promised to those taking part but in some cases the information may be used to inform their

³⁸³ On ‘techno-nationalism’ see Edgerton DEH (2007) The contradictions of techno-nationalism and techno-globalism: a historical perspective *New Global Studies* 1(1): 1-32, available at: https://workspace.imperial.ac.uk/historyofscience/Public/files/c_contradictions_of_technoglobalism.pdf.

³⁸⁴ Sir John Chisholm, Executive Chair of GeL and also a non-executive director of HSCIC. See: <http://www.nesta.org.uk/news/14-predictions-2014/great-whole-genome-race>. For a discussion of the link between biotech R&D investment and national economic growth, see Nuffield Council on Bioethics (2012) *Emerging biotechnologies: technology, choice and the public good* (esp. Chapter 7), available at: <http://www.nuffieldbioethics.org/emerging-biotechnologies>.

³⁸⁵ The actual number of patients will be about 75,000 as the remaining sequences will be tumour sequences (see: <http://www.genomicsengland.co.uk/the-100000-genomes-project/faqs/>).

³⁸⁶ Human Genomics Strategy Group (2012) *Building on our inheritance: genomic technology in healthcare. a report by the Human Genomics Strategy Group*: “We are currently on the cusp of a revolution in healthcare: genomic medicine – patient diagnosis and treatment based on information about a person’s entire DNA sequence, or ‘genome’ – becoming part of mainstream healthcare practice.”, available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/213705/dh_132382.pdf, at page 12.

³⁸⁷ See <https://www.gov.uk/government/publications/mapping-100000-genomes-strategic-priorities-data-and-ethics>.

³⁸⁸ “Although this [GeL] is a company, it is only formed as a company so it can move more quickly to do these things [help patients, the NHS], to bring maximum benefit at the fastest speed.” Mark Caulfield, cited in Martin P and Hollin G (2014) *A new model of innovation in biomedicine? A review of evidence relating to the changing relationship between the private and public sector in the use of human genomics and personal medical information*, available at: www.nuffieldbioethics.org/project/biological-health-data/evidence-gathering/, at page 14.

³⁸⁹ See: <http://www.genomicsengland.co.uk/about-genomics-england/>.

treatment. A specific set of known genetic predispositions will be looked for and it is likely that the protocol will allow this information to be fed back to participants with their consent.³⁹⁰ Patients who do not wish to enrol will continue to receive the best NHS treatment currently available for their condition.

The intention is to make data available to researchers in a secure setting where they can study it without extracting or storing it on a different infrastructure, so that their interactions with the data may be tracked and audited.³⁹¹ It will also contribute to making the NHS a 'learning health system' (see above) through Genomics England Clinical Interpretation Partnerships (GeCIP).

6.55 The decision to deliver the 100,000 Genomes project through a limited company (owned and seeded with £100M investment by the Department of Health, but with ambition to seek substantial additional investment) was to allow it to operate flexibly and responsively, to enter into contracts and relationships with businesses, and to seize opportunities as they appeared. GeL's main asset will be the 100,000 genomes database, from which it expects to make a financial return, either through direct payment for data access, or through royalty sharing or joint venture schemes with other companies.³⁹² Although GeL has responsibility for ownership and delivery of the project it will contract other UK companies, universities and NHS institutions to carry out sample collection, sequencing, annotation and storage. The company has a number of parameters for how it will work which include that it will "ensure the benefits of the investment flows from the company to a large range of companies and contractors including SMEs" and "use any surplus to benefit the public health community."³⁹³ Those benefitting are likely to be software developers, sequencing and annotation providers, and the life sciences industry, which will use the knowledge generated by the project to develop new products.³⁹⁴

Information governance

6.56 Although GeL's corporate governance arrangements follow a commercial model, information governance follows a pattern more familiar from biomedical research. A formal data access process will be established with advice from the Ethics Advisory Committee and wider consultation. The procedure will include a data access committee that will examine and make decisions about data access applications, and a formal data access agreement to be signed by researchers.

6.57 Access to information with linked patient identifiers will be restricted to the clinicians working with patients (and, indeed, findings relevant to treatment of individual patients will be fed back to clinicians). Other users will only have access to de-identified data. All data will be held behind an NHS firewall, a model that has been compared to a 'reference library': unlike HSCIC, GeL will not provide extracts of individual-level data. GeL's customers (including academic and industry researchers) will have controlled access to that environment to carry out data analyses on linked, de-identified clinical

³⁹⁰ The final protocol was not available at the time of writing.

³⁹¹ In the pilot phase, data is released to commercial organisations and annotation of the genome sequences is conducted within the annotators' infrastructure, outside the NHS firewall.

³⁹² See: <http://www.genomicsengland.co.uk/wp-content/uploads/2013/11/Annotation-Supplier-Event-November-5-2013.pdf>.

³⁹³ See: <http://www.genomicsengland.co.uk/about-genomics-england/how-we-work/>.

³⁹⁴ GeL peremptorily signed contracts with the sequencing provider, Illumina, to generate the genome sequences, thereby inflating a bubble in UK-based sequencing capacity and associated infrastructure. See: <http://www.genomicsengland.co.uk/uk-to-become-world-number-one-in-dna-testing-with-plan-to-revolutionise-fight-against-cancer-and-rare-diseases/>.

data and genome sequences. All operations within the environment will be traceable so that, for example, it will be possible to see who has reviewed what data, when and for how long and GeL will carry out regular 'penetration testing' to assess security.³⁹⁵ When their analyses are complete, users will only be able to export their results (and not the raw data) across the NHS firewall.³⁹⁶ There is also an intention to develop a function that allows users to build 'apps' within the environment to interrogate data.³⁹⁷

Consent

- 6.58 The project will seek patients' active and explicit consent to retrieve samples and generate genome sequences and to authorise clinical, research and commercial use of the data, clinical feedback, re-contact through the patient's clinician and lifelong access to patients' medical records to allow the continuing updating of the GeL database. In other words, there will be a 'broad' initial consent procedure (for which UK Biobank has been widely cited as a model – see chapter 7, below). The richness of whole genome sequence and other -omic data, its association with one individual, and the possibility of it revealing predictive information about biological relatives, raises familiar privacy concerns as well as important treatment/benefit options for family members.³⁹⁸ GeL states clearly that it cannot promise study participants that it will not be possible for users of the data to identify them, despite the security measures in place to prevent this.
- 6.59 Because many participants are patients or relatives of patients with serious disease the structure of their interests and motivations may be different than, for example, with an administrative or prospective research database involving apparently healthy people. Although they may have strong incentives (both self- and other-regarding) or dispositional vulnerabilities (owing to their status as patients) the thorough consent process suggests that their preferences can be respected, even if the options available do not correspond to their personal interests: even if, for example, the prospect of commercial use of the resource is a personal disincentive, it need not be morally unreasonable to offer people an option that allows them to trade this off against positive benefits as they see them. In other words, the circumstances of consent do not include coercion or unreasonable inducement, although there is potentially some uncertainty about the possible personal therapeutic benefit.³⁹⁹ In clinical genetic practice it has sometimes been the case that the boundary between clinicians as physicians and as researchers has become somewhat blurred. So, for example, having collected samples from patients for genetic aetiological research clinicians may have later been in a position to provide clinical benefit through feedback of information about genetic risks for patients and family members. In the 100,000 Genomes project clinicians will both recruit patients to the project and later may carry responsibilities for the feedback of clinically significant information to their patients or to use this to inform their treatment. Clearly, in such a situation, it is important that the patients being

³⁹⁵ See: Notes of the meeting on 24th March, 2014 between patient advocates and Professor Michael Parker and Ms. Vivienne Parry of Genomics England, available at: <http://independentcancerpatientsvoice.org.uk/consultations/>.

³⁹⁶ See <http://www.genomicsengland.co.uk/town-hall-engagement-event/>. The arrangements for the pilot phase are different from the main project – see: <http://www.genomicsengland.co.uk/annotation-supplier-event/>.

³⁹⁷ Described by Professor Tim Hubbard at Progress Educational Trust/GeL event "Genetic Conditions: how should your DNA be used in the 100,000 Genomes Project?" See: <http://www.progress.org.uk/geneticconditions>.

³⁹⁸ Lucassen A and Parker M (2010) Confidentiality and sharing genetic information with relatives *The Lancet* **375(9725)**: 1507-09.

³⁹⁹ On the 'therapeutic misconception', see paragraph 4.37.

recruited have clear and realistic expectations about the clinical benefits that may or may not accrue to them through participation in the project.

- 6.60 Patients and patient groups were, in fact, consulted and involved in the design of the consent process and materials. Although the consent given by participants is not open to additional conditions, and therefore does not offer an opportunity to express preferences that might shape the design of the resource or its use beyond their simple participation or non-participation, it is clear that candidate patients' receipt of the best currently available treatment is not dependent on their participation in the 100,000 Genomes project.
- 6.61 The long term nature of the resource is also relevant. It is possible that clinically significant findings might emerge at any point throughout the life time of the project. However, families also change over time and the salience of findings may change for family members (for example, as they face reproductive choices). Some mechanism for managing these uncertainties, such as setting a time limit for feedback of information, will therefore be of great importance. Many of the most significant ethical questions relate to how the use of the resource will also develop in the future and what this will mean in changing circumstances. The mechanism for responding to such changes relies on the governance provided through the institution, in which the Ethics Advisory Committee plays an important role, and the option for participants to withdraw from the study.

Elements of an ethical approach

- 6.62 Looking at the 100,000 Genomes project through the lens of the approach we set out in chapter 5 focuses attention on two aspects in particular: the design of the project (and how this incorporates the public and private interests involved) and the governance of the project (and how well this respects the interests of participants in changing circumstances).
- 6.63 Though there were a number of 'Town Hall' meetings in the development phase of the project, and the consent materials were developed through interviews and focus groups with patients and members of the public of diverse ages and backgrounds, there is no evidence of broader social engagement around questions of project design.⁴⁰⁰ Like NPfIT, the policy process was carried out in haste with a strong political (indeed, Prime Ministerial) impetus, and many of its key parameters and elements of infrastructure were locked in prior to determination of the governance systems. Without social accountability there is a possibility that political, commercial and health drivers may conflict (or, at least, that their relationship may appear ambiguous, with potentially adverse consequences for broader public trust). Especially given the context of ambitious public rhetoric about realising the promise of genomic medicine in the UK, special care needs to be taken to avoid overselling the prospects of therapeutic benefit to participants.
- 6.64 We know, from previous public engagement, that the issue of commercial involvement excites particular preferences for many, who treat this as a morally salient, rather than morally irrelevant, feature.⁴⁰¹ This may be due to considerations of justice (equitable sharing of benefits), trust (belief that commercial involvement represents greater privacy risk) or other reasons. Although the choice to participate is not coerced, the

⁴⁰⁰ See, however: <http://www.genomicsengland.co.uk/town-hall-engagement-event/>.

⁴⁰¹ See paragraph 5.18.

question arises whether the design of the project could have offered a morally preferable set of options than those offered by GeL. This is not a trivial question since, though a government-owned company, GeL exists to promote the public interest. Of course, the public interest may also be served by generating national income through the commercial activities stimulated by the resource. Furthermore, commercial input is an increasingly important part of contemporary academic research and ruling this out completely (as the 1958 birth cohort does, for example) may mean that the capacity for potentially desirable academic research is limited.⁴⁰² Given that there are other ways to achieve this, as the example offered by SHIP shows (see above), the case needs to be made out publicly that GeL represents the politically, scientifically and morally optimum resolution of the public and private interests at stake.

- 6.65 The second question relates to the possible implications of as-yet-undefined uses of data in what may be a technologically and informationally very different future environment, and how these will be governed for public benefit and the protection of individual interests. Given the broad consent model and the breadth of potential uses and users, once the 100,000 genomes resource is established, considerable reliance will be placed on the governance system. One way of ensuring broader accountability would be to set out transparently the set of morally reasonable expectations about data use within the powers that are available to decision makers and to use this as a focus for both governance and for more inclusive decision making about the future of this public resource. The key document will be the data access policy.⁴⁰³ Such mechanisms, should not limit the flexibility and ability to seize opportunities that were seen as advantages of constituting GeL as a private sector actor, but would provide greater clarity and assurance with regard to policy.
- 6.66 Genomic testing is likely to become more routine in health systems and GeL may form a bridgehead for new forms of data linking in the NHS. The first 100,000 genomes could well be merely the vanguard for a more substantial genomic and phenotypic database, especially as the value of such databases increases significantly in relation to their size. Whether the model offered by GeL represents the most appropriate model for securing the public interest in the ethical use of genomic information in health services is therefore an important question since this was not extensively or publicly debated prior to the initiation of the project.

Recommendation 9

We recommend that broader public consideration should be given to whether GeL provides the most appropriate model for the ethical use of genomic information generated in health services for public benefit before it becomes the *de facto* infrastructure for future projects.

⁴⁰² For information on the '1958 Birth Cohort' (the National Child Development Study), see: <http://www2.le.ac.uk/projects/birthcohort/1958bc>.

⁴⁰³ At the time of writing this is still under development.

Conclusion

- 6.67 In this chapter we have looked more closely at the formation of a number of data initiatives that represent different approaches to the linking and re-use of data in health systems. We have looked at this formation not only structurally, but also in terms of how they came about, what incentives and drivers pulled and pushed them in different directions, and in the light of our contention that data initiatives where public and private interests are at stake are social and political practices, as well as moral and scientific ones.
- 6.68 Each of the data initiatives resolves questions about centralisation and distribution of resources, how data are disclosed or accessed, the range of users and purposes, and how control is exercised in different ways. Thus the HSCIC and GeL models are more centralised than SHIP; the HSCIC model allows some disclosure of individual-level data whereas SHIP and GeL will only disclose the results of analyses carried out within their infrastructure; HSCIC and GeL allow direct access to data by commercial companies whereas SHIP is more guarded; GeL requires explicit, though broad, patient consent, whereas HSCIC and SHIP have (at least initially) rather different authorisation procedures in which individual preferences and values may figure (through a rather blunt opt-out or through a more constitutive participation); etc. Though there are lessons to be shared amongst these various experiences, perhaps the most salient lessons relate to our principles of participation and accounting for decisions through formal governance and wider social engagement. These strongly suggest that there are serious consequences for public trust and for the viability of data initiatives if they do not first take steps to identify the applicable moral norms that they must negotiate and put in place, in relation to these, well-supported measures to respect the interests engaged, supported by credible justification.
- 6.69 As we have argued a key question that faces data initiatives and the health systems as a whole is what uses of data should be 'expected' as part of delivering national health care with quality, safety, and cost-effectiveness with ongoing improvement in the standards of care. It is becoming increasingly evident that there are commercial drivers behind many high-profile initiatives that have been proposed in recent years and, as empirical studies show, this is of significant concern for the public. The issues go beyond individual privacy, especially as datasets start to be linked together, and there is a need to have governance structures in which all interests are enabled to participate and that involve continuing review and reflection on the societal implications of such initiatives. Unless there are trustworthy governance systems in place that can engage with and reflect reasonable expectations in continuously evolving circumstances, initiatives that could have wide public benefits may continue to be challenged and fail to secure public confidence.