NHS England

Privacy Impact Assessment: care.data

High quality care for all, now and for future generations

THE NHS CONSTITUTION
the NHS belongs to us all

Document Control

Document Purpose	Information
Document Name	Privacy Impact Assessment: care.data
Version	1.0
Publication Date	15/01/2014
Description	This document details the privacy impact assessment for the care.data programme.
Associated Documents	N/A
Issued by	Chief Data Officer, NHS England
Contact Details	england.cdoqueries@nhs.net

Document Status

This is a controlled document. Whilst this document may be printed, the electronic version posted on the NHS England website is the controlled copy. Any printed copies of this document are not controlled. As a controlled document, this document should not be saved onto local or network drives but should always be accessed from the website

Contents

Executive Summary	1
The purpose of a privacy impact assessment	3
What is care.data?	4
Privacy issues as a consequence of care.data	6
Business Case	16
Alternatives to identifiable data	19
What will we do to protect privacy?	20
Public Acceptability	22
Conclusions	23
Endnotes	24
Appendix A: Managing Privacy Risk	25
Appendix B: Examples of use	27
Appendix C: Definition of Terms	29

1 Executive Summary

NHS England (formerly the NHS Commissioning Board [NHS CB][1] was established on 1 October 2012 as an Executive Non-Departmental Body.

NHS England aims to ensure high quality care for all, now and for future generations. We are committed to transparency and to putting patients and the public at the heart of all decisions, grounded by the values and principles of the NHS Constitution. The responsibilities of NHS England can be divided into the following domains:

- Reducing mortality
- Improving care for patients with long term conditions
- Improving acute care
- Improving patient experience
- Improving patient safety

Care.data will bring together securely, health and social care information from all care settings in order to improve the quality, efficiency, and equity of services. For the first time, it will allow nationwide linkage of primary and secondary care data in order to identify any unwarranted variations in care across the country. Care.data will enable increased use of information that the NHS already collects with the intention of improving healthcare, by ensuring that timely and accurate data are made available to NHS commissioners, providers, and researchers.

Under the Health and Social Care Act 2012, NHS England can direct the Health and Social Care Information Centre (HSCIC) to collect data from every provider of care funded by the NHS. This is limited to where the data are necessary for exercising the functions of NHS England. These data are collated, stored, and disseminated or published by the HSCIC rather than NHS England. The HSCIC provides a secure data environment, which operates to the very highest technical and security standards. The HSCIC will disseminate data in three formats:

1) Anonymous or aggregated data will be published in line with Information Commissioner's Office (ICO) anonymisation code of practice, e.g. with small number suppression, to ensure that the risk of re-identification is very remote[2].

2) Pseudonymised[3] data will be made available to specific approved groups of users, initially for commissioning uses only and in line with ICO guidance (see section 4.3 box 2).

3) Identifiable data will only be made available where there is a legal basis for doing so (e.g. with patient consent or approval under Regulation 5 of the Health Service [Control of Patient Information] Regulations 2002 [commonly known as 'section 251 approval).

A privacy impact assessment (PIA) is a tool, process or method to identify, assess, mitigate or avoid privacy risks. This PIA describes how data will be collected, processed, disseminated and published for care.data. It explains what the programme will do to protect privacy and the solutions that have been identified and implemented to help safeguard privacy. This document will enable readers to assess for themselves what may be considered a potential impact on their privacy.

The HSCIC has undertaken a PIA for all the personal data it processes, which includes the data extracted for care.data[4]. The HSCIC provides many of the technical and information governance controls for care.data therefore this PIA draws upon the overarching HSCIC PIA. This PIA, however, specifically considers the privacy impact of care.data in greater detail.

In summary, the benefits for processing the data are

clear. The only way to determine whether the NHS is achieving its aim of universal, high quality care is through analysing detailed, high quality information about the care being provided to patients. By using data we can, for instance, identify examples of best practice so that these can be rolled out across the country or identify examples of substandard care, so that we can take swift action to improve services.

In order to achieve these aims, it is necessary for the HSCIC to link – in a secure environment – data from different parts of the NHS and from social care services. To ensure the accuracy of the linkage, the NHS number, postcode, date of birth and gender are extracted by the HSCIC. This PIA describes in detail the privacy implications of this extraction and how any risks are mitigated. Briefly, these details include:

- The ways in which patients will be made aware of how their health information is shared with the HSCIC and what choices are available to them;
- How patients will have greater control over the identifiable information held about them;
- How patients can object to the use of identifiable information beyond their direct care;
- How personal confidential data are processed in ways that reduce risk and increase security, viz.;
    - anonymisation (following the Information Commissioner's guidance)
    - data sharing contracts with all organisations that are approved to receive data
    - applying sanctions to organisations that do not comply with the terms of their contract

2 The purpose of a privacy impact assessment

Privacy impact assessments (PIAs) were launched in the UK by the Information Commissioner in December 2007 and were mandated by the Cabinet Office for information and communications technology (ICT) projects following the Data Handling Review of June 2008[5].

The Health and Social Care Act 2012 introduces legislative powers that enable NHS England to direct the HSCIC to obtain and process identifiable patient data in certain circumstances without the need for patient consent. This arrangement includes care.data.

Patients, and those people legally empowered to act on their behalf, must be informed about how identifiable data about them are used. Therefore, alongside other awareness-raising activities, NHS England and the HSCIC are informing patients about how care.data might affect the privacy of personal data. The privacy impact assessment:

- Describes the purpose and objectives of the care.data programme;
- Assesses the potential implications for privacy; and
- Explains what NHS England and the HSCIC will do to protect privacy.

The scope of this PIA will cover the whole of the care.data programme, including each of the domains of health and social care information that are currently planned to be included within the programme. The care.data programme has a number of phases relating both to the datasets to be acquired and the functionality offered. All of these phases will be encompassed by this PIA. The first phase is the linkage of GP data with hospital data so the emphasis, in this first iteration, is on this element. The PIA will be kept under review and revised as the detail for each phase is developed. We welcome feedback on this PIA.

3 What is care.data?

The NHS has some of the best information systems in the world. Since the 1980s, we have been collecting information about every hospital admission, nationwide. This information is brought together at the Health and Social Care Information Centre, where it is anonymised. The information has been invaluable for monitoring the quality of hospital care, for planning NHS services, and for conducting research into new treatments. While we have this type of information for some of the care provided outside hospitals, there are significant gaps, meaning that it is not possible to see a complete picture of the care that individuals receive.

NHS England has therefore commissioned a programme on behalf of the NHS, public health and social care services to address these shortcomings. Known as the care.data programme, this initiative will ensure that there is more rounded information available to citizens, patients, clinicians, researchers and the people that plan health and care services. Our aim is to ensure that the best possible evidence is available to improve the quality of care for all.

The care.data programme is designed to ensure that

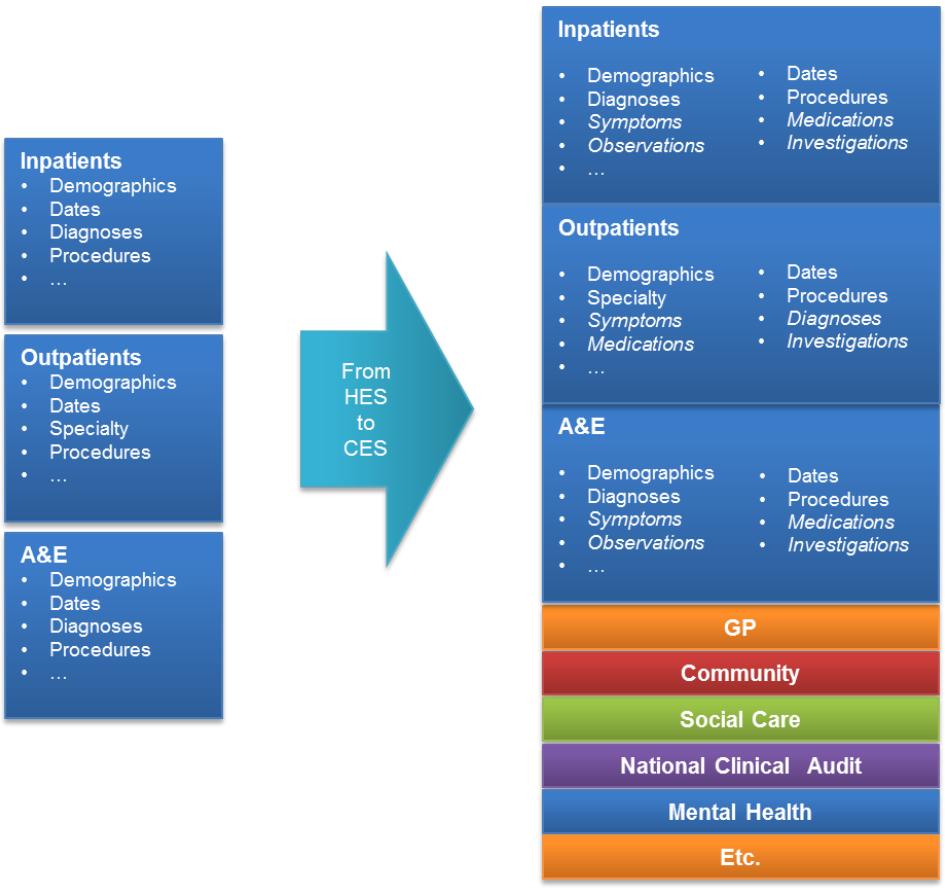

Figure1: HES to CES

commissioners have access to a dataset that contains linked information from all GP practices and all hospitals. Over the following years, data will be progressively added from other care settings, including community health services and social care, and the hospital dataset itself will become much richer and more complete. This transformation will see Hospital Episode Statistics (HES) evolve into a care episode service (CES).

The HSCIC will collect and process the data for care.data using its powers under the Health and Social Care Act 2012. The HSCIC is England's central, authoritative source of health and social care information. The HSCIC will process patients' confidential data in a secure environment and will only release confidential data where there is a legal basis for doing so.

The main functions of the HSCIC in relation to care.data are to:

i. collect and process, patient identifiable data extracted from patient records;
ii. assure the data quality of patient identifiable data;
iii. link and de-identify patient identifiable data;
iv. publish aggregate data;
v. disseminate potentially identifiable data to approved bodies and where strict controls exist so the likelihood of an individual being identified are very small;
vi. disseminate patients and specific bodies patient identifiable data (only where necessary, in exceptional circumstances and when lawfully authorised, such as under Regulation 5 of the Health Service [Control of Patient Information] Regulations 2002 [commonly known as 'section 251 approval']). This is a future aim and will be subject to independent approval.

The following diagram shows how the linked dataset will be made available in order to realise the benefits outlined above. The privacy impact of these data flows is considered in further detail below.

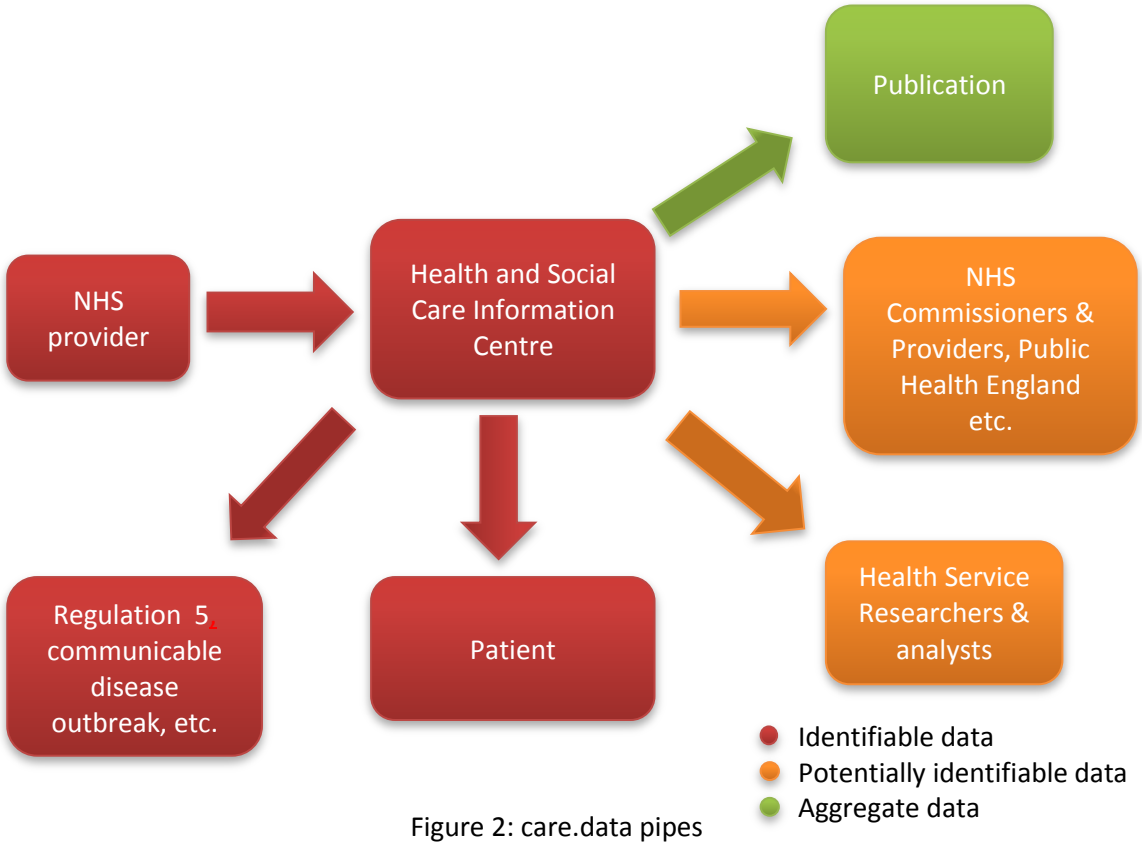

Figure 2: care.data pipes

4 Privacy issues as a consequence of care.data

This section assesses the potential impact on privacy of care.data. To assess the potential privacy impact of care.data, it is necessary to weigh up not only whether the impact is positive, neutral or negative but also to consider the extent to which any adverse impact on privacy may be acceptable if it brings other benefits. Safeguards to protect privacy are explained in section 7 – *What will we do to protect privacy?*

Care.data involves the extract of personal confidential data from health records, including *sensitive personal data* as defined in the Data Protection Act 1998. Identifiers (NHS number, date of birth, postcode and gender) are extracted from providers together with coded clinical information and sent to the HSCIC. As with any disclosure of personal confidential data, there are associated risks to privacy and confidentiality.

The privacy impact can be considered in three areas:

- The extraction of personal confidential data into the HSCIC
- The processing of the personal confidential data when held by the HSCIC
- The onward of disclosure of data from the HSCIC

4.1 The extraction of personal confidential data into the HSCIC

The Health and Social Care Act 2012 provides a legal basis for the extraction of personal confidential data in certain circumstances. The Act sets aside the requirement under the common law duty of confidence to seek patient consent[6]. Certain requirements under the Data Protection Act 1998 continue to apply – in particular, the fair processing principle, which means that patients must be made aware of how confidential data are processed for care.data.

The extraction of personal confidential data from providers without consent carries the risk that patients may lose trust in the confidential nature of the health service. This risk is two fold: firstly, patients will not receive optimal healthcare if they withhold information from the clinicians that are treating them; and secondly, that this loss of trust degrades the quality of data for care.data and other secondary uses of NHS data.

To mitigate against the risk, the NHS constitution gives patients the right to object to their personal confidential data leaving their GP practice. In line with the commitment given by the Secretary of State for Health in April 2013, patient objections will be upheld other than in exceptional circumstances such as a public health emergency[7]. Previously, there was no straightforward mechanism for patients to exercise this right. This is therefore a step forward in providing patients with greater control over the identifiable information held about them.

It is important to note that personal confidential data have been processed for many years where there has been a legal basis for doing so (e.g., where there has been special approval for medical purposes such as research). This new objection extends to all disclosures of personal confidential data from the GP practice, not just care.data.

At present, it is not possible for patients to prevent flows of confidential data from other care settings into the HSCIC, for example from hospitals. For this reason, we have ensured that patients can also object to the disclosure of confidential data from the HSCIC (see section 4.3).

In order to ensure that patients are aware of the changes to how data are processed for care.data, and to ensure they are aware they can object, a number of awareness raising activities are underway. Figure 1 summarises the awareness raising activities that are taking place.

Box 1 - Summary of supporting activities and resources

1. A patient leaflet and poster about information sharing made available in GP practices
2. Materials developed in accessible formats including Braille, large print and audio versions.
3. Materials for GP practices to support patient awareness raising including a how to guide and a template press release for local tailoring
4. Detailed FAQs for both GPs and patients
5. Testing of these materials in a limited number of GP practices with feedback incorporated into the national version of the leaflet and poster.
6. The mailing of a leaflet about information sharing to every household in England
7. Separate GP and patient information lines to support understanding
8. Regional events for GPs and NHS managers to encourage awareness raising at a regional level for example via regional press releases.
9. Social and digital media: dedicated web support pages for patients through NHS Choices and for professionals through NHS England.
10. Use of central social media channels to help raise awareness and to direct to particular FAQS such as objection process.
11. Engagement with a number of national patient groups, charity and voluntary sector organisations to enable cascade of messages through their regular and social media channels.

In order to evaluate the potential impact on their privacy, patients need to understand what data are to be extracted. As explained above in section three, the first stage of CES will involve linking GP data to hospital data. Data from other parts of the health and social care service will be linked over time and this PIA will be updated to reflect this. The dataset extracted from GP systems has been published and includes data such as referrals, prescriptions, symptoms, diagnoses, and treatments. Whilst all health data is classified as sensitive personal data under the DPA, a list of particularly sensitive items will be excluded from extracts[8]. The data extracted is in the format of a series of codes. Free text (i.e., words, sentences, and paragraphs) will *not* be extracted for care.data.

Extraction of the GP data will be on a monthly basis using the General Practice Extraction Service (GPES). This is a tool provided by the HSCIC, which extracts data from GP practices into the HSCIC. Extractions of those data items included in the published dataset will start in spring 2014. The first extract will include data recorded in GP records since autumn 2013. The analysis of historic data would bring much greater insight into the provision of care and increase the opportunities for valuable research. It is therefore the longer-term vision of the programme to extract historic data; however, we are adopting a phased approach so historic data will not be extracted initially.

In accordance with the Data Protection Act 1998, only the minimum necessary patient identifiable data will be collected. The GP dataset has been considered by an independent group of clinical informatics experts, which included representatives from the British Medical Association (BMA) and the Royal College of General Practitioners (RCGP). The group was satisfied that the dataset seemed appropriate for commissioning. Any future changes to the dataset or to its scope will be subject to further review by the group. In addition, the care.data proposal was also reviewed and approved by the GPES Independent Advisory Group (IAG), which includes clinical and patient representation. Any changes or updates to the dataset will be published on the NHS England website.

Table 1 below shows the reasons for processing and benefits, the impact on privacy and the controls and pledges.

Table 1: Reasons, impact and controls for the extraction of personal confidential data into the HSCIC

Reasons for processing and benefits	Impact on privacy	Controls and Pledges
• Data collected are fundamental to the NHS, and/or necessary to improving public health or health services. • Personal confidential data are extracted from healthcare providers to enable linkage.	• Some people may feel a loss of individual autonomy (no patient consent) • Some patients may not be aware of or understand their choices.	• Statutory basis for data collection required or permitted by law[9] • Identifiable data must be necessary to satisfy the purpose • Awareness raising activities will help patients understand how their data are used not only for care.data but other uses of healthcare. • Patients can object to the processing of the personal confidential data in GP records. • Control 1 (see Section 7 – "Information Governance Controls") • Pledge B, C and D (see section 7 – "Additional care.data pledges to protect information")

4.2 The processing of the personal confidential data when held by the HSCIC

Under the Health and Social Care Act 2012, the HSCIC is established as a 'safe haven' with powers to collect and analyse confidential information about patients. The HSCIC will process the personal confidential data for the care.data service.

The HSCIC's PIA[10] details the risks and responsibilities it has to protect the confidentiality of all the data it holds. This PIA is much broader than the care.data PIA because the HSCIC is the data controller for numerous datasets in addition to those collected for the care.data programme. As stated in the HSCIC PIA, 'the HSCIC like all organisations that process and store patient identifiable data, must protect the confidentiality of that data and must guard against risks and threats from inside and outside the organisation'. The risks described include threats associated with 'cyberspace' such as hackers attempting to access the data illegally. The HSCIC PIA describes in detail how these risks and threats are addressed and minimised by effective information governance controls.

Processing of data by the HSCIC has a potential impact on privacy because the HSCIC is an organisation to which patients have not disclosed information themselves. At a local level, personal confidential health data have been used for many years for the purposes of indirect care (e.g. for planning services, audit, and research). There have been incidents of local data breaches and also a misunderstanding of the complex legal and information governance framework for health data. Local processing can be difficult to monitor and audit, and the likelihood of an individual being identifiable when processing takes place locally is higher than when data are processed centrally (e.g. recognising the name of a neighbour). Whilst a centralised data collection has potential privacy implications, these risks can be balanced with a reduction in the requirements for local processing of personal confidential data and with assurances that data processing by the HSCIC is to the highest security standards. The technical expertise and detailed knowledge of information governance is very difficult to match across all organisations operating at a local level.

It is necessary for the HSCIC to receive identifiers so that it can link data from different healthcare settings to realise the benefits outlined in section 5. Data linkage involves matching together the records from two or more care settings about the same patient to provide a more complete picture of the patient's needs, experiences of care, and outcomes. For example, hospital records and general practice records could be linked in order to analyse the impact upon outcomes of different care pathways for a particular condition.

The privacy risks associated with the HSCIC are mitigated because the process of linking the record is automated. Occasionally, in a small number of cases, it is necessary for HSCIC analysts to check the data for data quality reasons. However, this human involvement is done following strict rules and processes, all of which are designed to protect the confidentiality of the individual. These include, for example, rules around retaining the data, destroying the data, disclosing the data and illegally matching data to identify individuals. Patient identifiers (NHS number, date of birth, postcode and gender) are held separately from clinical data and wherever practicable HSCIC staff are assigned access rights to either the patient identifiers or the clinical data not both.

It is important to reiterate that the data that is extracted from GP practices for care.data does not include patients' names and addresses. Furthermore, the data are presented in terms of clinical codes rather than free text (i.e., no words, sentences, or paragraphs). Once the record has been linked, the identifiers are removed so a new record is created that does not identify the patient.

Where patients have objected to the flow of their personal confidential data from the general practice record, the HSCIC will receive clinical data without any identifiers attached (i.e., anonymised data). The HSCIC will extract the fact that the patient has objected and the date of that objection; no other personal data will be extracted.

If a patient is (a) content for personal confidential

data from their GP record to be extracted into the secure environment of the HSCIC but (b) objects to flows of personal confidential data from the HSCIC (see section 4.3) then the HSCIC will extract the fact of the objection, the date of the objection and the individual's NHS number. The NHS number will be used internally within the HSCIC to match these data to other data held for that patient so that the data can be anonymised before release.

Where a patient objects to flows of personal confidential data (a) from their GP practice and also (b) from the HSCIC, then it is necessary for the patient's NHS number to flow to the HSCIC so that their objection to flows of personal confidential data leaving the HSCIC can be implemented.

Table 2 below shows the reasons for processing and benefits, the impact on privacy and the controls and pledges.

Table 2: Reasons, impact and controls for the processing of personal confidential data when held by the HSCIC

Reasons for processing and benefits	Impact on privacy	Controls and Pledges
• Accuracy has to be checked before data are de-identified (it is not possible afterwards) • De-identification so that the data can be used more freely for the benefit of patients. • Information used by the public to make healthcare decisions and by people inside and outside the NHS for activities such as medical research, public health and national clinical audit, has to be good quality. The HSCIC is responsible for ensuring this. • Linking data from different healthcare settings is a powerful means of increasing knowledge and can bring benefits to commissioning, in medical research and public health.	• Data collection, storage and processing creates a risk of confidential information being accessed without the knowledge or consent of patient • Risks in terms of changes to scope (e.g. to dataset) without patients being aware.	• Statutory basis for collection and analysis. • Identifiable data stored only where necessary and destroyed or aggregated, anonymised or pseudonymised as soon as possible. • A single national extraction reduces the need for local processing of personal confidential data where patients are more likely to be identifiable or where the safeguards in place, may not be as robust. • Patient identifiers are held separately from clinical data within the HSCIC. • De-identifying data reduces or eliminates the risk of a person's identity being revealed and thus helps protect privacy • Approval from an Independent Advisory Group for any changes to scope, e.g. to the GP dataset, and publication of the minutes and recommendations of this group[11]. • Controls 1, 2, 3, 4 and 7 (see Section 7 – "Information Governance Controls") • Pledges A, B, C, D, E and F - (see section 7 – "Additional care.data pledges to protect information")

4.3 The onward disclosure of information from the HSCIC

The law pulls in different directions where dissemination of information is concerned; human rights legislation, data protection legislation, and the common law duty of confidentiality all require us to protect information that could identify an individual. The Health and Social Care Act 2012, however, allows the HSCIC to obtain and disseminate information about patients when acting under direction from the Secretary of State or NHS England.

The data flows diagram in section 3 shows how data will be made available from the HSCIC. There are three categories of disclosure:

- Green flow – aggregate data
- Amber flow – potentially identifiable data
- Red flow – personal confidential data

Green flows of data will be published only in aggregated form with additional safeguards (e.g., small number suppression) so the risk of identifying an individual is very remote. This will be in line with the ICO code of practice on anonymisation[2] the Information Standards Board anonymisation standard for publishing health and social care data[12]. Transforming identifiable data into anonymised data protects personal privacy and enables published information to be used for public benefit.

Amber data are pseudonymised in line with ICO guidance and will only be disclosed by the HSCIC to approved users. There is a remote risk that a patient could be identified even though identifiers are removed (e.g. if you knew a patient with a rare disease lived in a particular area). However, the ICO advises that limited access allows the disclosure of richer data[2]. All amber disclosures will be in accordance with robust information governance controls listed in box 2.

Box 2 - The following robust safeguards must be in place in relation to disclosure of data by the HSCIC:

- purpose limitation, i.e. the data can only be used by the recipient for an agreed purpose or set of purposes;
- training of recipients' staff with access to data, especially on security and data minimisation principles;
- controls over the ability to bring other data into the environment, allowing the risk of re-identification by linkage or association to be managed;
- limitation of the use of the data to a particular project or projects;
- restriction on the disclosure of the data;
- prohibition on any attempt at re-identification and measures for the destruction of any accidentally re-identified personal data;
- arrangements for technical and organisational security, e.g., staff confidentiality agreements;
- encryption and key management to restrict access to data;
- limiting the copying of, or the number of copies of the data;
- arrangements for the destruction of the data on completion of the project; and
- penalties, such as contractual ones that can be imposed on the recipients if they breach the conditions placed on them.

Whilst there is a privacy risk that the analysts granted access to these pseudonymised flows could potentially re-identify patients maliciously by combining the pseudonymised data with other available datasets (a technique known as a jigsaw attack) such an attack would be illegal and would be subject to sanction by the ICO.

Red Flows – These flows involve the disclosure of personal confidential data from the HSCIC and are only permitted where there is a legal basis (e.g., explicit patient consent or approval under Regulation 5 of the Health Service [Control of Patient Information] Regulations 2002 [commonly known as 'section 251 approval'] or exceptionally where there is an overriding public interest in disclosure such as an outbreak of a new disease or a civil emergency). In order to establish trust in care.data from patients and healthcare professionals, personal confidential data collected for care.data will initially only be disclosed where there is an overriding public interest even though disclosures under Regulation 5 or with patient consent would be legally permissible.

If it is agreed in the future[13] that personal confidential data, collected as part of the care.data

programme, will be disclosed by the HSCIC e.g. where there is Regulation 5 approval, patients can object to this by informing their GP and such objections will be honoured. GPs can register such objections by entering a code into the GP record.

As stated in section 4, there is not a straightforward process for patients to prevent data flows from other care settings, e.g. hospitals, to the HSCIC. This code prevents personal confidential data derived from any healthcare setting leaving the HSCIC unless there is an overriding public interest such as a civil emergency.

Whilst it is possible for patients to object to the processing of personal confidential data under section 10 of the Data Protection Act 1998, this new code allows patients to exercise, to a large degree, choice more easily: they simply need to ask their GP to enter this code into their GP practice record. Put simply, patients who are concerned about their privacy can now control the flow of confidential data both out of their GP practices and out of the HSCIC.

Table 3 below shows the reasons for processing and benefits, the impact on privacy and the controls and pledges.

Table 3: Reasons, impact and controls for the onward disclosure of information from the HSCIC

Reasons for processing and benefits	Impact on privacy	Controls and Pledges
• Once data are de-identified they can be used without breaching confidentiality for a large number of "secondary purposes" that are fundamental to the operation of the NHS and/or necessary to improving public health or health and social services where jointly commissioned with the NHS. • Data are used to help plan and monitor services, understand the health needs of patients and improve the quality of health care provision. • Data are used to understand the outcomes that patients receive, as well as the patient experience and efficiency of the service. • Information used by the public to make health care decisions, and by people inside and outside the NHS for activities such as medical research, public health intelligence and clinical audit on a national scale. • Comparing the quality of care provided by different hospitals to identify outliers, using the lessons learnt from those performing exceptionally well and take urgent steps to investigate and address those organisations performing less well.	• In some cases, a small residual risk that identifiable data could be revealed[2] • Risks of jigsaw attacks increase as more effectively anonymised data are made available, to more organisations.	Green data: • Anonymisation techniques will be applied as described in the Appendix 2 of the ICO's anonymisation code e.g. small number suppression, rounding up or down of numbers etc. Amber data: • Robust information governance controls will be applied as detailed in box 2. Red data: • Disclosures of personal confidential data will be limited in the first instance to exceptional circumstances for example in the event of a civil emergency. • Disclosures of personal confidential data can only occur where there is a legal basis, for example under Regulation 5 of the Health Service (Control of Patient Information]) Regulations 2002 (commonly known as 'section 251 approval') • Patients can object to their personal confidential data leaving the HSCIC. • Controls 5, 6, 7 (see Section 7 – "Information Governance Controls") • Pledges A, B, C, D and E (see section 7 – "Additional care.data pledges to protect information")

4.4 Conclusion of privacy issues as a consequence of care.data

The main tension identified within this privacy assessment is the balance between the benefits of:

- using linked personal confidential data from health and social care records to improve the quality, efficiency, and equity of care provision through better commissioning of services with a focus on safety, outcomes and patient experience; versus
- the risks to patient privacy from the collection, linkage, storage and dissemination of the data in a variety of formats.

A key component of any assessment is the degree to which these risks are mitigated by the controls and security that will be applied. In this case, only the HSCIC will process data in identifiable form, an arrangement that markedly limits the risks to an individual that their privacy will be breached by this programme. Moreover, a potential positive impact of care.data is that more organisations should be able to use pseudonymised information where they currently use identifiable information.

As also referenced in the HSCIC PIA the potential risks to privacy from care.data are:

A. Loss of individual autonomy from use of patient identifiable data without consent
B. Risk of confidential information being accessed and viewed without knowledge or consent of patients
C. Linking and de-identification processes may not be reliable enough to achieve total anonymisation of data
D. Risk of data being accessed illegally and then sold or otherwise misused by commercial organisations, criminals or others; and
E. Risk of data being accessed legally and then the data being misused.

The actual mitigating controls that the care.data programme will use to safeguard these risks are summarised below with more details in section 7.

The HSCIC will:

A. Obtain and process only the minimum necessary patient identifiable data from other organisations
B. Store and process data in its capacity as "safe haven", under the Health and Social Care Act 2012
C. Keep to the absolute minimum the number of staff able to access and view patient identifiable data, and wherever practicable assign staff rights of access to either patient identifiers or clinical data but not both
D. Destroy data held in identifiable form as soon as they are no longer required, or in accordance with the HSCIC's retention policy
E. Disclose only anonymised data, unless there is a legal basis for the disclosure of confidential data.
F. When disclosing anonymised data, restrict the data disclosed according to the context in which the data will be used.
G. Monitor who accesses patient identifiable data by maintaining an audit trail to record, retain and report on system events as highlighted above (i.e., which staff members have been assigned access rights to view patient identifiable data).

5 Business Case

This section provides the business case for care.data. It begins by explaining why NHS England believes that these changes to the use of patients' data are so important and the value that the linked pseudonymised data and published aggregate data can bring to a wide range of people including patients, the public, health and social care providers, commissioners and researchers.

The NHS has some of the best information systems in the world. Since the 1980s, we have been collecting information about every hospital admission, nationwide. This information is brought together at the HSCIC, where it is anonymised. The information has been invaluable for monitoring the quality of hospital care, for planning NHS services, and for conducting research into new treatments. However, the information collected is incomplete, with areas such as prescribing and test results not currently included. Additionally, while we have this type of information already for some of the care provided outside hospitals, there are significant gaps. As a result, it is not currently possible for us to see a complete picture of the care that patients receive.

NHS England has therefore commissioned a programme of work on behalf of the NHS, public health and social care services to address these gaps. Known as the care.data programme, this initiative will ensure that there is more rounded information available to citizens, patients, clinicians, researchers and the people that plan health and care services. Our aim is to ensure that the best possible evidence is available to improve the quality of care for all.

The six aims of care.data are to enable:

i. Greater accountability
ii. Informed choices
iii. Greater efficiency
iv. Better outcomes
v. Improved customer services
vi. Economic growth.

A key piece of work that NHS England will undertake jointly with the HSCIC is to ensure that the benefits of care.data are assessed to ensure that the aims of the programme are being met.

1. Greater Accountability

Through the care.data programme, NHS England will help citizens to hold the NHS to account by making more information available about the quality, safety, and efficiency of the care provided. For example, we will make more information available about prescribing patterns so that citizens can see how equitable is the provision of drugs across England. We will do the same for waiting times, disease outcomes, and other metrics.

2. Informed Choices

Better access to higher quality and more complete information will give patients the opportunity to exercise greater control over their care and wellbeing. For example, the document '*Liberating the NHS: An information strategy*'[14] lists the kinds of information that people will use, including information about:

- suitable medicines and treatments, together with their risks, benefits and side effects;
- clinical outcomes and success rates, such as readmission or mortality rates;
- other indicators of quality and performance, such as infection rates.

An important role of the care.data programme is to provide such information to patients and clinicians, and enable patients to make more informed choices. By doing so, patients and clinicians will be able to play a more effective part in improving the quality and efficiency of the health service.

NHS England is committed to making more information available to patients and citizens through the care.data programme, and we will encourage people to make greater use of this information. We will also be responsive by providing information that people say they want in the formats they want it in.

3. Greater Efficiency

Detailed analysis of data can help improve efficiency through a variety of mechanisms, including:

- Ready access to high quality information can lead to improvements by allowing professionals to identify variances and inconsistencies in their practice compared to how other health care providers practice. For example, GPs may identify more efficient prescribing practices amongst their peers, and pathologists may identify practices in other laboratories that will help increase their productivity.
- As part of the care.data programme, NHS England will make more information available about the efficiency and performance of all parts of the NHS. Having ready access to this information will support NHS organisations to become more efficient and will help reduce the cost to individual organisations of obtaining and processing information about performance.
- Using risk models and decision aids can help ensure that care is provided to those who most need it or prevent unscheduled hospital admission eliminating waste and thereby improve efficiency.

4. Better Outcomes

Other than some local examples, there is little linkage of detailed level data across primary and secondary care settings, so there is little opportunity to monitor patient outcomes. Linking data provides a more complete picture of the care so that analysts can look at the effect of an intervention or a particular route a patient took from diagnosis, through to treatment and discharge and see what worked best so that all patients can benefit.

To give an example: 20 patients have the same surgical procedure and are discharged by a hospital. The hospital has no further contact with the patients and therefore hospital analysts consider that the procedure was successful. However, the hospital was unaware that 10 of those patients visited their GP with complications, which were managed by the GP practice. By looking at the linked data, the hospital analysts would have had a more complete picture of patient outcomes.

5. Improved Customer Services

We need to ensure patients are at the heart of the health and social care service and services are tailored to each individual's needs so they receive a first class customer service. Patients can use information about services to make informed choices about their healthcare. Eventually it is intended that patients can have access to their data including the data collected for care.data so they can share this for example, with healthcare providers, charities or their family and play an active role in shaping their own healthcare.

6. Economic Growth

In order to safeguard the fundamental philosophy of providing high quality care to all, free at the point of delivery, the NHS requires a strong economy. The care.data programme will support economic growth in a variety of ways. For example:

- Greater access to high quality health and care data will help reinforce the UK as a global centre for life sciences and health services research. Making de-identified data available at scale will help researchers discover and refine new treatments. It will also help epidemiologists and public health researchers to shed more light on the role of social conditions and lifestyle choices on health outcomes.
- Making comparative data available to app developers and website designers will support the development of a vibrant market place.
- Offering a range of data services at regional and local levels will supports economic growth by encouraging small and medium-

sized enterprises (SMEs) to provide of innovative, locally-tailored analytical tools and services.

- Better information will support the modernisation of services, which in turn will support economic growth

6 Alternatives to identifiable data

The fundamental purpose and benefit of care.data is to collate and link health and social care data from a wide range of care settings in order to provide a more complete picture of the care received by patients. Put simply, in order to ensure that it is providing joined-up care, the NHS needs joined-up data. For example, in order to gauge the quality of services for patients who have had a hip fracture, clinicians, commissioners, and researchers need information about:

- what happened to patients while they are being cared for
    - by the ambulance trust,
    - by the A&E department,
    - in the operating theatre and
    - on the hospital ward
- how well their care was coordinated after leaving hospital, and
- whether they maintained their independence.

In order to make a complete assessment of the outcomes for patients, we therefore need to link data from all of the settings at which they may receive care, including primary care, secondary care, tertiary care, community health services, and social care.

Clearly, it is essential that a patient's data from one care setting are linked accurately with their data from another care setting. We use four separate identifiers to ensure that GP and hospital data are linked for the care.data programme. These identifiers are the patient's NHS number, date of birth, gender and postcode.

We have rejected alternative data linkage techniques using fewer identifiers because the scale of the linkage required (i.e. patient records for the whole population from a wide range of health and care settings) means that there are more individuals who share similar identifiers. Using fewer identifiers would lead to more incorrect linkages, which would compromise data accuracy and bring into question the validity of care.data. Moreover, using fewer identifiers would lead to a lower proportion of linked records, which would diminish the usefulness of care.data especially because of a bias in the characteristics of patients whose records could not be linked.

Another option would be to de-identify the data at source in a consistent way that allowed individuals' data to be linked without revealing their "real world" identities. Known as *pseudonymisation-at-source,* this technique relies on the use of a common key across all care settings, which generates a unique pseudonym for each individual that allows their data to be linked. At the moment, the HSCIC considers pseudonymisation-at-source to be impractical because there is such a diverse range of care settings providing data to the programme (primary, secondary, tertiary, community, and social care) and such a diverse range of information systems used in each setting. However, the protection of patient confidentiality is a priority for the HSCIC and NHS England so a review of the use of pseudonymisation tools within the HSCIC is underway to ensure that the organisation is applying privacy enhancing technologies in the most effective ways.

7 What will we do to protect privacy?

The care.data programme is being delivered by the HSCIC, whose core purpose within legislation is to process patient records safely and securely. As stated in the HSCIC PIA 'The HSCIC has been processing patient records safely and securely since its inception. It has introduced strong security controls, published and implemented security policies and published information about its processing as required for compliance with the Department of Health's Information Governance Framework.

The HSCIC takes its responsibilities as a custodian of patient information extremely seriously and is also committing to a number of pledges to protect privacy as set out below'. In Appendix A, we describe how the privacy risks identified in section 4 are addressed by these controls and pledges.

7.1 Information Governance Controls

The HSCIC will collect, process, disseminate and publish data on behalf of NHS England for and care.data programme. The HSCIC provides assurances regarding Information Governance through:

- An Information Assurance Steering Group, with reporting lines to the Executive Board
- satisfactory completion of the NHS Information Governance Toolkit13, and compliance with ISO27001/2 Information Security Standards, which include:
  - Staff training and contracts
  - Information technology system security and audit trails
  - Robust management arrangements
  - Full compliance with legislative requirements
  - Provision of the “safe haven” for sensitive information

Specifically the HSCIC will:

1. Obtain and process the minimum necessary patient identifiable data from other organisations;

2. Store and process identifiable data securely, meeting or exceeding the standards required of NHS organisations, including technology to:
   i. De-identify data received as early as possible, and where records have to be linked, it will separate patient identifying data from clinical data, and assign a meaningless identifier (pseudonymisation)
   ii. Store data in its capacity as the "safe haven" under the Health and Social Care Act 2012.
   iii. protect against attacks from unauthorised individuals (e.g. hackers)
   iv. protect against inappropriate behaviour by staff;
   v. provide only legitimate personnel with access to HSCIC systems, and to no more access than they legitimately require;

3. Keep to the absolute minimum the number of staff able to access and view patient identifiable data, and wherever practicable assign staff rights of access to either patient identifiers or clinical data but not both;

4. Destroy data held in identifiable form as soon as they are no longer required, or in accordance with the retention policy;

5. Disclose only anonymised data, other than:
   i. with explicit patient consent;
   ii. where required by law, or
   iii. where allowed by law, with necessary support and approvals, and either:
   - the support of the Independent Advisory Group; or
   - where urgent, with the agreement of both the Senior Information Risk Officer and Caldicott Guardian for HSCIC;

6. When disclosing anonymised data, restrict the

data disclosed according to the context in which the data will be used:

i. When publishing statistics and other aggregated information, apply disclosure control standards[15] to ensure data are anonymised;

ii. When disclosing patient-level data to a trusted organisation:

- confirm the data are anonymised by carrying out a risk assessment
- maintain a written agreement with the recipient organisation that stipulates the permitted access to, and uses of, the data;

7. Monitor who accesses patient identifiable data by maintaining an audit trail to record, retain and report on system events highlighted above (i.e. the staff members who have been assigned access rights to view patient identifiable data).

7.2 Additional care.data pledges to protect information

In addition to the information governance controls outlined above, further safeguards will be put in place to protect information collected, processed and disseminated as part of care.data. The two organisations will be held to account against these pledges by the Department of Health.

A. The HSCIC will publish a Code of Practice to govern the use of confidential data supplied to the Health and Social Care Information Centre that encompasses care.data;

B. The HSCIC and NHS England will respect the wishes of patients who request that their data are not used by care.data, unless there is a statutory duty or an overriding public interest (e.g. public health emergency) to do otherwise;

C. The HSCIC will commission, at least annually, external information governance audit against information governance standards.

D. NHS England and the HSCIC will be transparent about their activities and communicate openly, fairly and lawfully through the NHS England and HSCIC public websites and other channels where appropriate;

E. The HSCIC will publish procedures for dealing with requests for information and operate effective policies and procedures to encourage good information governance by staff, with proportionate sanctions (e.g. dismissal) for inappropriate or negligent behaviour.

8 Public Acceptability

The HSCIC PIA describes how the government consulted stakeholders to inform the powers in the Health and Social Care Act[16].

Personal confidential data have been used for purposes beyond direct care for many years such as for healthcare planning and for research. It is important that patients are clear about what information is being shared, how it is being shared and why so that they can understand the risks and benefits to them and to the wider population. Some patients may have particular concerns and therefore NHS England has made it simple for patients to object.

8.1 Independent scrutiny

The Independent Advisory Group (IAG) of the General Practice Extraction Service (GPES), considered the specification of the GP data. The group, which includes lay representatives, approved the extraction of GP data in order for it to be linked to hospital data and made available to commissioners.

NHS England and the HSCIC have worked closely with the British Medical Association and with the Royal College of General Practitioners during the process. We have listened to, and incorporated views of, these professional organisations, which culminated in publishing joint guidance and materials for GP practices.

8.2 Patient Information Materials

The BMA's Patient Liaison Group (PLG) was involved in commenting on the patient materials for informing patients about care.data. These materials, which included posters and leaflets, were then tested in a small number of practices over the summer of 2013. GP practice staff and patients were invited to provide feedback on the materials and NHS England communicated with these first practices in order to gauge public acceptability at this preliminary stage. Dialogue will continue with these practices as care.data is implemented on a national scale.

Information about the programme has been sent to over 350,000 patient groups, charities, and voluntary organisations. These organisations are being asked to cascade information about care.data to their members through their usual channels, including social media. In addition, NHS England has been engaging with the *strategic partnership programme*[17], which enables voluntary sector organisations to work in equal partnership with the Department of Health (DH), the NHS and social care to help shape and deliver policies and programmes.

A series of meetings are being held with patient groups to discuss care.data. Meetings have been held with stakeholders for example with the Association of Medical Research Charities, Cancer Research UK and the British Heart Foundation. This is an ongoing exercise and further meetings will be held over the coming weeks and months with patient groups and charities to discuss their views on the design and implementation of care.data.

NHS England and the HSCIC have listened to stakeholders and a leaflet about information sharing will be delivered to every household in England. We are also implementing a patient information line to support patients who have questions or concerned and we will monitor feedback.

9 Conclusions

Any processing or storage of identifiable patient data introduces potential risks of data misuse and breaches of privacy. Although they can never be eliminated, such potential risks are significantly mitigated by the robust information governance controls as set out in section 7 which are all designed to safeguard patients' privacy. The centrality of information governance to the care.data to meet or exceed all information governance standards provides greater assurance about privacy than most organisations are able to provide. Moreover, there is also a positive impact on privacy resulting from care.data de-identifying data. Making aggregated, anonymous and pseudonymous data available, to commissioners, researchers and other approved bodies minimises their need to use identifiable data.

However, the processing of a person's information without their permission can be considered a loss of autonomy for that individual. For this reason, in addition to the extensive safeguards for the data, NHS England is supporting data controllers to raise awareness among patients and making it simple for patients to object to the disclosure of personal confidential data.

In summary, people who conclude that the net impact of care.data on privacy will be positive are very likely to be supportive of the programme. Even people who feel the impact will be detrimental to privacy may recognise that the potential benefits of care.data using data from patient records are great, and may therefore feel they are justified ethically on that basis. However, some people may believe that any use of patient identifiable data without explicit patient consent is unacceptable. These people are unlikely to be supportive of care.data whatever its potential benefits and may object to the use of personal confidential data for wider healthcare purposes.

The HSCIC will be processing data on behalf of NHS England and we have detailed the information governance and pledges in relation to care.data. The HSCIC PIA concludes 'While the HSCIC is new, its functions, including the safe and secure processing of data are well founded, tried and tested in previous constituent organisations. The patient, and therefore protecting patient confidentiality, is at the heart of everything we do'. NHS England is committed to working in partnership with the HSCIC and shares this view.

Endnotes

1. NHS England is the operating name of the NHS Commissioning Board as established by the Health and Social Care Act 2012. It is referred to throughout this document as NHS England.
2. Anonymisation: managing data protection risk code of practice http://www.ico.org.uk/for_organisations/data_protection/topic_guides/anonymisation
3. The process of distinguishing individuals in a dataset by using a unique identifier, which does not reveal their 'real world' identity.
4. http://www.hscic.gov.uk/media/12931/Privacy-Impact-Assessment/pdf/privacy_impact_assessment_2013.pdf
5. http://www.ico.org.uk/about_us/consultations/~/media/documents/library/Corporate/Research_and_reports/pia-executive-summary.pdf
6. Under Sections 259 and 256 of the Health and Social Care Act 2012
7. At the launch of the Caldicott Information Governance Review Report 26 April 2013
8. http://www.england.nhs.uk/wp-content/uploads/2013/08/cd-ces-tech-spec.pdf
9. The Health and Social Care Act provides powers for the Health and Social Care Information Centre to require organisations to submit data to it when data collection has been mandated by NHS England or Secretary of State, and in some circumstances, where requested by other bodies.
10. http://www.hscic.gov.uk/media/12931/Privacy-Impact-Assessment/pdf/privacy_impact_assessment_2013.pdf
11. The GPES Independent Advisory Group consider proposals to change the GP element of care.data and information is provided at: http://www.hscic.gov.uk/article/1858/GPES-Independent-Advisory-Group
12. http://www.isb.nhs.uk/news-folder/anon
13. Changes to scope are subject to independent review.
14. http://consultations.dh.gov.uk/information-revolution/informationrevolution
15. The Health and Social Care Information Centre's current policy is available at: http://www.hscic.gov.uk/media/1350/Publications-Calendar-Statistical-Governance-Policy/pdf/The_HSCIC_Statistical_Governance_Policy_v3.1.pdf
16. HSCIC PIA section 2.3 (http://www.hscic.gov.uk/media/12931/Privacy-Impact-Assessment/pdf/privacy_impact_assessment_2013.pdf
17. https://www.gov.uk/government/publications/the-department-of-health-voluntary-sector-strategic-partner-programme
18. http://www.ico.org.uk/for_organisations/data_protection/topic_guides/privacy_impact_assessment
19. https://www.gov.uk/government/publications/confidentiality-nhs-code-of-practice

Appendix A: Managing Privacy Risk

Types of privacy risk

The Information Commissioner's Office Privacy Impact Assessment Handbook[18] explains why privacy matters and identifies and describes four classes of privacy risk:

- privacy of personal information;
- privacy of the person;
- privacy of personal behaviour; and
- privacy of personal communications.

The care.data programme could potentially pose risks to the privacy of personal information (i.e. the first of these classes of privacy risk). The two sub-categories of risk to privacy of personal information are relevant:

A. Risks to individuals as a result of contravention of their rights in relation to privacy, or loss, damage, misuse or abuse of their personal information;
B. Risks to organisations providing and/or using data within care.data as a result of:

I. perceived harm to privacy;
II. failure to meet public expectations on the protection of personal information (causing damage to the organisation's reputation);
III. failure to comply with the law, leading to enforcement action from the Information Commissioner; or compensation claims from individuals.

Risk Mitigation Matrix

Section 4 identifies a list of potential privacy risks and their potential impacts, and section 7 identifies information governance controls and pledges to reduce the risks to privacy. The table below indicates which risks each of the pledges is intended to address.

Control/Pledge to reduce risk/impact	Loss of autonomy	Confidential information viewed without consent	Misuse following illegal access	Misuse following lawful access
1) Obtain only the minimum necessary identifiable data	✓	✓		
2) Store and process identifiable data securely		✓	✓	✓
3) Minimise which staff are able to view identifiable data	✓	✓		✓
4) Destroy identifiable data when no longer necessary	✓	✓	✓	
5) Disclose only anonymised data (other than where there are lawful exceptions)	✓	✓		✓
6) Restrict what data are disclosed according to the context e.g. whether or not published	✓	✓		
7) Monitor who accesses patient identifiable data		✓		
A. Publish a code of practice to govern the use of data			✓	✓
B. Respect patient objection	✓	✓		
C. Commission information governance audits		✓	✓	✓
D. Be transparent and communicate fairly and lawfully	✓	✓		
E. Operate good information governance amongst staff with sanctions for misconduct		✓	✓	✓

Appendix B: Examples of use

The following are examples of how care.data could be used:

Example 1 - Colorectal cancer
Colorectal cancer is the third most common cancer after breast cancer and lung cancer, and is a major cause of mortality. Patients' chances of survival at five years are far higher when colorectal cancers are detected at an early stage (93%) versus those diagnosed late (6%); however, one quarter of colorectal cancer cases are diagnosed during emergency presentations to hospital.

The North East London Cancer Network identified a marked variability in the timeliness and accessibility of diagnostic services for patients with colorectal symptoms. The colorectal pathway is highly complex, with patients being referred to diagnostic services from both the community and by many different secondary care services.

High quality data are required to guide the development and implementation of care pathways, and to support patients and clinicians in making shared, informed decisions about treatment options. At a local level, GP data linked to hospital data are being used in a study that is examining variations in routes to diagnosis among patients with colorectal cancer in outer North East London. The project is exploring existing diagnostic pathways and aims to identify those pathways that result in the best outcomes for patients including the fewest avoidable healthcare episodes.

What does this mean for patients?

Every patient wants to be confident that they are receiving the best possible care. However, poorly designed pathways of care tend to result in more fragmented, inefficient care, and poor patient experience. By linking GP and hospital data, analysts can help to define what constitutes optimal care along coordinated care pathways. This information can then be used by NHS commissioners to reduce the variability in current pathways of care, thereby helping to ensure equal access to the best treatment options across the country.

The North East London Cancer Network is an example of how the care.data dataset could be used to expand a local example of good practice into a service that could improve the quality, efficiency, and equity of care across the whole of England. By providing access to an individual-level, pseudonymised national dataset of linked GP and hospital data, researchers will be able to spend less time collecting and collating data and more time investigating patterns and trends, such as the time lags between diagnosis and treatment for a whole range of cancers and other diseases, together with any unwarranted variation in these lags across the country and between different patient groups.

Put simply, this means that the HSCIC would:

- link GP data to hospital data,
- then remove all identifiers and then provide these linked data to researchers in a way that does not identify individuals.

The researchers could then use the data to calculate the time a patient presented with symptoms to the time they were treated, and the subsequent outcome of the treatment. An example: if patients in Bristol experience a 12 week delay from presentation to treatment compared with a 3 week delay for patients in Bath, then further research can take place to understand why there is a difference. What is the NHS in Bath doing differently from Bristol? Where are the delays occurring on the patient's journey? Are patients in Bristol having more appointments? What can be learnt from Bath and applied to other areas? It is not necessary to know the identities of patients receiving care in Bristol or Bath to conduct this type of research, but it

is necessary to combine the information from the GP practice and the hospital to look at these types of scenarios in detail. The lessons learnt could then be used to improve the delivery of services and care across the country.

Example 2 - using data to identify patients most at risk

There is a clear benefit to patients and the public in ensuring that expert analysts can use data to develop models, which can identify patients that are most at risk, for example of unplanned hospital admissions, or patients who would most benefit from a particular treatment. Having more linked data available means that the accuracy of the models are improved.

What does this means for patients?

Where predictive modelling is accurate it prioritises care for those who most need it. This results in better care for patients and a more efficient health and social care service.

Example 3 - Public Health England

Public Health England (PHE) recently launched the world's largest single database of cancer patients. Containing clinical information on all 350,000 cancers diagnosed each year, this database will deliver near real-time cancer data to PHE analysts and epidemiologists. The dataset also offers exciting research opportunities (e.g. by allowing genome sequencing data to be linked to clinical data).

The registry currently receives data from all NHS Trusts. However, it is currently very difficult for PHE analysts to access national linked GP and HES data. They believe allowing them ready access to such linked data would be extremely beneficial, not least in understanding how best to improve cancer services across primary and secondary care.

The more information available to analysts and researchers about individual, pseudonymous, cancer patients, the greater their understanding of the disease epidemiology, treatment pathways, and outcomes. Therefore, nationwide linked GP- HES data is an extremely valuable addition to the PHE dataset.

What does this mean for patients?

Using the linked cancer dataset, researchers will be able to examine in detail the experiences of different sub-groups of cancer patients, for example to determine what treatments worked best. This information will then be used to inform the care of similar patients in future. The ultimate aim is to develop so-called stratified medicine (i.e., personalised treatment pathways for cancer patients based on their particular type of genetic mutation and other characteristics and preferences). In simple terms, this research could allow treatments to be more tailored to an individual (i.e. rather than medication for all patients with breast cancer, it would be medication for a subgroup of breast cancer patients, resulting in more effective treatment of the patient). In summary, this approach could deliver:

- increases in innovation and new treatment options,
- improved efficiency and better outcomes for patients
- increased patient understanding and choice of the different treatment options available to them.

Appendix C: Definition of Terms

This document uses a variety of terms of particular relevance to privacy, and which could be open to more than one interpretation. To avoid the risk of misinterpretation, the table below contains a set of definitions. Wherever possible, it relies on existing published definitions, and in particular those in the Data Protection Act 1998, in Regulation 5 of the Health Service (Control of Patient Information) Regulations 2002 (commonly known as 'section 251 approval'), and in Confidentiality: NHS Code of Practice 2003. Where a definition is a partial extract from a lengthy published definition, the convention "..." is used below to denote this fact.

Term	Definition (or extract from full published definition)	Source
Aggregate data	Data derived from records about more than one person, and expressed in summary form, such as statistical tables.	Anonymisation Standard for Publishing Health and Social Care Data Specification.
Anonymisation	Any processing that minimises the likelihood that a data set will identify individuals. A wide variety of anonymisation techniques can be used; some examples of such processing are explained in this specification. Also commonly referred to as "de-identification".	Anonymisation Standard for Publishing Health and Social Care Data Specification
Clinical Audit	The evaluation of clinical performance against standards or through comparative analysis, with the aim of informing the management of services.	Confidentiality: NHS Code of Practice[9]
Confidential patient information	"....patient information is "confidential patient information" where— (a) the identity of the individual in question is ascertainable— (i) from that information, or (ii) from that information and other information which is in the possession of, or is likely to come into the possession of, the person processing that information, and (b) that information was obtained or generated by a person who, in the circumstances, owed an obligation of confidence to that individual."	Regulation 5 of the Health Service (Control of Patient Information) Regulations 2002 (commonly known as 'section 251 approval')
De-identifying data	Any processing that minimises the likelihood that a data set will identify individuals. A wide variety of anonymisation techniques can be used; some examples of such processing are explained in this specification. Also commonly referred to as "anonymisation".	Anonymisation Standard for Publishing Health and Social Care Data Specification

Explicit consent	"This means articulated patient agreement. A clear and voluntary indication of preference or choice, usually given orally or in writing and freely given in circumstances where the available options and the consequences have been made clear."	Confidentiality: NHS Code of Practice
Identifiable information	A set of information from which a person (or persons) can be identified. Identifiable information is confidential, and so the definition for confidential patient information above also applies. Identifiable information can take a variety of forms, such as full patient records, extracts from records, and information not typically considered a record such as labelled laboratory samples.	Confidentiality: NHS Code of Practice
Information governance	Information Governance is to do with the way organisations 'process' or handle information. It covers personal information, i.e. that relating to patients/service users and employees, and corporate information, e.g. financial and accounting records.	Information Governance Framework Standard
Patient identifiable data	Key identifiable information includes: • patient's name, address, full post code, date of birth; • pictures, photographs, videos, audio-tapes or other images of patients; • NHS number and local patient identifiable codes; • anything else that may be used to identify a patient directly or indirectly. For example, rare diseases, drug treatments or statistical analyses which have very small numbers within a small population may allow individuals to be identified.	Confidentiality: NHS Code of Practice
Personal data	"Data which relate to a living individual who can be identified:- - from those data; or - from those data and other information which is in the possession of, or is likely to come into the possession of, the data controller and includes any expression of opinion about the individual and any indication of the intentions of the data controller or any other person in respect of the individual".	Data Protection Act

Processing	"Processing, in relation to information or data, means obtaining, recording or holding the information or data (which includes, in relation to personal data, obtaining or recording the information to be contained in the data) or carrying out any operation or set of operations on the information or data, including: • organisation, adaptation or alteration of the information or data; • retrieval, consultation or use of the information or data (which, in relation to personal data, includes using the information contained in the data); • disclosure of the information or data (which, in relation to personal data, includes disclosing the information contained in the data) by transmission, dissemination or otherwise making available, or • alignment, combination, blocking, erasure or destruction of the information or data." Note that a very similar definition for "processing" is used within the NHS Act 2006.	Data Protection Act
Pseudonym-isation	A technique that replaces identifiers with a pseudonym that uniquely identifies a person. In practice, pseudonymisation is typically combined with other anonymisation techniques.	Anonymisation Standard for Publishing Health and Social Care Data Specification
Public interest	"Exceptional circumstances that justify overruling the right of an individual to confidentiality in order to serve a broader societal interest. Decisions about the public interest are complex and must take account of both the potential harm that disclosure may cause and the interest of society in the continued provision of confidential health services."	Confidentiality: NHS Code of Practice
Safe haven	A bounded secure environment suitable for the receipt, storage, transmission and other processing of any confidential information, including the most sensitive personal information. It may be a physical space (such as a secure room), a configuration of electronic devices or a combination of the two, where secure processes are enforced.	

Regulation 5	Refers to Regulation 5 of the Health Service (Control of Patient Information]) Regulations 2002 (commonly known as 'section 251'). It provides the power to ensure that patient-identifiable information needed to support essential NHS activity can be used without the consent of patients. The power can only be used to support medical purposes that are in the interests of patients or the wider public, where consent is not a practicable alternative and where anonymised information will not suffice.	NHS Information Governance – Guidance on Legal and Professional Obligations - DH/Digital Information Policy Sept. 2007
Sensitive personal data	The Act defines categories of sensitive personal data, namely, "personal data consisting of information as to:- (a) the racial or ethnic origin of the data subject; (b) his political opinions; (c) his religious beliefs or other beliefs of a similar nature; (d) whether he is a member of a trade union (within the meaning of the Trade Union and Labour Relations (Consolidation) Act 1992); (e) his physical or mental health or condition; (f) his sexual life; (g) the commission or alleged commission by him of any offence; or (h) any proceedings for any offence committed or alleged to have been committed by him, the disposal of such proceedings or the sentence of any court in such proceedings."	Data Protection Act