Health and Social Care Information Centre

Information Governance Assessment

Customer: NHS Commissioning Board

Customer Requirement Reference Number: NIC–178106-MLSWX

Date: February 2013

Version 1.0

1 Information Governance Assessment: Summary Sheet

Customer requirement reference no	NIC–178106-MLSWX
Customer organisation(s)	NHS Commissioning Board
Data controller (if different)	Health and Social Care Information Centre
Purpose classification	Secondary purposes
Aggregated or individual-level data?	Individual-level data
Identifying or non-identifying?	Assessed as identifying
Legal basis (if identifying)	Health and Social Care Act 2012
IG Assessor Comments	
Assessed by	M Oswald

The full information governance assessment follows overleaf. A description and explanation for all numbered fields in the assessment can be found in *Information Governance Assessment Explanatory Notes* (to be published separately).

2 Information Governance Assessment

2.1 The Data Recipient(s)

1. Are the only data recipients the registered GP practices of the patients? No

   If yes, the practice is responsible for ensuring proper governance of the data output; go to section 2.6

2. *If no, who will receive the data provided?* Health and Social Care Information Centre (carrying out directions from the NHS Commissioning Board).

3. *What was their last published IG Toolkit score, or similar assessment?*

   For the last 2 years, the HSCIC has scored highly (97% and 98%) against the IG Toolkit.

2.2 The Data Requested

4. The data items to be provided are specified in section 4 and Appendix A of the Customer Requirement Summary. The purposes for which they are to be used are identified in section 2 and 3.

5. Is the dataset requested appropriate, and not excessive? Yes

1. Comment To consider whether any part of the processing of care.data GPES extract is excessive, the extraction process is considered below in three stages: 1. extracting the data from general practice systems; 2.processing and storing the extracted data within the Health and Social Care Information Centre; 3. disclosure of the extracted data by the Health and Social Care Information Centre.

2. 1. Extracting the data from general practice systems

3. Under the Health and Social Care Act 2012, the Health and Social Care Information Centre is empowered to require general practices to supply identifiable or anonymised information about patients, if directed to do so by the NHS Commissioning Board (as explained further in section 2.5). Patient consent is not required for disclosure. Under the Act, the Board may stipulate whatever data it considers "necessary or expedient", although it must seek advice first from the Health and Social Care Information Centre. The Board has made clear what data it considers to be necessary – demographic information from all patients, with associated clinical information for patient records that contain information relevant to commissioning (with data on events like sexual health diagnoses that might be considered especially sensitive excluded). In order to only extract relevant (and not excessive information), considerable effort has been made to identify the clinical codes that are relevant to commissioning, excluding irrelevant codes and those expected to be of poor quality and thus unreliable.

4. Section 35(1) of the Data Protection Act allows for disclosures of personal data that is required by statute. However, under Principle 3 of the Data Protection Act, processing (including storing) more personal data than necessary is unlawful. Under the Health and Social Care Act, the Board decides what information it requires to be necessary for its functions. Therefore, it could be argued that whatever data the Board requires should not be considered excessive under the Data Protection Act. Nonetheless, the responsibilities of the general practice (as data controller of its patient records) and/or the Health and Social Care Information Centre (which becomes data controller once the data are extracted) must also be considered.

   The GPES Information Governance Principles (which were written prior to the Health and Social Care Act) are based on the responsibilities of data controllers under the Data Protection Act to authorise processing of personal data, and state that no extractions will

be made from general practice records without the explicit authorisation of the practice. However, the Health and Social Care Act empowers the Information Centre, when acting on a direction from the NHS Commissioning Board, to require health and social care bodies including general practices to provide it with data in "such form and manner, and within such period, as the Centre may specify". Advice from Department of Health lawyers and the Information Commissioner's Office is that general practices are required to comply.

2. Processing and storing the extracted data within the Health and Social Care Information Centre

After extraction from GPES, the primary care dataset will be matched against Hospital Episode Statistics Index records, and where a patient is found to have a hospital episode index record, the associated HES Id (a pseudonym) will be added to the patient's primary care record. The match is made using four identifiers: NHS number, Date of Birth, Gender and Postcode. No match is expected in a sizeable minority of cases because some patients will not have had a hospital episode. The (unmatched) primary care record alone is still considered useful for commissioning. Whether or not a match is made, the four original identifiers are retained. The identifiers are required to enable the record to be matchaed and updated with new information about the patient in subsequent months.

Next, the HES Id is used to link the primary care record against Hospital Episode Statistics records (for a given period) which are identified using the HES Id. Where there is a match, a record is created in a new file comprising the HES Id, the primary care data, and the hospital data. Where there is no match (which is expected in a significant minority of cases where a patient registered in a general practice has no associated hospital episode), a record containing the four identifiers, plus the primary care data, is inserted in the new linked primary care/ secondary care file. The original primary care data set is then destroyed.

Through this processing, patient records are created which contain the four identifiers, and which may contain primary care clinical information, and which may also contain secondary care clinical information (such as diagnoses). The identifiability of the data is significantly reduced through the processing, and access to the data is strictly controlled. These records are then retained to enable access and use by customers, for both known, and as yet unknown, uses. Because the full extent of the uses have not yet been identified, it is unclear whether all of the clinical data in the file, both in the matched and unmatched patient records, will be used. On this basis, the processing might be considered excessive. The judgement is whether it is proportionate to collect and temporarily store some data that may prove useful for potential commissioning purposes.The Information Commissioner's Office guidance is that: "You should not hold personal data on the off-chance that it might be useful in the future. However, it is permissible to hold information for a foreseeable event that may never occur".

5. 3. Disclosure of the extracted data by the Health and Social Care Information Centre

Once the linked primary & secondary care file has been created, it will be made available to approved customers in one of two ways. Firstly, anonymised reports will be produced for customers by the Health and Social Care Information Centre. Secondly, some customers will be given authorisation to write queries and extract and download anonymised reports for themselves. For this phase of the care.data extraction, the NHS Commissioning Board has made clear it does not require any identifiable information to be released. The Health and Social Care Information Centre, as data controller for the linked primary & secondary care file, will be responsible for assessing and ensuring that only anonymised or pseudonymised extracts will be made available by these two methods. The Information Centre will judge that no excessive data will be released to customers and will ensure that any data released is covered by appropriate terms and conditions with customers. The Information Centre has a panel of senior statisticians and

other staff who consider such information requests/releases (and who will apply the recently-approved Anonymisation Standard for Publishing Health and Social Care Data).

On balance, the Health and Social Care Information Centre judges that the processing is not excessive.

2.3 Is the Dataset to be Provided Identifying or Non-identifying?

6. Does the dataset contain obvious identifiers such as name, address and/or NHS Number? Yes

7. *If yes, which items?* NHS Number, plus quasi-identifiers such as date of birth and postcode.

If yes, the dataset is clearly identifying, so go to section 2.4.

2.3.1 Risk Assessment: will the data provided be identifying?

Where the data provided to the customer contains obvious identifiers like name and address, it will be "identifying". However, there are many cases where it is not obvious whether data are identifying, and these require judgement based on a risk assessment. This section should contain that assessment.

8. Are they aggregated or individual-level data? Individual-level data

Re-identification risk assessment[1]		
9.	What is the threat level associated with the data and its release ("normal" or "high")?	High
10.	What is the risk of extra information being used to reveal identity ("normal" or "high")?	Normal
11.	**Conclusion: is the dataset to be provided assessed as "identifying" or "non-identifying"?**	**Identifying**
12.	Explain this risk assessment, referencing sources of standards, guidance or advice The data set being extracted from general practice systems includes NHS number, a unique identifier of the patient. It also contains date of birth and postcode, which are sometimes referred to as quasi-identifiers. Knowing somebody's NHS number will not automatically tell you who they are, as access to systems which link the names and addresses of patients to their NHS number is restricted, although there are thousands of health and social care staff authorised to access that information. Access to the patient records created through the processing by the Information Centre will be strictly controlled. Nevertheless, given the large volume of patient records, the use of the NHS number and quasi-identifiers, and the proposed processing of the data within the HSCIC after the patient records leave the general practice, the risk of re-identification is small but not negligible.	

[1] This risk assessment process, and explanatory notes, are based the draft "De-identification standard for publishing health and social care data", and the accompanying draft guidance "Drawing the line between identifying and non-identifying data". These are due to be published by the end of 2012.

	Therefore, on balance, the data extraction is assessed as identifying.
13.	*If non-identifying, could the data to be released be published?*[2] *Why?*

14. In light of this risk assessment, what specific de-identification techniques, if any, will be deployed?

These are described in section 2.2. The principle ones are: removal of NHS Number, transformation of date of birth into year of birth, aggregation and small number processing prior to release of data, and restricted access (e.g. through role-based access controls).

2.4 Healthcare or secondary purposes?

15. The purpose(s) of the extraction can be found in section 2 and 3 of the Customer Requirement Summary.

16. Is the data extraction for healthcare[3] or secondary purposes[4]? Secondary purposes

Provide a brief explanation: It is primarily to improve the commisioning of healthcare.

17. If "healthcare purposes", what efforts (if any) are being made to ensure that consent can be implied?

None - statute (namely the Health and Social Care Act) is the legal basis for this extraction, and not consent (see section 2.5 below). Under the Data Protection Act (Schedules 2 and 3), the condition for processing the personal data is that the processing is necessary because of a legal obligation. Sensitive personal data are being extracted, and the relevant condition for this under Schedule 3 of the Act is that the processing is necessary for governmental functions.

The legal basis for the disclosure from general practice systems is statute. As a result, there is no legal necessity to allow patients to opt out of the extraction. Given that only anonymised information is to be released by the Information Centre, the NHS Commissioning Board has decided that records for every patient should be extracted.

If "healthcare purposes", go to section 2.6.

If the dataset to be provided is classified as "non-identifying", go to section 2.6.

2.5 Using identifying information for secondary purposes

18. What is the legal basis for using identifying data? Health and Social Care Act 2012

19. If S251 is the justification, what is the status of the application?

[2] A dataset that is non-identifying when released into a controlled environment where access is restricted may be identifying when released into the public domain (because of the increased risks).

[3] i.e. Uses which "directly contribute to the diagnosis, care and treatment of an individual; or the audit/assurance of the quality of the healthcare provided". See the Confidentiality Code of Practice 2003. Sometimes referred to as "direct care purposes".

[4] i.e. Any other purpose than a "healthcare purpose". This includes epidemiology, health research, financial audit and the management of health [and social] care services.

If approved, attach evidence of approval

Provide reasons / backing / evidence for the above justification, including any advice or approvals received General practices should not disclose confidential information about patients without a lawful basis for the disclosure. This care.data extraction of identifying, and confidential, data for the NHS Commissioning Board from general practices by the Health and Care Information Centre after April 1 2013 is justifiable under the Health and Social Care Act ("the Act") on the following basis:

- section 254 of the Act empowers the NHS Commissioning Board to direct the Information Centre to collect the data which the Board considers "it is necessary or expedient for the Board to have", having first consulted the Information Centre;

- the Information Centre is consulting relevant representatives of the Board, the potential users of the information to be extracted, and the suppliers of the information (for example, through the IAG), as is required under section 258 of the Act;

- section 259 of the Act empowers the Information Centre to require general practices to supply it with any information it requires to carry out its functions (including the function of satisfying the Board's directions);

- the Information Centre will not be publishing the information because, as identifiable information, it falls under section 260(2) of the Act.

Note that had this been a request from another body, and not a direction from the Board, this collection of confidential information would not have been justifiable under section 259(3) of the Act.

General practices are obliged to comply with the care.data request for information under section 259 of the Health and Social Care Act "in such form and manner, and within such period, as the Centre may specify". The specified form and manner specified by the Information Centre is a GPES extraction. Therefore, under the Act, all general practices are obliged to provide the care.data through GPES, if that is the means specified by the Information Centre. However, the Information Centre is seeking the authorisation of practices before extraction, as is normal under GPES. If a practice prefers to supply the required information to the Information Centre in another way, then the Information Centre must first agree to the form and manner in which the practice proposes to supply the required data.

2.6 Compliance with GPES Information Governance Principles

20. Does the request comply with the GPES IG Principles? Yes

21. Comment This is an extraction of patient identifiable information, with statutory justification, so page 9 of the GPES IG Principles applies. Although there is a statutory basis for the extraction, IAG support is sought, and practice authorisation will be required before the data are extracted. According to the GPES IG Principles, the data will be "deleted from GPES data stores as soon as the data have been released" (see page 7). This will be done for the primary care data extracted through GPES. On this basis, the extraction complies with the GPES IG Principles. However, in the case of care.data, the data extracted will be processed and stored (in "pseudonymised" form) within the Information Centre, rather than passed to an external organisation (as is normally the case for GPES). The Information Centre, acting on directions from the NHS Commissioning Board, is the data recipient for the GPES extract. The data are to be transferred, processed, stored and used securely within the Data Management Environment (DME) at the Information Centre for one month, after which they will be updated or replaced.

2.7 Information Governance Risks of Extraction and Disclosure

There are risks with every GPES extraction and release of information to customers. However, the following table identifies any areas where there are particular conditions that may give rise to information governance risks for this extraction.

Information Governance Risks		
22.	Risk of privacy breach during extraction processing	
23.	Privacy breach whilst data is under recipient's control	See notes 1 and 2 below.
24.	Privacy breach following recipient's disclosure of data	
25.	Other information governance risks	

Notes

1. The Information Centre will effectively be maintaining a new data store of linked data, refreshed monthly, containing records of the population of England. The Information Centre has in place secure systems for storage and access of these data stores. Nevertheless, no system is 100% secure, and so there is some small risk given the many millions of potentially identifiable, patient records.

2.The governance process for GPES stipulates that each request is scrutinised by an Independent Advisory Group (comprising general practitioners, lay members and others) which advises the Health and Social Care Information Centre on whether the benefits of the extraction merit the risks involved. The need for the information requested has to be justified to the group. The group's advice is published, and can be reviewed by general practitioners to help them decide whether or not to participate in an extraction.

The governance controls around the release of, or access to, the data within the file created through the data linkage process is not open to the same external scrutiny. Decisions on access will be the responsibility of the Health and Social Care Information Centre. As with GPES, there will be a data sharing and re-use agreement that a customer will be required to sign. It will set out the data to be made available, and require customers to stipulate a purpose or range of purposes, before they are provided with data extracts, or granted direct electronic access to "views" (i.e. selected elements) of the primary care/secondary care linked file. Providing users with the ability to query views of the linked file introduces potential risks of re-identification from inference. This risk will need to be taken into account in the Information Centre's judgement when determining what views of the data are made available to customers.

A thorough assessment of re-identification risk to ensure that the data released are anonymised is required according to the Anonymisation Code of Practice (recently published by the Information Commissioner's Office (ICO)), and the Anonymisation Standard for Publishing Health and Social Care Data (recently approved by the Information Standards Board). The governance arrangements described above and the scale and potential sensitivity of the data to be extracted monthly and stored, are small but relevant risks for consideration by the Independent Advisory Group.

2.8 Assessors

Assessment made by:

Name: Malcolm Oswald

Role: GPES Information Governance Advisor

Date: 08/02/2013

For the Health and Social Care Information Centre

Assessment Checked by:

Name: Clare Sanderson

Role: Director of Information Governance

Date: 08/02/2013

For the Health and Social Care Information Centre

The persons above confirm that to the best of their knowledge the information governance assessment is fair and accurate.

Do either of the two people above have any caveats or other comments to state in relation to the information governance assessment provided?