Pseudonymisation Implementation Project (PIP)

Reference Paper 2

Guidance on Business Processes and New Safe Havens

Final v1.0 - 20 November 2009

Guidance on Business Processes and New Safe Havens			
Programme	NPFIT	Document Record ID Key	
Sub-Prog / Project	Pseudonymisation Implementation Project (PIP)	**NPFIT-FNT-TO-BPR-0024.01**	
Prog. Director	J Thorp	Version	01
Owner	.	Status	Final
Author	Wally Gowing	Version Date	20 November 2009

Document Status:

This is a controlled document.
Whilst this document may be printed, the electronic version maintained in FileCM is the controlled copy. Any printed copies of the document are not controlled.

Related Documents:

These documents will provide additional information.

Ref	Doc Reference Number	Title	Version
1	NPFIT-FNT-TO-BPR-0022.01	PIP Implementation Guidance	FV1
2	NPFIT-FNT-TO-BPR-0023.01	Reference Paper 1 - Terminology[1]	FV1
	NPFIT-FNT-TO-BPR-0024.01	Reference Paper 2 – Business Processes and New Safe Havens[2]	FV1
3	NPFIT-FNT-TO-BPR-0025.01	Reference Paper 3 – De-identification[3]	FV1
4	TBA	Reference Paper 4 – Technical White Paper[4]	FV1
5	dh_4069254	NHS Code of Practice on Confidentiality[5]	
6	NA	PIP Planning Template and Guidance[6]	

[1] **http://www.connectingforhealth.nhs.uk/systemsandservices/sus/delivery/pseudo**
[2] **http://www.connectingforhealth.nhs.uk/systemsandservices/sus/delivery/pseudo**
[3] **http://www.connectingforhealth.nhs.uk/systemsandservices/sus/delivery/pseudo**
[4] **http://www.connectingforhealth.nhs.uk/systemsandservices/sus/delivery/pseudo**
[5] [5]

www.dh.gov.uk/en/Managingyourorganisation/Informationpolicy/Patientconfidentialityandcaldicottguardians/DH_4100550
[6] **http://www.connectingforhealth.nhs.uk/systemsandservices/sus/delivery/pseudo**

Contents

Final v1.0 20 November 2009

Guidance on New Safe Haven and Business Processes

1 Introduction

1.1 Purpose and scope

This paper is one of a set of documents produced as part of the Pseudonymisation Implementation Project (PIP) and provides a reference document for local organisations implementing NHS wide guidance on local NHS data usage and governance for secondary uses.

The purpose of this paper is to:

- Set out those factors, which have been taken into account in the development of guidance on the implementation of new safe havens and the modification of local business processes enable the guidance to be justified
- Provide informatics staff with the background and rationale for proposed new safe haven regimes by acting as reference document
- Provide specific guidance for users on the basis of their organisation and local circumstances.

The scope of this paper is limited the new safe haven and business process aspects of the implementation to support secondary uses.

1.2 Related papers

This guidance builds on the earlier PIP documents

- PIP Implementation Planning Guidance
- PIP Maturity Model

Other supporting reference documents are:

- Reference Document 1 – PIP Terminology, Ref 2
- Reference Document 3 – De-identification, Ref 3
- Reference Document 4 – Techniques, Ref 4

1.3 Background and context for implementing Local Data Usage and Governance

The background and context for implementing local data usage and governance for secondary uses is covered in depth in Sections 1 and 2 of the Implementation Guidance on Local NHS Data Usage and Governance for Secondary Uses. This should be consulted to provide the overall context for New Safe Havens and impact on NHS business processes.

Of particular significance is the requirement for a secure environment to exist in implementing New Safe Havens and modifying business processes.

1.4 Business process guidance - Work in Progress

The guidance on business processes has been developed from applying principles and pragmatism to NHS business processes. Whilst this guidance reflects the knowledge and experience of the PIP Team at the middle of November 2009, it is not feasible to provide exhaustive guidance on the modification of all NHS business processes in one pass. This is because there are too many local variations in operational and organisational arrangements. Therefore the guidance relating to business processes should be regarded as work in progress and will be extended via the SUS websites as issues are raised and as more knowledge is gained.

1.5 Patient label

Throughout this and the related papers, the term ‘patient label’ is used to describe the data item(s) that distinguishes one patient from another in a set of data; please note that this is purely for the purposes of clarity within the papers.

For identifiable data, the patient label may be the NHS Number (if present) but could be within an organisation, its Local Patient Identifier; another data item, or combination of data items, that can be used to uniquely identify one patient from another; in de-identified data sets, patient labels will vary but, for example, can be pseudonyms or table row numbers (the latter may not be unique as activity relating to the same patient may appear more than once in a table.)

2 Secondary Uses

2.1 Secondary Uses Definition

A high level definition of secondary uses is developed in the PIP Reference Paper 1 Guidance on Terminology. Secondary uses equates to non-healthcare medical purposes use within medical purposes as set out in Confidentiality: the NHS Code of Practice, Ref 6. In effect, secondary use of patient data is the use for purposes that do not directly contribute to the safe care of the individual concerned.

Purposes that directly contribute to the safe care of the patient are classified as primary uses and include care, diagnosis, referral and treatment processes together with relevant supporting administrative processes, such as clinical letters and patient administration, patient management on a ward, managing appointments for car; as well as the audit/assurance of the quality of the healthcare provided.

Confidentiality also clearly states that use of patient data for non-healthcare medical purposes must be 'effectively anonymised', that is in de-identified form unless it is with the patient's consent or otherwise covered in law, such as with approval under Section 251 of the 2006 NHS Act given by the National Information Governance Board (NIGB) Ethics and Confidentiality Committee (ECC).

It is necessary to distinguish between the two types of use in order to determine what data a user can see. Examples of secondary use of patient data are performance management, commissioning, contract monitoring; all of which do not require the identity of patients. There are functions within the NHS that use the same data sources for both secondary and primary uses. An example at PCT level would be for performance monitoring of Referral to Treatment (RTT) which should use de-identified data, but for organising care provision within 18 weeks, access to identifiable data is a primary use and therefore permissible by a suitably authorised member of staff.

2.2 Implications

There are implications that emerge from applying the definition of secondary use. These include:

- When a patient record is used for a primary purpose, the user needs to know who the patient is. All primary purposes require identifiable patient information to ensure patient safety
- Identifiers must be removed from data before it is used for additional or secondary purposes.
- Processing and transformation of patient records to enable data to be transitioned into a suitable state for a secondary use is considered as "an additional purpose" and in itself a secondary use.

The processing involved in the transformation of data into an acceptable form for secondary use to be made of it includes:

- Data quality (DQ) checks: an example is checking that attributes on the record are accurate such as ensuring the patient is attributed to the correct practice and PCT.
- Undertaking derivations of identifiable data items: an example is deriving age at the start or end of an episode of care, which relies on the patient's date of birth.
- Undertaking record linkage: an example is to link records from different datasets or over time, usually based on NHS Number and cross checking dates of birth for corroboration.
- Applying de-identification processes, such as pseudonymisation.

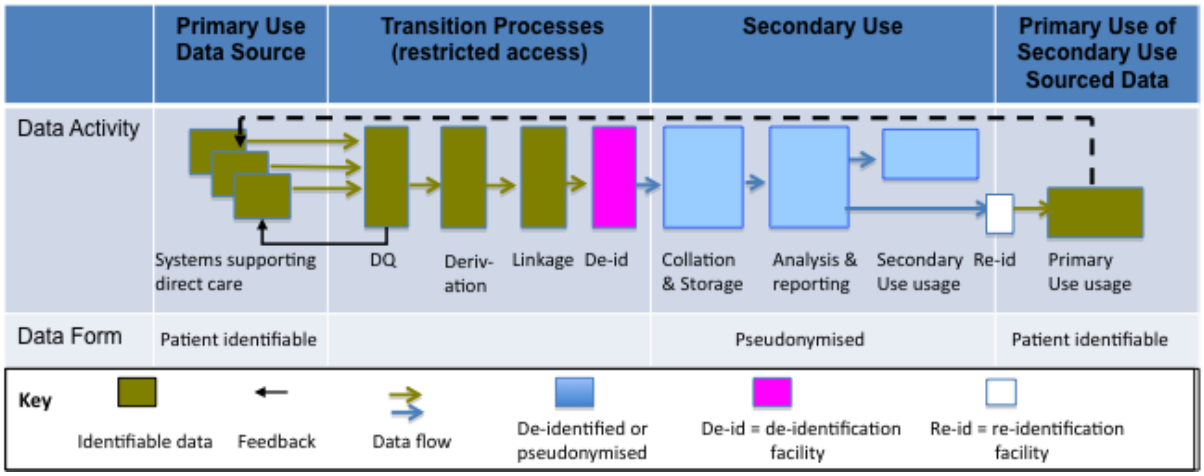
The DQ work requires access to identifiable data by appropriate nominated authorised staff, whilst the subsequent items are undertaken through computer processing and without disclosure of data. These steps should only require access to identifiable data if something goes wrong with the processes or as a result of inconsistency of data. A range of DQ activities is considered in Section

In order to be able to undertake such processing, these must be permitted under Section 251 (S251) of the 2006 NHS Act to allow the access and use to proceed. This in turn requires the approval of the NIGB Ethics and Confidentiality Committee (ECC) or incorporation in specific S251 regulations. The relevant approval is being pursued by the PIP Team.

2.3 Data Transformation to support Secondary Use

Figure 1 illustrates the flow of data from primary use systems to enable secondary use and subsequent primary use of secondary use sourced data. The steps in the transition processes reflect those described in Section 2.2 and are the ones needed to enable patient level data to be used in de-identified form for secondary use purposes.

Figure 1 Generic Secondary Use Data Flow Model



The data flow from Primary Use Source may originate as, for example:

- Electronic and digital as in a SUS extracted CDS or a direct flow from a provider or a local authority to a commissioner for activity outside that covered by CDS as a dataset or as an email
- Electronic and paper based, such as a fax for out of area treatment
- Paper based, such as an invoice for non-contracted activity.

but will always become computer based within the receiving organisation. The implication of the above is that paper-based communications of data for secondary use purposes, even though potentially between safe havens, should cease in order to be more secure and encrypted. This requires further investigation in order to determine levels of flows, whether it is feasible to cease all paper flows or not, impact of change, practicality of implementation and timescales over which implementation is carried out.

3 New Safe Haven

3.1 Safe haven concept

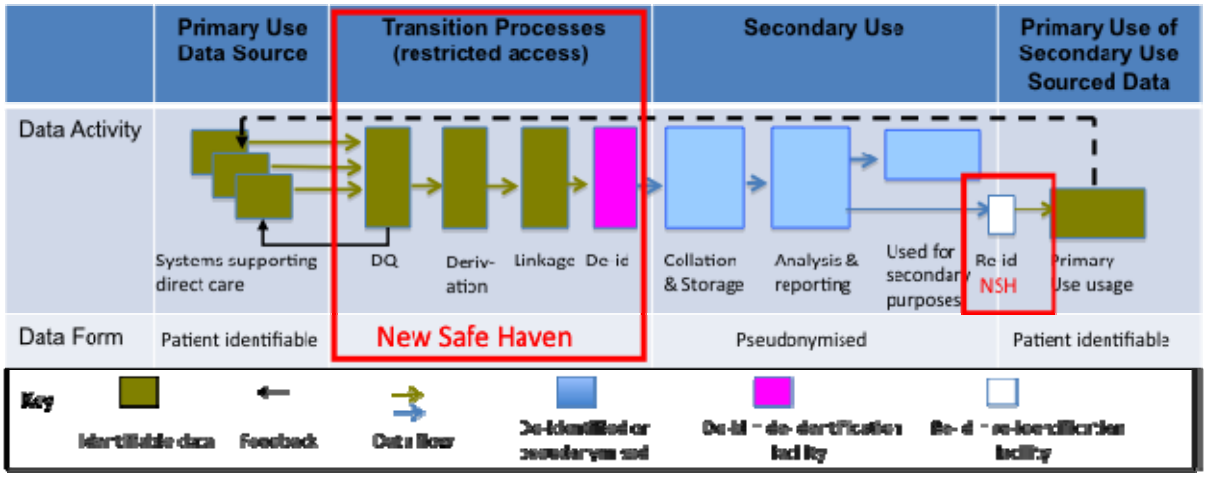
The NHS has used safe havens for over 20 years to ensure the safety and secure handling of confidential patient identifiable information. The first use was to provide security when faxes were used to transmit patient data between providers and purchasers. In that case a physical location of a locked room was used to restrict access to fax machines and hence patient identifiable data.

The same concept of restricting access to identifiable data, albeit in a logical context, is required to support the process that enables de-identified records to be created, hence the term New Safe Haven.

3.2 New Safe Haven for de-identification purposes

The New Safe Haven will exist to provide the means of restricting access to authorised users to identifiable data for the purposes of receiving and sending identifiable data that is expected to be used for secondary purposes and for supporting de-identification of the identifiable data. In **Figure 2,** the New Safe Haven encompasses the activities to support the Transition Processes component and the enabling of the final component, which is making primary use of data.

Figure 2 New Safe Haven



This means that the New Safe Haven comprises the facilities to restrict access by authorised users to identifiable data for the purpose of supporting de-identification, which in turn means that:

- The facilities can only be used by a small number of authorised staff sufficient to perform the functions and provide cover and back-up to ensure continuity of service
- Authorisation of the staff performing roles in the New Safe Haven should be through the Caldicott Guardian and the equivalent of local Registration Authority processes for accessing Spine based applications
- The systems (or sub-systems) used for the data transition processes must have appropriate access control mechanisms to restrict access to authorised users for the specific purpose of supporting de-identification processes.
- The New Safe Haven may have a physical location, but it is only essential in the case of relevant paper based data flows, such as faxes.
- The New Safe Haven can be defined in terms of
    - the activities to be undertaken to support de-identification
    - posts/people authorised to access identifiable data for the purpose of supporting de-identification
    - posts/people authorised to access identifiable data for the purpose of supplying identifiable data to authorised users

- The facilities necessary to support the activities.

The New Safe Haven can also be defined in terms of access control and data management arrangements as these indicate which data can be accessed by what means and by whom.

3.3 New Safe Haven Security

New Safe Haven security must conform to NHS good practice concerning the handling of identifiable data, as predicated by ISO 27001 and 27002 and the CFH Good Practice Guidance. Adherence to relevant good practice is particularly important as the New Safe Haven may be virtual in nature and New Safe Haven staff may be distributed throughout an organisation.

The additional good practice for a New Safe Haven that must apply is:

- Access to safe haven functionality, such as accessing databases for DQ purposes must be restricted to registered, authorised users.
- Access to the safe haven functionality should be at least password controlled by individual user accounts and passwords. Accounts must not be shared between users. Password management must meet CfH good practice requirements for strength of passwords, refresh frequency, etc.
- Only authorised and registered staff should have access to the core storage of identifiable and linked data.
- If paper based flows involving patient identifiable data still occur, then post delivery and equipment (such as fax machines) should be operated in secure areas in order to function effectively for New Safe Haven purposes. Access to such areas should be restricted and equipment should have a code password and be turned off out of office hours.

4 Data Quality

4.1 Overview

The accuracy of patient records is vital for the effective support of provision of services to patients and to support the operation of effective commissioning. It is important for patient safety that when data quality checks are undertaken as part of processing data for a secondary purpose and any errors and discrepancies are corrected and reported back, to sources where primary use is made of the same data.

Data quality checks can be undertaken on receipt of data, but issues around quality of data will arise during use of the data. Issues that are discovered later in data that has been pseudonymised should be handled through the Safe Haven mechanism by referring the issue back to the relevant New Safe Haven.

4.2 DQ of Commissioning Data Sets

Commissioning Data Sets (CDS) are subject to basic data quality checks at the point of submission via the use of XML schemas and at the loading stage in SUS. These DQ checks are concerned with completion of fields and simple logic. One DQ issue raised by commissioners concerns what is perceived as incorrect assignment of patients to practices and responsible PCTs and commissioners.

There is a data feed from Personal Demographic Service (PDS) to SUS so that practice, responsible and residence PCT data for patients will be used from that feed as the derived data items held in SUS – these replace the items previously derived via NACS.

Reports have been developed to indicate where there are discrepancies between the practice, responsible and residence PCT as entered by the provider and as derived from PDS. The aim of these reports is to provide a focus for which activity records need checking for patient attribution, with the intention of reducing the number of records that need to be checked. Despite using data from PDS, there will always be discrepancies due to the time lag as patients move and change practice and PCT.

The ‘discrepancy’ reports will be available to commissioners, PCTs and Providers so that checks can be made at both source and user of the CDS. The reports use the Local Patient Identifier (LPI) as the patient label, which can be used for inter-organisation communication, avoiding the need for use of NHS Number at the PCT/commissioner end Ref 2 Para 2.3. The degree to which these reports will assist DQ work will be governed by their timeliness in relation to the monthly reporting cycles.

4.3 Operational DQ

There is a range of activities that frequently lead to data quality checks being undertaken. These activities, mainly starting at the commissioner/PCT end, include

- Establishing correct attribution of patients for CDS records, i.e. “is this our patient?”
- Responding to invoices for non-contracted activity and out of area treatment, again the issue is: “is this our patient”, but also “is the procedure reasonable?”
- Aiding the development of disease registers - by identifying potential members from other information sources, e.g. CDS (note inclusion of patients onto a disease register can only be undertaken with express consent).
- Cross checking non-CDS flows with CDS records - in order to optimise information about the population in an area meeting specific criteria, e.g. Checking paper records received from the NHS Trust for LDP returns around smoking in pregnancy and breast feeding against the CDS data.
- Pathway analysis – an example involves de-duplicating community hospital admission lists in order to determine actual admission and discharge dates by ignoring multiple discharge and admissions that occur with off-site outpatient appointments.

All such activities are concerned with accuracy of data in relation to the provision of healthcare services and should be undertaken within the New Safe Haven.

5 Business processes involving secondary use of patient data

5.1 Business Process Types

The rationale for the New Safe Haven has been set out in Section 3 and the Safe Haven's role in undertaking data quality activity has been set out in Section 4. It is assumed from here onwards that if DQ matters arise in business processes, then the resulting DQ steps will be undertaken under the auspices of the New Safe Haven. Working on this basis, business processes can be considered in relation to primary and secondary purposes and the use of identifiable and de-identified data.

There are three main types of business process or component steps within an overall business processes that need to be considered. These are examined below:

A - those processes using patient data involved in the direct care of patients;

B - those processes using patient data not involved in the direct care of patients;

C – combination of Types A and B; where one of the purposes of use of patient data may be to support the direct care of patients, such as an intervention by clinicians with using non direct data to identify individual patients.

Type A - Business processes that are involved directly in the care of patients must be undertaken with patient identifiable data in order to ensure accuracy and patient safety. An example of processes that can be undertaken with identifiable data is patient administration, such as booking appointments, managing waiting lists.

Type B - If business processes are not involved directly in the care of patients, then they must be undertaken with de-identified data; if necessary the business process must be modified. Examples of processes that should be undertaken with de-identified data include:

- Analysis of waiting lists, such as numbers waiting by varying time periods.
- Monitoring RTT, such as numbers of patients waiting by weeks.

Findings from the Pseudonymisation Pilots indicated that analyses of sets of data can be undertaken effectively using pseudonymised data. The major problem impacting on the process concerned the attribution of patients to practice or commissioner/PCT. The question of 'is this our patient' is a matter of data quality and the issue should be resolved in the New Safe Haven.

There is also a specific Type B process not involving the direct care of patients for which a piece of identifiable data is required, i.e. spatial analysis of patient records. This is considered separately in Section 5.4.

Type C – combination; here the purpose of the overall business process is to process and analyse patient records in order to understand and report on specific issues, which may subsequently require interaction with patients. A simple example is the use of the PARR algorithm, see Section 6.3.

Types B and C are considered further below with some examples.

5.2 Type B - Processes not involved in the direct care of patients

The Pseudonymisation Pilots provided evidence that analyses of sets of data can be undertaken effectively using pseudonymised data. The major problem impacting on the process concerned the attribution of patients to practice or commissioner/PCT. The question of 'is this our patient' is a matter of data quality and the issue should be resolved in the New Safe Haven.

5.3 Type C – Combination of Types A and B

World Class Commissioning (WCC) involves a greater understanding of the health of local populations and circumstances and the proactive targeting of healthcare services. To achieve this requires bringing together data from disparate sources to provide the basis for linking patients with services.

Risk Stratification is a technique that involves stratifying a group of patients through retrospective analysis of their use of services (activity) and other characteristics. This stratification is used to identify a subset or cohort of the patients that might most benefit from certain interventions and pro-active care.

An example of this technique is the use of the PARR algorithm. In approving the use of PARR, the Patient Information Advisory Group (PIAG) required that the analysis be undertaken with pseudonymised data and the results, requiring the identification of patients, be seen and used by community matron on behalf of a practice.

Using the example above, the cohort selection can be undertaken in one of two ways:

- A clinician can initiate/undertake the analysis themselves using standard analyses and reports on the basis of legitimate relationships with the patients;
- The analysis can be undertaken as a secondary use with de-identified data and for the selected cohort to be made available to the relevant clinicians in identified form.

5.4 Spatial Analysis

The spatial analysis of patient level data is important in identifying and addressing concentrations of health problems, which may be associated with environmental factors or the social and economic circumstances of populations. It is also important in determining the access that patients have to healthcare facilities and services. Robust spatial analysis can only be undertaken if the full postcode of a patient or service user is available. As postcode is an important potential contributor to identification of patients, it is necessary to make data and analytical facilities available to end-users without revealing postcodes of individual patients.

Analysis is based on deriving areas from postcodes. The areas can be standard geographies', either national or local such as Census wards, electoral wards, neighbourhood renewal areas, SRB areas, community clusters or non-standard local geographies, such as district nursing areas and social service team areas. There is a need to ensure that the full range of analyses, such as point and cluster analyses, can be undertaken in order to support activities, such as analysing hospital use in relation to cancers by co-ordinate points to see what correlation they have to flight paths into and out of a major airport.

Analysis should therefore be undertaken either:

- On area derived from full postcodes and files provided to end–users. File content suitable for end-user access to undertake the spatial analyses would be pseudonymised patient level records with relevant subject matter (e.g. patients filtered by condition) and derived areas, but no postcodes or
- Through analysis being undertaken within the New Safe Haven using patient data with clear postcodes. Output should be provided in suitable mapped plots and if patient level data is required by end-users, then it should be provided in pseudonymised form with modified postcodes, such as post-code sector or blurred forms.

Care must be taken with non-standard geographies that they are not defined in such a way as to provide information that would lead to identification of individuals.

5.5 Inter-organisational communications

Communication will need to take place between organisations when patient activity data is being used for a secondary business purpose. For communications between commissioners and providers, commissioners will have data in a locally pseudonymised form and the New Safe Haven of the originating provider will have data in identifiable form. Communications should be pursued as follows:

- If the issues concern aspects of the activity, such as episode start and end dates, then the label used to describe the records concerned should be those used to describe the activity, such as hospital spell number or generated record identity, and not those based on the patient.
- If the issues concern aspects about a patient, such as incorrect attribution to a practice and PCT, then and only then should a patient label be used. A patient label sufficiently meaningful to both organisations to support the communication should be used, for example the provider's Local Patient Identifier (LPI).
- If the communication is between New Safe Havens or a New Safe Haven and a service supporting direct care, then it is appropriate to use identifiable data, such as NHS Number and date of birth as patient labels if they are available.

A particular implication of the above concerns support for billing processes between NHS organisations. Details of the activity undertaken by a provider on a patient will be contained in CDS through SUS or a locally agreed MDS flow between the provider and the commissioner of the activity. These flows should be between the organisations' new safe havens and are expected to contain patient identifiable data, namely NHS Number, date of birth and postcode.

There may also be invoices sent from the provider to the commissioner for the activity undertaken and referencing individual items of activity. Such invoices must not contain the generally accessible patient identifiable data items, namely NHS Number, date of birth and postcode, as the invoice is about the activity undertaken and about the patient. As indicated above, activity should be identified by data items, such as hospital spell number. Invoices may contain less accessible identifiable data items, such as the LPI if the commissioner does not have access to the providers' LPI (which can happen in PCTs which have not separated commissioner and provider functions).

6 Organisational Implications of using de-identified data in Business Processes

6.1 Review business processes

In order to assess the impact of de-identification on business processes, all business processes should be reviewed and considered as below:

1 Undertake data quality and other transition processes through the New Safe Haven – this should be common to all business processes involving patient level data. Therefore if there are data quality steps as outlined earlier, such as in the OATS business process, then the DQ steps should be transferred to the auspices of the New Safe Haven.

2 If the process does not involve a secondary use (Type A), proceed with identifiable data.

3 If the process only involves secondary uses (Type B), then proceed with de-identified data.

4 If the process involves spatial analysis (Type B spatial analysis), then proceed as set out in Section 5.4.

5 If the outcome of the process involves secondary uses followed by primary uses, i.e. identifying patients to clinicians and those with legitimate relationships, then the process should be separated into its secondary and primary components. This should be undertaken either sequentially using de-pseudonymisation prior to the primary use or by enabling direct access to the identifiable data by a relevant clinician, as indicated in the example in Section 5.3.

6.2 Business Processes Guidance

A wide range of business processes has been raised with the PIP Team in relation to the potential impact of using de-identified data. Guidance has been developed for some issues through discussion with individual sites and the PIP Advisory Group (a representative group of staff from NHS provider and PCT organisations). At this stage, not all issues have been worked on to provide guidance, as the guidance needs to be checked with a wider range of potential users to ensure that it is practical.

Guidance for major issues for Commissioners and Providers is set out separately below. An appendix to this paper contains this information, together with guidance on generic issues. The information in the Appendix will be published as a separate document on the SUS websites and be updated as new issues arise and further guidance material is developed for use.

6.3 Commissioner Business Processes

Guidance has been developed for some of the major issues raised in relation to commissioner business processes and this is set out in **Table 1.**

Table 1 Commissioner Business Process Guidance

Issue	Guidance
Practices need identifiable patient data from CDS for PBC	Data should be supplied to practices in identifiable form other than where there is a legal requirement to fully anonymise respecting or where the patient has indicated dissent. Where PCT staff provide direct support to practices in implementing PBC and require access to identifiable data, these staff should be designated as having that responsibility and should access relevant data through suitable auditable access control functionality
Practices need identifiable data from CDS for Active Case Management (ACM),	Data should be supplied to associated clinicians in identifiable form, respecting guidance on sensitive data in relation to STD type data and patient dissent. Where PCT staff provide direct support to practices in

Issue	Guidance
as GPs identify patients suitable for Active Case Management	implementing ACM then these staff should be designated as having that responsibility and should access relevant data through suitable auditable access control functionality
World Class Commissioning (WCC) requires identifiable data	WCC requires linked data about patients and the selection of patients whom it may be appropriate to clinically intervene, in relation to particular conditions or circumstances to be identified. The record linkage should be undertaken within a New Safe Haven; records should be provided in pseudonymised form for analysis; the resulting patient cohort should be provided in identifiable form to relevant clinicians.
Out of Area Treatments (OATs) invoices are for named individuals and Finance need access to identifiable data	OATs invoices should be received in the Commissioner's New Safe Haven (NSH). The NSH should verify that the patient subject of the invoice 'belongs' to the commissioner and a de-identified version of the invoice should be provided to the Finance Department to enable financial processes to proceed.
Cost per case issues are usually raised and pursued by Finance Departments using identifiable data	Determining that there are cost per case issues does not rely on the identity of the patient; the cases can be picked out by Finance staff from pseudonymised patient labels and then be referred to the commissioner New Safe Haven for pursuit on data quality issues (e.g. is the pricing correct, is it the correct procedure and length of stay) via the provider New Safe Haven in order to identify the cause of genuinely high cost cases.
Dual roles of staff within departments that have legitimate need to access de-pseudonymised data as well as pseudonymised data.	The business processes should be reviewed to ensure that access to identifiable data is necessary in any roles. Where possible, work should be reorganised so that there are some staff that undertake activities with identifiable data only and some with de-identified data only. If staff still have roles that require access to both identifiable and pseudonymised data, then the roles of the staff should be split into two (i.e. those involving identifiable data separated from those accessing pseudonymised data) and system accesses set up appropriately in order that access to identifiable data can be logged and audited. However, this may not be practical in all situations; if this is the case, then the access by relevant staff to both identifiable and pseudonymised data should be signed off by the Caldicott Guardian.
Shared services - Use of a single pseudonym and safe haven process per organisation where 2 or more PCTs share an informatics service and data warehouse. Is a shared pseudonym and safe haven across the 2 or more PCTs permissible?	The use of the same pseudonymisation algorithm and key or seed for 2 or more PCTs is permissible. The patients for which the two PCTs are responsible will be different and therefore have different NHS Numbers for which different pseudonyms will be generated, meaning that there will be no confusion on pseudonyms. Some patients may appear in two neighbouring PCTs as resident and responsible populations will overlap. Just as those patients' NHS Numbers and records will appear in the activity data for both PCTs, so will their pseudonyms. (See PIP Reference Paper 3 on De-identification on implementation of pseudonyms in Shared Services) Sharing a single New Safe Haven (NSH) needs the relevant Caldicott Guardians to sign off the handling of identifiable data by the Shared Service NSH on behalf of the PCTs and requires the PCTs to ensure their Data Protection Act registration reflects that the NSH is acting as a Data Processor on their behalf.

Outstanding Issues

At the date of publication of this version of the guidance there are a number of issues which have been raised for which guidance has not yet been produced. These topics, listed below, will be initially addressed via the publication of Frequently Asked Questions (FAQs) and will then be incorporated into subsequent versions of this guidance document. Topics still under review are:

- Contracting – upcoding, such as hypertension and diabetes
- Screening
- PbC referral review and audit
- PMS practices – this is the subject of work with a PCT in order to develop guidance
- Small numbers in reports
- Links to Research – this subject is outside the scope of the current guidance.

6.4 Provider Business Processes –

Guidance has been developed for some of the major issues raised in relation to provider business processes and this is set out in **Table 2.**

Table 2 Provider Business Process

Issue	Guidance
OATS Notification and Invoicing	OATs documentation should not be sent by paper and fax; it should be electronic and encrypted, can be emailed (via NHS secure email) and should be sent to the relevant commissioner's New Safe Haven. This suggests that an NHS wide directory of addresses of new safe havens would be helpful; feedback on this would be useful.
Dual roles of staff within departments that have legitimate need to access de-pseudonymised data as well as pseudonymised data.	The business processes should be reviewed to ensure that access to identifiable data is necessary in all roles. Consideration should be given to re-organising the work so that there are some staff that undertake activities with identifiable data only and some with de-identified data only. If staff still have roles that require access to both identifiable and pseudonymised data, then the preferred solution is that the roles of the staff should be split into two (i.e. those involving identifiable data separated from those accessing pseudonymised data) and system accesses set up appropriately in order that access to identifiable data can be logged and audited. However, this may not be practical in all situations; if this is the case, then the access by relevant staff to both identifiable and pseudonymised data should be signed off by the Caldicott Guardian
Clinical team members who make primary and secondary uses of identifiable data in stand alone systems	This is a special case of the 'dual roles' case referenced above, especially in relation to stand-alone systems. Where clinicians have active interactions and legitimate relationships with the patients concerned, the clinicians are likely to be aware of the identities of patients after de-identification has taken place. In such cases the requirement for de-identification becomes an artificial device and counter productive to the business process and should not be implemented. However, access should still be subject to suitable auditable access control functionality.

Output from the PAS system is used for secondary purposes, but it is not feasible to pseudonymise the output from the PAS system	A solution is to use middleware to transform reports that are not needed in identifiable form into pseudonymised versions**.**
Waiting Lists	Waiting lists are an example of the same data being used for different purposes. In ensuring the delivery of services for individual patients, a primary use is being made of personal data. However, the management of waiting lists – how many people waiting for services, where and when, etc - is a secondary use and does not need identifiable data. The analysis of a problem overlaps with RTT, where an analyst may be involved in determining the cause and resolution of a problem with de-identified data and the decision to contact individual patients and making that contact should lie with relevant case managers making primary use of identifiable data.
Clinical audit	Only local clinical audit following the care pathway or quality assuring the care given may use identifiable data. National / regional comparative audits must use either pseudonymised data or have their own Section 251 support.

Appendix - Business Process Guidance

Business Area	Business example/Guidance request	Guidance
Generic issues		
Spatial / geographic analysis		
PH/PBC/WCC	Aggregating data by postcode to enable geographical analysis. NON-STANDARD GEOGRAPHIES e.g. district nursing areas, social service team areas, police beats. In these cases it is not possible to identify the geography other than from full postcode as they do not follow administrative boundaries or other tiers of standard geography	Analysis should be undertaken either On area derived from full postcodes and files provided to end –users. File content suitable for end-user access to undertake the spatial analyses would be pseudonymised patient level records with relevant subject matter (e.g. patients filtered by condition) and derived areas, but no postcodes. Through analysis being undertaken within the New Safe Haven using patient data with clear postcodes. Output should be provided in suitable mapped form plots, and if patient level data is required by end-users, then it should be provided in pseudonymised form with modified postcodes, such as post-code sector or blurred forms
PH/PBC/WCC	Aggregating data by postcode to enable geographical analysis, such as POINT ANALYSIS or PLACE ANALYSIS Full unit postcodes are used to assign data to a geographic co-ordinate that is then mapped and analysed to see where clusters occur. For example analysis of hospital use in relation to cancers by co-ordinate points to see what correlation they have to flight paths into and out of a major airport.	
	Aggregating data by postcode to enable geographical analysis. STANDARD GEOGRAPHIES. Data is aggregated to standard geographies (OAs, LSOAs, MSOAs, Census wards, electoral wards, neighbourhood renewal areas, SRB areas, community clusters etc.) to enable mapping analysis. Some of these geographies are nationally standard others are a local standard (e.g. across a city partnership).	

Commissioner Issues		
Confirming a patient is assigned to the correct PCT		
Activity outside contracts	Out of Area activity – how to handle Non-contracted activity (NCA) and Out-of-Area (OOA) or Out of Area Treatment (OATs)? Patients who receive treatment at distant hospital are cross-charged to their responsible PCT by an end-of-month invoicing process. Since there is no contract in place, what may happen is that the Hospital e-mails or faxes a list of NHS numbers to the responsible PCT (as they see it), summarising clinical information, like Specialty Code, and a cost. First thing the accounts staff at the receiving PCT do is check that the patient is indeed one of theirs - in one PCT they use the Summary Care Record system.	These communications are significant as the process does not fit in with the usual type of electronic data flows (i.e. paper, fax, email etc) in being codified and structured, but will need to fit in with the local data usage and governance regime. The OOA , OATs and NCA need to be checked in order to confirm the patients in terms of their practice/PCT. This is a data quality issue and should be handled within the New Safe Haven, and it should use NHS Number in communication in order to ensure the correct patient/practice/PCT.
Supporting primary use of secondary use data (e.g. linking SUS data with other data sources)		
PH/PBC/WCC	Linking for intervention activities	Undertake within New Safe Haven using identifiable data
PBC	Provide patient level data to GP Practices. This allows GPs to identify patients suitable for Active Case Management, review whether patients need to attend multiple outpatient follow up appointments in secondary care.	Data should be supplied in identifiable form, respecting guidance on sensitive data in relation to STD type data and patient dissent. Where PBC does require access to identifiable data by PCT based staff, then these staff should be designated as having that responsibility and should access relevant data through suitable auditable access control functionality
Active Case Management	Practices need identifiable data from CDS for Active Case Management (ACM), as GPs identify patients suitable for Active Case Management	Data should be supplied to associated clinicians in identifiable form, respecting guidance on sensitive data in relation to STD type data and patient dissent. Where PCT staff provide direct support to practices in implementing ACM then these staff should be designated as having that responsibility and should access relevant data through suitable auditable access control functionality
PH/WCC	WCC requires identifiable data as the PCT have a number of requirements and commitments to research and audit certain events, diseases and care pathways.	WCC requires linked data about patients and the selection of patients whom it may be appropriate to clinically intervene, in relation to particular conditions or circumstances to be identified. The record

		linkage should be undertaken within a New Safe Haven; records should be provided in pseudonymised form for analysis; the resulting patient cohort should be provided in identifiable form to relevant clinicians
Supporting data quality issues with other systems (e.g. linking records to find the data on SUS)		
PH/PBC/WCC	Developing disease registers (e.g. to form a basis of a health screening service). GP registration and other system information are the main source of this data. However inpatient data can be used as a further check on the data quality, e.g. for diabetic retinopathy GP data is used to form the basis of the register but by identifying inpatients with diabetes from the CDS it is possible to cross reference and maximise completeness	Undertake necessary actions within New Safe Haven
Active Case Management	Practices need identifiable data from CDS for Active Case Management (ACM), as GPs identify patients suitable for Active Case Management	Data should be supplied to associated clinicians in identifiable form, respecting guidance on sensitive data in relation to STD type data and patient dissent. Where PCT staff provide direct support to practices in implementing ACM then these staff should be designated as having that responsibility and should access relevant data through suitable auditable access control functionality
PH/PBC/WCC	Checking paper records received from the NHS Trust for our LDP returns around smoking in pregnancy and breast feeding against the CDS data. We use CDS as a quality check on this, identifying new mothers on the snapshot to form our expected population and comparing this against the NHS numbers on the returned forms. We can then provide an estimate of the recording rate (necessary as part of the LDP return) and also provide the NHS Trust with details of mothers for whom no forms have been received. The NHS Trust can then examine the patient notes to find the form and return it ensuring data is as complete as possible.	Undertake necessary actions within New Safe Havens
WCC/Risk Stratification	Use of PARR++ system, which uses practice based data as well as CDS data	Undertake linkage of data within New Safe Haven; undertake analysis in pseudonymised form and supply to relevant authorised clinical end user in identifiable form

Provider Issues		
Out of Area Treatment invoicing	OATS Notification and Invoicing	OATs documentation should not be sent by paper and fax; it should be electronic and encrypted, can be emailed and should be sent to the relevant commissioner's New Safe Haven
Examining outliers such as high cost cases		
Cost per case	Cost per case issues are usually raised and pursued by Finance Departments using identifiable data	Determining that there are cost per case issues does not rely on the identity of the patient; the cases can be picked out by Finance staff from pseudonymised patient labels and then be referred to the commissioner New Safe Haven for pursuit on data quality issues (e.g. is the pricing correct, is it the correct procedure and length of stay) via the provider New Safe Haven in order to identify the cause of genuinely high cost cases.
Organisational and staffing issues		
Staffing	Dual roles of staff within departments that have legitimate need to access de-pseudonymised data as well as pseudonymised data.	The business processes should be reviewed to ensure that access to identifiable data is necessary in any roles. Where possible, work should be reorganised so that there are some staff that undertake activities with identifiable data only and some with de-identified data only. If staff still have roles that require access to both identifiable and pseudonymised data, then the roles of the staff should be split into two (i.e. those involving identifiable data separated from those accessing pseudonymised data) and system accesses set up appropriately in order that access to identifiable data can be logged and audited. However, this may not be practical in all situations; if this is the case, then the access by relevant staff to both identifiable and pseudonymised data should be signed off by the Caldicott Guardian.
Shared Services	Shared services - Use of a single pseudonym and safe haven process per organisation where 2 or more PCTs share an informatics service and data warehouse. Is a shared pseudonym and safe haven across the 2 or more PCTs permissible?	The use of the same pseudonymisation algorithm and key or seed for 2 or more PCTs is permissible. The patients for which the two PCTs are responsible will be different and therefore have different NHS Numbers for which different pseudonyms will be generated, meaning that there will be no confusion on pseudonyms. Some patients may appear in two neighbouring PCTs as resident and responsible populations will overlap. Just as those patients' NHS Numbers and records will appear

		in the activity data for both PCTs, so will their pseudonyms. (See PIP Reference Paper 3 on De-identification on implementation of pseudonyms in Shared Services) Sharing a single New Safe Haven (NSH) needs the relevant Caldicott Guardians to sign off the handling of identifiable data by the Shared Service NSH on behalf of the PCTs and requires the PCTs to ensure their Data Protection Act registration reflects that the NSH is acting as a Data Processor on their behalf.
Provider Issues		
Out of Area Treatment invoicing	OATS Notification and Invoicing	OATs documentation should not be sent by paper and fax; it should be electronic and encrypted, can be emailed and should be sent to the relevant commissioner's New Safe Haven
Staff roles	Dual roles of staff within departments that have legitimate need to access de-pseudonymised data as well as pseudonymised data.	The business processes should be reviewed to ensure that access to identifiable data is necessary in all roles. Consideration should be given to re-organising the work so that there are some staff that undertake activities with identifiable data only and some with de-identified data only. If staff still have roles that require access to both identifiable and pseudonymised data, then the preferred solution is that the roles of the staff should be split into two (i.e. those involving identifiable data separated from those accessing pseudonymised data) and system accesses set up appropriately in order that access to identifiable data can be logged and audited. However, this may not be practical in all situations; if this is the case, then the access by relevant staff to both identifiable and pseudonymised data should be signed off by the Caldicott Guardian
Clinical Teams	Clinical team members who make primary and secondary uses of identifiable data in stand alone systems	This is a special case of the 'dual roles' case referenced above, especially in relation to stand-alone systems. Where clinicians have active interactions and legitimate relationships with the patients concerned, the clinicians are likely to be aware of the identities of patients after de-identification has taken place. In such cases the requirement for de-identification becomes an artificial device and counter productive to the business process and should not be

		implemented. However, access should still be subject to suitable auditable access control functionality.
PAS systems	Output from the PAS system is used for secondary purposes, but it is not feasible to pseudonymise the output from the PAS system	A solution is to use middleware to transform reports that are not needed in identifiable form into pseudonymised versions. If it is not feasible to create such a suitable output file for the middleware, then the PAS system needs to be treated as a Legacy System.
Multiple purpose use of the patient data	Waiting Lists	Waiting lists are an example of the same data being used for different purposes. In ensuring the delivery of services for individual patients, a primary use is being made of personal data. However, the management of waiting lists – how many people waiting for services, where and when, etc - is a secondary use and does not need identifiable data. The analysis of a problem overlaps with RTT, where an analyst may be involved in determining the cause and resolution of a problem with de-identified data and the decision to contact individual patients and making that contact should lie with relevant case managers making primary use of identifiable data.
Clincial Audit	Clinical audit	Only local clinical audit following the care pathway or quality assuring the care given may use identifiable data. National / regional comparative audits must use either pseudonymised data or have their own Section 251 support.