Section two (to be completed by the project lead and or team)

Privacy Impact Assessment Template

GP Repository for Clinical Care (GPRCC) Project – Phase 2

Reference number: v2.3

Date PIA completed: 28 March 2017 (Updated February 2018)

The CCG **MUST** comply with the Data Protection Act 1998 and other legal requirements. The Privacy Impact Assessment (PIA) process assists by evoking a privacy by design approach to all projects/activities. PIAs are a tool which can help organisations identify the most effective ways to comply with their data protection obligations and meet individuals' expectations of privacy.

The PIA should be completed clearly and accurately as they may be published on the CCG's website (unless they contain commercially sensitive information) after being approved.

Stage 1

Project Summary:

Nottinghamshire CCGs support a centralised repository that allows a small subset of data to flow from GP, community, acute, and Local Authority systems, into the NHIS data warehouse. Collectively, these data form the "GP Repository for Clinical Care (GPRCC)". The scope of this PIA concerns Phase 2 of the project, which considers the use of these data for secondary purposes (i.e. other than direct patient care). A separate PIA considers direct care usage (Phase 1).

The typical scenarios for secondary use of the data are in the planning and evaluation of health services by the CCGs and Vanguards, understanding local population health, predicting future demand, analysing patient pathways etc. All of these areas of work support the CCGs in meeting statutory duties/functions in commissioning healthcare services. Although analysts can often query pseudonymised datasets from each of these sources independently, typically obtained from a different route e.g. as SUS (Secondary User Service) data, they cannot ask questions that require the data to be linked. For example 'How many admissions to secondary care do patients with severe COPD looked after by community teams have in comparison with those who are not cared for by such a team?'.

Data within the GPRCC is collected principally for the purposes of direct patient care, and therefore contains personal identifiers including name, NHS Number, date of birth, address, postcode, telephone number and date of death, which are stored in pseudonymised form in a secure SQL environment (NHIS data warehouse). The purpose of data sharing is outlined within information sharing or data processing agreements. Data collection and retention is governed by direct care needs, as assessed in the Phase 1 PIA. However, it is technically possible to link together these datasets for secondary analysis purposes, as they are pseudonymised using the same key. For data linking of pseudonymised datasets a key IG control will be to ensure the mitigations put in place against risk of re-identification and ensuring that when data is being processed for a 'secondary' none direct purpose that this remains none personal data.

As a first step we have asked CCGs and Public Health to provide us with a list of questions they would like to ask of such an integrated data set. From that we have explored what kinds of data structures would be needed to support such queries. At one end of the spectrum, we might release an aggregated dataset for analysts to use with columns for each parameter they would want to measure, and the final column simply showing a number of patients in each category.

Typical requirements from Public Health include:

- For councils to predict the number of children (under 5) in need of a school place.
- For Public Health to provide needs assessments (number of people in each area with severe COPD, heart failure, who are smoking, have alcohol related admission).
- For Public Health and CCGs to determine how many patients there are in each Care Home, to monitor quality (number of urgent admissions, falls).
- Are some care homes better able to cope with patients with dementia than others?.
- For Councils and CCGs to determine the relative costs to the Health and Social Care system of care home residents versus non care home residents? Total costs to include

healthcare (admissions, outpatient visits, ED attendances, community service costs, GP costs) as well as care home costs (or costs of providing services at home including help dressing, prompting medications, meals)

Data categories that would support these requirements include:

- How many patients are there:
  ○ In each year 5 year age-sex band.
  ○ In the 15-17 female year band for pregnancies.
  ○ Of school age.
  ○ In the over 65 population.

  …. and split these by:
  ○ LSOA (more detailed postcode would allow Mosaic analysis).
  ○ Care Home flag.
  ○ Ethnicity.
  ○ Carers flag (identifies if someone is a carer).
  ○ Care Delivery Group.
  ○ Practice, CCG.
  ○ Ward, LSOA, MSOA, parliamentary constituency.

  … with counts of
- Admissions for COPD, Heart failure, Falls, UTIs, with Dementia, Cancer, etc.
- Outpatient visits.
- A&E visits.
- Number and type of Social Services patients are receiving.

Typical requirements for CCGs and Vanguards include:

- How effective are Care Delivery Groups (CDGs - joint health and social care teams), PRISM system (care teams integrated with secondary care) in terms of improving a range of health and social outcomes? Outcomes might include mortality, anxiety, depression, mobility, self-rated health, ability to perform Activities of Daily Living, loneliness.
- Can CDGs/PRISM change the way healthcare is delivered so that more occurs in the community than in hospitals?
- Can CDGs/PRISM remove some of the workload from primary care (GPs)?
- Which of PRISM/CDGs is the more efficient system?
- What kinds of healthcare professionals, education will we need in the future (predicting
- and managing workforce requirements)?
- Is there sufficient capacity in primary care to deal with today’s / tomorrow’s workload?
- How many appointments are available in primary care (on any one day, bookable over
- next few days)?
- Is their variation in referrals to secondary care between practices? If so which areas are these in? Might education, community service provision help plug skills gaps?
- How good is the quality of data recording in primary care? Are key diseases flagged as Summary items so they are picked out in correspondence to hospitals or community referrals?
- How well do we deal with the problems of BME and hard-to-reach groups (homeless, young people)

Data categories that would support these items include:

- Number (and cost) of GP consultations.
- Number (and cost) of Community services and visits.
- Number (and cost) of Admissions, ED attendances.
- Number (and cost) of 111 calls, ambulance calls, Out Of Hours visits, Walk In Centre (WIC) visits.

...all split by:
○ Age, gender.
○ Long Term Conditions (COPD, Heart Failure, dementia, diabetes, etc).
○ Outcomes including mortality and scores for anxiety, depression, mobility, self
rated health, ability to perform Activities of Daily Living, loneliness.

- Number of appointments available in each GP practice (split by staff group) and hospital specialty, slots for visits for community teams. Number of Did Not Attends (DNAs) in GP, OP attendances.
- Post codes for patients to show geographical distribution of patients by GP practice (for workforce planning)
- Dates of residency in care home (to analyse which admissions occurred while patients resident in care home).
- Dates of referrals, reason for referral (to analyse variations in referral patterns).

This is a broad range of questions. Covering all of these with a single aggregated dataset would reduce the size of each category to very small numbers, which may be as identifiable as if we were using patient-level pseudonymised data. We could use a bundle of aggregated sets, some split by practice, some split by residential area, but as more detailed questions are asked, the categories could become very small again.

Given the wide range of questions that will need addressing we suggest a solution with two tracks. Firstly, we propose building a pseudonymised dataset at event level, containing no personal identifiers and substituting age for date of birth. The pseudonymised dataset would cover primary, secondary, mental health and community care data (but not Social Care data, initially). This would be used to explore data models that would support the range of queries indicated. No event level data should be removed from the data warehouse, but aggregated data (numbers, totals with suppression of small numbers <6) can be removed.

Secondly, we propose building a cube that analysts could query containing the same data, but always aggregated. Using this technology we can surface the ability to extract numbers/rates and any other kind of aggregate totals but not to see any numbers below a threshold value e.g. 5.
While the second solution is highly complex to develop as a data model and will take longer to build, it will be informed by queries run on the first solution and is likely to eventually cover the majority of analyses that need to be run. At that point, we would propose withdrawing the first solution from regular use. If an analyst poses a question that cannot be answered by the cube, we can either attempt to augment the cube to address the issue, or explore use of the pseudonymised dataset for that particular question only.

Data Governance Requirements for processing data for ‘secondary uses’:

As a health and care community the Records and Information Group has agreed guidance and principles on how data can be used within the legal framework for secondary use purposes. For ease of reference the community wide agreed guidance is below:

Nottinghamshire
Data Sharing Non Dire

The key requirements that should be followed when processing data for a secondary purpose include:

1. Authority sought from all the ‘data controllers’ who have shared data in the central repository to support direct care. This can either be on an ad hoc basis or outlined in revised data sharing agreements or data processing contracts. This is a fundamental requirement to ensure that data shared for direct care is not used for another purpose inconsistent with the initial purpose for sharing (potentially breaching principle 2 of the Data Protection Act 1998). Additionally the Joint Data Management Team (hosted by Rushcliffe CCG acting in a data processor capacity) cannot make decisions about ‘determining the manner and purpose of data processing’ as this instruction needs to come from the ‘data controllers’.
2. Specific to social care data as the processing for ‘secondary use’ has explicitly been excluded unless clear instruction received this data should not be linked with any other data for the purposes of

processing for 'secondary use'.

3. For all data processing for 'secondary uses' it must be identified to which organisational statutory function the output will support (see appendix 2 of the RIG guidance 'Nottinghamshire Data Sharing for None Direct Care/Secondary Purposes'. This process will strengthen the legitimacy of processing data for a secondary purpose.
4. An **assessment of compliance with the Nottinghamshire Records and Information Group's (RIG) guidance** which outlines the systems, processes and controls for 'secondary uses' of data. In reference to the RIG guidance 'Nottinghamshire Data Sharing for None Direct Care/Secondary Purposes' this will include Pseudonymisation at source and or Pseudonymisation on landing as outlined below.

Pseudonymisation at source supports record linkage but does not support the re-identification of individuals and requires a shared identifier (e.g. NHS Number). Consideration should be given to the quality of the data to enable accurate matching.

Pseudonymisation on landing requires compliance with Data Protection legislation and government policy regarding objections as data is transferred in the clear. Re-identification may be supported but where the body processing the data has access to the re-identified data confidentiality may be breached unless an additional legal basis e.g. explicit consent or s251 support has been obtained.

5. When secondary use does not comply with the RIG guidance or is not in support of an organisational statutory function or involves sharing outside the NHS or local authority, agreement from the relevant Caldicott Guardians should be sought (for the organisations who legally own the data within the central repository).
6. If the processing involves a significant activity change outside phase 1 or phase 2 of the GPRCC PIA scopes a revised or separate PIA assessment should be undertaken.
7. At the point when outputs of data processing for secondary purpose are identified this should by default be aggregate outputs and a risk of re-identification or a 'jigsaw attack' to be completed before disclosure of the analysed data. If data output/s is shared outside the NHS or Local Authority partners approval should be sought to ascertain if there is any commercially sensitivity issue of sharing the output which could damage working relationships or pose a risk to the reputational damage of an organisation.

System engineers in the Data Management Team at Rushcliffe CCG administrate eHealthScope and act as a Data Processor on behalf of GPs. Within the scope of Phase 2 of the GPRCC, they will only see pseudonymised data, a control put in place to guard against patient privacy; with the overriding principle being that any processing for a none direct purpose does not have any requirement to identify data to an individual patient.

Data for secondary use reporting will be stored separately from the direct care GPRCC data to support minimisation principles and reduce even further any risk of unintentional access to identifiable data.

Please see the supporting data flow diagram in Appendix B, illustrating the high level data flows, organisational responsibilities and the secure pseudonymisation and delivery mechanisms.

List of attachments: *(e.g. project initiation document or proposal)*	**Page**
GPRCC Phase 1 PIA RIG guidance 'Nottinghamshire Data Sharing for None Direct Care/Secondary Purposes'	

Brief description of the data affected *(is this personal confidential data e.g. health information, criminal records or other information people are likely to consider as private?):*
All data included under the GPRCC Phase 1 PIA, with the exception of Social Care due to the more complex data sharing governance. However, data will be de-identified prior to secondary use (pseudonymised, or aggregated with small number suppression).

Details of data being processed:					
	Whole records/referrals		Local identifier only		NHS Number
	Name		Date of birth		Postcode (full)

✓	Postcode (LSOA)[1]		Age (exact or <1 year)	✓	Age bands 5 years
	Age bands 10 years		Ethnicity	✓	Gender
	Religion		Disability	✓	GP practice
	Other (please describe):				

Will data be:					
✓	Anonymised	✓	Pseudonymised[2]		Fully identifiable (PID)
If processing is for secondary purposes (i.e. not related to direct care) and fully identifiable information is to be used, please explain why anonymised or pseudonymised data will not meet the project objectives?					
N/A					

Frequency of transfers: (delete as applicable)
One off / daily / weekly / monthly / quarterly / annually / other (please state):
Daily extractions from each provider, potential daily access to analysis data.

Please provide details on how long data will be retained by any organisation involved with processing, and destruction arrangements *(attach supporting documents where appropriate)*
Data will be collected and retained in line with the purposes and principles described in the GPRCC Phase 1 PIA, i.e. for so long as patients remain registered with a Practice supported by the Data Management Team, plus a period of 12 months thereafter for clinical audit.

Organisations involved and stakeholders:	
Organisation	**Contact Name and Details**
Data Management Team (part of Rushcliffe CCG)	Exemption Section 40(3A) 3rd Party Personal Information applied. To disclose the information would contravene one of the data protection principles set out in article 5 of the GDPR
NHIS (host of the IT environment)	
All GP Practices across Nottinghamshire	
Nottingham West Health Ltd (aka Primary Integrated Care Services, PICS) (8HY59)	
Nottingham University Hospitals Trust (RX1)	
Nottingham CityCare Partnership (NR3)	
Sherwood Forest Hospitals Foundation Trust (RK5)	
County Health Partnerships (RHA20)	
Nottinghamshire Healthcare Trust	

You do not need to complete stage 2 if the data involved in the project/activity is anonymised and the sharing is between organisations who have a legitimate* reason to receive the data or you are acting in a commissioning capacity which does not involve the CCG sending or receiving any personal confidential data or data which the CCG is the data controller. **However please highlight how you will keep the data secure and mitigate any risks.**
Data is initially collected and pseudonymised as described in the GPRCC Phase 1 PIA. For secondary use, an additional process will transform the pseudonymised GPRCC data (excluding Social Care) into a separate reporting database, applying any business rules and minimisation principles in doing so. This will be patient / event level pseudonymised data and access to it will be limited to the Data Management Team and a small number of named Analysts working alongside the DMT for development purposes. Queries on the reporting database will inform the development of a SQL Analysis Services “cube”, which is a powerful way of structuring data so that it can be aggregated in different ways, very quickly in real-time. It is possible to apply permission restrictions on a cube at different granularities. However, outputs are in the form of dynamically calculated values (counts, sums, averages etc.), so it is possible to suppress any results

[1] Lower Layer Super Output Area: relates to first half of postcode and number only of second half. Much public health reporting and published Indices of Deprivation are based on LSOA and this is the standard that is widely accepted and expected for large scale research/statistical reporting.

[2] A pseudonym is used to replace identifiable data so that patients cannot be identified without a pre-defined code/key.

which are considered small enough to pose a risk of patient identification. We propose to follow the ICO Anonymisation Code of Practice and NHS Digital HES guidelines, and to suppress results which involve fewer than 6 individuals. By ensuring that data is adequately anonymised, it will be possible for it to be used more widely with fewer restrictions and privacy risks.
Have you considered if an information sharing agreement, data transfer agreement or other contract is required?
Information Sharing Agreements and/or Data Processing Contracts have been established as part of GPRCC Phase 1. These will be revised so that each Data Controller agrees to the use of their data for secondary purposes, in the manner described within this PIA. Data Processing Contracts will also be amended, where necessary (e.g. those between GP Practices and the DMT). This requirement will ensure data is processed in accordance with the purpose for which it was provided.

Stage 2

Describe and map the data flows and who will have access to the data *(who is collecting, receiving, transferring or storing the data):*

Data sources	*→ transfer →*	GPRCC Reporting Database (DMT)	*→ transfer →*	Data recipients
GPRCC clinical database (excluding Social Care) inc: PICS CHP CityCare CHC NUH SFHFT NHCT NEMS	Pseudonymised transferred and transformed securely within a SQL Server environment.	Data now segregated from re-identifiable data and structured in a form that's better suited to secondary use analysis, with minimisation principles applied.	Data from all providers delivered by eHealthScope or SQL Server Analysis Services cube, both using AD / RBAC to authenticate individuals. Permissions may be set with different granularity, depending on role.	CCG and Public Health Analysts – possibly access to all users of eHealthScope where data is sufficiently anonymised.

Will the project involve the collection of new personal confidential information about individuals? *(if yes, please describe)*

No.

Will the project compel individuals to provide personal confidential information about themselves? *(if yes, please describe)*

No. Data will be excluded from the secondary use reporting database for patients who have given express dissent in their GP clinical information system to their data being used for secondary purposes. This status is updated daily as part of GPRCC Phase 1.

Will information be disclosed to organisations or people who have not previously had routine access to the information? *(if yes, please describe)*

Yes, but only in anonymised / aggregate form. During the development phase, or in circumstances where a query cannot be answered using the cube, Analysts may work alongside the DMT to query the pseudonymised secondary use database, provided that results involving fewer than 6 patients are suppressed.

Are you using information about individuals for a purpose it is not currently used for or in a way it is not currently used?

All data is collected primarily for direct care use. However, we will respect the wish of patients who object to their data being used for purposes beyond this, and keep this status updated daily.

Will explicit consent be obtained from data subjects?[3]

	Yes – verbal, recorded in record
✓*	Yes – written, scanned into record
✓	No – not required
	No – other reason

If explicit consent is not to be obtained, please state reason(s) below or give details about the legal basis you are relying on the process personal confidential data:

[3] The individuals who the records are about.

As the focus of the project is to deliver direct patient care, we are not seeking explicit patient consent to share out information from GP clinical systems. We are using implied consent. However any patients who have explicitly dissented to the sharing of their record for direct care purposes will have their record discarded automatically by the ‘black box’ loading process.

Community providers seek out explicit consent to share information with GPs when they first see their patients.

Communication packs have been issued to all practices so that posters appear in waiting rooms advising patients that a subset of their GP records are being shared in the community for direct patient care. This is part of a joint communications approach run with the MIG project. Practices are also advised to update their websites to describe electronic record sharing.

Before any data is used for secondary use, it will be pseudonymised and therefore considered none personal data. Strict IG controls will be put in place to reduce the risk of re-identification and ensure data being processed for a secondary use does not become processing personal data (which would then require compliance with DPA requirements). Patients who have recorded an objection with their GP Practice to data being used for secondary purposes will be excluded from the reporting dataset. All outputs shared outside of the data processing environment for purposes other than direct care will be anonymised/aggregate in line with the ICO Anonymisation Code of Practice so that explicit consent is not required.

Will the project require you to contact individuals in ways which they may find intrusive? *(if yes, please describe)*

No. This phase of the project does not involve contacting patients.

What security arrangements will be put in place to ensure the confidentiality and information security of personal confidential data? *(consider any residual threats to data security)*

The primary security arrangements are as described in the GPRCC Phase 1 PIA. In addition to this, there are further measures to improve security for secondary use. Reporting data is held in a separate database, which further reduces any residual risk of data being linked inappropriately. Data items (especially identifiers) that are not required for analysis are not carried across to the reporting dataset. Values which indicate the same individual differ from the pseudonyms in the clinical GPRCC database, and are non-reversible. Access to the data for the purposes of reporting will utilise technology to suppress results for small numbers of patients (<6) to minimise the risk of indirect re-identification.

The Data Management Team, in their role as Data Processor on behalf of General Practices (and Nottinghamshire County Council), can access this data but only see it in pseudonymised form (whether they access the data directly via SQL queries or via eHealthScope). No other CCG staff have access to this data, but a small number of named Analysts may work alongside the DMT from time to time to develop queries which can be supported by the cube or to answer particular questions which are too complex to resolve via the cube. Small number outputs from this environment would be suppressed for reporting purposes in all cases

What will be the impact of decisions brought about by the project or activity and processing of PID? *(please highlight both positive and adverse impacts either directly or indirectly)*

PID will not be processed for this work.

The key benefit is to patients, particularly at the local population level. Enabling CCGs, Vanguards and Public Health teams to gain new insight into demand, future need, pathways and inequality of care is essential in order to improve patient outcomes and respond to the intense challenges facing the health service as a whole.

Summarise the risks of this processing? (NB: these 3 can be merged as appropriate)
Please attach full risk assessments. *(refer to your risk management policy for risk RAG scoring)*

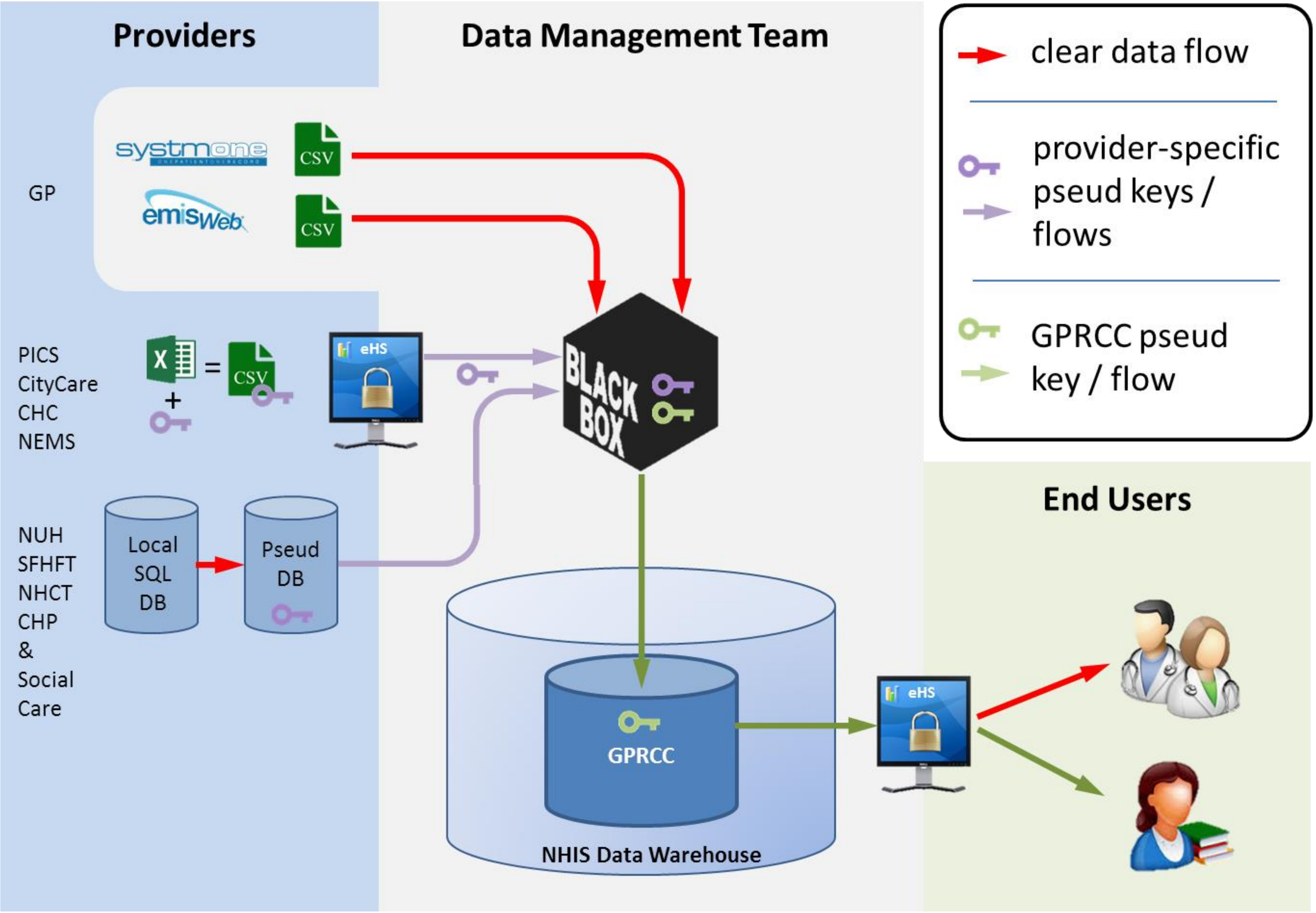
a) To Patients, providers, Practice, CCG

1*5=5	Data is lost in transfer	See details in the PIA
1*4=4	Access to data gained by third party	See details in the PIA
2*2=4	Patients object to implicit consent	Widely discussed with patient representatives as part of MIG project and in

		CCG clinical cabinets. Risks felt to be significantly less than risk to patients of not delivering effective care.

If applicable, please give details of any service user/staff/public consultations that are going to take place, or have taken place in relation to this processing? *(include internal and external stakeholders)*
The MIG project communication strategy was developed in such a way as to include GPRCC. However, the GPRCC board will need keep the communication materials provided to patients under review to ensure that they remain sufficient to meet duty of confidentiality requirements and assess whether further communication should be provided to patients.

Appendix B: Data Flow Diagram



Author:
Date of Issue: