Fair Shares for All
Sharing and protecting electronic patient healthcare data

A Report written for the British Computer Society Primary Healthcare Specialist Group

Ian Herbert
March 13th 2012

share 1 *v.t*. Give away part of

The New Shorter Oxford English Dictionary,
Fourth Edition 199

This report has benefited greatly from the work of others, and the comments provided by those who read one or more of the various drafts. It started life as a note by Ewan Davis on the use of Privacy Enhancing Technology produced by the BCS Primary Healthcare Specialist Group, following a meeting of its Clinical Information & Consultation Special Interest Group (CLICSIG). My thanks to all who contributed.

As ever the views and errors in this report are the responsibility of the author.

Ian Herbert 13th March 2012

Contents

Conventions used in this document

All quotes are in italics and enclosed in "”

Explanatory additions made to quotations made by the author are shown within {}

Publication titles and key terms are shown in *italics.*

Bold is used for titling and emphasising important words or phrases

References to external sources give the full or abbreviated title of the document or its number in the bibliography in Annex A, e.g. [Data Protection Act 1998] or [DPA] or [1]. References to other sections in this document are shown as e.g. see 2.3, see D5 (referring to section 5 of Annex D).

he, *him* or *his* should be taken to mean persons of all genders, except they refer to a named individual.

1 Summary

1.1 This report by the BCS Primary Healthcare Specialist Group is about the sharing and use of patient data collected by or on behalf of the NHS. We look at opinions and practice in the light of the relevant law and guidance in order to arrive at a set of principles and measures that support our ambition to see the data held in electronic health records used as fully and responsibly as possible in a way that respects patient privacy. This will help ensure that individuals receive the best care, that the NHS is administered and managed efficiently and effectively, and that clinical audit, research and development flourish. These functions cannot be performed without comprehensive good quality patient data, and sharing patient data is becoming easier as care providers increasingly store patient data in electronic patient records (EPR). The Information Revolution requires such data if it is to play its proper part in transforming care delivery and making the NHS sustainable.

1.2 Respect for the confidentiality of patient data is the foundation of the trust at the heart of the patient – clinician relationship. When data is shared in a way that respects patient privacy in line with patient it creates a virtuous circle and encourages patients to share their data for care and secondary purposes. If it is not it undermines public confidence in the NHS and discourages patients from providing data to their clinicians and others, clinicians from recording what they are told, and in extreme cases patients from seeking treatment. It will lead some patients to hold records to which they alone control access.

1.3 Patients provide data to their clinicians to support the delivery of safe, defensible personal care. While there is good evidence that most patients are happy for their data to be used for research purposes when they believe that their privacy is respected and they can exclude particularly sensitive data, such uses are "by permission" and not "as of right". Although the use of patient data in administration activities such as provider self-audit and provider payment is implicit in accepting NHS treatment, the evidence shows that the majority of patients are unaware of this and would be alarmed if the found out that their identifiable data were circulated widely for such purposes, especially beyond the NHS. For all secondary uses most patients believe that their identifiable data should only be used with their consent, or in anonymised or pseudonymised form, see 4.7.1. A substantial minority believe that their anonymised or pseudonymised data should only be used with their consent.

1.4 The evidence consistently indicates that retaining public confidence against a background of high profile data loss and privacy breaches requires that the health community applies the highest standards to ensure that patients' wishes are respected and their data protected by robust physical security and governance arrangements. Such standards need to exceed current legal requirements in order to satisfy patient and clinician expectations, and to be applied to both identifiable and de-identified data.

This is particularly important in the light of clear evidence of public (and expert) doubts about the efficacy of de-identification[1]

1.5 We recognise that requiring many patients to actively opt-in to a secondary data use is onerous, and may on occasion introduce sampling bias that reduces the statistical value of the data. We take a pragmatic view and want to remove unnecessary barriers to secondary use. We believe that where steps are taken to adequately reduce the opportunity and motivation to re-identify patients in de-identified data sets and the ease of doing so, it is acceptable to operate on an opt-out basis. What constitutes adequate steps depends on the nature of the data involved, its sensitivity to the patient, the context of use and the value of the use proposed. Decisions about adequacy should be taken by a trusted independent entity able to provide rapid approval with the minimum bureaucracy compatible with good governance, such as the National Information Governance Board's (NIGB's) Ethics and Confidentiality Committee (ECC).

1.6 The risk of de-identified data being re-identified depends on the way it is shared and with whom, as well as its intrinsic content, see Annex H2. It is greatest with rich data shared via 'the bird table', where anyone may take it and use it for any purpose(s). In such cases the risk may be so high that 'de-identified data' should be treated as if it were identifiable data. In the light of this, we advocate that, where possible, data are collected for specific purposes. Applying the "Least Principle" - the least data, copied the least number of times, held for the least time and used by the least number of people necessary for the purpose - substantially reduces the privacy risk However, we believe that rich and long-term general purpose databases should be permitted where they are the only way to answer important questions in a timely and cost-effective way. They can dramatically reduce the need to copy patient data for secondary purposes. The corollary is that they should operate under higher standards of security and governance with heavy penalties for data misuse in order to minimise its risk of occurrence.

1.7 Record linkage is essential for a range of valuable purposes and there is a misconception that it requires identifiable data. This is not so, and those wanting to link records from different sources should seek advice about how to do this with pseudonymised data. Linkage is relevant to a range of extremely valuable techniques in the area of disease management and risk stratification which analyse large numbers of patient records to identify the small number of patients who might benefit from additional support or a particular intervention. These can and should be conducted using privacy enhancing technologies (PET) so that neither identifiable data nor patient identifiers leave the care providers involved.

1.8 We are aware of excellent work being done by bodies inside and outside the NHS to improve patient information governance, including examples in the Department of Health, NHS care providers, other government agencies, academia and elsewhere. On the other hand we are

[1] However the author knows of no unauthorised case where a patient has been re-identified from de-identified data

also aware of examples where the way identifiable patient data is used would come as an unpleasant surprise to patients and the public.

1.9 In summary we want to encourage patients and their clinicians to provide their data for laudable research purposes, and acknowledge the need to use it to administer and manage the NHS, but we must seek to retain public confidence while doing so. Patients accept the electronic processing of their health data for primary purposes, but should have reason to feel confident that it is protected and used properly. Equally their clinicians should not find the data governance process unduly onerous. Except in very limited circumstances determined by the ECC, patients should be able to stop the use of their data for research in any form. Where privacy is adequately protected by the researcher, identifiable data (or de-identified data where the re-identification risk in secondary use is high) may be used via an informed opt-out, but where such protection is not available an informed opt-in should be used. At the same time all secondary users should take full advantage of the latest developments in PET and use de-identified data for their purposes: there should be very few if any secondary uses where identifiable data is needed and patient consent cannot be readily obtained. As now there should be one independent and trusted body, such as the current ECC, to determine proposed uses of identifiable patient data for secondary purposes without consent.

2 Why this report now?

2.1 The desire of the government (see 7.5) and researchers (see 6) to make the UK a world leader in medical research through the greater use of NHS patient data and the advent of a Health and Social Care Bill 2011[12-13] that proposes making identifiable patient data (see 7.3-4) more accessible without patient consent mean this is a particularly apposite time to examine patient data use and governance in the NHS. The latest report by the Academy of Medical Sciences on research governance [94] has influenced the Bill and the subsequent proposal by the Government to permit the use of anonymised patient data for life sciences research by default unless the patient opts out. We do not yet know enough about the life science research proposal to assess its impact on patient data confidentiality, but the Bill itself is seriously at odds with the assurances about patient control of their data given in the preceding White Papers [9-10] and abolishes the independent body (NIGB) charged since 2008 with ensuring sound personal information governance across the NHS. The reform mantra "no decision about me without me" does not appear to apply to sharing patient information. Other topics are dealt with in 8.

2.2 Patient data disclosure must of necessity increase to support the wider coverage of Payment by Results (PbR) in the NHS including sensitive areas such as mental health, see 7.2. As private provision of NHS care grows, bodies other than service requesters and providers will enter the billing process and require individual patient data. More sharing among more organisations makes it more necessary still to share the minimum data needed for openly-declared purposes in an effectively de-identified form. Some current central returns do not appear to match this specification. We are also aware of considerable use of identifiable data for analytical purposes at PCT level, which although arguably legal, could we believe use pseudonymised data.

2.3 Behind our concerns is an evident and major mismatch between the published ambitions of some of those using patient data for research purposes and the wishes of patients and their clinicians about such uses. The impact of this discontinuity is muted by low public and patient awareness of the secondary purposes for which their data is used, see 4.4.

2.4 Privacy-enhancing technology (PET, see H2) is improving significantly. This is raising the bar imposed by s251(4) of the NHS Act 2006 that an application to use identifiable patient data without consent must clear before the ECC can approve it. Not all secondary users are aware of this and make the most of PET. Recent work shows that to be effective PET must increasingly be accompanied by other constraints, see H2 [73,102,103].

2.5 Our conclusions come next. Chapters 4-8 show the evidence and analysis on which they are based, starting with a review of stakeholder attitudes and moving on to our concerns. En route we compare researcher and Government proposals with public and patient attitudes, making full use of the work done by the BCS Primary Healthcare Group work on recent

advances in PET. Annexes A-I provide extra detail on key topics for those who wish to delve deeper.

3 Our Conclusions

We have presented and analysed the evidence and the roots of our concerns in order to stimulate the generation of a consensus on how and in what form patient data can and should be shared for primary and secondary uses. We believe that our recommendations go a long way towards meeting both patient expectations (see 4) and the assurances given on patient information governance in the White Papers and consultations that preceded the current Health and Social Care Bill (see 7.1) Unfortunately the Health and Social Care Bill undermines rather than supports the assurances given in the White Papers and Consultation document. While we sympathise with researchers' desire to simplify the over-complex research governance process we find that the claim by the research lobby that legislation is a significant obstacle to accessing the patient data needed by researchers is not supported by the evidence in their reports or by other sources (see 6). It is too soon to assess the impact on patient data confidentiality of the proposals for providing patient data for life science research, but we have generated a set of questions (see 7.5.2) which when answered would enable us to do so. Although there is an inevitable tension between personal privacy and sharing patient data for the public good, we believe our approach will ultimately enable society to have its privacy cake and eat it, given open discussion in good faith among all the stakeholders. Many of the recommendations below are not new.

3.1 Our basic principles – see 1

3.1.1 We believe that patients provide data to their clinicians for their care in confidence as the Common Law acknowledges. In consequence we believe that:

(a) the processing of patient data should conform to, or better, the law
(b) patients, clinicians, secondary users, informaticians, information governance experts and IT experts must work together to ensure that secondary purposes can be satisfied with patient data in a form that poses the least risk to patient privacy
(c) That patients, as the ones at risk, should be:
    I. encouraged to supply data for secondary purposes
    II. able to choose whether to accept a significant risk of the disclosure of their data posed by a secondary use, or not.
    III. involved in deciding whether safe havens and care provider organisations should provide data for particular secondary uses
(d) Any patient information governance arrangements should
    I. not significantly impede those providing care for patients.
    II. conform as far as possible to patient expectations.

3.1.2 We believe that all people and organisations holding data about patients in any form must ensure that it is held and processed in a way that prevents its corruption, loss and unauthorised use. If this is not the case it will damage confidence in care providers and the NHS and reduce people's willingness to provide data to the NHS in any form for any purpose, besides risking harm and distress to patients.

3.1.3 We believe that all users and data controllers of patient data should abide by the 'least' principle, as the Data Processing Act suggests, as part of the measures they take to ensure that patient privacy is maintained. They should:

(a) use the least amount of data (in terms of both the number of subjects and the amount and kinds of data per subject)
(b) use (and retain) the data for the least amount of time[2]
(c) restrict access to the least number of authorised people
(d) copy it the least number of times (ideally none)

commensurate with the purpose(s) involved.

3.1.4 We believe, subject to the restrictions given in the DPA, that patients have a right to know:

(a) who is processing their data in any manner or form
(b) when they intend to process it and how long they will retain it
(c) what data are involved
(d) how they process it for (and if not obvious, why)
(e) what other organisations the processor may share it with, under what conditions and what data is involved
(f) how the processor ensures that patient privacy is maintained
(g) how to contact someone - typically the data controller - who is processing their data.

It applies to both identifiable and de-identified data. The information listed above comprises the bulk of the content recommended by the Information Commissioner for a privacy notice, see[3] It should be supplied when any form of consent for data processing is sought or available as part of any opt-out, and when a patient asks an organisation about one, some or all their processing of NHS patient data.

3.1.5 We believe that all users of patient data must be under a legal obligation of confidentiality similar to that of a clinician treating the patient.

3.1.6 We believe that subject to 3.1.1(d), using an individual's data for research is a privilege for the patient to grant, not a right that researchers have[4]. The NHS exists and is paid for by citizens to deliver healthcare. Its role as a data source for research is desirable and laudable, but secondary. The provision of an opt-out in Government proposals to permit access to anonymised patient data for life sciences research supports this view[105]. More data would become available if the opt-out was available study by study as well as in all-or-nothing form, but we acknowledge that this may prove difficult to implement[5].

---

[2] The retention periods for various types of records (including care provider patient records) are set out in *Records Management NHS Code of Practice Part 2* (2009). Electronic GP records should be retained indefinitely. Data for statistical and historical research may be retained indefinitely, see DPA 33.

[3] http://www.ico.gov.uk/for_organisations/data_protection/topic_guides/privacy_notices.aspx The DPA only applies to personal data, see definition in footnote to D3. The major problem here would be informing people about the use of their **de-identified data** other than by non-person specific notices.

[4] For example they may object to the researcher, the organisation or kind of organisation doing the research, the purpose of the research or on the grounds of the sensitivity of the data to be used. Objections could be based on deep-seated religious or cultural beliefs, see following footnote.

[5] The major challenge here is to constrain the use of de-identified data.

3.1.7 We believe that there is a case for permitting patients to qualify a consent where it is being sought in advance for more generic purposes, typically for the use of warehouse data controlled by a safe haven. We appreciate that it may be difficult to persist such qualifications in a computable form but if such qualification is rare this may not prove a serious deficiency. One can envisage that a qualification could be so widely applicable that a patient may wish to provide it in advance of any consent being requested[6]. The lack of a facility to qualify consent would probably increase the demand by patients to opt out of research as a whole.

3.1.8 We are aware that some uses, such as analysis of linked patient data to determine the risk of re-admission to hospital, are difficult to categorise as a primary or secondary use, as they may be used to determine the individual patient's management as well as for general health service management. Unless the analysis is done by one or more care provider(s) (or a data processor working for a care provider) **we believe** that it should be regarded as a secondary use.

3.1.9 We are aware that while most patients are content for their identifiable data to be used by and within their care providers for their health service audit, administration and management without consent, this is not the case where it is used by organisations that didn't provide care to them, such as central clearing houses, other providers and commissioning groups. The Information Commissioner and NIGB share this view.

3.2 Publicising NHS information governance policy

3.2.1 We recommend that much greater effort is made to publicise :

(a) patient information governance arrangements in the NHS, including how de-identified data may be used. These are recorded in the *NHS Care Record Guarantee*, in greater detail in *Confidentiality: NHS Code of Practice* and in a much scantier form in the *NHS Constitution* and its handbook, but the vast majority of patients have never heard of any of these publications, let alone read them.

(b) any proposals to change them, and to involve the public in the change process.

(a) the information governance policies of organisations processing NHS patient data. This should be done in a way that the public and care providers can understand

NIGB would be the obvious agent to manage implementation of these recommendations.

3.3 Patient data sharing during healthcare provision – see 4.5.5

There is anecdotal evidence that in practice role-based access control (RBAC) and legitimate relationships do not always restrict the use of patient data as patients would expect during care[7], for example that access to clinical data is not always constrained to clinicians, and further to those clinicians who are actively involved in the care of the patient.

---

[6] For example, someone who is a Roman Catholic would be likely to object to their data being used for any research into biochemical birth control.

[7] There are situations outside the provision of care when others should see identifiable patient data, e.g. when handling a patient complaint

3.3.1 We recommend that the CfH role-based access control implementations are examined, to see whether the information filtering of patient data that it generates conforms to the NHS Confidentiality code of practice, and if not, why not.

3.4 Additional patient control over their sensitive data – see 4.6

There is a clear public wish for such a facility (the sealed envelope facility in NHS Connecting for Health) and it would resolve some significant issues, but it is not clear that it has ever been fully implemented, and if not, why.

3.4.1 We recommend that the work on sealing done by CfH is retrieved, and work done to find out whether the same or similar facilities have actually been implemented in systems, and if not, why not.

3.4.2 We recommend exploring the requirements in detail with patients, their clinicians and IT specialists in order to arrive at a specification that offers the desired benefits to patients while being usable during the care process. This should include consideration of sealing sensitive information by default or proxy on behalf of people who would be unable to do it for themselves, such as those with severe mental illness.

3.4.3 We recommend that sealed information of any type should not be shared for secondary purposes of any type without patient consent, other than by a court order or statute.

3.5 Keeping patient data secure and private– see 4.8

There is strong evidence that although patients generally trust the NHS with their data, they have serious concerns about hospitals' and researchers' abilities to prevent its loss or illegal or inappropriate sharing. It has undoubtedly improved in recent years but there is still some way to go to ensure that best practice is universal.

3.5.1 We recommend that the security measures applied to all patient data held by the NHS and safe havens and researchers outside it should conform to best practice; at a minimum this should meet the recommendations in the Information Governance Tool Kit, including encryption. Regulation of sources, use and users, and where appropriate PET, should be deployed to reduce the risk of unauthorised sharing and re-identification.

3.5.2 We recommend that unauthorised processing of patient data, including but not restricted to re-identification, accessing and sharing in any form, should attract substantial penalties, and consideration should be given to make some or all of these activities criminal offences.

3.6 Piloting and using privacy-enhancing technology – see 4.8, Annexes H & I

3.6.1 We recommend that secondary users should devote more effort to

(a) exploiting, extending and publicising the opportunities offered by modern privacy-enhancing technology,

(b) exploiting, extending and publicising data handling protocols and other measures that reduce the risk of data misuse.

The work on open pseudonymisation by researchers at the University of Nottingham [98] and the piloting of sophisticated commercial PET software by THIN and the Information Centre (see I2) are examples of what can be achieved using these techniques.

3.7 Behaviour and governance of large rich patient databases – see 6.2.2 and Annex F

Safe havens run by trusted third parties that hold patient data make life easier for data sources and secondary users, reducing the repeated storage and transfer of data and usually offering more effective information governance and securer storage. They are the natural home for patient data warehouses housing patient records covering long periods and including linked data from multiple sources if required which may be used for multiple purposes by different teams.

However warehousing patient data breaks the first three 'least' principles (see 3.1.3), especially where it involves linking comprehensive data from different sources and persisting it indefinitely. Although usually more professionally secured, warehouses are equally attractive to unauthorised users and users that can access them by statute or with a court order for purposes unconnected with healthcare or the life sciences. The richer their content the easier it is for a determined person to re-identify de-identified records. They are more liable to stimulate 'mission creep' amongst potential users.

3.7.1 We recommend that NIGB should be asked, in conjunction with care providers, patients and secondary users, to establish definition of a safe haven, and any sub-varieties of it as are found necessary.

3.7.2 We recommend that safe havens are strongly encouraged to collect patient data in response to specific secondary user requests than prospectively and then warehouse it.

3.7.3 We recommend that safe havens wishing to warehouse data, particularly linked data covering substantial fractions of patients' lives, should only be permitted after independent assessment by NIGB or a body with an equivalent remit has shown there to be no viable alternative.

3.7.4 We recommend as do the latest AMS report [94] and the Information Revolution consultation[10] that safe havens should have more stringent governance standards, see F, to reduce the likelihood of unauthorised or undesirable processing and to avoid ad-hoc 'scope creep'.

3.7.5 We recommend that patient data is only collected prospectively (i.e. for warehousing) with subject consent[8] and very well publicised or individual fair processing (aka privacy) notices for patients, unless it is in de-identified form for specific known purposes only and it is agreed that there is an insignificant risk of re-identification in the circumstances in which it will be used[9]. The weakest form of consent - for use only when collecting de-

[8] By this we mean consent **for the collection and warehousing of the data**. Separate consent would be required for using identifiable warehoused data in a particular project or for a specific purpose and may be required for using de-identified data where the risk of re-identification is high, e.g. for the use of rich data in its original form in some contexts for or less rich data to be shared via 'the bird table'. Where identifiable data is collected by a safe haven for a specific purpose(s) and not warehoused, the consent covers both its collection and use for the stated purpose(s).

[9] E.g. it is used in situ.

identified data - is a comprehensive, very well publicised and widely available privacy notice accompanied by an opt-out that is easy to use. Identifiable data should only be collected with an opt-in, i.e. explicit consent.

3.7.6 We recommend that where data is to be transferred to a safe haven for very generic or unspecified purposes, patients should be offered facilities to specify constraints on the kind of data that may be collected, who may use it and/or what it may be used for[10].

3.7.7 We recommend that safe havens should be strongly encouraged to only collect data that has been de-identified at source.

3.7.8 We recommend that safe havens should strongly encourage secondary users to use safe haven data in situ.

3.7.9 We recommend that consideration should be given to making use in situ obligatory for rich warehoused data.

3.7.10 We recommend that the governance of each safe haven should be managed by an advisory group that consists of researchers, IG experts, service providers and a substantial patient element, e.g. a third or more of the total membership. It is this advisory group, aided by such technical advice as is necessary, that should make the decisions about research proposals implicit in 3.7.2.

3.7.11 We recommend that the NIGB, in conjunction with existing data warehouse operators, should produce model governance rules for safe havens. See Annex F and in particular F5 for more detail on what the rules should cover.

3.7.12 We recommend that NIGB should be asked to investigate the consequences for patient information governance of having several multiple safe havens involved in the collection and/or processing of patient data before it is supplied to a secondary user, see F3 para 3.

3.8 Changing the research culture - see 6.1.3 and 6.3

The reports generated in whole or in part by the research lobby [55,67,70,94] do not reflect the respect for patient data confidentiality shown by many individuals and organisations that use it for research purposes. Patients involved in medical research in various ways and very supportive of it made it clear that they found researcher attitude to patients and others unsatisfactory for a variety of reasons, and that consent was fundamental to patient involvement.

3.8.1 We recommend that the research lobby move to a paradigm where patients are considered participants in research, even where studies are not interventional. This involves demonstrating respect for patient expectations and rights acknowledged in some of its reports [55,67,94] and described in the survey evidence, see 4. This may involve bettering the legal obligations where desirable and feasible. Surveys also indicate that the public are largely unaware of the actual and potential use of their data for research

[10] This recommendation applies to all research databases, not just safe havens. Presumably where patients control the sharing of especially sensitive data (see 4.6), the requirement for constraints on the kind of data that may be collected would diminish substantially or disappear. We acknowledge that implementation will be difficult or impossible where data is de-identified, and secondary users must be able to find the constraints, e.g. by attaching them to the extracted data or holding them in a national demographic database such as PDS.

and how its disclosure is governed, see 4.4. This requires genuine dialogue between researchers, patients and their clinicians.

3.8.2 We recommend that patients, care providers, their managers and researchers work together to arrive at a shared understanding that encourages them all to support and play their part in healthcare research, as recommended by [55]. All should play significant roles in research governance. Care providers have a key role in identifying people who are suitable for a research study and putting them in touch with the researcher where a study is interventional and/or uses their identifiable data. Researchers must expect to pay any additional costs incurred by care providers in doing so.

3.8.3 We recommend that as one of the first tasks for the new culture the research lobby catalogues the kinds of research that can only be done with identifiable data and explains why, so that care providers, patients and the public can understand the need.

3.9 Extended data flows in the reformed NHS – see 7.2

In additional to the major expansion of patient data flows already planned to support the extended scope of commissioning and payment by results (PbR), the Information Revolution propose new flows for patient-reported outcome measures, patient and staff complaints and quality metrics. There will be a great temptation to use unconsented identifiable personal data for these flows when it is not necessary to do so.

3.9.1 We recommend that, in line with the promises given in the Information Revolution White Paper that preceded the Bill, all such flows should be examined with a view to ensuring that they comprise pseudonymised, anonymised or aggregated data, or identifiable data processed with patient consent. For example, as time goes on, more bodies are likely to become involved in patient service ordering and payment, such as referral centres, specialised payment agencies and clearing houses, but only the service requester and provider need to know the identity of the patient who received the service.

3.9.2 We recommend that where the DH or NHS Commissioning Board considers that a flow should be of identifiable data, the proposal should as now be submitted to the Ethics & Confidentiality Committee for determination under section 251 of the NHS Health and Social Care Act 2006.

3.10 Health & Social Care Bill 2011: general – see 7.3.1

If enacted as it now stands, the Health and Social Care Bill [12,13] would enable the whittling away of patient data confidentiality without further reference to Parliament [101] and runs counter to the promises given in the White Papers, see 7.1, and the views expressed by patients and the public, see 4. It would enable the Secretary of State and Commissioning Board to direct that identifiable patient data be collected and processed or permit its collection and processing without any checks or balances, other than possibly by appeal to the Data Processing and Human Rights Acts. It would weaken the statutory requirement that now requires that all proposals to use identifiable patient data without consent must be examined by the ECC and a recommendation made by them to the Secretary of State. The powers

proposed for collecting NHS patient data are broader than those in the Coroners & Justice Bill 2009 which were withdrawn in the face of vigorous opposition by the BMA and others.

3.10.1 We recommend that after considering our report, the BMA reopens negotiations with the Government to have the Bill amended along the lines suggested below. We would be pleased to assist the BMA if they would like us to do so.

3.11 Health & Social Care Bill 2011: new data gateways – see 7.3.2-3

3.11.1 We recommend a review of all the new data sharing gateways proposed in this Bill which can be used to share identifiable patient data, to determine which

1. can be performed satisfactorily with de-identified data,
2. cannot but it is practical to seek consent and
3. cannot but the secondary user(s) considers that it would be impractical to seek consent,

including a justification for each determination of type 3.

3.11 2 We recommend that where 1 or 2 is found to be true, the Bill should explicitly preclude the sharing of identifiable patient data without consent

3.11.3 We recommend that where 3 is found to be true, the Bill should state that applications to use identifiable data should be submitted to the Ethics & Confidentiality Committee for determination on similar grounds to those used for assessing applications under section 251 of the NHS Health and Social Care Act 2006 (i.e. the current process for allowing the use of identifiable data without consent or a legal or statutory justification)

3.11.4 We recommend that the ability to direct (i.e. to require without further scrutiny by any other body) that identifiable patient data be shared is not given to the Secretary of State, Monitor, the Commissioning Board or any other health service body or body providing services to the NHS. Instead applications to do so should be assessed by the ECC and a recommendation made to the Secretary of State.

3.12 Information governance management in the Health & Social Care Bill 2011 – see 7.4

The proposal to split NIGB into components within the Care Quality Commission, itself a significant user of patient data, is equivalent to making a poacher the gamekeeper. If the message from the DH about giving other NIGB functions to a Health Research Authority is true, this would be doubly so. There is no evidence that such a change is needed while there is abundant evidence that a strong independent body is need to maintain and evolve good practice during the Information Revolution.

3.12.1 We recommend that in line with the commitment given in the White Papers preceding the Bill, see 7.1.1(e), the National Information Governance Board (or a single organisation of similar independence and status) should be retained with the brief to oversee and advise on the use of NHS patient information in any form for any purpose, and that it should house the Ethics and Confidentiality Committee (ECC) as now.

3.12.2 We recommend that the remit of the ECC should, as now, cover the determination of any proposal for the secondary use of identifiable patient data that is not:

- consented to by the patient
- authorised by a court order
- already authorised by an existing statute.

3.13 Patient data use for life sciences research – see 7.5

While we welcome the general thrust of the proposals in this area published in December 2011 by the Government they are not yet detailed enough to enable us to assess the implications for patient data privacy. For example it is not clear whether the proposals involve the **collection** of identifiable data and / or the **warehousing** of long period individual patient data collated from various NHS and other sources.

3.13.1 We recommend the use of the questions in 7.5.2 to guide the production of the relevant draft governance arrangements and to answer them during the process. This will assist us and others to assess the adequacy of governance arrangements proposed and help ensure that they are (or become) sound.

3.14 Processing patient data outside EU legislation – see 8.1-2

3.14.1 We recommend that the DH, NHS staff and patients check that any service used to process and/or store patient data provides data protection equivalent to that provided in the EU, to ensure that it does not become subject under alien legislation to what would be regarded in the EU as improper disclosure. This check should be done whenever data processing or storage is proposed to be done abroad, in the Cloud or by foreign companies with a presence in the UK. This check may become unnecessary if the current changes proposed for the European Data Protection Directive and Regulations come into force.

3.15 Data that should not be shared – see 8.3

In most current systems, it will be time-consuming to remove third party data that should not shared without the party's consent or to obtain that consent before a record containing it is shared.

3.15.1 We recommend that in future system users move to checking whether the third party consents to his data being shared with others at the time the data is collected, and record this fact on the system at the time of data entry. This can only apply where information is provided by the third party involved or another third party. Third party data also needs to be flagged as such in patient records.

3.16 Shared electronic patient records – see 8.4

Shared electronic patient records (SePRs), where more than one care provider, usually of different types, uses the same care record implementation as the system of prime entry, are being introduced across much of England. This major opportunity to provide more coherent care is being hampered by the lack of suitable data (and clinical) governance arrangements. At the same time, governance becomes difficult to manage and data quality ends up as nobody's baby.

3.16.1 We recommend that the work of the RCGP on the SePR should be extended to include staff from other relevant care sectors (e.g. mental and community health), patients and carers to establish the better models of care that SePRs enable and develop the appropriate clinical and information governance arrangements to support them..

3.17 Sharing and data quality – see 8.5

While sharing of patient data will rapidly become more necessary in the future, sharing poor quality data is likely to propagate the degradation of the uses to which it is put.

3.17.1 We recommend that facilitation of data quality improvement along the lines developed by PRIMIS should be extended urgently into all care domains, and allied with the work on recording standards now going on in the RCP, OpenEHR and elsewhere.

3.18 Governance of patient-entered data – see 8.6

There are many kinds of data that patients and their carers could provide that would help transform the care process, but it is not obvious where and who should hold it in order to maximise the benefit it gives. Neither is it clear how and when it should be shared with others involved in the patient's professional and non-professional care, and how and who should decide this.

3.18.1 We recommend that patients, carers and clinicians should explore these issues together, with a view to drafting requirements in this area and where necessary experimenting with various options before piloting the preferred ones

3.19 A catch 22 with s251 permissions to process identifiable data - see D5

Section 251 approval does **not** remove the need for the processing to comply with relevant elements of the Data Protection Act 1998, of which the most important is probably the data subject's (i.e. patient's) right to object to processing of his personal data on the grounds of the distress or damage it would cause / is causing. As s251 was expressly created to avoid having to contact each patient and seek consent, it is very difficult to ensure that the patient is aware of the processing and so his right to object to it.

3.19.1 We recommend that NIGB is asked to examine the problem in the light of the views of patients, data controllers and secondary users and recommend a resolution.

3.20 New system facilities required

Implementing some of the recommendations above will require a number of additional facilities in patient record systems used by care providers[11]. These include, among others:

– Handling (e.g. entering, recording and displaying) information about:
  1. consent & dissent- for what, who and how long, any constraints, and date given

[11] Some facilities might also be required in PHR systems.

2. corrections to record items, including provenance and date made
3. annotations to record items, including provenance and date made
4. identification of:
    a. 3rd party data (include any 3rd party consent to disclosure provided, and whether the 3rd party is a carer)[12], see 8.3
    b. data which it is considered would be harmful to the patient or others if he was aware of it, including who decided and when. This status may vary over time.
    c. people with whom a particular piece of data should not be shared, e.g. patient contact details with a previous partner
5. facilities to implement the processing and storage of especially sensitive identifiable patient data, see 4.6

- means to interoperate with any PHR(s) possessed by the patient
- preventing & logging improper attempts to access identifiable patient data
- logging access to data designated as especially sensitive by the patient, including when, who by and their role – see 4.6
- logging record access, including when, who, their role and whether they were involved in the care of the patient at the time of access
- logging data copying, including when, who by and their role
- providing pseudonymisation and patient re-identification at source
- central reporting of data security breaches, with mechanisms to inform the patient if the breach poses a significant risk to him, and find and deal with any offender and system issues.

It is doubtful whether any current systems provide all these facilities.

3.20.1 We recommend that the full set of requirements should be established, and be the subject of the minimum necessary standardisation, prioritisation and phased introduction across the UK.

[12] Other clinicians providing care to the patient are not treated as third parties, see [40]

4 Public and patient attitudes to sharing

Patient data should be shared and used in ways that patients expect– there should be *"no surprises"* [15] due to patient lack of awareness of how the NHS operates or by the way that bodies using NHS patient data have interpreted the relevant regulations and guidance on the topic[13]. This is fundamental to maintaining the trust needed for a satisfactory relationship between a patient and the clinician(s) who care for him and maintain records about him. It is also recognised that there is a major public interest in ensuring that patients and the public believe that the NHS respects patient data confidentiality [21 1.2 page 3]. Such trust encourages patients to provide and share the data needed for their care, and they and their clinicians to provide it to others for legitimate secondary uses. But while patients generally trust their clinicians to keep their data secure and respect their privacy, many have serious doubts about the ability of others in the the NHS, and the NHS as a collection of organisations, to do the same, see 4.8.

4.1 The survey evidence available

There are sufficient observational data to reliably indicate public and patient views on the sharing of healthcare data about them. Annex A describes 17 relevant studies [47-63], including six national surveys of sizeable population samples. Annexes B and C summarise the methods they used and their conclusions respectively. One of the older studies, *Share with Care* of 2002 [47] covers primary and secondary use, [51-2,57-60,63] deal with general topics and the rest explore attitudes to secondary use, particularly research. Although the studies do not cover the same ground there is a surprising consistency between their findings where they do and the whole set give a stable and reasonably coherent picture.

So we now have the *"empirical evidence on public and patients' awareness and attitudes"* that the AMS report *Personal Data for Public Good* [67] saw as necessary to inform the *"development of good practice"*. It is therefore possible to establish whether an existing or proposed data sharing practice would surprise patients and if so to what degree.

4.2 Summary of survey conclusions

In general people are content for relevant elements of their identifiable data to be shared for their care on a need to know basis decided without further ado by their clinicians – indeed anecdotal evidence suggests that many are surprised to discover that this does not routinely happen now. However the majority wish to be asked for consent before sharing for non-care purposes.

While the majority is happy for its aggregated, anonymised or pseudonymised data to be shared without consent for purposes other than care, a substantial minority is not, see 4.7.2. Some wish to be able to refuse sharing their data for all or specific research purposes, or for

[13] The NHS reform transition IG documentation of 2011 also uses this as its basic principle, see *NHS Information Governance: Effective Management of Records during a period of transition or organisational change* DH, September 2011.

research by particular organisations, or kinds of organisations. However the studies also show that people's awareness of secondary uses of their data is minimal, see 4.4.

4.3 What is most sensitive varies

It is worth remembering that patients have different kinds of concerns about their privacy. Some may be more worried about the use of their clinical record by ill-wishers to discover their current name or address and telephone number than the unauthorised disclosure of their clinical history.

4.4 Patient awareness of how their data is used

Other than for their care people "*had low awareness of how the NHS uses patient information*"[47]. This has been noted by most studies of public and patient views on the uses of patient data and is always lowest when it comes to secondary uses. Most people are unaware of the existence [50], let alone the role, of the NIGB Ethics and Confidentiality Committee (ECC) or of the central returns that they appear in and the data that they contain. Some assume that their data is used in the running and management of the NHS but know little about what data is used and how. Coupled with trust in their clinicians - especially GPs - this ignorance may go some way to explaining why there are relatively few complaints about the use of their data from patients. There is no doubt that for the use of personal data to be considered fair in terms of the DPA 1998, patients should be given more information about what it is used for and how and access to more detailed information if they so wish. This includes giving clear and easily accessible information about the presence of any opt-out, and how to use it [14]. It would be courteous and helpful to extend this to the use of their de-identified and aggregate data.

Increasingly aware of their legal obligations, many care providers now seek explicit patient consent to collect patient-related output measure (PROM) data and other information and explain what it will be used for as they do so. Even so, much more needs to be done to make the intended use of their data visible to patients: [60] summed it up well in its conclusions:

"*Valuable and socially useful forms of data sharing cannot be protected through obscurity. Doing so cedes the discussion to the most vociferous privacy activists. Instead a new settlement on the use of medical records must be constructed, through a genuine dialogue with the public on the benefits and risks of the uses of medical data*"

A variety of channels are needed to reach patients where the use does not involve explicit consent, with more refined and reliable targetting where an opt out is offered for a secondary use.

4.5 Using patients' data for their healthcare

An electronic patient record can now be very rich, and include data generated by healthcare professionals, the patient and third parties

[14] For example, much of the furore created by the Summary Care Record was due to the obscure way in which the opt out was provided. The patient opt out from GP data collection offered by the Health Research Support Service pilots undoubtedly existed, but was not mentioned in any of the most obvious seeming documentation, even the leaflet created for patients [85]

including carers and relatives. It can comprise data about the patient's state and social context, conclusions made about the patient, treatment requested, planned and provided, and data about others, such as carers, where they affect or could affect the patient's health. It will refer to, or have attached, messages, documents and multimedia material used and created during the process of care. For the clinician it also acts as a legal record of what they observed, concluded, did and learned from others.

4.5.1 Currently patient information governance by clinicians during care in the NHS is generally sound and conforms to the considerable body of guidance available [see UK material in 14-46]. While fair processing notices (also known as privacy notices) and explicit consent are notable by their absence, there are relatively few patient complaints about the use and sharing of data collected about them by the health care professionals involved in their care[15]. However patients have significant concerns about the NHS's ability to keep their data securely, see 4.8.

4.5.2 The storage and use of the patient's data for their healthcare is considered by clinicians and (almost all) patients as an integral part of their treatment. The act of seeking treatment – and registration in the case of General Practice – is assumed by clinicians and almost all patients to imply the use of data already present in their patient record and the collection and recording of data considered relevant to their care by their clinicians. However [47] qualified this with "… *a third of the quantitative group wanted to be asked every time* {identifiable} *information was used, including for treatment. The discussion groups considered this option but rejected it as not feasible".*

4.5.3 Patients generally trust the clinicians caring for them to do the 'right thing' with their data [47] and are aware that clinicians have a duty of confidence towards them. This duty flows from Common Law, and is reflected in current clinical ethics which strongly support the need to protect patient data confidentiality. "*Most people trusted the NHS to protect patient confidentiality Overall, people were more concerned about who used their information and whether it was anonymous than how it would be used"* [42]. *"On the whole people were comfortable with their GP, hospital doctors and emergency services having access to their data, though they reserved the right to limit access to very sensitive information (via the "virtual sealed envelope{see below}). People felt that all others treating them should be allowed access to relevant information at certain times on a "need to know" basis"* [47]. Other studies [51-2,57] also emphasised the 'need to know' basis for sharing data.

4.5.4 For almost all patients the belief that their clinician will do the 'right thing' extends to the selection of the relevant data to share with other clinicians currently or prospectively involved in their care, as indicated by commitment 2 of the NHS care Record Guarantee [42]. Patient may insist if

[15] For the NIGB view of carer's concerns see [90, answer to Q31]

they wish that specific information should not be shared[16] with other clinicians involved in their care, a right guaranteed - with some cautions - by the NHS Care Record Guarantee and mentioned in the NHS Confidentiality Code of Practice[15]. The venereal disease regulations permit the sharing of information about infectious venereal disease cases with other clinicians in order to find, check and if necessary treat contacts of the patient.

4.5.5 For privacy to have much meaning during care, the patient's clinical data should only be available to the clinicians caring for the patient and then only the data relevant to that care. There is anecdotal evidence that the complexity of CfH's role based access control (RBAC) system, the difficulty of ensuring that it reliably reflects staff changes and variations in people's roles and case loads in near real time and the detailed interaction needed with existing systems to share the data classifications that drive what a person in a role can see mean that the patient's record content is shared more freely than should be the case.

4.6 Patient control of the sharing of very sensitive data.

As [47] indicates, there is public appetite for the ability to control the sharing of bits of their recorded healthcare data that they regard as very sensitive. It reported that "*if given a "virtual sealed envelope" which they controlled* {access to}*, about 60% of respondents would put none of their health information in it. Around 25% would put a little bit in it, and 8% would want to put a lot or all of it into it."*

4.6.1 Sealing facilities are mentioned as available in the NHS Care Records Guarantee [42 p14] in the *"newer electronic record systems"* [17] but do not appear to have ever been implemented under the CfH banner. NHS Connecting for Health refined the requirement into two levels, *sealing* and *sealing & locking*[18]. The presence of a sealed item (but not its contents) would be visible to people outside its origin, so that the viewer could if he wished ask the patient for consent to see the content. The intention is that a sealed item can be broken into by a patient's clinician where he feels that access is in the patient's vital interests and he is unable to obtain consent, for example if the patient is unconscious and critically ill in A&E. Sealed & locked items would only be visible to the person who recorded it, or others caring for the patient in the unit concerned, as the patient wishes; no-one else would be aware of its existence.

4.6.2 However a number of practical issues need to be considered by those who use and design sealing facilities and the devil is very much in the detail. A key one is what sealing means when it is applied to an item being

---

[16] This puts the GP in an interesting position, as one could (not unreasonably) say that a GP practice has a continuing relationship providing care to the patient throughout the patient's period of registration, even if the patient never sees a member of the practice staff. The public might well agree.

[17] At present comprises the Summary care Record and (all?) other systems installed by LSPs under the CfH banner. If implemented, such facilities should apply be available in all electronic record systems.

[18] 'Sealing and locking' was mentioned in the v4 Care Record Guarantee but is not in the current one, v5.

copied into the record from elsewhere, such as a test result or discharge letter arriving at a general practice. There is also the challenge of needing to seal all the items from which a concealed fact can be inferred, as well as the fact itself. For example a test request, test result, defining symptoms and prescription may more or less imply a concealed diagnosis, especially when considered in conjunction. Retrospective sealing is problematic if the subject matter has already been shared (which may itself be very difficult to determine where the request to seal arrives long after it the data was recorded by the clinician): unsharing isn't an option. Sealing may have to apply to much more than the data considered sensitive by the patient where the latter is part of a much larger and indivisible item, such as a scanned document. The Care Record Guarantee now states[19] that the patient's right to seal an item may be overridden by the patient's clinician if he feels that it is not in the public interest to do so. There are also questions about how patient control over sealed material could be implemented[20].

4.6.3 Patient-controlled sealing recognises that what is considered very sensitive varies from patient to patient, and may change over time. It would give the patient greater control of what they regard as most private,[21] and so takes the steam out of one of the most serious objections that patients have to sharing their data. In this way it relieves the patient's clinician of making some of the most difficult decisions about whether to share relevant information with colleagues or not. In some cases the clinician may be in a position to seek the patient's consent to share sealed data. All major GP systems have a sealing facility of some sort but it is not common in systems elsewhere, and it is not known whether or how thoroughly care record systems forming part of the National Care Records Service implement the facility.

4.6.4 The ultimate sealed envelope is a personal health record (PHR) to which the patient alone controls access. He may share all or part of the contents with whom he pleases when he pleases as the information technology context and standards permit. For more on PHRs and the possible privacy concerns associated with them, see 8.6.2-4..

4.7 Using patient data for secondary purposes

4.7.1 While all surveys that asked the question found that the majority of the public and patients were were in principle willing to take part in patient data-based research, the surveys found that attitudes towards secondary uses are very different to those towards primary uses.

"*In general, people felt that information released outside of the NHS, or used inside the NHS for purposes other than treatment, should be*

---

[19] This constraint appears for the first time on page 14 of v5 of the Care Record Guarantee .

[20] It is not clear how (or even if) the patient would control access to sealed items, or whether it would based on trust, all accesses by someone other than the content author being reported to the patient for checking after the access had occurred.

[21] Presumably other than access by statute or court order. The Health Select Committee endorsed the need for sealed envelopes and stated that sealed data should only be used for secondary purposes with patient consent, see 2007 report on EPRs, p98, www.publications.parliament.uk/pa/cm200607/cmselect/cmhealth/422/422.pdf.

anonymised – or patient permission sought to use identifiable data. Once information was anonymised, a majority in the qualitative group were happy not to be asked for consent to share it. Some would like to be informed as a courtesy" [47].

The output from [49] was more demanding: "*We have seen .. that consent is a key issue in securing the general public's acceptance of the use of personal health information in medical research. When given a variety of scenarios in which consent might not be essential, only a maximum of a third of people agree with them. Indeed just over one in five (21%) do not find any of the scenarios acceptable, indicating that consent should always be sought"*.

The observations from [50] tended to confirm this view "*There was strong agreement across all the groups that explicitly being asked for their consent to take part in biomedical research was a good thing. Even those that were very positive about taking part in biomedical research, and would readily give consent, stressed the importance of the consent seeking process. There was some variation in how stringent the consent requirements were for different types of research: the most minimal consent procedures were required for routine compilation and analysis of statistics"* {i.e. aggregate data}".

In [54] when asked whether identifiable data could be used for secondary purposes without patient consent[22], 5% of the public and 4% of patients said yes, 53% and 46% respectively said no, and 30% of both said only after approval by a group such as PIAG.

All studies that raised this question indicate that the commonest wish is to be asked for consent before identifiable data is used for secondary purposes. [54] appears to assume that identifiable sealed (and sealed & locked) data would never be used for secondary uses. There was some antipathy towards the use of patient-identifiable data outside the NHS, partly because of greater fears that it might be at greater risk of unauthorised processing [47, 57].

4.7.2 Despite the fact that de-identified data may be legally processed without consent of the data subject[23] a majority of studies reported that a substantial fraction of people[60 p4 para 4] – in one large survey[61] a majority of young people - thought that de-identified data should not be processed without their consent. In [54] 29% of the public and 19% of patients thought that effective anonymisation was not possible and only a minority of the public and patients thought that truly anonymised data should be used without consent. [50] noted similar doubts about anonymisation, but found that consent was not required, as did [47].

---

22 Unless of course the sharing is required or enabled by law, a court order, by the substantial public interest (a ground currently only used by the ECC).

23 Court of Appeal case R v Department of Health, ex parte Source Informatics (2000) .Anonymised data is not considered personal data under the DPA, and therefore not subject to it. However de-identified data may still be considered personal data in some circumstances , see H1.

[54] was the only study to look at the use of anonymised data from sealed envelopes[24]. It found that:

Question & answer	Public	Patients
Should anonymised sealed envelope data be used for secondary purposes?		
Yes	30%	26%
Yes, if patient consents	43%	52%
Never	26%	22%
If people are asked for consent , when should this be done?		
Every time the data is used	73%	77%
Just once	19%	16%
Only in specific circumstances	7%	7%

The results for the equivalent question about pseudonymised sealed envelope data were similar. As s251 approval is only permissive, it would not override the patient's concealment of data by sealing of either sort.

4.7.3 Besides concerns about privacy potential research subjects may not wish their data in any form to be used for certain kinds of research. For instance, Jehovah's Witnesses would object if their data was used for blood product transfusion research and most Catholics could be expected to do likewise for studies into biochemical contraception. Patients may also object to who processes their data. While the NHS was generally acceptable academia was less so and commercial organisations the least acceptable [54,57].

4.8 Keeping shared data secure

People's attitudes to permitting the sharing of their data, the need for consent and the kind of consent required are heavily coloured by the fear that data collected for their care by the NHS may be shared without proper authority in both its original home and at any other sites it is subsequently legitimately copied to[51, 52, 54]. This may be by accident or by design and hospitals figure prominently in one survey[63]. Several recent large losses of NHS data have exacerbated people's concerns, and the NHS figures prominently in the list of data protection actions taken by the Information Commissioner's Office in the period Jan 2010-Dec 2011[25], see http://www.ico.gov.uk/what_we_cover/taking_action/dp_pecr.aspx . In some cases patients are so worried that they have considered withholding data, seeking treatment at NHS providers with better data security records or even occasionally going outside the NHS [63] for care. It also makes people less willing to share their data for secondary uses.

In the light of this, all holders of electronic patient data, whether identifiable or not, should be firmly encouraged to adopt information security best practice, q.v at least as stringent as required by the NHS Information Governance toolkit [41]. The obvious first steps to securing the data from the patient's point of view is to encrypt it and where appropriate to de-identify it, see Annexes H & I. However for many the same fear extends to de-identified data as they doubt that anonymisation or pseudonymisation will

[24] Whether the study was concerned with sealed & sealed & locked data, or just sealed data, is not clear from the report of the results.

[25] In that period, the NHS was involved in 1 of the 4 prosecutions, 0 of 9 monetary penalty notices, 32 of 108 undertakings and 0 of 4 enforcement notices. The prosecution concerned an NHS staff member supplying patient contact details to her boyfriend for commercial gain.

effectively hide their identity [44,48] when de-identified material is used in conjunction with material from other databases.[26]. The same high standards of security should therefore be applied to databases of both identifiable and de-identified data, with the addition of extra identity-inference countermeasures to uses of the latter.[27] . Making unauthorised processing (including sharing and re-identification) a criminal offence attracting substantial penalties would help deter potential malfeasants.

Rigorous application of the 'least' principle, especially data copying and restriction of users to those who have a legitimate reason to use the patient record (typically clinicians caring for him) will play a major role in keeping the data secure. There is considerable anecdotal evidence that confidential emails and faxes about individual patients are sometimes routinely handled by staff who are not entitled to see them, and that printed copies are not kept securely out of the view of those not authorised to see them.

If these concerns were allayed, it is highly likely that public and patient concerns about the sharing of their data, particularly de-identified data, would relax considerably[54]. Wider public awareness of the thorough information governance procedures and precautions taken against accidental loss or theft of data in place at a number of significant healthcare data centres[28] would help damp down these concerns.

4.9 Patient rights and obligations over their data

At this point it is worth summarising at this point what a patient's rights and obligations about sharing his identifiable data appear to be. That patients and the data controllers of their healthcare data must share their personal data as required by statute, orders of a court or when it is decided that the public interest overrides the duty of confidentiality to the individual and the need to maintain trust in a confidential health service is taken as a given. At law the patient has no right to prevent the use of his de-identified data, although a substantial minority of the patients and public do not agree with this. The list takes into account the results of the patient and public opinion surveys and the contents of official guidance, in particular version 5 of the Care Record Guarantee[29]. It comprises:

1. the right to access records about him (subject to the clinician's right to withold material which he considers would be harmful to the patient or others, and third party data not contributed by the patient, see 8.3) [30]
2. the right to control of the sharing of identifiable data recorded about him for the purpose of his care (although a patient should take his clinician's advice before exercising this right, and ideally should agree what to share with him). NIGB [66 3.9] qualifies this right with *"unless*

[26] Including data provided by the data subject or friends to social media applications such as Facebook.

[27] For more on these see [73 chapter 9.3]

[28] E.g. the research safe havens run by THIN, QResearch, GPRD, BioBank and the central data servers run by EMIS, INPS and TPP for clinical record keeping systems that support healthcare provision

[29] Items 1-6, 9-12 come from version 5 of the Care Record Guarantee, although it only mentions sealed data, not sealed and locked data as version 4 did.

[30] *"The Information Commissioner has made it clear that having online access to medical records does not replace formal rights of access under the Data Protection Act (DPA) and patients can still make subject access requests in the usual way".*[45 5.3.3]

{the data} *is part of a direct clinical communication such as a referral*."

3. the obligation, subject to the preceding right, to allow the NHS to share as much information about him as it needs to provide him with healthcare or to check the quality of care provision
4. the right to more direct control of the sharing of data that he designates as particularly sensitive (viz sealed and sealed & locked envelopes[31]).
5. the right to an informed choice over whether his data should shared for his care with organisations outside the NHS, should it be considered in his best interests to do so
6. the right to control sharing his identifiable data for secondary purposes, subject to override by the NIGB ECC. The legal basis for this is the Common Law right to confidentiality, Article 8 of the Human Rights Act 1998 and the data subject's right to object to the processing of their data under the DPA 1998 II 10
7. the obligation to provide as much relevant information as he can to his clinicians to enable them to provide safe, appropriate care to him, in the knowledge that the clinician may record it. This a moral obligation only and has no legal force
8. the obligation to allow information about him to be shared where it is required by clinicians for the care of others[32]
9. the right to have his records kept securely and only made available to people who have a right to see them
10. the right to have any belief he has that his record is being viewed inappropriately investigated and to receive the results, and the right to have inappropriate access discovered by the NHS reported to him, and action taken by the NHS
11. the right to notify his clinician(s) of material in his EPR(s) which he discovers to be factually incorrect and to request its correction. Other than in extremely exceptional circumstances, erroneous data (annotated as such) should still be present in the record as others may have relied on it before it was identified as such.
12. the right to ask for opinions or comments in his EPR(s) which he believes to be untrue to be annotated with his views. For the same reason as in 9 the item being annotated should not be deleted, but if he feels that the information in his record is causing distress or harm, he can apply to have it amended or deleted.
13. The opportunity to be "*notified of of opportunities to join in relevant ethically approved research and* [the right] *to choose whether they wish to do so*" . This appears on page 53 of the latest version of the NHS Constitution Handbook [38] and applies to both interventional and observational studies. [33] [34]

---

[31] But see footnote 19.

[32] Clinicians may share a patient's data with others without his consent where they consider that other patients' interests override the confidentiality owed to the patient [25]. For sexually transmitted diseases treated by specialist services, this right is enshrined in the law to assist contact tracing & treatment.

[33] While laudable it is not clear how *relevance* is determined. For patients yet to be diagnosed with one or more "*relevant*" conditions, it may be done by seeking patient consent at or after the time of

14. the opportunity to consent or refuse to provide his identifiable or de-identified data as safe haven content if he so wishes, providing that if he consents any wishes he has about non-use of specific kinds of data / for specific kinds of purpose / by specific bodies / kinds of bodies are respected and it is made clear to him how the safe haven will operate.

diagnosis. For those already diagnosed it implies access to the patient record, typically the GP record. This should only be done by a clinician caring for the patient and not by researchers. The right to choose whether to participate or not appears in [55] and is cited in the latest AMS report [94 6.4.4 para 2].

[34] This is related to the finding of Q19 in survey [54], see 5.

5 Secondary user's opinions on using patient data

Study [54] was carried out by Connecting for Health and shows that in general researchers are significantly less concerned about patient privacy than patients and the public, see the tables below[35].The views of NHS and Social Care staff fall between the two.

Q15 Should identifiable patient data be used for additional purposes without explicit patient permission?	**% of patients & public**	**% of NHS & s. c[36]. staff**	**% of researc hers**
Always	4	6	9
Yes if approved by PIAG or a similar group	30	42	54
Depends on researcher or type of research	13	23	25
Never	50	29	11
Q11 Should anonymised patient data be used for additional purposes without explicit patient permission? (only partial data published)			
Not possible to anonymise data	25	19	13
Always	7	14	23
Yes if potential benefit to patients & public	2		16
Q8 Should anonymised sealed envelope data be used for additional purposes?			
Yes	28	53	76
Only with consent of the patient	47	33	20
Never	25	14	4

Q18 Given approval by an NHS Research Ethics Committee & the organisation holding the patients data, should researchers be able to use patient records to find suitable study subjects?	**% of patients, carers & public**	**% of NHS & s.c. staff**	**% of researc hers**
Yes without seeking PIAG support	7	10	29
Yes with PIAG approval	44	52	49
Depends who the researcher is	4	5	7
Depends what the research is about	6	8	6
No	39	23	6
Q19 Should it be possible to flag patient records to show if patients are willing to be contacted directly by researchers?			
Yes for all types of research	50	61	66
Yes for certain types of research or researchers	32	26	25
No	17	11	7

Q18 shows a substantial regard by all groups for PIAG (the Patient Information Advisory Group, now the Ethics and Confidentiality Committee of the National Information Governance Board). The major concern among many patient, carer and public respondents to Q19 was that flags should only attached to patient records with the consent of the patient, which begs the question of how that consent is obtained.

[35] In all questions the small numbers of 'don't know' answers have been ignored, so the column totals do not add up to 100%. The survey data also included the views of the category: research councils, ethics committees and professional bodies. These are not shown in the extract above.

[36] s.c = social care

6 Researcher access to patient data

Researchers have been seeking easier access to patient data since the passing of the Data Protection and Human Rights Acts in 1998, and the subsequent strengthening of the legal status of data protection and the right to privacy and their enforcement. Patient identifiable data is now only accessible to researchers with the patients' consent, a s251 approval from the NIGB's Ethics and Confidentiality Committee (ECC) or a very substantial public interest claim (which is extremely rare). Researchers are the most frequent customers of the ECC. In this connection it is interesting to note that the latest AMS report[94] makes no complaints about the performance of the ECC. The major publications putting forward researcher's views are:

- *Personal data for public good: using health information in medical research* The Academy of Medical Sciences Jan 2006 [67]. This claimed that the law already permitted researchers to use patient identifiable data without consent. However as far as the author is aware no researcher has obtained access to identifiable patient data on this basis, and certainly not since 2006[37].
- *Data Sharing Review* Richard Thomas & Mark Walport, Department of Justice, 2008 [70]. This was a report about sharing all forms of data and not just patient records, but Sir Mark Walport was (and is) Director of the Wellcome Trust and a clinician whose major interest is research.
- *Towards consensus for best practice Use of patient records from general Practice for Research* Wellcome Trust June 2009 [81]
- *UK e-health records research capacity and capability* Medical Research Council Jan 2011 [93]
- *A new pathway for the regulation and governance of health research* The Academy of Medical Science Jan 2011 [94]

The messages from the major publications [67,70,94] are consistent where they cover the same ground, although their recommendations have evolved somewhat. The latest report is the most relevant for our purposes and we look at it in detail below.

6.1. Comments on the body of the 2011 AMS report

A new pathway for the regulation and governance of health research is concerned with making the overall regulation and governance of health research in the UK easier, a mission which BCS Health thoroughly endorses. Our comments on the body of the Report are represented below in descending order of approximate significance:

1. While the studies mentioned in box 6.6 on page 68 clearly indicate public support for research[38], the fourth study[50] warned that: "*Even those that were very positive about taking part in biomedical research, and would readily give consent, stressed the importance of the consent seeking process".*, Two of the reports cited and others mentioned in Annex C

[37] The author does not know if an attempt to use identifiable data on this basis has ever been attempted.

[38] As do all the studies listed in Annex C that covered the question.

reached a similar conclusion. Several of the surveys listed in Annex C found that the same view also applied to the use of de-identified data. The evidence that we now have indicates that patients and the public would not support the Report's implication that identifiable personal data should be available for researchers to use without patient consent.

2. The Report makes no mention of the advances in privacy enhancing technology that now enable pseudonymised data to satisfactorily perform most of the functions that previously mandated the use of identifiable data, see Annexes G, H and I. With proper controls on its use and users, it renders most of the complaints made about patient data access obsolete: effectively de-identified data are not considered personal data and so fall outside the scope of the DPA and the common law duty of confidentiality, and do not impinge on the Human Rights Act right to privacy. They can be legally used without patient consent (although as noted above in 4.7.2 a substantial percentage of patients think that consent should be requested, and/or would like to be able to choose whether to become a research participant).

3. The INVOLVE-AMRC workshop [55] was commissioned as an adjunct to the production of the AMS report and attended by 30 participants involved in research as patients or in a more formal capacity. While very supportive of research, the following views emerged:

   - "*We don't want researchers to be regulating themselves"*
   - "*Patients are partners in care and want to work with medical professionals to get the best from healthcare. We need to shift away from the paternalistic approach of the old days …."*
   - *"The group … emphasised that the key consideration for patients would continue to be whether their data is anonymised or identifiable. Views about the use of identifiable data varied but consent was seen as a given, supported by simple guidelines and clear information about how data is accessed, by whom, how it is managed, and how it is kept secure*."
   - *"…some doctors and health care managers saw research as a burden and not as a routine part of healthcare"*

   Considered alongside the other observations this indicates that there is significant culture problem within the research community itself.

4. The proposal for "*a Health Research Agency "*[39] is welcome. However proposing that it should "*encompass the responsibilities for both general ethical approval (including the functions of the ECC)"* would make one of the leading proponents of the uses of patient data for research purposes an arbiter of whether it should be used in this way and generate an obvious conflict of interest. It would also fragment the mechanisms for approving the use of identifiable patient data for secondary purposes (which also include NHS finance, management, administration and planning) and remove them from the overall oversight of the single independent body charged with dealing with patient information governance in the NHS (currently the NIGB).

[39] As the HaSC Bill 2011 currently stands - Jan 10 2012 – this proposal has not been incorporated into the Bill, and a patient information governance committee and the ECC will become part of the CQC.

5. While under the heading "*inappropriate constraints on access to patient data*" on page 4 it goes on to say "*..access to patient data for research is currently hampered by a fragmented legal framework, inconsistency in interpretation of the regulations, variable guidance and a lack of clarity among investigators, regulators, patients and the public*", the report gives no information about how much of the problem described in 1.2.2 is specifically due to the protection currently given to patient data, and how much to the regulatory mechanisms themselves.
6. Although cited in the report, the UK Clinical Research Collaboration has refused to publish *Attitudes and awareness amongst General Practitioners (GPs) and patients about the use of patient data in research* [56], which it commissioned as background for the AMS report [40]. No reason has been given.
7. The Report omits mention in box 6.6 on page 68-9 of two of the larger opinion surveys [47, 54], the second of which also demonstrates the contrast between public, patient & carer and researcher views, see 5 above. It also misses several of the smaller studies shown in Annex B. For these reasons it cannot be taken as a representative picture of patient and public views.
8. That there has never been a legal action against a researcher for abuse of a patient's right to confidentiality could equally well be interpreted as implying that the current arrangements are sound, not excessive as p99 10.3 principle 1 suggests[41].
9. While the example used to demonstrate bias caused by seeking explicit consent shows that there is a difference in patient properties between the unconsented and consented samples[42], no evidence is presented that demonstrates that the difference is caused solely or even in part by the seeking of consent, or that the work necessary to show this has been done.

6.2 Comments on the AMS Report 2011 recommendations

The recommendations directly relevant to patient data access are given on page 5 in section iii:

"We urge the Government to evaluate progress in implementing the recommendations of the 2008 Data Sharing Review. Specifically we recommend that:

---

[40] A request from the author for a copy in June 2011 was refused. For more on this, see http://www.ukcrc.org/aboutus/boards/boardsubgrouponpa/ and follow 'Useful Link' to Minutes of 30th June 2010 and look at Item 6. The minutes file is named 3272_Item+2+-+Minutes+-+UKCRC+Board+Sub-Group+on+Public+Awareness+Meeting+-+30+June+2010+(2).pdf.

[41] There has in fact been a High Court case involving the sequelae of the leakage of a clinician's confidential healthcare data from a Cancer Registry (which the plaintiff won). The clinician was suspended from work by her NHS employer for five years prior to the High Court hearing. *BMJ 336: 63 (Published 10 January 2008)*

[42] On consent & bias, the PRIVIREAL 2005 report to the EU Commission [17] by the Article 29 Working Party said on page 19 in end note 3 *"Arguments that doctors are too busy to get consent or that 100% samples are needed are rarely plausible, and are almost always made in too general a way without attending to all the factors needed for a derogation from consent*."

- *"'Safe havens' are established as a matter of urgency to allow access to data for approved research.*
- *Accredited investigators and research team members should be considered part of a clinical care team to enable identifying patients eligible for approved studies.*
- *The UK Data Protection Act should be reviewed to identify and amend aspects requiring clarification and to inform proposed revisions to the EU Data Directive."*

Each is examined in more detail below.

6.2.1 The opening general recommendation makes no mention of the fact that the Data Sharing Review led directly to the controversial data sharing proposals embedded within the Coroners & Justice Bill 2009. These were vigorously opposed by the BMA, the BCS [8], Privacy International and others, and were removed from the Bill before its enactment. This is not something any government would be wise to repeat

6.2.2 'Safe havens' already play a major role as data sources for researchers and this recommendation is welcome provided their behaviour and governance recognises their heightened responsibilities for ensuring that patient confidentiality is respected. As with patient information governance generally, the devil is in the detail. Safe havens are explored in greater detail in Annex F, q.v.

6.2.3 Most patients do not wish researchers to use their records without consent to see if they are suitable to take part in a research study as suggested by the second recommendation [48,50,54,60] and see the response to questions 15 and 18[43] in 5. One survey [54] found that the biggest single group would accept the process if the NIGB ECC approved it, as happens now. NIGB itself endorses this approach and explicitly rejects researcher access to patient records for this purpose [43.44]. The selection of patients as study subjects can be done using pseudonymised data, with the pseudonyms of potential subjects passed back to the source care provider for re-identification and contacting by provider staff. Surveys of patients and researchers indicate majority support for recording the results of asking patients whether they are willing to allow researchers to contact them and minority support for allowing patients to exclude certain types of research or researchers from such a consent, see 5, q19 in the table. Honouring any patient wishes not to use their data where they regard the data involved as particularly sensitive, or object to the purpose of the research or the organisation or kind of organisation carrying it out is desirable but difficult to achieve algorithmically unless explicit consent is used to obtain study subjects. It is only feasible for data collected after such wishes have been

[43] This Report recommendation may be aligned with a recent change to the NHS Constitution Handbook [38 page 53], which now says that; *"Procedures to ensure that patients are notified of opportunities to join in relevant ethically approved research and are free to choose whether they wish to do so. Research is a core part of the NHS. It enables the NHS to improve the current and future health of the people it serves. The NHS will do all it can to ensure that patients, from every part of England, are made aware of research that is of particular relevance to them"*. It is difficult to see how this could happen without access to information in patient record, which is implied by the life sciences initiative, see 7.5.1 bullet 1 2nd sub-bullet.

expressed and recorded by source systems, ideally in a standard format. However it would be difficult to implement in retrospect where the data is already in a safe haven, and impossible where the data has been anonymised. It would involve designating a standard location where such consent constraints could be found.

6.2.3 As for the third recommendation, while the current research governance framework is undoubtedly far from ideal, the report does not establish that the law and regulations governing access to identifiable patient data are significant obstacles to research, and in particular fails to indicate how the DPA could be changed in a way acceptable to the public that would lead to a notable improvement. Neither did the preceding Data Sharing Review. One step that researchers cannot remove is the consideration of any new request for access to patient data by its data controller, e.g. a GP, or a Caldicott guardian, before a secondary use can take place.

6.3 Conclusions on the Report

The Report is a view from the research establishment of research governance issues. While we agree with much that it says, we do not see evidence in the Report or elsewhere that supports its recommendations on patient data access. Its recommendations:

1. run counter to the
    a. Confidentiality NHS Code of Conduct [15],
    b. Care Record Guarantee [42],
    c. conclusions of NIGB [43,44] on the selection of research participants,
    d. the views of patients and the public, q.v. that headline results of the surveys listed in Annex D, over half of which were commissioned by research bodies or bodies such as the Academy of Medical Sciences which are part of the research lobby.
    e. the tone and some of the substance of the Involve-AMRC report [55] quoted above,
    f. the undertakings given about patient control of the sharing of patient data in the White Papers [9,10] that preceded the Health and Social Care Bill 2011.[44]
    g. The activities and beliefs of much of the research community, see Appendix I for an example of such views.
    h. The views of the Information Commissioner [14,15]
    i. The proposed changes to the European Data Protection Directive[101]
2. fail to . fail to recognise the conservation of patient privacy as part of any of the principles the Report contains, even principle 1 "*safeguard the well-being of research participants"*.

[44] For more on the relevant elements of the White Papers, see 7.1

3. indicate the need for a change of culture in the research lobby, see 6.1 bullets 1, 3, & 4.
4. suggest that there is less awareness of and commitment to research among clinicians caring for patients on behalf of the NHS than is desirable. However withholding the UKCRC report on GP and patient attitudes [56] does not help the report make the case for this. Anecdotal evidence suggests that care provider enthusiasm for research is diminished by the fact that researchers do not routinely reimburse care providers for work they do on their behalf
5. do not take into account the current capabilities of PET

We believe that properly exploiting modern privacy-enhancing technology, encouraging greater awareness of the need for, and benefits of healthcare research amongst care provider staff, being prepared to pay them for the work they do to assist researchers and developing a more symmetrical relationship with patients, their clinicians and health managers would go a long way towards resolving the difficulties in obtaining patient data and study participants. Where this is already happening, notably in some of the safe havens now in operation, the problems identified by the research lobby have shrunk greatly, and in some cases ceased to be a matter of concern. In the (few) cases where the researcher believes that identifiable patient data is needed and that it would be too onerous to obtain patient consent, the ECC should be approached as now (but see the complication mentioned in last paragraph of D5).

7 Issues raised by the current reform of the NHS

7.1 The White Papers and on patient data confidentiality

We welcome recognition in *Equity and excellence: Liberating the NHS* [9] and *An Information revolution: a consultation* [10] of the importance of patient data confidentiality and the measures promised to maintain and strengthen it, including confirming the patient control over the sharing of their data already given in the Care Record Guarantee. However we do not see any reason given or obvious elsewhere for unpicking the current arrangements for the development of standards for data "*safety, security, reliability and resilience"* for health and social care by splitting it between the Commissioning Board and the DH rather than making them the remit of one independent body such as NIGB.

7.1.1 The Equity & Excellence paper [9] proposed that:

(a) "*Shared decision-making will become the norm: no decision about me without me.*" [Executive Summary 4a]

(b) "*Patients … will have increased control over their own care records.*" [Executive Summary 4b]

(c) "*the Department{of Health} is committed to evidence-based policy making and a culture of evaluation and learning*" [1.23]

(d) "*The patient will determine who else can access their records ….. We will consult on arrangements, including appropriate confidentiality safeguards, later this year*" [*2.11*]

(e) "*there will be safeguards to protect personally identifiable information"* [2.13]

(f) "*We will clarify the legal ownership and responsibilities of organisations and people who manage health data. This may require primary legislation and we will consult on arrang*ements *later this year*" [2.16]

7.1.2 The Information Revolution Consultation [10] said:

(a) "*The information revolution depends on a 'presumption of openness', which will mean routine publication of aggregate datasets built-up from data held securely in people's records. This … will not mean releasing data that enables individuals to be identified.*[45] *Personal information will, of course, continue to be subject to strict security arrangements."* [1.11]

(b) "*Control of their records gives patients and service users a clearer understanding of their health needs, their treatments, their care and other options available to them and will help make shared decision-making a reality. Providing patients with greater control of their records is also symbolic of a new relationship between individuals and services.*" [2.3]

[45] This necessary constraint will severely restrict what can be published, see Annex H. For instance, only very simple individual data, if any, could be published on a 'bird table' basis, and even then it should only be for random samples of the study population. Only limited aggregate data could be published unless the study population is very large. These limitations are well described in the draft de-identification standard being developed by the Information Centre for Health & Social Care.

(c) *"Mindful of the responsibilities involved, people will need to make informed choices about the extent to which they want to take control of their records."* [2.6]

(d) *"Opening up access to records and placing control firmly in the hands of patients and service users represents a significant and positive shift in the basis of the relationship between people and their care professionals."* [2.11]

(e) *"Ensuring confidentiality and security of data will be a key concern for service users, and, consequently, a fundamental issue for the success of our information revolution. Where organisations hold patient or service user records electronically, the systems used must meet appropriate standards of safety, security, reliability and resilience. The NHS Commissioning Board will be responsible for centrally developing and maintaining these standards for the NHS. Equivalent standards set by the Department of Health will also be required for social care and for public health services"* [2.17].

7.1.3 '*Control*' according to *Liberating the NHS* and *The Information Revolutio*n comprises facilities for patients to:

(1) access the clinician's electronic healthcare record[46]

(2) control access to their record [Executive Summary 4b,8 2.11]

(3) download a copy [9 2.12,10.2.7]47.

(4) "*interact with them {their GPs}through their records, … This might mean people recording their symptoms, health status, self test results …. and medication they have taken*" [9 2.9]

With some qualification, patient control of the sharing of their records is already provided under the Care Record Guarantee[42] commitments 3, 5 and 6. Commitment 8 gives the patient the right to request correction of erroneous data in the record and to have his annotations about the opinions or comments of others recorded.

Patient access to their records and the right to have a printed copy have been guaranteed to patients since 1990, although under the DPA a copy may attract a fee (which should now be abolished). A downloading option would be novel, and is as yet extremely rare. On-line patient data access is supported by most GP system suppliers and is already happening in a very small but growing number of general practices. Patients can give on-line access to others such as carers or relatives by sharing their password with them, which is not ideal. There is as yet no national standard format for a printed copy or downloading and it is not easy to provide a totally granular standard when record keeping styles and structures vary as much from system to system as they do. However the intermediate record structure used by GP2GP to enable records to move between different GP systems provides a good starting point for a standard electronic interchange format.

---

[46] Even at Haughton Thornley, where patient record access has been encouraged and very well supported over the last five years and those who have are generally very pleased with the result, only about 14% of patients have opted to use it to date: they tend to be those with long-term illness. Only a much smaller number of them went on to use the CfH Healthspace data entry facilities. It would be interesting to know what % of the Wells Park practice patients have opted to access their records, and if they have similar profiles.

[47] Facilities 1 - 3 are commitments in the current version (5) of the NHS Care Record Guarantee [42] published earlier this year.

It should also be borne in mind that the vast majority of ePRs have hitherto been created by the clinician for his purposes and may not be fit for sharing, research, or unaided use by the patient. Clinicians now need to ensure that their records are suitable for the range of purposes for which they might be used, including access by patients, some of whom will need help to understand their content, but should not compromise the record's fitness for their own purposes in the process. Some patient access systems already provide automatic hyper-linking to relevant knowledge sources and so facilitate patient understanding and reduce the input required from the clinician. Third party data in patient records raises issues which are dealt with more fully in 8.3.

Patients need to be aware of the potential dangers to their care of sharing an out-of-date copy of their record with other clinicians caring for them. For this reason, on-line access is preferable where available. Measures also need to be in place to make patients aware of, and reduce the risk of, inadvertent or coerced sharing of their data which might harm them or those for whom they are responsible. Marketing people would also be delighted to get access to patient healthcare data, in order to better target advertising material for health-related goods and services. Moves are afoot by the DH to commission a project to develop patient guidance on this topic from BCS Health.

7.1.2 It is not clear what facility 4 means, and it is discussed further in 8.6.

7.1.3 The truth of the matter is that the record describes the interaction of patient and clinician, and the clinician - often the record's data controller under the Data Protection Act - also has rights and obligations associated with it, as acknowledged in [9 2.5]. In many senses the clinician could be considered as much the subject of the record as the patient and the record as a co-production of a patient and his clinician(s). Being able to access data held by the clinician, suggest corrections and annotations, and have a copy of it are key steps in making the patient a more equal partner in their own care, but do not put the patient in 'control' in the usual sense of the word, or make him a 'data controller' as understood by the DPA.

'Patient control' is therefore a misnomer when applied to the EPR, and creates false expectations. It should not be used. Neither are patients demanding it. Talking of greater patient control of the **sharing of their data** held by or on behalf of the NHS, as proposed by White paper and implicitly in the consultation document, would be much nearer the mark and is of concern to patients.

7.2 How do non-Bill-specific data flows measure up against these promises?

As far as is known the increased flow of patient data required to support the planned expansion of payment by results (PbR) to cover 75% of NHS expenditure (including sensitive areas such as mental health) rather than the current 33% will not be controlled by patients in any way. Neither will the commissioning dataset central returns, some of which contain more detail than the stated purpose would suggest is necessary, leading to

speculation that their sponsors have wider purposes in mind for them than those made public so far. As time goes on more bodies are likely to become involved in patient service ordering and payment, such as referral centres, specialised payment servicing agencies such as debt factors and clearing houses. The temptation for all such flows to be of identifiable or weakly de-identified patient data, e.g. including NHS number, must be resisted, as was noted by CfH as recently as 2010 in [86] [48]. Only the provider and the referrer/service requester need to know the identity of the patient who received the service. There is also a burgeoning amount of analytical work going on in primary care at PCT level or thereabouts, often using linked GP and hospital data. In some cases it uses identifiable data where it is not necessary to do so, even if it is arguably legal. NIGB has stopped one attempt to incrementally build a warehouse of identifiable GP data. We believe that in the light of the advances in PET, claims that dataflows and uses such as these need to comprise identifiable data should be subject to professional scrutiny, e.g. by a body such as the ECC, before being approved.

7.3 How does the Bill itself measure up against these promises?

7.3.1 Despite the statement in the Bill's[49] Explanatory Notes [13 5] that "*The Bill is intended to give effect to the policies requiring primary legislation that were set out in the White Paper Equity and Excellence: Liberating the NHS*", the Bill itself currently takes us in the opposite direction to that promised by the White Papers [9,10]. It makes a nonsense of "*no decision about me without me*" when it comes to data sharing[50]. It rides roughshod over the clear evidence we have of patient expectations about the use of their confidential data, see 4.4-7 and their rights and obligations, see 4.9. It does not stand up as an example of "*evidence-based policy making*", see 7.1.1 bullet 3. For the view of the Bill's proposals from an expert in data protection legislation, see [93].

7.3.2 The Bill[12] proposes new powers for sharing identifiable data by direction and regulation by the Secretary of State or NHS Commissioning Board. Regulation involves mimimal parliamentary scrutiny and assent, which may not take place before the full House: direction involves none. Many of the new statutory gateways involved can be further extended by the Secretary of State by regulation. These powers are broader than those proposed in the original Coroners and Justice Bill 2009, which proposed creating data sharing orders by regulation. Dr Chris Pounder, an expert on data protection legislation, said in [101] "*Perhaps even, in future, an*

---

[48] Section 4.4 on page 10 noted "*Invoicing requires changes in working practices in that Finance Departments routinely handle identifiable data for which there is no legal basis, nor any real need. Invoices are based on paying for activity and the question of whether the activity is for the correct commissioner is a data quality issue and must be handled as such. This is being pursued with the SUS PbR User Group and DH Finance Directorate.*"

[49] This refers to the version sent from the Commons to the Lords on the 8th Sept 2011. The patient data sharing arrangements mentioned in it do not appear to have changed since the Bill entered the Commons.

[50] NIGB considered that "no decision about me without" should apply to patient data sharing, see answer to q4 in[91].

intelligent Minister might justify the creation of a mega-medical database of patient data in terms of efficiency! Perish that thought but look at Clause 250(1)(a) if you want a chapter and verse."

The new Health and Social Care Information Centre (HaSCIC) is the main vehicle selected to carry out this work and would be empowered to:

- establish information systems as directed by the Secretary of State or Commissioning Board [12 s251], or as requested by *any other person*. Monitor, NICE, CQC or others defined by regulation [12 s252] may make *mandatory requests,*
- require [s255] any health or social care body to provide it with the information it needs for its functions in the form and when required, irrespective of any duty of confidence owed by the person providing it
- publish [s256] and disseminate [s257] the information it collects, excluding identifiable patient information, subject to well-perforated constraints and no external checks or balances other than from the Secretary of State if he suspects the HaSCIC is not performing as it should. It seems from other sources [91] that the IC is intended to become a major "safe haven", collecting and linking patient data, before distributing it to would-be users. The Bill does not say whether the Information Centre will be accumulating patient-identifiable data[51] over lengthy periods of time. The Bill would permit the dissemination and publication of identifiable data about care providers or clinicians under certain conditions.
- delegate any of its functions, including data collection, publishing and dissemination, to any other body [s265]

The Bill intends to make the Information Centre and National Centre for Clinical Excellence Health Service bodies in DPA 1998 s69 (3) and the Access to Health Records Act 1990[52]. As yet it is not clear from the Bill's explanatory notes and other published material what this is intended to achieve.

7.3.3 Other gateways proposed that involve or may involve identifiable patient data would allow:

1. the Secretary of State to force disclosure of personal information et al by the Commissioning Board or care commissioning groups (CCGs) [s17]
2. The Commissioning Board to force disclosure of personal information et al by CCGs [s23].
3. The Commissioning Board [s20] and CCGs [s23] to disclose information under certain circumstances, "*notwithstanding any common law duty of confidentiality"*
4. Monitor to require information from CCGs or providers [s96(1)(e), s97 (1)(a), s102], and via regulations enabled by [s72(1)(c)]. The information

[51] Identifiable data covers data only containing unique identifiers such as NHS number, as well as data including one or more of name, address, date of birth, telephone numbers, e-mail address and date of death,

[52] http://www.publications.parliament.uk/pa/bills/lbill/2010-2012/0092/lbill_2010-20120092_en_44.htm

is not expected to identify living individuals, but the proposal permits this.

5. Local Healthwatchs to disclose information to organisations such as Healthwatch England [s180], and be able to require information from service providers [sch 15]. It is not clear whether this would involve identifiable patient data or not.
6. Healthwatch England to report annually on people's needs and experience of health services to the Care Quality Commission [s178] This is not likely to involve the use of identifiable patient data, but it is not precluded from doing so.
7. Local authority health and wellbeing boards and local authorities to obtain information from local authorities and board members [s196] Again this is unlikely to concern identifiable living individuals, but is not precluded from doing so.
8. Local authorities to request private providers of commissioned NHS and public health services to answer questions
9. The Health Service Ombudsman to share investigation reports and reasons for declining to investigate as she considers appropriate [s198]. The complainant has no say in the matter.
10. [s35] the Secretary of State and others concerned with the approval of professionals to carry certain functions under the Mental Health Act 1983 to share information used in connection with an approval function.

Proposals 1-3 in particular could, as they stand and fuelled by the increasing competition amongst healthcare providers and for customers (i.e. patients) that the Bill is intended to provoke, lead to a situation somewhat similar to the USA. There a deliberately weakened HIPAA Act has severely damaged patient data confidentiality and public belief in the confidentiality of health care professionals and the services they provide.

7.3.4 As already noted in 6.2.2, there is a good case for looking at tighter regulation of the behaviour of those organisations responsible for large and comprehensive databases for secondary uses, such those run by the Health Research Support Service, in other existing and proposed safe havens and at the Information Centre, as these pose a more significant risk of patient re-identification than single-source data collections. Several existing and proposed safe havens known to the author already operate to standards that significantly exceeed the legal requirements, and only use pseudonymised data. Additional primary legislation that weakens the current protection of identifiable personal data would run counter to the pledges given in the Information revolution White Paper [10]. is undesirable and in the light of the advances in PET unnecessary.

7.4 Top-level NHS patient information governance

7.4.1 As well as proposing major new statutory gateways to require or permit the sharing of identifiable patient data the Bill also proposes to remove the top-level mechanism dealing with patient information

governance, the National Information Governance Board for Health & Social Care (NIGB). NIGB and its Ethics & Confidentiality Committee have worked well, and it is worthy of note that although the majority of applications to the ECC are by researchers, the latest AMS report on research governance [94] makes no complaints about the performance of the ECC. No reasons or evidence have been supplied to support the case for doing away with the NIGB, although it is obvious from published material that it does not see eye-to-eye with the research lobby about the use of patient data.

7.4.2 The proposal to place NHS patient data governance policy generation and oversight under a significant customer for personal health data[53], the Care Quality Commission, will create a dangerous conflict of interest. Doing away with the single most effective bastion of patient privacy further devalues the commitments to enhance the protection of patient data confidentiality given in the White Papers, see 7.1. It is essential to retain NIGB as it is, or replace it with a body of with equivalent independence and standing. Such a body is particularly necessary during the Information revolution promised to accompany the major reform of the NHS that is imminent.

7.5 Using patient data in life sciences research

7.5.1 In response to the research lobby, the government has announced major measures to make it easier for the life science and healthcare industries to use data from patient records. These were first made public in the Prime Minister's speech of the 5th of December 2005[105], and are elaborated in three subsequent papers [105-108]. The proposals include:

1. "*We will support patients to have access to novel treatments, and be part of the development of wider patient benefits by consulting on an amendment to the NHS Constitution so that, whilst protecting the right of an individual to opt out, there is a default assumption that:*
   - *data collected as part of NHS care can be used for approved research, with appropriate protection for patient confidentiality; and*
   - *patients are content to be approached about research studies for which they may be eligible, to enable them to decide whether they want a discussion about consenting to be involved".* [107].
2. "*launching a new secure service to link primary and secondary care data at an unidentifiable patient level, and investing £60m in a secure Clinical Practice Research Datalink (CPRD) to provide researchers with access to patient data for clinical trials recruitment and observational studies*" "*There will be the provision of secure data linkage services by the Health and Social Care Information Centre by September 2012…*" [106]
3. "*Our capacity to link patient data to biological samples is also being strengthened. The NIHR* [National Institute for Health Research] *is investing £2.5m pump-priming this year in a new national Bioresource [that] will help companies to recruit patients for stratified experimental*

53 *"The Department of Health Advisory Non-Departmental Public Bodies Review has indicated that the functions of the NIGB will transfer in part to the Care Quality Commission and part to the proposed Research Regulator"* http://www.nigb.nhs.uk/pubs/annualreport2010.pdf letter from NIGB to Andrew Lansley accompanying the annual report. The second part of this arrangement is not mentioned in the Bill, and it is not clear whether it is still intended to apply.

medicine studies as well as providing the potential to study the molecular basis of disease, identify the most appropriate biomarkers for diagnosis and drug discovery, and to test the mechanism of action and effects of new drugs. This resource will complement the UK Biobank, led by MRC and the Wellcome Trust." [106]

4. *"London's three AHSCs* [Academic Health Science Centres]*, (Imperial, Kings Health Partners and UCL Partners) will explore the potential to develop information systems that build on the NHS record and pull together patient level data for London's population. This will enable large groups of patients to be engaged in world-class clinical research on disease-specific and personalised biological therapies, regenerative medicine and medical devices*." [99]

5. *"In order to maximise opportunities for utilising patient data to support research, we have launched a crossfunder call for Centres in e-health, which will commit £15m to Centres. These aim to build and sustain a vibrant health informatics research capability in the UK. Outline proposals are being considered at present and awards will be made in mid-2012*." [98]

7.5.2 These initiatives and the patient opt out facility are welcome, but more information is needed before the soundness of the patient information governance arrangements can be assessed. Phrases such as "*The end result would be that every willing patient is a research patient*" [97] are disingenuous and do not inspire confidence: every patient who does not opt out becomes a research patient, whether willing or not. Opt-out consent is certainly not explicit in EU[46 III.A.3] and DPA[36 B9,22-23] terms , and therefore not suitable for collecting sensitive personal data such as patient data. We also know that sharing de-identified data about individuals in its original form via the 'bird table'[54] is the highest risk way of sharing such data as far as personal privacy is concerned, so much so that it may need to be treated as sharing personal data. Among the information needed soon are answers to the following questions:

(a) why do we need both the CPRD and the HaSCIC?
(b) how will the potential data subjects of CPRD and HaSCIC be made aware of the existence of the opt-outs to both defaults, and the full consequences of not using them, i.e. how their data will be shared, with whom and what for?
(c) Will the opt-outs apply to the initiatives mentioned in 2, 3 & 4 above?
(d) Will a qualified opt in be available to allow someone to bar use of his data for a particular purpose, or use by particular organisations or kinds of organisations?
(e) How will patient data be made "unidentifiable"?
(f) Will the CPRD and HaSCIC provide aggregate, anonymised or pseudonymised data for secondary uses, or any combination of these?

[54] i.e. for anybody to use without let or hindrance

(g) If data is pseudonymised, will the re-identification of data be available from CPRD and HaSCIC, the data sources or not at all?

(h) Will the CPRD and/or the HaSCIC collect data reactively or prospectively, i.e. are they intended to collect and link patient data as and when a customer requires it, or to be data warehouses that accumulate increasingly comprehensive records of linked patient data as time goes on, or both?

(i) Will CPRD and/or HaSCIC collect and link data from non-healthcare sources, such as ONS and social care?

(j) Will either or both of the CPRD and HaSCIC collect identifiable personal data?

(k) If it is collected, will either or both of the CPRD and HaSCIC retain identifiable personal data, e.g. after any necessary linking and/or pseudonymisation is done?

(l) Will the CPRD & HaSCIC ever provide identifiable personal data to secondary users?

(m) What kind of organisations and people will be able to request data from the CPRD and/or HaSCIC?

(n) How will CPRD and HaSCIC make data available for secondary uses:

I. on-site only (i.e. the CPRD & HaSCIC act as safe havens)?

II. as a copy to be taken away by the secondary user?

III. either, as the customer wishes?

(o) If secondary users can take copies of data from the CPRD and HaSCIC, what restrictions will be made on how they may use it? For example, will they be able to give it to others or attempt to re-identify patients, and/or use it for any purposes, e.g. other than those notified to the CPRD or HaSCIC when the data was requested?

(p) How will such restrictions be enforced? By the criminal law, as part of a contract or licence or a mixture of all three?

(q) What penalties are proposed for the misuse of data about people provided to secondary users by the CPRD and HaSCIC?

(r) Will secondary users serviced by the CPRD and HaSCIC be allowed to retain any data provided, and if so for how long and under what conditions?

(s) What measures will the CPRD and HaSCIC take to assess and minimise the disclosive power of any secondary use data / results supplied to secondary users?

(t) Will the CPRD and HaSCIC provide data only for research, or for all secondary purposes?

(u) Will the CPRD and HaSCIC themselves use the data they collect for secondary purposes (other than linking, quality control and distribution to third parties)?

(v) Can data held by the CPRD and HaSCIC be requested and used by organisations or people outside the UK?

(w) Could data provided in this way become subject to the national laws of another country, and therefore move beyond the control of the NHS (and the UK)?

(x) Will patient data be provided free by the CPRD and/or HaSCIC? If not, what tariff arrangements are proposed?

(y) In 3 above, is it proposed to make patient sample material itself available to researchers? Exactly what service(s) is the "*new national Bioresource*" to provide?

(z) Will the London AHSCs (see 4 above) make use of the data linking facilities described in 2 above, or have their own? If the latter is the case, then questions (c)-(x) are also relevant to them.

(aa) What is the role foreseen for the "*Centres in e-health*" mentioned in 5 above?

(bb) How will patients and their care providers be involved in the governance of the CPRD and HaSCIC[55] provision of data to third parties?

[55] The governance arrangements for the General Practice Extraction Service run by the current NHS Information Centre are already published, see http://www.ic.nhs.uk/gpes. While we cannot yet be certain that they will cope satisfactorily with all the uses of primary care data likely to arise in the reformed NHS, they form a good and clear starting point for others to use. They include an Independent Advisory Group that examines all requests for primary care data that come to GPES. Of 10 members, 4 are lay members

8 Other concerns about sharing patient data

8.1 Non-EU legislation governing personal data

8.1.1 *"Considering the elevated risk to the personal data in an EHR system in an environment without adequate protection, the Article 29 Working Party wants to underline that any processing – especially the storage – of EHR data should take place within jurisdictions applying the EU Data Protection Directive or an adequate data protection legal framework".* [EU Article 29 working party, WP 131 *On the processing of personal data relating to health in electronic health records (EHR*), Feb 2007 [22]

8.1.2 "A*n adequate data protection legal framework"* means whatever the user wishes it to mean and offers no significant protection. . Health data is sent abroad for processing, e.g. dictated letters for computer entry, and documents for scanning, and some of the data involved has appeared on the open market (cf the Irish hospital data breach described in http://www.mxsweep.com/blog/bid/65375/Irish-Hospital-Admits-Encryption-Neglect-After-Data-Breach). Three pieces of research in the last 9 years have concluded that the implementation of the 'Safe Harbor' regulations intended to preserve the privacy of European data in the USA is at best poor[56,] Patients and care providers planning to send identifiable data for processing outside EU jurisdiction should ensure that strict, enforceable safeguards and significant penalties for breaches are in place to prevent unauthorised disclosure, in the last resort through binding contractual arrangements. They also need to be certain that disclosure cannot be authorised in totally unexpected ways by legislation that applies to the chosen data processor. For example, US 'safe harbour' arrangements will not prevent use of the Patriot Act by the USA government to access data controlled by and/or processed by US companies, even where it is stored on territory outside the USA: presumably this would apply to patient data in Microsoft's Health Vault.

8.1.3 But help may be at hand [100, 109]. The revised data protection directive submitted to the EU Parliament in January 2012 states that [109]- "*EU rules must apply if personal data is handled abroad by companies that are active in the EU market and offer their services to EU citizens*"]. If approved, the revised directive and regulations must be implemented within two years.[57]

8.2 Cloud services

The Cloud is an extreme case of the above, where a data controller uses one or more services available over the Web. Data storage would be one such service. Which national law applies to Cloud activities, especially data storage? Is the Cloud just a data processor in DPA terms? Can you be sure that deleted personal data really is deleted? Contracts with Cloud providers should ensure that these features are stated to the satisfaction of the

[56] The last, in Dec 2008 by Galexia Pty, [78], indicated that the situation was, if anything getting worse rather than better. For example, of the 1597 entries on the Safe Harbor List, only 348 were complied with the Safe Harbor regulations and of these only 54 were compliant for all categories of data. 206 entries were false.

[57] There is major concern in the USA over the impact on US businesses of these proposals [110].

would-be service user and his clients, e.g. NHS patients. "*The data controller using it {the Cloud} can delegate the responsibility for ensuring privacy, but not his accountability"* [84]. Again the Patriot Act applies if the client data involved is held by a company domiciled in the USA, and as with 8.1 the recent proposals of the EU Justice Commissioner[109] would provide an EU-wide solution to the problem.

8.3 Sharing third party data

8.3.1 EPR content[58] provided by or about third parties may be shared only if [DPA 7(4), 45 5.3.3, 7.5, 12.2]:

1. it does not reveal the third party's identity, or
2. the information is only being provided to the patient and he provided the information in the first place
3. identifying details are redacted before the data is shared or
4. the third party consents to the sharing, or
5. none of the above apply, but the data controller considers, in the context, that the need to share outweighs the third party's right to confidentiality

Other clinicians who are caring for the patient do not count as third parties [DPA 7(4)(c)]. De-identification becomes more difficult when the identity may be inferred from the information provided by a third party, which is often easiest for the patient to do

8.3.2 The ideal solution is to ask providers[59] of third party data when they provide it if they are content for others (especially the patient) see what they have said. It may be worthwhile to anonymise the data, although this can be difficult to do without removing much of its meaning, and is difficult or impossible where third party information is part of a more extensive and indivisible data item, such as a scanned document. Systems need to have facilities to indicate data as about and / or provided by a third party, to record if consent to share has been given, and whether that consent extends to the patient or not. The recorder, e.g. GP, can of course choose to attempt to record it without explicit or implicit attribution, but this is not a desirable practice.

A common proxy to avoid this problem is to only share coded and certain types of structured data, in the knowledge that (almost[60]) all third party data is in text form. The rub is in 'almost': there are specific Read and Snomed CT codes that provide information about third parties, such as spouses.

8.3.3 A carer is a particularly important example of a third party. "*A key concern that has been expressed to NIGB is that carers often feel unable to be completely honest in their interactions with professionals because of*

[58] Third party data is much more common in mental health than GP records, and NHS mental health care providers have staff whose job it is to remove it (and items that might led to harm to the patient or others) before allowing others (including the patient) to see the record

[59] Where they are not the patient

[60] E.g. the 'spouse' Read codes such as 13HD. *violent spouse* . These should also be removed from the record before it is shared with the patient and others unless one of the conditions listed in 8.3 applies.

concerns that such information will be written down and then inadvertently disclosed to the patient themselves when the carer does not want the patient to know that this is how they feel. Considering carers as a separate service user and needing their own support will help with this. Staff training to be mindful of and sensitive towards these issues is important to ensure carers' views are taken into account" NIGB, Response to the consultation on the Greater Choice paper, Jan 2011 [90 answer to Q31]

8.4 Shared electronic patient records (SePRs)

8.4.1 Detailed care records used by multiple care domains, such as general practice, community health and others, as their record of prime entry are currently being implemented in the UK. In theory this should lead to better integrated care, however it raises the following question[61].

Is domain A (e.g. community health) entitled to see data held by another, B (e. g. general practice) without patient consent

1. if A is providing care for the patient or has been asked to do so?
   (a) even if A is providing care, should A be entitled to see **all** the information held by B, much of which may not be relevant to the care currently being provided by A?
2. if A does not provide, or been asked to provide care to the patient?

Putting the Good Practice Guidelines (GPG) v4 [45] to one side for a moment, the current guidance indicates that A's element of an SePR should not be shared with other providers using the SePR unless they are also caring for the patient, and even then that only relevant information from A should be shared with other SePR providers caring for the patient for the duration of that care, unless in either case the patient gives their consent[62]. If sharing is by browsing other's records, it is not clear how sharing is restricted to the relevant information, which may well change over time, and will very likely be different for each of the other care providers sharing the SePR.

8.4.2 All this essentially treats the SePR database as being run by a DPA *Data Processor*, with each contributing healthcare provider acting as the *Data Controller* of his own data. This would seem to match patient expectations: most patient patients would, one suspect, be very surprised to discover that their GP record was available to a set of care providers with whom they have so far had no contact (though the converse might be more palatable). Unfortunately it stultifies the potential that the SePR offers for joining up care, once clinical practice and patient expectations get to grips with it.

---

[61] Or put another way, how would sharing only the relevant data with other clinicians, a la [1, 15, 42], square with giving others using the same shared detailed care record browsing rights when the patient is referred to them , or worse still, as soon as the other clinician creates his patient sub-record?

[62] One could (not unreasonably) say that a GP practice has a continuing relationship providing care to the patient throughout the patient's period of registration, even if the patient never sees a member of the practice staff other than (almost always) at registration. Patients might well agree. This would imply that the GP should be able to see all other users' data for his patient. The only question might be whether he should see data prior to the date the patient registered with him, some of it might well be relevant to the patient's future care.

8.4.3 The GPG v4 principles are rather different [45 5], and attempt to realise at least some of the opportunities that the SePR offer for joining up care. It therefore raises somewhat different issues. Many of these have major clinical governance implications, but the sharing challenges appear to include:

- how are all patients with a provider that moves to an SePR informed of the new governance arrangements, and thereby given a chance to consent or otherwise to sharing their data with other providers using the SePR?
- the statement in GPG v4 5 2.5, penultimate paragraph that "*these other organisations may view and copy this data* {i.e. another provider's data} *and use it for their own purposes*" is on the face of it rather alarming
- the fact that a care provider can make an entry in another provider's 'record' (as evidence by GPG v4 5.2.6 principle 9)is even more so
- who (if anyone) is responsible for the overall quality of the SePR? Presumably the author of each entry in a 'record' assumes responsibility for it, but not for copies made by others and appearing elsewhere. Someone who copies someone else's data into his 'record', and/or copies his data into someone else's record is presumably responsible for ensuring that sufficient context accompanies it to ensure its correct interpretation, but not for any content that he did not author. Lack of clarity about who is responsible for what, and the power to enter data into someone else's record will reduce a clinician's ability (and probably also his willingness) to control the quality of his part of the SePR.

One option (if the system has the facility) is to set every patient's status to *don't share* at the outset, changing it only as each patient is seen during the normal course of events and asked his preferences[63]. In the meantime sharing through explicit messaging, e.g. on referral and discharge, would continue as normal among the providers sharing the SePR.

8.4.4 The principles in GPG v4 5.2.6 are at a high level and the promised DH generic data sharing agreement has yet to appear. As GPG v4 notes, we are not far down the (new) road of learning how to use the SePR, but it obvious that to make the most of it, care providers sharing an SePR and their patients need to evolve a way of working[64] that makes the most of the opportunities it presents and then to create a patient information sharing model that supports such a modus operandi.

8.5 Sharing & data quality

Sharing data for a purpose for which it was not collected means that the data will usually almost always be less suitable and complete for that purpose than is ideal. While sharing may produce worthwhile results, they may be less accurate, or at worst, occasionally incorrect.

---

[63] On average GPs see c.80% of their patients within 2 years, and those include almost all those who need most care.

[64] For example, given use of a common problem-orientated framework across all domains involved in the care of a healthcare issue, the SePR could offer an unrivalled opportunity to have an integrated view of the care being provided for a healthcare issue.

Sharing data makes it all the more important that the source data is accurate, up-to-date and comprehensive enough for the purpose for which it is being shared. It is very unlikely that corrections to source data will ever be propagated to any secondary databases that include it[65], and unsharing is never totally possible (if at all)[66]. Sharing poor quality data may lead to the provision of unsafe and/or inappropriate care, or harm patients in other ways. Tacit knowledge that helps interpretation of data, and provides awareness of known errors, rarely travels with the data being shared. However not sharing data for personal care can embody even greater risks, unless the patient makes his clinician aware of what he needs to know by other means.

Patient access to their information[67], data quality facilitation such as that provided by PRIMIS +, peer comparison, clinician training in record keeping, the generation and use of clinical standards / guidelines[68] for record keeping, the deployment of standard information structures and good system design where data entry properly reflects the business process and semantics are all known to play an important part in ensuring the quality of information in electronic patient records. Central facilitation of these initiatives is therefore now more important than ever.

Data extractions for secondary use should mirror the way the business they are reporting operates, as well as the purposes they are serving. Their data structures should be well defined, and extraction guidelines for those providing the data need to be clear and comprehensive. Validation of data extracted should be comprehensive. Sadly this is true of all too few centrally-collected datasets. Patient records are sometimes considerably less useful for their purposes than researchers believe them to be.

In spite the risks, exposing poor quality data to others can be a powerful force for its improvement, but only if appropriate and timely feedback loops to the original data sources are in place. Responsibility for ensuring the quality of patient healthcare data ultimately lies with those who generate / first record it, and the attitudes of their management to ensuring good data quality. The best guardian of data quality is its use for their own purposes by the person who recorded it, which is not always true of data collected for secondary purposes, as [9 5.7] acknowledges.

---

[65] Inaccurate data was the cause of the Helen Wilkinson case, and a preceding case where an error in a pathology laboratory message resulted in a patient being recorded as having a significant mental health problem and was consequently denied a mortgage.

[66] Which makes it all the more desirable to share by referring to the original rather than copying it. In the electronic world this requires stable references, (typically. URLs) of the material being shared and a means of restricting the extent of the material referred to.

[67] As [10] noted, the value of this is illustrated by a study undertaken in a general practice in 2004, in which 70% of patients found at least one error or omission in their electronic medical record, and 23% found an error or omission that could be described as important, q.v. *Patients' experiences when accessing their on-line electronic patient records in primary care*. Pyper, C., Amery, J., Watson, M., & Crook, C. (2004) British Journal of General Practice, **54**, 38-4

[68] E.g. the RCGP Good Practice Guidelines [45], and the work of the RCP on record keeping & headings.

8.6 Sharing patient-entered data

At present the patient's / carer's observations and views in the clinician's record have been filtered and summarised by the clinician before entry. Patient / carer data entry as envisaged – however loosely – by the Information Revolution White Paper [10] is very different. Besides annotating annotations entries made by their clinicians in their EPRs, suggesting corrections to erroneous data and recording their own observations and measurements, patients could:

- describe the history of their presenting complaint(s)[69], or
- give their view of their quality of life with a long-term illness since their last annual review, or
- complete a professionally created questionnaire at the behest of their clinician

before attending the consultation with the clinician. They could add informal material for use by their professional and non-professional carer(s) – who keeps the house key, that they have pets which need looking after, their other responsibilities as a carer, and so on. They could interact with their clinician(s) to get a seek advice, request a service, answer a query, tho' this is much more than just sharing data. Although it is not a magic bullet, there is no doubt that patient data entry and electronic interaction of patients with their clinicians and informal carers could play a major role in transforming the care process.

8.6.1 Other than that solicited by a clinician, e.g. via an electronic questionnaire, sharing patient entered data raises many options and questions.

Patients could enter data into:

- their EPR(s) if the clinician(s) agreed,
- a record which they alone control – a personal health record (PHR)
- or a mixture of the two.

Whichever is used:

- agreement will be required about how patient-entered data is to be shared with others involved in the care process, e.g. in referrals
- does patient-entered data have the same credence in clinicians' eyes as that entered by clinicians? Should it be used for automated decision support? Who decides?
- what should other users of the record do when confronted with material entered by the patient and the clinician that conflicts?

If patients enter data into the EPR:

- who is responsible for its governance? Should it be decided on a patient by patient basis with the clinician involved?

---

[69] This has been used to some effect by a hospital urogenital department to reduce the embarrassment of women attending for the first time, as well improving the quality of the consultation. Commercial patient interrogation software is also available to take general patient histories. The first example was the MICKIE system developed by Chris Evans of the NPL and used in the '70s to guide patients with psychiatric problems through a structured questionnaire with heuristic abilities. Interestingly patients appeared to give more honest information about smoking, drinking and sexual behaviour to MICKIE than they did in a face-to-face encounter with a GP.

- should it go into a specific portion of the record for which the clinician becomes the data processor, but the patient is the data controller?[70]
- is either solution practical?

A very small percentage of GPs may already copy data recorded by patients into their EPRs where they think it would be useful, but as far as is known, none permit the patient to enter unsolicited data directly into the EPR maintained by the clinician. It is not clear whether unsolicited data entry into the clinician's EPR is even desirable, and its introduction would have a major impact on clinical practice and the governance of the clinician's patient record. It is a topic for discussion and experimentation, but not national roll-out.

8.6.2 If the patient chooses to have a personal health record (PHR):

- The PHR could include elements of the patient's EPRs selected by the patient or sent to him by his clinicians, e.g. care plans, future appointments and test results, as well as data entered by him (and carers if he so wishes).
- The patient could use it to hold data he regards as very sensitive, as an alternative to using sealed envelopes, see 4.6.
- This paper has so far considered patient data recorded by clinicians under a duty of confidence to the patient. If patients have PHRs on systems and technology outside the NHS, the situation changes radically, see 8.1 and 8.2. The PHR could be hosted as a service e.g. by Microsoft Health Vault, held by an app on the patient's smart phone or provided by a Cloud-based service (or both[71]).
- there should be a logical means of using any PHR with an EPR during encounters with clinicians, and for the patient to show others all or parts of his PHR, the parts shared being likely to vary from occasion to occasion. If the most is to be made of the PHR data, this implies federating the EPR and PHR electronically on such occasions. At the same time clinicians and patients will wish to ensure that using an associated PHR does not corrupt the EHR or its technical environment and vice-versa.
- If patients and their clinicians wish to use a PHR alongside their EPR(s), the two (or more) records will require some common or mappable structure & semantics. Crucially there will need to be a means of relating data in the PHR and EPR, e.g. so that the patient's and clinician's observations of the side effects of using a particular drug can be considered as a whole. The difficulty of achieving this commonality has been blogged by a member of the Google development team as a major reason why Google Health was withdrawn.

---

[68] Essentially the record becomes a species of shared electronic patient record (SePR, see 8.4), with all content shared by patient and clinician, but with control of further sharing divided as the contributors see fit.

[71] In this context it is interesting to note that Google announced the impending withdrawal of their PHR in June 2011 – see http://googleblog.blogspot.com/2011/06/update-on-google-health-and-google.html

8.6.3 There's no such thing as a free lunch, and patients wishing to use a commercial system to hold their PHR will need to read the vendor's small print very carefully to ensure that their privacy expectations are met.[72]. The potential perils about the applicability of foreign legislation mentioned in 8.1 and 8.2 may apply and it is possible in some systems that the information entered may be shared widely with other parties (although this rarely includes commonplace identifiers such as name and address without the patient's explicit consent, more often sharing IP addresses and leaving cookies instead).

8.6.4 The BCS and others have published relevant generic guidance for the public, and the Information Commissioner has draft guidance on data sharing out for consultation. The NHS needs to provide clear guidance, drawing on these sources and others, and BCS Health have been commissioned by the DH and the BCS Policy Advisory Board to produce it

8.7 Is it a secondary use or not?

Risk stratification currently uses identifiable data from multiple sources to forecast likely service use, particularly re-admission to an acute unit. It involves linking GP patient data with information from a variety of other service providers. As Annex I shows, doing the task with pseudonymised data from different sources, and re-identifying patients at risk in their GP practice in order to intervene is not a problem. It also (quite properly) ensures that the linker &/or analyst does not see or hold any identifiable linked patient data.

But the clinician (typically the GP) doing the planning may require the linked data from other care providers. If this is a primary use, can the linked hospital and community health data about the patient be revealed to the GP caring for the patient (but who may not have been when some or all of the non-GP care took place)? Is patient consent necessary before this can occur? As risk stratification is being used to make decisions about an individual's care, DPA s33 [1] or 2000 SI [3] cannot be used to justify its invasion of privacy. What patients think should happen is not yet known.

[72] E.g. social media applications such as Facebook, YouTube, Linked In and Twitter. How the law should treat their personal content is also currently unclear although the revised EU DPD [109] does recognise these issues.

Annex A References

(shown by type of reference and in ascending order by publication date)

A1 Government Bills, Legislation & Responses

1 *Data Protection Act* HMG 1998
2 *Human Rights Act* HMG 1998
3 *Data Protection (Processing of Sensitive Personal Data) Order 2000* uksi_20000417_en.pdf
4 *Health Service (Control of patient information regulations 2002* [SI 1438/2002]) http://www.legislation.gov.uk/uksi/2002/1438/pdfs/uksi_20021438_en.pdf
5 The NHS (General Medical Services Contracts) Regulations 200432, the NHS (Personal Medical Services Agreements) Regulations 200433 and the APMS Directions 34
6 *NHS Act 2006, sections 251-252* HMG 2006
7 *Coroners & Justice Bill,* HMG 2009
8 *BCS response to the Coroners & Justice Bill* BCS 2009 http://www.bcs.org/upload/pdf/coroner-justic-bill.pdf
9 *Equity and excellence: Liberating the NHS* HMG, July 2010 http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_117353
10 *Liberating the NHS: An Information revolution: a consultation* HMG, October 2010 http://www.dh.gov.uk/en/Consultations/Liveconsultations/DH_120080
11 Karen Thompson *NIGB response to the Information Revolution consultation*, NIGB, 2010
12 *Health and Social Care Bill 2011* HMG, January 2011 onwards http://services.parliament.uk/bills/2010-11/healthandsocialcare.html
13 *Health and Social Care Bill Explanatory notes* HMG, January 2011 onwards http://services.parliament.uk/bills/2010-11/healthandsocialcare.html

A2 Information Governance Guidance

14 *Use And Disclosure Of Health Data* Information Commissioners Office 2002
15 *Confidentiality: NHS Code of Practice* DH 2003 http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_4069253
16 *Joint Guidance On Use Of IT Equipment And Access To Patient Data* DH, GMC & ICO. 2004
17 *Recommendation from PRIVIREAL to the European Commission (d) the research exemption* PRIVIREAL 2005 http://www.privireal.org/content/recommendations/ Page 19 end note 3
18 *Confidentiality and Disclosure of Information: General Medical Services (GMS), Personal Medical Services (PMS), and Alternative Provider Medical Services (APMS) Code of Practice* DH, 24 March 2005 http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/@dh/@en/documents/digitalasset/dh_4107304.pdf
19 *Dealing with subject access requests involving other people's information,* Information Commissioner's Office, July 2007
20 *The Information Commissioner's view of NHS Electronic Care Records* Information Commissioners Office, Jan 2007
21 *NHS Information Governance Guidance on Legal and Professional Obligations* DH, 2007 http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/@dh/@en/documents/digitalasset/dh_079619.pdf
22 *Working Document on the processing of personal data relating to health in electronic health records (EHR)* EU Article 29 Working Party, Feb 2007 http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2007/wp131_en.pdf
23 *Standards of conduct, performance and ethics* Health Professions Council, 2008
24 *SRPGG guidance* RCGP 2009

25 *Confidentiality and disclosure of health information tool kit* BMA 2009
26 *Guidance on: The interpretation of commitment 12 in the NHS Care Record Guarantee for England* NIGB, February 2009 http://www.nigb.nhs.uk/pubs/guidance040209.pdf
27 *Confidentiality* General Medical Council 2009
28 *Confidentiality: disclosing information for education and training purposes* General Medical Council 2009
29 *Confidentiality: disclosing information for insurance, employment and similar purposes* General Medical Council 2009
30 *Confidentiality: disclosing information about serious communicable diseases General Medical Council 2009*
31 *Confidentiality: disclosing records for financial and administrative purposes* General Medical Council 2009
32 *Confidentiality: reporting concerns about patients to the DVLA or the DVA* General Medical Council 2009
33 *Confidentiality: reporting gunshot and knife wounds* General Medical Council 2009
34 *Confidentiality: responding to criticism in the press* General Medical Council 2009
35 *Record Keeping Guidance for Nurses and Midwives* NMC, July 2009 http://www.nmc-uk.org/Documents/Guidance/nmcGuidanceRecordKeepingGuidanceforNursesandMidwives.pdf
36 *The Guide to Data Protection* Information Commissioners Office 2009
37 *NHS Constitution* DH 2010
38 *The Handbook to the NHS Constitution* DH 2010
39 *Enabling Patients to Access Electronic Health Records Guidance for Health Professionals* RCGP, September 2010
40 *Confidentiality: NHS Code of Practice Supplementary Guidance: Public Interest Disclosures* DH November 2010
41 *NHS Information Governance Toolkit, version 9* DH Informatics Directorate, 2011 https://www.igt.connectingforhealth.nhs.uk/Home.aspx?tk=407707739475285&cb=81bfc4d1-a12b-428e-8c0d-40c2befb3ec5&lnv=7&clnav=YES
42 *NHS Care Record Guarantee version 5*, NIGB January 2011 http://www.nigb.nhs.uk/pubs/nhscrg.pdf
43 *Identifying and contacting research participants* NIGB, 2011 http://www.nigb.nhs.uk/advice/identifying
44 *FAQs about identifying research participants* NIGB, 2011 http://www.nigb.nhs.uk/advice/indentfaqs
45 *Good Practice Guidelines for GP electronic patient records v4 (2011)* DH/RCGP/BMA, March 2011 http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/documents/digitalasset/dh_125350.pdf
46 *Opinion 15/2011 on the definition of consent* EU Article 29 Working Party, July 2011 http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2011/wp187_en.p df

A3 Empirical data on opinions & expectations

47 *Share With Care*, NHS Information Authority and Which?, October 2002

48 *Public attitudes towards the use of primary care patient record data in medical research without consent: a qualitative study,* M R Robling, K Hood, H Houston, R Pill, J Fay, H M Evans, J Med Ethics 2004;30:104–109. doi: 10.1136/jme.2003.005157

49 *Use of Personal Health Information in Medical Research*, IPSOS Mori, 2007

50 *Public Attitudes to Research Governance: A qualitative study in a deliberative context*, University of Surrey, 2007, commissioned by the Wellcome Trust

51 *Using patient information in the NHS 2008* NHS Connecting for Health 2008 http://www.connectingforhealth.nhs.uk/patients/consultation/reportfeb08.pdf

52 *Using patient information in the NHS* NHS Connecting for Health 2009 http://www.connectingforhealth.nhs.uk/patients/consultation/reportmar09.pdf

53 *Monitor 1 report* Sept 2009 Wellcome http://www.wellcome.ac.uk/About-us/Publications/Reports/Public-engagement/WTX058859.htm

54 *Summary of Responses to the Consultation on the Additional Uses of Patient Data* CfH,27th November 2009 http://collections.europarchive.org/tna/20100509080731/http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/documents/digitalasset/dh_110715.pdf

55 *Patient perspectives on the regulation and governance of medical research* Association of Medical Research Charities & INVOLVE December 2010 www.amrc.org.uk/news-policy--debate consultation-responses externalconsultations-by-year

56 *Attitudes and awareness amongst General Practitioners (GPs) and patients about the use of patient data in research* UK Clinical Research Collaboration (2010).– a study by the UK Clinical Research Collaboration Board Sub-Group on Public Awareness. UCKRC, London

57 *Using patient information in the NHS survey 2010* http://www.connectingforhealth.nhs.uk/patients/consultation/usinginfo.pdf

58 *Using patient information in the NHS Conclusions 2010* http://www.connectingforhealth.nhs.uk/patients/consultation/conclusion.pdf

59 *Using patient information in the NHS Executive Summary 2010* http://www.connectingforhealth.nhs.uk/patients/consultation/exec.pdf and http://www.connectingforhealth.nhs.uk/engagement/public/consultations/hsreport.pdf .Several CfH public consultations are listed at the URL below, but a URL is only given for one of their outputs. Most significantly for this report, no link is given to the output of the consultation on 'sealed envelopes' (or is this [54], which contains a lot on the topic?). http://www.connectingforhealth.nhs.uk/engagement/public/consultations/

60 *Who sees what Exploring public views on personal electronic health records.* New Economic Foundation, October 2010 http://www.neweconomics.org/sites/neweconomics.org/files/Who_Sees_What.pdf

61 *Privacy & prejudice: young people's views on the development and use of EPRs*. Royal Academy of Engineering 2010 http://www.raeng.org.uk/news/publications/list/reports/Privacy_and_Prejudice_EPR_views.pdf

62 Ipsos-Mori, *Public support for research in the NHS* AMRC, June 2011 http://www.ipsos-mori.com/researchpublications/researcharchive/2811/Public-support-for-research-in-the-NHS.aspx

63 Fairwarning: *How Privacy Considerations Drive Patient Decisions and Impact Patient Care Outcomes*, New London Consulting, Oct 2011 http://www.fairwarningaudit.com/documents/2011-WHITEPAPER-UK-PATIENT-SURVEY.pdf

A4 Other references

64 William W Lowrance PhD *Learning from Experience Privacy and the Secondary Use of Data in Health Research* The Nuffield Trust, Dec 2002 http://www.nuffieldtrust.org.uk/sites/files/nuffield/publication/learning-from-experience-nov02.pdf

65 *A Handbook of Ethics for Health Informatics Professionals* E-H W Kluge, BCS 2003

66 *Handbook of privacy and Privacy-enhancing Technology EU PISA Project, 2003* http://www.andrewpatrick.ca/pisa/handbook/Handbook_Privacy_and_PET_final.pdf

67 *Personal data for public good: using health information in medical research* The Academy of Medical Sciences.. January 2006

68 *Data Protection Technical Guidance Note: Privacy Enhancing Technologies* Information Commissioner, 2006 http://www.ico.gov.uk/upload/documents/library/data_protection/detailed_specialist_guides/privacy_enhancing_technologies.pdf

69 *Report of the Care Record Development Board Working Group on the Secondary Uses of Patient Information* CRDB, 2007 http://webarchive.nationalarchives.gov.uk/20081105162013/http://www.connectingforhealth.nhs.uk/crdb/workstreams/secusesreport.pdf

70 *Data Sharing Review*, Richard Thomas & Mark Walport, DoJ, 2008

71 Macleod, U. and Watt, G.C.M. *The impact of consent on observational research: a comparison of outcomes from consenters and non consenters to an observational study*. BMC Medical Research Methodology, 8 (15) 2008 pp. 1-6. ISSN 1471-2288

72 *Response to the Data Sharing Review* NIGB 2008 http://www.nigb.nhs.uk/pubs/datareviewnigb.pdf

73 Ross Anderson *Security Engineering 2nd edition* Wiley, April 2008 By arrangement with the publishers, *Chapter 9 multilateral security and others are* available free at http://www.cl.cam.ac.uk/~rja14/Papers/

74 *Research Capability Programme Pseudonymisation Study(PD14)* August 2008, http://www.nihr.ac.uk/systems/Documents/RCP_Programme_Documents/PD1420Pseudonymisation20study.pdf

75 *Research Capability Programme Patient Consent Approach (PD15)*August 2008 http://www.nihr.ac.uk/systems/Documents/RCP_Programme_Documents/PD1420Pseudonymisation20study.pdf

76 *Research Capability Programme Operating Model for "Honest broker" services to support research (PD19)* Aug 2008 http://www.nihr.ac.uk/systems/Documents/RCP_Programme_Documents/PD1920-20Operating20Model20for20Honest20Broker20Services20to20Support20Research.pdf

77 *Response to Consultation on Public, Patients and other interested parties views on Additional Uses of Patient Data* NIGB, December 2008 http://www.nigb.nhs.uk/pubs/rcpnigb.pd

78 Chris Conolly *The US Safe Harbor - Fact or Fiction?* Galexia, Dec 2008 http://www.galexia.com/public/research/assets/safe_harbor_fact_or_fiction 2008

79 Sir Michael Scholar *National Statistician's Guidance: Confidentiality of Official Statistics* UK Statistics Authority, Jan 2009 http://www.statisticsauthority.gov.uk/national-statistician/ns-reports--reviews-and-guidance/national-statistician-s-guidance/index.html

80 NHS CFH RCP *Programme PD214 - Output Based Specification* - OBS2 5th March 2009 http://www.nihr.ac.uk/systems/Documents/RCP_Programme_Documents/PD2120RCP20output20based20specification2028OBS29202.pdf

81 *Towards consensus for best practice Use of patient records from general Practice for Research* Wellcome Trust, June 2009 http://www.wellcome.ac.uk/stellent/groups/corporatesite/@policy_communications/documents/web_document/wtx055660.pdf

82 *NIGB Response to GMC Research guidance: a draft for consultation* NIGB, Sept 2009 http://www.nigb.nhs.uk/pubs/gmcconsultation.pdf

83 *NIGB response to the EU consultation on 'The legal framework for the fundamental right to protection of personal data'* NIGB 2009

84 *Cloud Computing Security Risk Assessment* ENISA Nov 2009 – http://www.enisa.europa.eu/publications/studies/reports/act/rm/files/deliverables/cloud-computing-risk-assessment

85 *Information on the Health Research Support Service*, NIHR 2010 http://www.nihr.ac.uk/files/Research%20Capability%20Programme/HRSS%20Leaflet%20FINAL%20150211.pdf.

86 W Gowing *Summary of Pseudonymisation Implementation Guidance Final FV1.3* CfH, 15th June 2010 http://www.connectingforhealth.nhs.uk/systemsandservices/pseudo/pipguidev1_2.pdf

87 *NIGB response to the MoJ call for Evidence on Current Data Protection Legislative Framework* NIGB, October 2010 http://www.nigb.nhs.uk/pubs/mojdpaevidence

88 *NIGB Annual Report 2010* NIGB, 2010 http://www.nigb.nhs.uk/pubs/annualreport2010.pdf

89 *HRSS Pilot Programme: Overarching Governance Framework*, NIHR 20th Oct 2010 http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/index.htm - refers to a document (DH_4008777) that doesn't exist.

90 *Response to Liberating the NHS: Greater choice & contro*l NIGB 2010 http://www.nigb.nhs.uk/pubs/LiberatingNHSgreaterchoiceNIGBresponse.pdf

91 *Response to Liberating the NHS: An Information revolution* NIGB 2010 http://www.nigb.nhs.uk/pubs/InformationRevolutionNIGBresponse.pdf

92 *GSS/GSR Disclosure Control Policy for Tables Produced from Surveys* ONS,2010? http://www.ons.gov.uk/ons/guide-method/best-practice/disclosure-control-policy-for-tables/index.html

93 *UK e-health records research capacity and capability* MRC, January 2011

94 *A new pathway for the regulation and governance of health research* Academy of Medical Sciences, Jan 2011 http://www.acmedsci.ac.uk/p47prid88.html ".

95 BCS Health *Preparing the NHS for an information revolution*, BCS Jan 2011 http:// http://www.bcs.org/upload/pdf/liberating-the-nhs.pdf

96 *Pilot Health Research Support Service Frequently Asked Questions* NIHR 2011 http://www.nihr.ac.uk/files/Research%20Capability%20Programme/HRSS_Pilot_FAQ.pdf http://www.bcs.org/upload/pdf/liberating-the-nhs.pdf

97 Fiona Barr *BMA fears Health Bill will erode privacy* Ehealth Insider, 24th Feb 2011 http://www.ehi.co.uk/news/EHI/6670/bma_fears_health_bill_will_erode_privacy

98 Prof J Hippisley-Cox *Open Pseudonymisation* University of Nottingham 2011 http://www.openpsedonymiser.org

99 Daloni Carlisle *Interview with Tim Straughan* EHealth Insider, 24th October 2011 http://www.ehi.co.uk/insight/analysis/815/ehi-interview:-tim-straughan

100 *Stronger data protection rules at EU level: EU Justice Commissioner Viviane Reding and German Consumer Protection Minister Ilse Aigner join forces* 7th Nov 2011 http://europa.eu/rapid/pressReleasesAction.do?reference=MEMO/11/762&format=HTML&aged=0&language=EN&guiLanguage=en

101 C Pounder *Does the Health Care Bill permit medical research without patient consent?* Amberhawk Training Ltd, November 2011 http:// http://amberhawk.typepad.com/amberhawk/2011/11/does-the-health-care-bill-permit-medical-research-without-patient-consent.html

102 Paul Ohm *Broken promises of Privacy: responding to the surprising failure of anonymisation* 57 UCLA Law Review 1701 (2010)

103 Jane Yakowitz *Tragedy of the Commons* http://ssrn.com/abstract=1789749, 2011

104 Yun Shen, Siani Pearson *Privacy Enhancing Technologies: A Review* HP Laboratories, HPL-2011-113, 2011 http://www.hpl.hp.com/techreports/2011/HPL-2011-113.pdf

105 *Full Text of David Cameron's speech of December 5th 2011 on the NHS and the Life Sciences* http://www.huffingtonpost.co.uk/2011/12/05/david-cameron-on-life-sciences-speech_n_1129316.htm

106 HMG *Investing in UK Health and Life Sciences* Department for Business Innovation and Skills, December 2011 http://www.bis.gov.uk/assets/biscore/innovation/docs/i/11-1428-investing-in-uk-health-and-life-sciences.pdf

107 *Strategy for the UK Life Sciences* HMG Office for the Life Sciences and the Department for Business innovation & Skills, December 2011 http://www.bis.gov.uk/assets/biscore/innovation/docs/s/11-1429-strategy-for-uk-life-sciences

108 Sir David Nicholson *Innovation, Health and Wealth, Accelerating Adoption and Diffusion in the NHS* DH, December 2011 http://www.midtech.org.uk/wp-content/uploads/2010/05/InnovationHealthandWealth.pdf

109 *Commission proposes a comprehensive reform of data protection rules to increase users' control of their data and to cut costs for businesses* Viviane Reding, Jan 25 2012 http://europa.eu/rapid/pressReleasesAction.do?reference=IP/12/46

110 *EU comprehensive data protection reform announcement this week* (update) Mintz Levin Cohn Ferris Glovsky and Popeo PC Jan 2012 http://www.lexology.com/library/detail.aspx?g=47047283-3184-469f-8d27-4d5186daeb20

Annex B Opinion studies on sharing patient data

Titles of studies on research uses only are printed in black, broader studies in green.

Reference number publication title	Commissioner	Method used
47 Share With Care, NHS Information Authority and Which?, 2002	NHS National Implementation . Programme	Survey of 2,087 age 15+, 4 public member focus groups and 36 interviews with members of special interest groups
48 Public attitudes towards the use of primary care patient record data in medical research without consent: a qualitative study JME 2004,	Universities-based research project	53 interviews, 49 of members of the public and 4 of members of local (Welsh) community health councils
49 Use of Personal Health Information in Medical Research, IPSOS Mori,2007	Medical Research Council	Survey of 2,106, plus 6 interviews & 3 public workshops
50 Public Attitudes to Research Governance: A qualitative study in a deliberative context, University of Surrey, 2007	Wellcome Trust	12 discussion groups, each meeting twice – 4 general public, 2 public & patients, 2 public & biomedical, 2 patients & 2 biomedical-
51 Using patient information in the NHS 2008 NHS CfH, 2008	Dept of Health	Deliberative event involving 109 members of the public
52 Using patient information in the NHS 2009 NHS CfH, 2009	Dept of Health	As for ref 45: 49 of the 2008 attendees, & 47 new recruits. All attendees asked to complete a questionnaire
53 Monitor 1 report 2009 (will be repeated every 3 years)	Wellcome Trust	Survey of 1,179 adults & 374 14-18 yr olds interviewed
54 Summary of Responses to the Consultation on the Additional Uses of Patient Data NHS CfH, November 2009	NHS Research Capability Programme	1,598 survey responses, 1,555 consisting of completed questionnaires: 105 responses from organisations. 100 attended 9 stakeholder events for the public and 3 for researchers & NHS staff
55 Patient perspectives on the regulation and governance of medical research Association of Medical Research Charities & INVOLVE, 2010	Academy of Medical Sciences	Workshop for 30 participants with a background of public involvement in research, as patients or in more formal committee or board roles
56 Attitudes and awareness amongst General Practitioners (GPs) and patients about the use of patient data in research UKCRC, 2010	UK Clinical Research Collaboration	Unknown, as UKCRC have declined to make the report publicly available: it has however been cited in [86] & ??
57 Using patient information in the NHS survey 2010 NHS CfH, 2010	Dept of Health	As for ref 47, including a rerun of the questionnaire: 72 previous attendees plus 38 new recruits
58 Using patient information in the NHS Conclusions 2010 NHS CfH	Dept of Health	*See 52*
59 Using patient information in the NHS Executive Summary	Dept of Health	*See 52*

2010 NHS CfH		
60 Who sees what Exploring public views on personal electronic health records. New Economic Foundation, 2010	Wellcome Trust	c.6000 over two years, including adults and young people
61 Privacy & prejudice: young people's views on the development and use of EPRs... Royal Academy of Engineering	Wellcome Trust, EPSRC, ESRC & MRC	e-poll results of 2,900 school pupils who attended an event, 31 young people at a 2 day conference, 6 focus groups (5 of pupils and I of adults)
62 Public support for research in the NHS Ipsos-Mori, June 2011	AMRC	990 adults interviewed across Great Britain at home using laptops. Results weighted to reflect known profile of adult population.
63: *How Privacy Considerations Drive Patient Decisions and Impact Patient Care Outcomes*, New London Consulting, Oct 2011	Fairwarning	1001 patients surveyed across the UK – no idea what kind of patient, although most had GCSEs so were mainly over 16.

Annex C Conclusions of opinion studies

Titles of studies on research uses only are printed in black, other studies in green.

Reference number & title	Main conclusions
47 Share With Care	1. People trust the NHS to look after their data 2. People have a low awareness of how the NHS uses their information 3. People more concerned about who used their info, and whether it's anonymised or not, than how it is used 4. If given a facility to control access to their data, 60% wouldn't use it, 25% would use it a bit and 8% a lot 5. People happy for GPs, hospital doctors & emergency services to access all their data: all others should only have access to relevant information when and if they need to know it. 6. Information used outside the NHS, or in it but not for care, should be anonymised or patient consent sought 7. Majority happy for anonymised information to be used without consent, although some would like to be informed
48 Public attitudes towards the use of primary care patient record data in medical research without consent: a qualitative study,	1. This exploratory study highlights public concerns when research uses medical records without patient consent 2. Public acceptability regarding the use of medical records in research cannot simply be assumed 3. Patients were concerned that helping researchers could compromise their GP's ability to provide care to them 4. Further work needed to see how widespread such views are and to inform those advising on confidentiality issues.
49 Use of Personal Health Information in Medical Research	1. Public awareness of the use of personal health information for the purposes of medical research is low 2. Confidentiality and consent feature highly in the debate and are central to building public trust 3. If the public is informed about what medical research entails, they are generally positive towards it 4. Views of people with long-term conditions are generally more positive 5. Attitudes to medical research are generally positive and if communications are handled well, this might increase propensity for agreement to use personal health information for medical research purposes
50 Public Attitudes to Research Governance: A qualitative study in a deliberative context	1. Participants, even previous biomedical research participants, realised their awareness of biomedical research issues was low and wanted to know more 2. Only individual identifiable data was seen as personal data and to be protected: aggregate data was not. 3. Sensitive data, e.g. about sexual & mental health & sexual behaviour, was always seen as personal data. 4. Participants were not unwilling to provide personal data for research if they understood why it was wanted and had confidence in the integrity of the research process 5. Anonymity becomes problematic where participants believed that health information pertinent to them could emerge from the research and not be conveyed back to them 6. Where personal data were required for biomedical research, anonymity was generally seen as important, but there was scepticism about guarantees of anonymity based on people's experiences 7. Understandings of consent in the research process are borne out of general conventions of courtesy in

	interpersonal interactions. 8. Even those very positive about taking part in biomedical research, and who would readily give consent, stressed the importance of seeking consent. There was variation in how stringent the consent requirements were for different types of research: the most minimal needed was for routine compilation and analysis of statistics 9. Implied consent was seen as no consent: routinely varying an agreed data use was generally unacceptable 10. People recognised there are many views of consent, implying difficulties for a 'one size fits all' consent process. 11. Considering potential consequences to oneself and others of giving consent is an important heuristic in judging whether to grant consent 12. People valued the personal approach in seeking consent, with the GP clearly the most trusted mediator, even though' it was realized to be an unrealistic ideal, 13. There was general awareness of the Data Protection Act but not of the Patient Information Advisory Group 14. Self-regulation was seen as positive because it includes doctors and experts. Participants also valued lay involvement in self-regulation bodies such as the GMC, BMA and ethics committees
51 Using patient information in the NHS 2008	1. The most popular requirement among attendees was to keep patient records secure 2. Closely followed by ensuring that access is restricted to those who should have it, i.e. maintain confidentiality 3. The quality of the data in electronic records was recognised as very important 4. Data needs to be shared as required amongst the clinicians caring for the patient 5. Patients wanted greater patient awareness of what the NCRS was doing, and planning to do
52 Using patient information in the NHS 2009	1. Top four topics in 2008 were also raised in 2009, the top 2 being the same in both years, see above (ref 46) 2. For the first time the issue of patient consent to the sharing of their data was brought up.
53 Monitor 1 report	1. When asked about willingness take part in research involving access to their anonymised medical records 28% would be very willing, 46% fairly willing, 12% fairly unwilling & 13% unwilling 2. When asked about whether they would have confidentiality concerns in different kinds of research project, 19% of those giving tissue / blood samples, 72% of those giving access to their records and 16% of those testing a new drug or treatment said they would.
54 Summary of Responses to the Consultation on the Additional Uses of Patient Data	The report grouped replies by the categories public, patients, NHS & SC staff, researchers and organisations and showed significant differences between the attitudes of researchers and the public & patients. 1. C 50% of public & patients said researchers should always seek patient consent before using identifiable data: 29% of NHS staff and 11% researchers, thought the same 2. 78% of respondents said this should be done each time anyone wants to use the data 3. Given ethics committee & patient data controller approval, 46% of respondents said that researchers could search patient records for suitable subjects if NIGB ECC approved it. 43% of public & 33% of patients said never 4. Majority of groups & 55% of all respondents said it should be possible to put flags on EPRs to indicate they could be contacted by researchers. Only 44% of public & 46 % of organisations agreed. 5. c 62% of respondents agreed an Information Custodian would be useful, 29% did not (37% of public &33% of patients) 6. 44% of respondents questioned the utility or accountability of an Information Custodian 7. on the research use of anonymised data, 25% of respondents thought it was not possible to truly anonymise

	data, 15% said it was OK to use it and 13% said consent should be sought every time 8. 45% of respondents considered that sealed envelope content could be used for secondary uses if anonymised. 69% of public & 74% patients thought this should never happen, or only with patient consent 9. 69% of respondents (73% of public & 77% of patients) said that if patients to be asked for consent to use anonymised sealed envelope data, it should happen every time anyone wants to use it. 10. 42% of respondents (c 25% of public & patients) agreed that linked anonymised sealed envelope data could be used for secondary purposes. 39% of respondents said only with patient consent, and 16% of respondents (29% of public & 24% of patients) said it should never be used. 11. majority (64%) of respondents, and all groups (except researchers who thought a one-off consent enough), agreed that where consent was sought to use linked sealed envelope data, it should be sought for each use.
55 Patient perspectives on the regulation and governance of medical research	1. Patients should have a central role in the shaping, conduct, regulation, governance & scrutiny of research activity 2. Creating trust is key in engaging with patients and making them feel confident to participate in research 3. Good communications, transparency and professional attitudes are key to creating the right research culture, including the big issues of consent and anonymisation. Current paternalism must become a thing of the past. 4. Regulation & governance should support patient involvement in research, not hinder it 5. Public involvement in the regulation & governance of research must be robust, well-informed & properly resourced 6. Many regarded healthcare professionals as poor communicators, not well 'clued-up' on research relevant to their patients, and viewing research as a burden rather than a routine part of healthcare *7.* Concerned that bringing all the regulation for health research into one body could lead to a perception that 'researchers are regulating researchers' thus losing the wider focus of the original bodies involved in it.
56 Attitudes and awareness amongst General Practitioners (GPs) and patients about the use of patient data in research	UKCRC have decided not to publish this research, see http://www.ukcrc.org/aboutus/boards/boardsubgrouponpa/ and follow 'Useful Link' to Minutes of 30th June 2010 and look at Item 6. The minutes file is named 3272_Item+2+-+Minutes+-+UKCRC+Board+Sub-Group+on+Public+Awareness+Meeting+-+30+June+2010+(2).pdf. A request from the author in July 2011 for a copy was refused. Content from it is referred to in [86] and [??].
57 Using patient information in the NHS Appendix 2010	1. Participants saw opportunities and benefits of using health records for medical research. 2. Many expressed an interest in taking part in research– if their data were used anonymously and sensitively. 3. Many said they would be happy for the NHS to use their data, but some disliked the idea of other organisations using their data. They said that in order for people to have confidence in the system, it was very important for an appropriate person, such as a GP, to have the job of accessing and handling health records. 4. Participants could see the benefit of health and social care services sharing patients' information. 5. However, perhaps because at a time social workers were being heavily criticised in the press, many weren't sure that giving social workers access to people's medical records was a good idea. 6. Some people were also worried about the possibility of non-NHS organisations being able to access data, e.g. housing associations, private companies &, in some cases, charities. 7. Participants agreed that access to data should be on a "need-to-know basis" only.

	8. Some even disliked sharing basic health information with staff who didn't need it directly for their job, such as meals on wheels staff.
58 Using patient information in the NHS Conclusions 2010	The attendee survey showed that: 1. more (75% vs 66% in 2009) believe their electronic health records will be available in the right place & time 2. fewer (52% vs 57% in 2009) believe the wrong people will be able to access their records if they are shared electronically 3. fear of data loss by system failure, accident, hacking, possibility of transfer to third parties outside the NHS 4. but more (52% vs 41% in 2009) agreed that their records would be kept confidential if kept electronically 5. but there was also a 10% increase to 40% in those who believed the same to be true of paper records 6. fewer (52% vs 59% in 2009) believe it will be easier for them to restrict who sees their records electronically
59 Using patient information in the NHS Executive Summary 2010	See above, refs 52 and 53.
60 Who sees what Exploring public views on personal electronic health records.	1. >80% believe their consent should be sought before creating a new EPR for them (e.g. the SCR) 2. ≥60% believe the consent process should be more robust than the opt out used for the SCR 3. 74% of adults believe that EPRs should be used for research 4. ≥74% believe that consent is required before sharing identifiable data with researchers 5. 34% adults & 56% of young people believe that consent is required if anonymous data is used for research 6. 92% adults & 97% of young people believe they should have access to their own EPRs, but - only 35% believe access should be allowed at home, and - c 12% believe that patients should be allowed to add information to their records 7. 57% of young people agree that the NHS should support medical research.
61 Privacy & prejudice: young people's views on the development and use of EPRs	1. >60% would be happy for their anonymised record to be used for health and medical research 2. >50% believe a researcher should seek consent each time they want to use their anonymised record# 3. Both these %s increased during the deliberative workshops
62 Public support for research in the NHS	1. 97% think it important for NHS to support research into new treatments for patients 2. 92% think it's important for the NHS to support research funded by medical research charities 3. 46% think NHS should be required to support research, 48% that it should be encouraged to do so 4. .If you were affected by condition such as heart disease or cancer would you like your doctor to tell you about research that you could take part in if the research involved .. a. trialling a new medicine or treatment 72% yes b. taking a sample of blood for testing in the lab 88 % yes c. talking to researchers about your family history 88% yes d. allowing a researcher to access your medical records only 80% yes
63 How Privacy Considerations Drive Patient Decisions and Impact Patient Care Outcomes, 2011	1. 54% would withhold info from care provider based on privacy concerns 2. 38% would postpone seeking care for a sensitive condition due to privacy concerns 3. 45% would seek care outside their community due to privacy concerns, 37% prepared to travel 30 miles + 4. 81% never been worried about security of their data at a care provider treating them

	5. 62% said personal information breeches would make them want to seek treatment elsewhere 6. 4% had been alerted to, or themselves discovered a breach affecting them 7. in 22% of cases, the breacher was a family member, 15% a co-worker, 15% unknown health provider employee, 10% a friend, 7% a neighbour and 32% not known 8. almost all patients think provider management should aggressively protect patient data confidentiality

Annex D Summary of major relevant UK legislation

Personal privacy is a fundamental right of UK citizens. It is recognized by the United Nations, the European Union (EU) and in the UK Human Rights Act, q.v. D1. Processing of personal healthcare data is governed by the common law duty of confidentiality which clinicians owe their patients, see. D2, as well as by legislation and the associated regulations, see D3-5. This duty of confidence sustains the trust essential between individual patients and the clinicians caring for them, and the "*public interest in maintaining trust in a confidential service*" [21 page 3].

The UK Data Protection Act 1998 (DPA) was introduced to bring general UK data protection legislation into line with the EU Data Protection Directive 95/46/EC (DPD). Both the EU and UK legislation are acknowledged to have flaws, and are currently being reviewed[73,[83,109]].. Several authorities have pointed out that the DPA does not reflect the DPD as well as it should. For example, the DPA has a slightly different but tighter definition of 'personal data', does not define '*consent*' and uses '*medical purposes*' in a way that agrees with neither the DPD or UK clinicians' understanding of the term. Among the consequences of this is a plethora of guidance that is difficult to assimilate, and strongly held but differing interpretations of what the law permits.

A comprehensive but brief summary of all legislation prior to September 2007 is given in *NHS Information Governance Guidance on Legal and Professional Obligations* [18]. What follows only gives the essential details of key pieces of legislation: it is not exhaustive.

D1 The UK Human Rights Act 1998 (the HRA)

At the head of every Bill presented to Parliament, the sponsor has to declare if, and how, the contents of the Bill impact on the Human Rights embedded in this Act. Article 8 is most relevant to this paper

"Article 8 Right to respect for private and family life

1. Everyone has the right to respect for his private and family life, his home and his correspondence.

2. There shall be no interference by a public authority with the exercise of this right except such as is in accordance with the law and is necessary in a democratic society in the interests of national security, public safety or the economic well-being of the country, for the prevention of disorder or crime, for

---

[73] Directive 95/46/EC was created to harmonise data protection practice across the European Union. The major reason for considering revision of the Directive is that it has failed to do this as well as intended, each nation interpreting it differently. The Article 29 Working Party was set up under 95/46/EC to monitor its implementation, suggest changes, and advise on external proposals for change.

the protection of health or morals, or for the protection of the rights and freedoms of others."

D2 The English and Welsh common law

When push comes to shove, the major bulwark protecting identifiable patient data is the duty of confidentiality that care professionals owe to patients. This is recognised under the common law, and is why the legislation described in D5 is necessary.

Other than by consent, confidential information may be used where[21 1.2]:

- statute requires or permits it
- where there is a legal duty to do so, e.g. a Court orders it
- the data controller of the confidential information believes, on a case by case basis, that the public interest / protection of the public to be achieved by its use outweighs both the obligation of confidentiality to the patient concerned **and** the broader public interest in the provision of a confidential health service.

However where confidential personal information is processed, it should in some circumstances still be carried out in accordance with the DPA principles. So for example the patient may be able to exercise the right recognised under DPA Part II to object to the processing where it causes, or will cause him harm or distress.

D3 The UK Data Protection Act 1998 (the DPA)

What follows summarises the UK Data Protection Act 1998 (DPA), and subsequent regulations made under it. The DPA only applies to the processing[74] of *personal data*[75] about living people; it does **not** apply to data about those who have died, or data that does not identify, directly or indirectly, the person it is about.

D3.1 The Data Protection Principles

Processing must be in line with the following principles (DPA Part I):

	Data must be:
1	Processed fairly & lawfully, ≥1 Schedule 2 condition is true, & for sensitive data ≥1 Schedule 3 condition is also true
2	Only obtained & processed for specified lawful purpose(s)

[74] *Processing* includes obtaining, recording, holding, or carrying out any operation(s) on the data, including organization, adaptation, alteration, retrieval, consultation, use, disclosure, alignment, combination, blocking, erasure or destruction of the data or information.

[75] *Personal data* is data that either identifies the person (i.e. the data subject) it relates to, **or can do so when combined with other information in the possession of, or likely to come into the possession of, the data controller** (DPA 1998 Part 1,1): recital 26 of the EU DPD adds "*or by any other person*" at the end. A person's health and healthcare data is a kind of personal data known as *sensitive personal data* (DPA 1998 Part 1,2) and the DPA protects it more tightly than personal data. Note that the DPA does not consider social care data to be sensitive personal data (which doesn't agree with the practice of most social care professionals).

3	Adequate, relevant & not excessive for those purposes
4	Accurate, and where necessary, kept up-to-date
5	Kept for no longer than necessary for the purpose(s)
6	Processed in accordance with the DPA rights of data subjects (see DPA Part 2)
7	Protected against unauthorized or unlawful processing, loss, destruction or damage
8	Not moved to a non-EU destination unless it offers an *adequate level of protection*

D3.1.1 Principle 1 needs further explanation.

For processing to be fair:

- the data must be obtained without deceit, particularly about the purpose(s) it is to be used for
- The data controller[76] must ensure *"that the data subject has, is provided with, or has made readily available to him"* the identity of the data controller, his representative if he has one, the processing purpose(s), and any further information needed in the light of the specific circumstances to enable the processing to be fair. This implies that the controller does not have to supply the information if he considers the subject already knows it, e.g. for a routine transaction the subject is familiar with. The two explicit exceptions are where disclosure is necessary for a non-contractual legal obligation (e.g. for performing a public function, e.g. maintaining the electoral register), or where providing the information would require a *disproportionate effort* from the controller.

For the processing to be lawful (principle 1), adherence to the DPA is necessary but not sufficient if other statutes or the common law forbid it. The common law is particularly relevant for medical data, as health professionals owe a common law duty of confidence to their patients.

The final conditions for principle 1 to be satisfied – the Schedule 2 and 3 conditions, also known as the necessities – are shown below in the order in which they appear in the DPA. Text in italics highlights the difference between otherwise similar Schedule 2 & 3 conditions.

Condition under which information may be processed:	**Sch 2**	**3**
With the consent of the data subject	1	
With the *explicit* consent of the subject		1
To fulfill a contract the subject is party to	2	
To satisfy data controller's (non-contractual) legal obligation(s) re employment	3	2
To protect the vital interests of the subject	4	
To protect the vital interests of the subject *or another person*		3
For the administration of justice, the law, a government function, *in the public interest*	5	
For the legitimate interests of the data controller or parties to whom data is disclosed, unless it prejudices subject's rights, freedoms or legitimate interests	6	
By not-for-profit political, philosophical, religious or trade union bodies, with due regard to the subject's rights, freedoms & legitimate interests & no disclosure		4

[76] The *data controller* is the person, or one of the persons, who decide the purpose(s) the data may be processed for, and how it may be processed.

If already made public at the behest of the subject	5
For use in legal proceedings	6
For the administration of justice, the law, a government function	7
For medical purposes[77], by a health professional, or a person who in the circumstances owes an equivalent duty of confidentiality to the subject - Health professional is defined in DPA s69	8
For monitoring ethnic or racial equality of opportunity or service use	9
As specified in an order made by the Secretary of State	10

The difference between the 'vital interests' condition in Schedule 2 (4) and Schedule 3 (3) would, for example, allow a care professionals to trace the contacts of subjects with communicable diseases without the subject's consent.

D3.1.2 In principle 6, the data subject's rights granted under the DPA comprise being able to:

Access a copy to the personal information held about them by the data controller - II 7
Object to processing that is , or will, cause the subject damage or distress – II 10
Prevent processing for direct marketing – II 11
Object to decisions made using the subject's data by automatic means – II 12
Claim damages for a breach of the Act – II 13
In some circumstances, have inaccuracies rectified, blocked, erased or destroyed – II 14

All these rights, as with many other rules in the DPA, are qualified by material elsewhere in the Act, and in regulations made under it since 1998.

It is worth mentioning here that disclosure of subject data to the subject which involves data about a third party, or that identifies a third party as a provider of subject information can be denied under section 7(4) unless

- the third party(s) have consented to its disclosure to the subject, or
- the data controller can anonymise the third party information[78].
- it is reasonable to comply with the request without the consent of the third party(s)[79].

The same restriction applies if the subject data is to be disclosed to anyone else.

D3.2 The research exemption

Section 33 of the DPA relaxes the principles when personal data (including sensitive personal data) is processed for research (which includes statistical and historical purposes), as long as it is not processed:

- to support measures or decisions about particular individuals

[77] *"medical purposes" includes the purposes of preventative medicine, medical diagnosis, medical research, the provision of care and treatment and the management of healthcare services* DPA Sch 3, 8(2). Note that the EU DPD 95/46/EC equivalent (Art 8(3)) does not include *medical research.*

[78] Although the DPA does not explicitly say that the same rules apply to subject data disclosed to anyone else, the RCGP guidance on record access for patients [34] states this is so for health data. Section 6.3 of [39] and 7(4)(c) of the DPA also makes it clear that data in the patient's record from or about third parties who are care professionals caring for the patient are not treated as third party data, and may be disclosed to the patient and others.

[79] E.g. the data about the third party was provided by the data subject himself

- in a way that causes, or is likely to cause substantial damage or distress to any data subject.

The relaxations exempt data processing for research from:

- the subject's access rights to it enshrined in principle 1, as long as the research results are not in a personally identifiable form
- principle 2 - use only for the specified purpose(s) for which it was collected, and
- principle 5. Research data may be kept indefinitely

but not from the other DPA principles. So any common law duty of confidentiality involved still applies unless it is overriden by another legal means, fair processing is still required, and the subject may still object to the processing.

D4 Data (processing of sensitive personal data) Order 2000

This permits the processing of sensitive personal data without consent in ten specific situations in England & Wales. Two have relevance to personal health & health care data, provided that they are *in the substantial public interest*:

> 4 – counselling & advice services, to facilitate the collection and processing of relevant data about third parties involved with the person being counselled.
>
> 9 – research, where *research* has the same meaning as in the DPA section 33, providing that it is not used:
>
> - to support measures or decisions about particular individuals, unless the individuals concerned provide explicit consent
> - in a way that causes, or is likely to cause substantial damage or distress to any data subject.

Note that this order does not explicitly provide the other relaxations afforded by the DPA in section 33 for personal data, see D3.2 above, but does allow the research to support measures or decisions about particular individuals **if** they provide explicit consent[80]. As with section 33, the order does not provide an exemption from the common law duty of confidentiality owed by a professional carer to his or her patients.

[80] This is consent to the processing, and not to implementation of the decisions the processing supports. Patient consent from all patients involved would therefore be required for PARR analysis using identifiable data, which is intended to support decisions about patients found to be at risk of readmission.

D5 The Health Service (Control of Patient Information) Regulations 2002

The Health Service (Control of Patient Information) Regulations 2002 were made under section 60 of the Health and Social Care Act 2001, and apply to England & Wales only. The regulations set aside a care professional's duty of confidence so that identifiable patient data can be, or required to be, collected and processed for *"medical purposes"*, which is not defined, but which from its mention elsewhere in the Regulations, is intended to include medical research. The Regulations explicitly legalise uses for cancer registration, and to establish the epidemiology and management of *"communicable diseases and other risks to public health"*. Additionally patient data *"may be processed for medical purposes in the circumstances set out in the Schedule to these Regulations"* where approved by the Secretary of State. These currently include:

- making patient data less identifiable, e.g. anonymising or pseudonomising it
- patient location-based medical research
- identifying & contacting patients to obtain consent
- processing of data from multiple sources in order to:
    - link data for the same patient
    - validate it
    - avoid incorrect linkage and the inclusion of duplicate data
- audit, monitor & analyse health service provision
- granting access to confidential patient information for one of the above.

The Regulations also state in s7(2) that *"No one shall process patient data under these regulations unless he is a health professional or a person who….owes a duty of confidentiality which is equivalent to …. a health professional"*, and provide an enforcement procedure and civil penalty for non-adherence to the Regulations.

The Regulations continue in force under section 251 of the Health & Social Care Act 2006. Section 251 may require or permit the lawful processing of prescribed patient information

"(a) in the interests of improving patient care, or

(b) in the public interest"

It may not be used to require data to be used *"solely or principally for the purpose of determining the care and treatment to be given to particular individuals."*

Processing may only take place in certain circumstances where carefully constructed safeguards are in place, including the granting of an application to do so by the NIGB Ethics and Confidentiality Committee. However the regulations (s251.4):

"*may not make provision requiring the processing of confidential patient information for any purpose if it would be reasonably practicable to achieve that purpose otherwise than pursuant to such regulations, having regard to the cost of and the technology available for achieving that purpose.*"

Examples of *achieving that purpose otherwise* would be by obtaining patient consent, or by using privacy enhancing techniques to anonymise or pseudonymise the data at source.

D5.1 S251 and the DPA 1998

Section 251 approval essentially absolves the data controller of the data to be processed from liability for breaking the duty of confidentiality he owes the patient. It does **not** remove the need for the processing to comply with relevant elements of the Data Protection Act 1998, of which the most important is probably the data subject's (i.e. patient's) right to object to processing of his data on the grounds of the distress or damage it would cause / is causing. This creates a conundrum, as it is not possible for the patient to object to the processing unless he is aware of it (ideally before it starts). As s251 was expressly created to avoid having to contact each patient and seek consent, such awareness is not possible unless (a) the researcher publicises the proposed project in a way that is likely to be noticed by all would-be subjects, and/or (b) the subjects are provided with a way to find out what research projects are under way, and whether they are involved.

Annex E What does consent mean?

E1 Introduction

Patient consent to the processing of their data is seen by patients and the public as a feature of courteous discourse, and a mark of respect [49,50], . It is a defining characteristic of the social etiquette involved in sharing something that you have with others. Consent must be *freely given, informed, specific and unambiguous*[81] and made in response to a **real choice**, i.e. the person involved is not coerced to give or withold consent.. *Informed* means that the patient should be made aware of:

3. what kind of data is being shared and for what purpose(s)
4. when and how often it will be collected
5. who will use it
6. who is responsible for ensuring that it is only used as per 1,2 & 3, and
7. how long it will be retained.

and corresponds to the Information Commissioner's fair processing information. The information must be reasonably specific if consent is to be meaningful, i.e. so that the subject can readily judge what the impact of giving or refusing consent would be for himself and others.

Significant changes to these, particularly the data sought and the purpose(s) for which it is to be used, should trigger a request for re-consent, although patients and the public recognize that this may be onerous [50].

E2 Varieties of consent

Consent comes in two major variants:

E2.1 Explicit consent[82] for each sharing request. This is seen as the gold standard by the public, but involves contacting all potential subjects, which may be impossible[83] or expensive[84] to do for large research cohorts. Explicit consent may be sought by:

(a) contacting each potential consenter, providing the necessary information and asking for consent. Personal contact is preferred by

---

[81] From the EU Article 29 Working Party Opinion 15/2011, WP187 of July 2011 [46], which describes what it considers consent to mean, and how it may be legitimately obtained.

[82] The DPA states that any consent provided under Schedule 3 for processing sensitive data (which includes health data) must be *explicit*.

[83] For example, the subject may be dead, or untraceable at the time his or her consent is sought.

[84] The public attitude survey in [49] found that *People are less accepting of financial constraints on seeking consent, with only 5% saying that cost factors are a viable reason for not doing so.*

the public, but it may be done by letter, e-mail, telephone, etc. Any consent given should be recorded.

(b) informing as many of the potential consenters as possible of the choice they have and the consequences, by forms of public announcement, and asking them to notify the project if they consent – the 'opt in' approach

E2.2 Implied consent, the most contentious variety. In the public's eyes *implied consent was equated with no consent* [50], and good cases can be made that that neither variant of it constitute consent at all. Historically consent has been implied by

(a) something the intended consenter does **not** do, the 'opt out' method, where non-receipt of an opt out from someone is taken as consent. Making sure that all potential subjects are aware of the opt-out, and what it means, are challenging and the opt out must be easy to use[85] if it is to be valid. However the public [50], NIGB [69 p15] and the EU Data Protection Article 29 Working Party [46] do not consider that this form of consent is valid.[86]. The current proposals put to the European Parliament for revising the European Data Protection Directive would implement this view - "*Wherever consent is required for data to be processed, it is clarified that it has to be given explicitly, rather than assumed*"[110].

(b) by something the intended consenter **does,** which is taken to indicate consent to an associated data processing activity involving the subject. This is an implicit 'opt-in'. Again, making it clear to the subject that doing one thing (e.g. seeking treatment by a GP) will result in the associated data processing being performed (e.g. the recording of information about subject's condition, its history and the resulting actions by the GP[87]) and that the former cannot be done without the latter may not be easy to do, although in the treatment example this is not the case. It has strong similarities with the notion of a contract: in effect the data processing is an essential concomitant of the primary action the patient wishes to happen, obtaining healthcare. While this is definition is acceptable to some (including the Information Commissioner) it is doubtful whether the Article 29 Working Party would agree [88].

---

[85] The original opt-out offered for the Summary Care Record failed on both counts

[86] The EU Article 29 Working Party said in Opinion 15/2011, III.A.1 [46]*... The words "indication" and "signifying" point in the direction of an action indeed being needed (as opposed to a situation where consent could be inferred from a lack of action).*

[87] GPs have had a legal obligation to keep records about patients since 1906.

[88] The EU Article 29 Working Party took the view in Working paper 131 2007, II.4(a) [22]: *that where as a necessary and unavoidable consequence of the medical situation a health professional has to process personal data in an EHR system it is misleading if he seeks to legitimise this processing through consent.*

E2.3 The seeking of generic consent is relevant where consent is being sought in advance of any particular use of a person's data for a class of purposes, such as research or administrative purposes. There is debate about how generic a purpose can be before the associated consent becomes so meaningless as to be invalid. Consent for all secondary purposes would be regarded as too generic, and for all healthcare-related research probably so. Broad purposes may also encourage refusal where a more detailed request or the ability to provide a qualified consent, e.g. 'all healthcare research other than into biochemical birth control', would not. Seeking consent for intermediate generic purposes where the ultimate intentions are not specified (and may require separate consent(s) in their own right) would probably be considered meaningful, for example:

(a) for repeated examination of the patient record to see if the patient is a suitable candidate for a secondary use, typically a research study

(b) for extraction of data from the patient record for warehousing, maybe then linking it with data from outer sources and/or pseudonymising it, for subsequent secondary purposes.

E3 Opportunities for collecting consent.

Other than specifically contacting patients to obtain consent at the time when it is required, there are other opportunities when the patient is in contact with the health service or government that could be used to collect consent. These include:

1. during screening (e.g. for cervical and breast cancer, bowel cancer)
2. during maternity care
3. during treatment
4. via the annual electoral register return each household has to make
5. when applying for a driving licence or passport
6. when voting
7. when filling in a council tax demand.

As noted in E2.3, the rub is providing something to consent to that would be considered to be specific enough to ensure the vailidity of the consent.

Annex F Safe Havens

F1 Introduction

Safe havens figure prominently in the recommendations of the research lobby reports. They currently play a significant role as a source of patient data for secondary uses, particularly research, and this is likely to grow rapidly in the near future. We have the English Health Research Support Service (HRSS)[89] being delivered by the Research Capability Programme, the Scottish Health Informatics Programme (SHIP) and the Welsh Health Information Research Unit (HIRU) with its Secure Anonymised Information Linkage System (SAIL). Outside the NHS we have several substantial databases built around a core of general practice data, including the GP Research Database (GPRD) run by the Medicines and Devices Health Research Agency (MHRA), THIN run by Cegedim and Qresearch run by the University of Nottingham.

It is therefore very important to have a shared understanding of what is meant by a safe haven in the context of the secondary usage of patient data.

F2 What is a safe haven?

The term *'Safe haven*' is used here to describe a facility run by a '*trusted third party*' that provides a secure environment to hold individual data that it receives about patients and citizens from one or more sources for subsequent secondary use by others. The integrity of the 'trusted third party' and their staff is critical to successful operation. A 'safe haven' may not be an **end-user** of patient data for secondary purposes, although it may link, quality control and pseudonymise data on behalf of its end-users, which are themselves kinds of secondary use.

F3 How do safe havens operate?

A safe haven may hold collect identified and/or de-identified data[90] but the majority hold and provide access to de-identified data. Safe havens may also process the data they collect for limited secondary purposes on behalf of their clients, such as anonymisation / pseudonymisation, linking and quality control. A safe haven may collect and briefly hold patient identifiers solely in order to pseudonymise the data they collect.

Data made available for external secondary purposes should for only be for use in an approved way by authorised users, with severe penalties for data misuse. Some safe havens, like SHIP, will only permit the data to be used in situ or via thin clients linked to the safe haven via its own software, so that safe haven staff have a unique ability to monitor the

---

[89] This was initially sponsored by the UK Clinical Research Collaboration and is being developed by NHS CfH. The sponsor is now the National Institute for Health Research.

[90] The definition given in [67 8.78] only covers *"identifiable personal data"*

behaviour of their clients. Others such as THIN and GPRD, provide copies of the required data to their clients under contracts that regulate its use and retention.

A safe haven may operate in one of two modes, or both. It may operate **reactively** by collecting and (where required) linking data in response to secondary user requests where use is authorised by the source data controller(s). It may also operate **prospectively**, accumulating data in a '*data warehouse*' prior to any specific request from a secondary user. The warehouse may accumulate linked data from many sources that covers long periods, sometimes much or all, of peoples' lifetimes, or enable linkage[91] but defer it until a researcher requests it. Secondary users select the data they wish to use from a catalogue published by the safe haven and apply for permission to use / copy it. By retaining data collected reactively after the studies concerned have finished, reactive operation can morph into prospective operation.

One can foresee that safe havens may wish to specialise in certain kinds of data, and/or particular kinds of processing. Safe havens may wish to take advantage of the specialist data services provided by others, so that data might be processed by more than one safe haven before being made available to the ultimate external user. This will make patient information governance more complex and create territory that appeals to the more commercially-minded operators, which in turn will further complicate governance. Such situations raise the question of what the patient can meaningfully be asked to consent to at the time of data collection.

The advances in PET now enable safe havens to perform all the operations expected of them (linking data from multiple sources, quality control, providing access to pseudonymised subsets of data for researchers, requesting a source to identify and contact the patient), without having to collect, hold or process personal data.

F4 Pros and cons of safe havens

The introduction of an intermediary between multiple data sources and secondary user(s) can reduce the amount of data transfers and processing that data sources and secondary users have to perform, particularly where data is collected prospectively, i.e. warehoused. Where safe havens are run by professional staff with statistical, IT and data processing expertise as well as clinical experience, they are much better able than most care provider data sources to ensure the security of their data collections, that the data that they release has an insignificant risk of patient re-identification in the proposed secondary use context, and to ensure that their clients do not misuse the data.

But safe haven data warehouses that persist large amounts of rich patient data for multiple, often ill-defined or unknown, secondary uses, pooling data from different sources and covering long periods thoroughly

[91] Typically by adding the same unique pseudonym to all records warehoused for the same patient.

contravenes the "Least principle" concerning data richness (see 1.7 and G2(a)) and makes them much more attractive targets for would-be authorised and unauthorised users. Data linked and/or added to over time by safe havens may become so rich that the risk of re-identification means that it should be treated as identifiable data in one, some or all of the contexts in which safe haven clients wish to use it. This is further complicated by not sometimes knowing the circumstances of its use by the safe haven's clients at the time the data is collected or even who the clients will be. In the (likely) event that their intentions and the data being used were not described in the privacy notice made available to patients and care providers when the data was first collected, additional explicit consent should be sought for its use if the data is, or is likely to be equivalent to, personal data in terms of the DPA or there will be a significant risk of re-identification during use.

Rich data makes it significantly easier to triangulate the identity of patients in conjunction with other data, and encourages 'mission creep' amongst potential users: this can be an advantage or a disadvantage, depending upon the suggested extra purpose(s). Indeed the Information Commissioner coined the term "toxic asset" for warehouses containing rich personal data about individuals in his Annual Report for 2007-8. A safe haven holding a data warehouse should only be authorised where there is no other practical way of satisfying the secondary purpose(s) which it is to serve.

F5 Safe haven governance.

For the reasons above, safe haven governance therefore be of a higher standard than normal, and higher still where data collection is prospective. Governance

(a) should involve representatives of the contributing data source controllers and data subjects in:
   I. the approval of uses on a case-by-case basis
   II. generation of the management structure and processes
   III. the audit of governance arrangements.

(b) require both much better informed recruitment of subjects including
   I. explicit consent for uses involving data that is identifiable or considered as equivalent to it,
   II. and at least an opt out where de-identified data is collected and / or used for secondary purposes outside the safe haven
   III. easy availability of clear and comprehensive privacy notices for subjects

(c) if it is not made a criminal offence, impose heavy penalties on any attempt at unauthorised re-identification or sharing of patient data by safe haven staff and clients as the AMS Report recommends[86 6.2.2].

(d) require logging of all processing taking place at the safe haven

(e) should require contracts open to public inspection with users that take data for use outside the haven. The contracts should clearly state clear how the data copied may and may not be used, how usage will be verified and the penalties for data misuse.

Access other than by safe haven staff must be restricted to the personnel, purposes and periods for which the data was obtained and as recorded in the original request for data or privacy notice for patients. Safe haven staff and users of safe haven data should be bound by the same duty of confidentiality as clinicians providing healthcare.

A safe haven should only provide de-identified data for secondary uses and not collect or process (e.g. link) identifiable patient data . Where patient identifiable data is collected and processed, there should be further controls, which will include the need for further consent for uses by the safe haven clients.

There should be a standard model code of conduct for safe havens which covers the various kinds of data that they can provide and the different ways in which they can share it, produced by NIGB. The authorisation of safe havens and the approval of codes of conduct for individual safe havens would be a proper function for the NIGB to undertake with assistance from the Information Commissioner where required.

Annex G Generic research patient data use

G1 Basic uses

Researchers would like to be able to use patient data in two ways:

(b) to select potential research subjects, both for interventional and observational research. Sometimes researchers need to select groups of subjects, such as families of various types, partners, twins or siblings

This is itself a secondary use (as are anonymisation, pseudonymisation, data linking and quality checking). Where an organisation caring for the patients identifies eligible patients and contacts them for permission to put their names forward to the researcher, most patients are content, but this can involve the data sources in work and expense that care providers may not consider a top priority. It cannot be done by the research project unless it approaches every potential member of the research cohort first for permission to check their record to see if he is eligible – the 'consent for consent' problem. Even then the patient may ask how the researcher decided that he was a potential member of the research cohort

(c) to use the data in the research subjects' records for the research itself. Almost always it is the clinical content of the record that are used, and the commonplace identifiers of patients, such as name, address and date of birth are not required. However sex and age are common discriminants in research, and the subject's address and/or postcode may be used to derive other location-related information, such as a deprivation index value, other environmental information or distance from a service provider. GP practice may also be required, although usually a meaningless practice identifier is sufficient. Very rarely the family name itself may be relevant, for example in genetic studies.

G2 Additional uses

In some cases researchers will also need:

(d) to link data for a patient from different sources and / or for different periods, bearing in mind that identifiers for an individual, such as address and name, may vary from source to source, and / or over time. This may include data about people who are dead, where the data may be on paper as well as or instead of in electronic form

(e) quality control data to ensure that it consistent, and to see how comprehensive it is. This includes the detection of duplicate patient records.

(f) to enable potential / actual research subjects to:

i. give consent for the use of their identifiable data, and/or

ii. agree to participate in a clinical trial,

iii. provide additional data direct to the research team

(g) contact a patient where use of their de-identified data indicates that he needs additional clinical care, e.g. as a result of a risk stratification study or by chance during the research.

(h) on occasion, re-use the data in subsequent studies, either on its own or with additional linked data.

Researchers would also like to minimise, ideally avoid:

(i) the expense and effort needed to seek consent

(j) the bureaucracy involved in holding patient-identifiable data

(k) the risk of subject re-identification / identifiable data loss from any secondary database that they hold.

Annex H Using de-identified data

H1 Types of de-identified data

De-identified data includes:

- aggregated data, where data is not held at the level of the individual person. Typically such data consists of counts of individuals with particular properties and/or who have received particular interventions, for example all those within an age band in a particular location that were diagnosed with a particular disease within a specific period.
- anonymised data where individual personal data is identified by a single unique identifier that is intrinsically meaningless, that is to say it the identifier cannot be used either directly or indirectly to identify the person to which it refers to. All commonplace identifiers, NHS number and all unique patient identifiers allocated by care providers have been removed or encrypted so that they are meaningless to the user(s).
- pseudonymised data (also known as key-coded or - rather ambiguously - linked data[92]). Here a unique personal identifier is generated, e.g. by picking them randomly from a list of integers or by applying an irreversible algorithm to one or more common place identifier such as NHS number, practice number, hospital number or name and date of birth. Again all commonplace identifiers and other unique identifiers in the source record are removed or encrypted. But unlike anonymised data a method is available at the data source that enables the pseudonym to be used to re-identify the person concerned. Typically this involves a table linking the pseudonymous identifier with a unique common place source identifier, such as NHS number, the source record identifier, or a combination of identifiers such as date of birth, postcode and full name.[93]

Aggregated, pseudonymised and anonymised data about living people are not regarded as personal data under the Data Protection Act 1998, unless the person holding it also has, or could readily have, access to data and/or techniques that enables an individual person to be re-identified by other means in combination with the de-identified data. . Under such circumstances **de-identified data becomes identifiable and so personal data in terms of the DPA 1998, even though it is not intrinsically so.** A data controller must therefore assess the likelihood of

[92] This term is never used in this sense elsewhere in this document, but only to describe data for a single individual gathered from more than one data source, e.g. a record that comprises an assemblage of a person's hospital, mental health and primary care data.

[93] It is entirely feasible to generate a unique pseudonym that is not based on any source record content, but if this is done it will not be possible to generate the same pseudonym for records for the same patient from different sources. The key thing about a pseudonym is that a means is provided that links it to the source data (e.g. via the table mechanism) that is not available to the secondary user.

this happening before deciding whether or not to disclose de-identified data to one or more other person.

H1.1 There is a pressing need to agree what constitutes the set of commonplace identifiers, none of which should appear in de-identified data. There is an emerging view – see I3.1 - that this comprises name, date of birth, address, full postcode, telephone number, e-mail address, NHS number and care provider-allocated unique identifiers (e.g. patient numbers allocated and openly used by staff in GP practices and hospitals).

H2 Risk of personal re-identification

The risk to patients that we wish to reduce is the chance of anyone learning something from the de-identified data about an identifiable patient (or a very small group of patients) that he did not already know. As noted above in H1 it depends both upon the content of the data itself **and** the circumstances in which it is used. All forms of de-identified data carry a risk of re-identification when used in combination with other information that someone knows, or can access, although the risk may be vanishingly small. The risk also rises as the number of people with the appropriate motivation and skill to access the data increases, i.e. it is a function of opportunity, motivation and technology. As the risk depends on many factors, some of which are difficult to quantify, such as the prevalence of a motive and the technical ease of access, it is usually not easy to precisely estimate the risk involved in a given secondary use, but the objective is always to minimise the opportunities and motivation and maximise the technical skill required.

The last few years has seen a keen debate in the UK and USA over the effectiveness of de-identification to protect personal privacy in the light of rapidly growing popular computer power and the increasing availability of personal data on the Net in public agency outputs[94], commercial databases that have been made public and social facilities such as FaceBook, LinkedIn and YouTube. Professor Ross Anderson and others in the UK have produced much material on this over the last 30 years or so, and Professor Anderson has produced one of the standard works on the subject[73]. In the USA Paul Ohm has predicted the end of the road for de-identification [102] after analysing some spectacular successes in re-identifying de-identified data in the USA and noting that repeated tracker queries can usually identify a subject given sufficient unimpeded access to a de-identified database. His work and that of others of a like mind is making significant inroads into what Jane Yakowitz has called the 'information commons' and government and public agency trust in the utility of de-identification as **the** tool to enhance data subject privacy. In a subsequent paper Yakowitz [103] has shown that some of the examples

[94] In this context the establishment of the UK Government's Open Data initiative is significant.

cited by Ohm were not properly de-identified[95] and also makes the point that no case of unauthorised re-identification in a research environment using statistical techniques has come to light yet. She says that re-identification of research data is unlikely because it is much easier to obtaining additional information about a specific person or persons by other means, such as using published identifiable data[96], hacking into identifiable data, bribing an insider, masquerading as someone who is entitled to know what is sought and combing through the target's refuse. She also points out that not knowing whether the target subjects are in the data sample used is a powerful obstacle for those attempting re-identification. While not rejecting all of Ohm's points she sees control of secondary users and their activities, including penalties for unauthorised uses and criminalising certain types of misuse such as attempting to re-identify subjects as key to ensuring the proper use of de-identified data.

For an excellent summary of inference techniques that can lead to re-identification of de-identified and aggregated records, and the corresponding counter measures see [73 Chapter 9.3] . The same UK reference says: "*Medical systems also teach us about the limits of some privacy enhancing technologies, such as de-identification. While making medical records anonymous in research databases can help mitigate the consequences of unauthorised access and prevent mission creep, it's by no means bulletproof. Rich data about real people can usually be re-identified*". Thus knowing a person's gender and age, the dates of attendance at particular clinics, that he suffers from one or more long term conditions, and his GP practice, may define a very small set of people, possibly only one, although not per se providing direct identification except to those who know him. Even with aggregate data, which in theory poses the lowest risk, small counts in cells can lead to the detection of individuals with unique combinations of the properties recorded in the database especially when many other personal properties are needed for cell membership, and at least some are unique and/or unusual, such as living in a small defined population (viz. at a particular full postcode), or having a rare disease.[97] There are also cases where the just the number of variables and the sample size between them can result in a very high probability that one or more individuals will be found with a unique set of variable values.

As always, the risk rises with the richness of the data and falls if the data is processed according to the 'least' principle, i.e. the data is:

(a) the least amount (in terms of both the number of subjects and the amount and variety of data per subject)

---

[95] In one case the data included date of birth and the USA equivalent of post code, and in another subject surname

[96] This is certainly true in the USA but maybe less so in the UK.

[97] The Office of National Statistics has rules about publishing aggregate data that includes small cell counts and limits on the number of variables to use in samples of various sizes which are intended to reduce the risk of re-identification [79,92].

(b) held for the least amount of time

(c) authorised for use by the least number of people

(d) copied the least number of times and ideally used at source

commensurate with the purpose(s) for which it is to be used.

The ideal is secondary use at the data source itself, the secondary user only receiving aggregated or de-identified results that have been checked to ensure that they are not significantly disclosive. Unfortunately this is not possible where the purpose involves linking patient data from several sources[98]. Conversely the riskiest situation for de-identified data would be an unencrypted rich dataset of linked records for many individuals with a known sampling frame or for the entire study population covering long periods of time available on a 'bird table' basis, i.e. it may be copied and used by anyone anywhere any number times for any purpose.

Risk minimisation involves a combination of de-identification and using the 'least' principle, a random sampling frame, inference countermeasures such as k-anonymity[99] and data blurring, and encryption of the data, checking the disclosive power of the data, and regulating data users and uses, with significant penalties for data abusers. Unfortunately inference counter measure such as k-anonymity and data blurring have the side effect of reducing the information content of the data involved, and so its utility for analysis. An appropriate set of these should always form part of the implementation of databases of individual's data, whether identifiable or not.

The table below ranks the risks of patient re-identification implicit in different ways of sharing de-identified data by aggregating the scoring of the major variables contributing to the risk. It does not pretend to be precise, but does provide a breakdown of the factors used to rank each alternative.

[98] It would also be impossible to do where the raw data could be amended before its use for secondary purposes, including any verification of the results by peers, was complete (unless the secondary use is based on a static copy provided by the data source).

[99] K anonymity is defined as *"the suppression or generalisation of an attribute so that its value is identical to that in k-1 other rows"*, Williams & Blum 2007. An example would be replacing date of birth with membership of the corresponding age band, e.g."12 Jan 1983" with "aged 25-30 yrs"..

	TOTAL SCORE – lower the score, less risk there is to patient confidentiality	Assessment of disclosivity of results likely, 1= done well, 3=probably not done	Susceptibility of data to inference attacks, 1= least, 3=greatest	Opportunity for unauthorised access, 1= min, 3=max	Technical ease of unauthorised access, 1=hardest, 3 = easiest	Likelihood of access by data misuser, 1= least likely, 3 = most likely	Assessment of data disclosivity, 1=d one well, 3=done poorly	
Individual data used at data source by authorised personnel	7	2	1	1	1	1	1	Results depend on type of data source, and their IT & IG skills. Could be any care provider, e.g. hospital, MH trust, GP practice, etc
Individual data used at safe haven by authorised personnel	6	1	1	1	1	1	1	High score due to expected high professionalism of safe haven operation and their likely oversight of the whole process of use.
Individual data copied from safe haven & used by authorised personnel	11	2	2	2	2	2	1	User IG skills a relative unknown, and user may not present results back to safe haven for checking
Individual data copied from data source(s) & used by authorised personnel	12	2	2	2	2	2	2	Source may not be able to do disclosivity checks on results & user may not present results for checking
Individual data published for anyone to copy & use as they wish – the 'bird table' model	17	3	3	3	3	3	2	Anyone can use data repeatedly so misuse likeliest to occur & unlikely to be detected unless 1+ data subject discovers it & complains. Data security depends critically on the skill of publisher in controlling the disclosivity of what is published. Method will preclude publishing of most individual data

H3 Pros and cons of using de-identified data

H3.1 Using data de-identified at source has major general benefits for secondary users. It means that:

(a) there is no need to seek patient consent to use their data or apply to the ECC for approval to use the data without consent under s251 of the Health and Social Care Act 2006, provided that they do not have, or have easy access to, other data or techniques that could re-identify the data subjects. However it is down to the data controller releasing the data to assure himself that the data recipient does not

have, or have access to, other data or might use techniques that de-identify the data.

(b) a secondary user does not have to become a controller of personal data under the DPA 1998, with all the obligations which that entails

(c) primary data sources are much more likely to allow their data to be used for secondary purposes

(d) no commonplace identifiers or identifiable patient data leave the source, minimising the risks of accidental loss, theft or mal-use of identifiable data. Better still, data that doesn't require linking with data from elsewhere can be processed in situ at the data source.

H3.2 In addition, pseudonymisation:

(a) enables the linking of data from different sources. As long as the same pseudonymising algorithm is applied to the same unique identifying data (such as NHS number) from all sources being used for a project, pseudonymised records can be linked in the same way as identifiable ones. If other commonplace identifiers are retrieved and encrypted using the same algorithm, material not used to generate a pseudonym can be used to suggest linkage or corroborate / refute linkage derived using the pseudonyms. The only difference is that it will not be possible to use manual methods to sort out non-matches, e.g. looking for obvious name misspelling. However by also pseudonymising derivatives such as the soundex equivalent of a surname it is possible to do more than seems possible at first sight. Both deterministic (e.g. using an encrypted unique identifier such as NHS number) and probabilistic linkage (e.g. using a combination of pseudonymised commonplace identifiers) are possible. Fuzzy matching currently produces less matches than does using identifiers in the clear – 87% versus 90-95%, but matching using more prevalent and accurate uses of unique identifiers such as the Scottish Community Health Index or NHS Number should improve this over time[100].

(b) if done at source, ensures that re-identification is only possible at the data source(s) although it may be proposed by the secondary user

(c) can be done as and when the data is needed for a secondary purpose rather than using an existing pseudonymised data base

H3.3 Examples where identifiable data may still be required for a secondary use include cases where:

(a) the user needs very fine grained geographic data about where patients live, full postcode. However attaching location-derived data

[100] 10 years ago a surprisingly large number of NHS patients (thought to have been well over a million) did not have NHS numbers

such as deprivation indices at the data source will remove a common cause of the need for this kind of geographic data

(b) research is family-tree based (although this is likely to require contact with members of a family involved, and so amenable to the collection of explicit consent).

(c) as a last resort, where there is a requirement for 100% linking of patient data from different sources **and** it has been demonstrated that pseudonyms cannot produce high enough matching rates **and** there are no means of raising the matching rate to the required level other than by using identifiable data. Non-matches in a research project are likely to be little more than an inconvenience where the study population can be increased to attain the target, and non-matches do not differ in a way that bias the sample. 100% matching is the target during patient-based administration or care provider payment activity, but given the advances in PET and the attachment of other unique identifiers to requests for service / referrals, such as a unique booking reference number and the need for the service requester and provider to sort out such queries, such a requirement is no justification for routinely sharing identifiable patient data.

H3.4 See Annex I for more on pseudonymisation and current initiatives in the field.

Annex I Privacy-enhancing technology

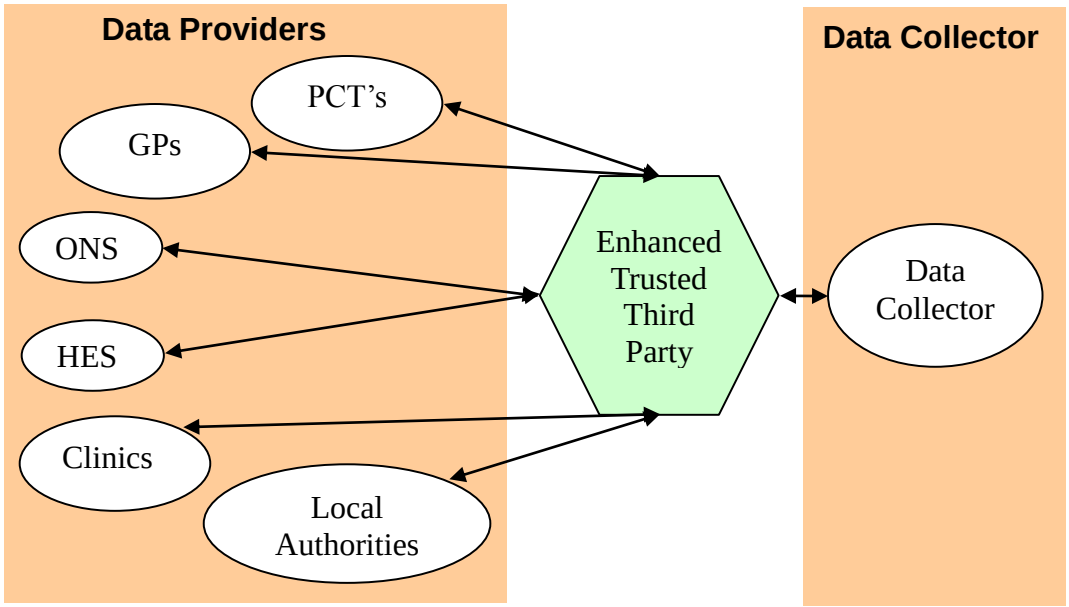
I1 The requirements[101]

There is an increasing need within Health Informatics to get a fuller view of a patient's activities by linking data from the various providers of care. Sometimes it is a care provider that wants to see that information, sometimes it is an academic researcher, sometimes it is government and sometimes it is a commercial body responsible for the effectiveness of its products. Who is to be entrusted with seeing all the collected data prior to its linking? One current approach is to introduce an extra organisation, ship the data there and have it do the linkage. This new intermediary is often called a "Trusted Third Party" or TTP. This is the approach being trialled by the NHS Research Capability Programme in its Healthcare Research Support Service pilots. It should be emphasised that this intermediary ***has*** to be trusted, but there is no guarantee it is trustworthy. This is the nub of the problem: a concern over provider trustworthiness is currently 'solved' by introducing yet more people to the sensitive data who have no need to otherwise see it. It also creates a new pool of more extensive potentially personally identifiable data which can be accessed under certain circumstances.

I2 A Specific Solution

It is possible to solve the data collection and linking problem without any patient names, addresses, postcodes, etc leaving the facilities in which they are normally protected. Because there is a need to store intermediate results somewhere and consistently pseudonymise across all providers, an intermediary is still required. However the intermediary does not have to see any readable health information. Ensuring this provides the guarantee of trustworthiness missing from the established TTP model.

[101] Sections I1-I2 were kindly provided by Rob Navarro of Sapior Ltd



Ensuring the intermediary cannot see any health information is a two step process, Setup and Operate.

I2.1 Setup

The setup stage begins with the data collector and data providers receiving invitation emails. The data collector acts on the email by clicking a URL and authenticating to the electronic Trusted Third Party's (eTTP's) servers. This causes a small program to be run in the data collector's browser which sets up a cryptographic framework and uploads a public key.

The setup for the data providers has two steps. One requires access to a sample of the data that will be uploaded to configure a cleansing, standardisation, enrichment and de-identification process specific to that data structure. This configuration information is then packaged into a small program and sent to the intermediary. The final setup step involves the data provider also responding to their invitation email by clicking on a URL and authenticating to the eTTP's servers. This causes a small program to be run in the data provider's browser that completes the cryptographic framework setup by downloading the "data collector's" public key.

I2.2 Operate

To operate the data collection and linking process, data providers simply authenticate to the eTTP's servers and navigate to their local data file that is to be uploaded. Because the structure of the data to be uploaded has already been analysed, the upload process is very simple.

Data collectors pick up their data by authenticating to the eTTP and checking to see if there is anything available for download. When sufficient

data providers have uploaded their data then the data collector can retrieve the linked information. The retrieval process automatically unpacks, decrypts and formats the data ready for use.

I2.3 Characteristics

6. Data providers can guarantee their patients that no names, addresses, postcodes, etc. ever leave the facility. This reduces the risk of a privacy breach and will maximise provider participation.
7. Both direct and fuzzy linkage is possible within the eTTP service.
8. High specification pseudonymisation can be applied at the eTTP. This ensures that any pseudonymised fields sent to the data collector are both consistent and robust.
9. The data collector can upload the results of their analyses for action and re-identification by the appropriate data providers. This facilitates secure communication and perfect re-identification without worrying about key management and security breaches.
10. If any eTTP intermediary staff or systems are ever compromised, all that is exposed is encrypted data. This is the lowest level of data breach possible. The encryption keys are only held at the end-points (data providers and collectors).
11. The trustworthiness of the eTTP intermediary is dependent on its encrypted data being separated from the encryption keys. This requires a strict and verifiable separation of eTTP from the data provider or data collector end points.
12. There is no technical restriction on what data fields the collector can receive. The data collector must get Information Governance approval as usual. The eTTP simply ensures no-one else can see the sensitive data in the process of transferring it from the providers.

eTTP fuzzy linkage accuracy will never be as high as for systems that pool all sensitive data in the clear. External studies have shown that the best linkage can result in accuracy of around 90-95% (against a gold standard exemplar). eTTP linkage quality (based on research using similar methodologies) is estimated to be around 87%[102]. So less good, but not by much.

I2.4 Where is the technology now?

The service described above is live with its first commercial customer. It has been reviewed by the NIGB ECC, and it has been determined that the pseudonomysed data collected from any source is not subject to the DPA, and does not therefore require data subject consent before collection and processing. The recipients may still need permission to provide data to their clients, but not to pseudonymise / collect / link the records via the eTTP.

---

[102] Analysis of Identifier Performance using a Deterministic Linkage Algorithm. Grannis S, Overhage J, McDonald C. AMIA 2002 Annual Symposium Proceedings.

There is now no need for special legislation for the NIS Information Centre or other bodies to collect and link patient data in the clear (the Information Centre is applying for accreditation now). Accuracy of the underlying fuzzy linking described above is currently being independently tested by University of Surrey

I3 Open Pseudonymisation

Under Professor Julia Hippisley-Cox, the Department of Clinical Epidemiology and General Practice has developed open-source pseudonymisation software that is available free under an open source licence to those wanting to use pseudonymised data patient data for secondary purposes. The University of Nottingham have made such software available for a .NET environment, and other versions are under consideration. Three of the major GP system suppliers have agreed to interface their patient care systems to the software.

I3.1 Desirable properties of the technology

The Department at Nottingham has also produced a series of statements about pseudonymisation, q.v. the table below, and asked attendees of a workshop held in September 2011 whether they agreed / disagreed / were unsure about each principle. Facilities were also provided to make free text comments on them. The survey results can be found on the open pseudonymisation web site.

#	Property description
1	Pseudonymisation is a key process which can be applied in order to limit the flows of identifiable data not used for direct patient care
2	All organisations should seek to maximise utility and application of robust pseudonymisation techniques in a consistent way across the NHS in order to maximise the privacy of individuals and maintain public trust.
3	Pseudonymisation of strong identifiers must take place BEFORE patient data leaves their source NHS computer systems
4	There must be a published standard list of what constitutes strong identifiers and how these should be recorded. The suggested list is name, address, postcode, date of birth, NHS number, hospital number, GP surgery patient ID, telephone number, email address.
5	There must be reliable mechanisms to allow re-identification of the patient at the source NHS site.
6	The data controller for the identifiable data (eg GP practice) must have (a) full knowledge and control over what data is extracted from their clinical computer system and how it is pseudonymised;(b) clear understanding of the use and subsequent disposition and governance arrangements of the extracted data and where responsibilities lie;(c) be able to switch the extraction on or off.

7	NHS clinical suppliers need to integrate software which implements pseudonymisation within the clinical system so this can be applied BEFORE data are extracted.
8	There should be an agreed one way hashing algorithm which should be used at standard across the NHS within pseudonymisation processes. The algorithm should the best available one the time bearing in mind the need for widespread accessibility. The current best hashing algorithm is SHA2-256 as this is collision resistant, available as standard in commonly used software platforms and is the US standard from 2010. If a better hash is developed, this will be reviewed.
9	Use of a project specific salt code appended to the identifiers before the hash algorithm is applied is useful mechanism to ensure that the resulting digest is (a) unique to a data sharing agreement and (b) can be consistently applied to data from multiple settings

The consensus list of principles can then be verified / improved by mapping them against various use cases developed by secondary users.

I3.2 Further reading on PET

More information is available on this topic in [68, 73, 66 and 104].