Queen Mary
University of London

Data Protection Impact Assessment

QResearch

Prepared by: Julia Hippisley-Cox (j.hippisley-cox@qmul.ac.uk)

Version: 1.3

QResearch DPIA

Document Owner:	**Julia Hippisley-Cox**
Status:	**Final**
Version:	**1.3**
Current Version Date:	**02.10.2026**
Review date:	**01.10.2027**

Revision History			
Author	Description	Version	Date
Julia Hippisley-Cox	1st version following review and approval by Paul Smallcombe, records and information compliance manager, Queen Mary University of London.	1.0	02.10.2025
Julia Hippisley-Cox	Update to step 4 - section 251 approval and step 2- key management	1.1	04.08.2026
Julia Hippisley-Cox	Expanded Step 5 risk assessment to address key management, unauthorised access, re-identification and inappropriate onward disclosure, with consequential updates to Steps 2 and 6	1.2	11.08.2026
Julia Hippisley-Cox	Step 1-4:updated data flow diagram and associated data processing and classification, opt out; data minimisation specification and CAG reference added Step 2 & 5 Updated section on children and other vulnerable groups;	1.3	01.10.2026

Final Approval		
Name - Position	Version	Date
Paul Smallcombe – Records & Information Compliance Manager	1.1	04.08.2026
Paul Smallcombe – Records & Information Compliance Manager	1.2	19.08.2026
Paul Smallcombe – Records & Information Compliance Manager	1.3	02.10.2026

Introduction

You should fill out this document at the start of any major project involving the use of personal data, or if you are making a significant change to an existing process.
To learn more about the process of completing a DPIA, please refer to the accompanying guidance document provided on the following webpage:

https://www.qmul.ac.uk/governance-and-legal-services/governance/information-governance/data-protection/data-protection-impact-assessments/

Step 1: Identify a Need for a DPIA
Explain broadly what the project aims to achieve and what type of processing it involves. You may find it helpful to refer or link to other documents, such as a project proposal. Summarise why you identified the need for a DPIA.

Project Aim

QResearch (www.qresearch.org) is an effectively anonymised medical research database which is used to answer key research questions relevant to the delivery and care of patients in the NHS including the development and validation of widely used clinical prediction tools such as QRISK, QDiabetes, QFracture, QAdmissions which are recommended by NICE for use across the NHS.

QResearch GP data is linked to NHS England secondary care data and QMUL provides secure access to effectively anonymised linked data for academics employed by UK universities who are undertaking research projects of benefit to the NHS. Research is also undertaken to evaluate the uptake, safety and effectiveness of medicines, risks and benefits of interventions, natural history of disease and health inequalities. QResearch is also used to undertake health inequalities research to determine whether there are differences between different groups (e.g. by age, gender, ethnic group, deprivation, geography) in natural history of disease, survival from different conditions and also to evaluate access to, uptake, safety or effectiveness of health care interventions. QResearch is also used to undertake commissioned research. For example, the NHS or Department of Health and Social Care (DHSC) or its arms lengths bodies can commission a research project to answer policy questions. Research data is only accessed by researchers employed by UK universities.

Type of processing involved

Encrypted identifiable GP data are obtained via the IM1 platform. Direct identifiers are either transformed or removed before coded research data extracts are saved. This is done in memory. The NHS number is scrambled using a one-way hashing algorithm to create a linkage code which cannot be reversed. The NHS number is then removed. The postcode is converted to deprivation scores and then removed. All free text comments, names, addresses, email addresses, contact phone numbers are removed from the data. Dates of birth are rounded to years of birth. Only coded data with a de-identified code which cannot be reversed, is then stored in the QResearch database. All personal data are removed/deleted at the first opportunity at the point of extraction after de-identification. The linked QResearch core is assessed as effectively anonymised within its controlled setting.

Need for DPIA

The DPIA is required to support the processing of personal data by QMUL from the GP IM1 system.

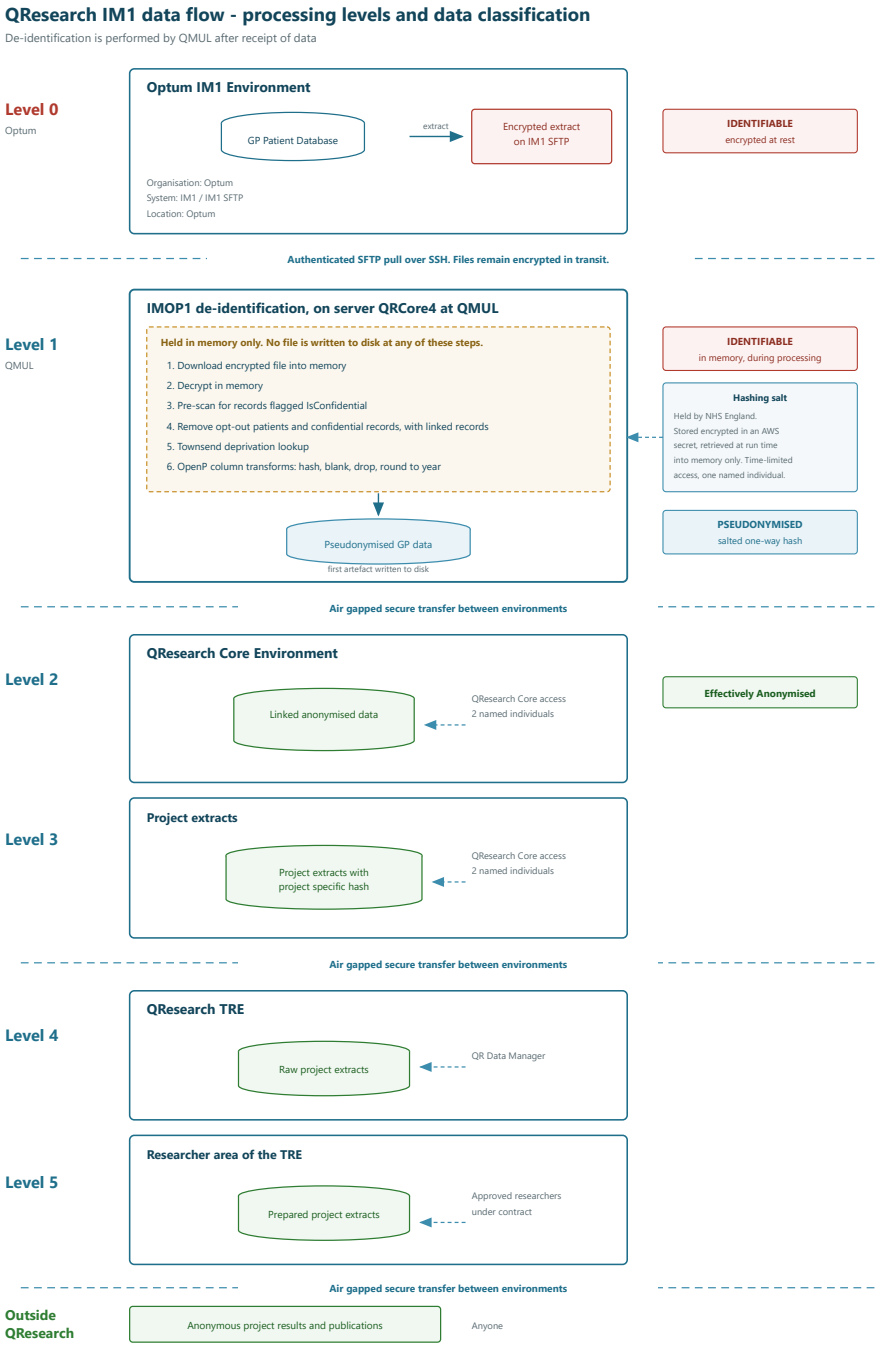
Step 2: Describe the Processing

Describe the nature of the processing: how will you collect, use, store and delete data? What is the source of the data? Will you be sharing data with anyone? You might find it useful to refer to a flow diagram or another way of describing data flows. What types of processing identified as likely high risk are involved?

Data Source

The source of the data is medical health records from general practices in England which contribute to the QResearch database. Queen Mary University of London (QMUL) is the legal entity which processes confidential data items (NHS number, date of birth, date of death, postcode, address, GP registration) under s251 support. Data sources include both TPP and Optum. Figure 1 shows the flow of data and the different levels of access.

Figure 1 Data Flow Diagram (example Optum)



Data Collection and processing

Encrypted identifiable data are collected from the GP practices via the IM1 platform provided by the GP system suppliers (Optum and TPP). Encrypted identifiable GP data are obtained via the IM1 platform. Direct identifiers are either transformed or removed before coded research data extracts are saved. This is done in memory.

- Unique codes, known as GUIDS, are generated using one-way industry-standard cryptographic hashing techniques that enable records to be linked without identifying individuals
- Only NHS England approved key holders have access to the cryptographic keys, ensuring that patient identities remain protected. NHS England remains responsible for key governance. One named QMUL individual is granted time-limited, purpose-specific access to the salt value solely for the extraction process. That individual does not have simultaneous access to the QResearch environment or NHS England datasets. Approved key holders do not have access to the QResearch database.
- Postcodes are removed and replaced with deprivation scores.
- All free text comments, titles, names, addresses, emails, mobile numbers, telephone numbers, IP addresses, booking references, scanned images, medical drawings, letters and other attachments are removed from the data
- Dates of birth are rounded to year of birth.
- Only effectively anonymised coded data with unique GUIDS are then stored in the QResearch database. The tables and fields which are retained are all required for research purposes and are described in the Excel spreadsheet QResearch IMOP Specification.xlsx for each supplier.

To further enhance security, QResearch applies a second layer of hashing that is project-specific when data are shared with researchers for each research project. This means that even if the same effectively anonymised data is used across multiple research projects, each dataset is uniquely hashed for its intended purpose. This additional cryptographic transformation ensures that data cannot be cross-linked between projects, significantly reducing the risk of re-identification and reinforcing the principle of data minimisation

Data Storage and Use

Once the data has been effectively anonymised, the data are stored in a Microsoft SQL database on Queen Mary University of London owned servers and used for observational medical research.

Data Deletion

Personal data are deleted immediately the datafiles have been effectively anonymised.

Sharing of Data

No personal data are shared with other organisations. Only effectively anonymised data from the medical research database are shared with academic researchers that have ethical approval to undertake research projects and who are based in the UK.

High risk types of processing

No processing has been identified as presenting a high residual risk. Potential risks associated with key management, unauthorised access, re-identification and inappropriate onward disclosure are assessed in Step 5, taking account of the controls described in this DPIA.

Describe the scope of the processing: what is the nature of the data, and does it include special category or criminal offence data? How much data will you be collecting and using? How often? How long will you keep it? How many individuals are affected? What geographical area does it cover?

Nature of Data

The data is health record data from patients registered with GP practices in England that contribute to the QResearch database. It includes special-category health data about patients including children, former patients and deceased patients. Whilst the GP data extract includes special category data, the method of processing means the data stored in the QResearch database can't be linked to an individual person. There are no criminal offence data or genetic data. Whilst there may be genetic diagnostic codes in the GP record, for example, stating that someone has a disease with a genetic component (e.g. Down's syndrome) but there is no access to an individuals' complex genome or actual gene sequencing as these are not stored in GP records. Similarly, there are no criminal records except where the GP has used a SNOMED-CT code to record an offence.

Ethnic origin data is stored in the GP records and this is required for analysis and research purposes.

Amount of Data Collected and Used

The minimum data is collected for the purpose stated above. QResearch IMOP Specification records the proposed action and research or processing purpose for each table and field. Fields are marked for removal, transformation or keeping. The source postcode is used for lookup and then dropped. The date of birth is reduced to year. The specified history starts on 1 January 1990 to permit longitudinal analyses of earlier exposures, diagnoses and outcomes, including follow-up into later life; inclusion of all ages permits research across the approved population, including childhood onset and later outcomes. The final rationale, field actions and implementation have been reviewed and approved by QMUL. Only the de-identified anonymised data is stored. Any personal data is deleted as soon as the anonymisation has been done, again at the first opportunity.

Data Collection Frequency

The data will be collected once or twice each year.

Number of Individuals Affected

The QResearch database includes approximately 40 million patients of all ages including current, past and deceased patients.

Geographical Area

England.

Describe the context of the processing: what is the nature of your relationship with the individuals? How much control will they have? Would they expect you to use their data in this way? Do they include children or other vulnerable groups? Are there prior concerns over this type of processing or security flaws? Is it novel in any way? What is the current state of technology in this area? Are there any current issues of public concern that you should factor in? Are you signed up to any approved code of conduct or certification scheme (once any have been approved)?

Nature of Relationship with Data Subjects

The GP practices are the original data controller for the personal data. Optum and TPP are data processors for the GP practices to make the data available on the IM1 platform. QMUL is the data controller for anonymisation process following extraction from the IM1 platform. QMUL is also the data controller for the resulting research database QResearch. QMUL has no relationship with the individual patients. QResearch data are effectively anonymised; residual re-identification risk depends on content, linkage and access context and is assessed for each approved use, with additional consideration of children and increased-risk groups.

Individuals' Control

Individuals are able to opt out of their GP sharing their health records for research and planning by completing a type 1 opt out form, a recorded National Data Opt Out or by requesting their GP to add a QResearch specific opt out code to their records. National data opt-outs, type 1 objections and QResearch specific objections recorded in the available GP data are applied during transformation before the research extract is stored. A later opt-out applies prospectively to a subsequent extract but does not trigger a retrospective deletion to historical data.

Individuals' Expected Use of Data

There is a privacy notice which is available either from the GP practice or via the QResearch website which means people will know about the research uses of their data. https://www.qresearch.org/information/patient-information-and-privacy-notice/

Children and Other Vulnerable Groups

QResearch includes children and others who may face greater confidentiality harm, particularly where rare conditions, sensitive records or linked data could make someone identifiable. The QResearch team assess these risks before access, taking account of the data, proposed linkage and access setting. Each approved project uses only the fields needed, study-specific identifiers and trained researchers in the TRE, with restrictions on re-identification and onward sharing. The QResearch team manually checks outputs, including small numbers and risks, before release. These controls operate under the approved Protocol, Confidentiality Advisory Group support (25/CAG/0147) and applicable opt-out arrangements.

Prior Concerns and Security Flaws

There have been no prior issues of concern or security flaws

Novelty of Processing

There is no novel of processing. Other academic institutions and organizations undertake similar activities.

Relevant Processing Technologies

Industry standard cryptographic techniques are used to encrypt the NHS number to create an identifier which is unique to QResearch and cannot be reverse engineered and hence does not identify an individual. Only NHS England approved key holders have access to the cryptographic keys, ensuring that patient identities remain protected. NHS England remains responsible for key governance. One named QMUL individual is granted time-limited, purpose-specific access to the salt value solely for the extraction process. That individual does not have simultaneous access to the QResearch environment or NHS England datasets. Approved key holders do not have access to the QResearch database.

Potential Current Issues of Public Concern

There are no current issues of public concern. We have patient and public representatives on the QResearch Advisory Board and QResearch Scientific Committee and these have not raised any concerns. Indeed, the public are likely to be reassured about the care we are taking to ensure that users meet the criteria for database access.

Details of any Approved Code of Conduct and Certification

All staff accessing data need to have completed information governance and research training. QResearch is also governed by the NHS England Data Security Protection Tool Kit
https://www.dsptoolkit.nhs.uk/OrganisationSearch/8HX86

Describe the purposes of the processing: what do you want to achieve? What is the intended effect on individuals? What are the benefits of the processing – for you, and more broadly?

Desired Outcome of Processing

Benefits of Processing

The purpose of the processing is to effectively anonymise data efficiently and securely so that only effectively anonymised data are stored in the medical research database. The mechanism for effectively anonymising the NHS Number is also applied to other NHS data sets by NHS England which are supplied to QResearch such as hospital and mortality data. This mechanism enables the GP data to be linked to hospital and mortality data.

The benefit is then that the resulting QResearch data can be used for medical research to generate new knowledge to improve understanding of disease or the management of patient care in the NHS.

Effects on Participants

Whilst there is no direct effect or benefits on participants (as participants cannot be identified from the resulting database), patients can benefit from the results of medical research using their effectively anonymised data.

Step 3: Consultation Process

Consider how to consult with relevant stakeholders: describe when and how you will seek individuals' views – or justify why it's not appropriate to do so. Who else do you need to involve within Queen Mary? Do you need to ask any processors to assist? Do you plan to consult information security experts, or any other experts?

Consultation with individuals Involved within QMUL

We have consulted

- Gerry Collins, Head of Contracts at JRMO, QMUL
- Maria Frovola, Associate Commercialisation Manager, Queen Mary Innovations
- Paul Smallcombe, Records & Information Compliance Manager, QMUL
- Mr Michael Garvey Eckett, Data and AI Research Governance Lead

Organisations outside of QMUL

We have consulted with

- QResearch Advisory Board which has patient and professional representation.
- Richard Langley - Head of Information Governance, NHS England
- The IM1 team at NHS England
- Dr Tom Nichols, RCGP Joint Chair of the Joint Committee on IT, representing the GP profession.
- Optum (GP system supplier)
- UseMyData patient organisation

Consultation with Information Security Experts

We have consulted with information security experts Dancing Houses Consulting Ltd who provide IT software and database support for QResearch under contract with QMUL.

Step 4: Assess Necessity and Proportionality

Describe compliance and proportionality measures, in particular: what is your lawful basis for processing? Does the processing actually achieve your purpose? Is there another way to achieve the same outcome? How will you prevent function creep? How will you ensure data quality and data minimisation? What information will you give individuals? How will you help to support their rights? What measures do you take to ensure processors comply? How do you safeguard any international transfers?

Lawful Basis for Processing

The lawful bases for processing personal data for this study are **Article 6(1)(e) of UK GDPR**, which states the processing is necessary to perform a task in the public interest and **Article 9(2)(j) of UK GDPR**, which states the processing is necessary for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Article 89(1) based on the Data Protection Act 2018. This is supported by the University Charter which states research as a purpose of the organisation.

In addition, because the research requires access to identifiable NHS patient information without consent at the point of extraction in order to de-identify it before storing the data in the QResearch database, section 251 support is in place to set aside the common-law duty of confidentiality (CAG reference 25/CAG/0147, dated 17/11/2025). The relevant NHS England data sharing arrangements are governed through the applicable Data Sharing Agreement. Any personal data will be handled in accordance with the UK GDPR and the Data Protection Act 2018

Describe how Purpose Achieved

The purpose is achieved by extracting medical data from GP practices, removing or transforming identifiable data during processing then storing the data in a research database so that the data can be analysed by researchers.

Function Creep

The QResearch database has defined ethics committee approval, governance processes (including an oversight Advisory Board) and data sharing agreements which limit the purposes for which the data can be used. These measures will prevent function creep.

Data Quality

Data gathered by GP practice as part of routine care. There are some data quality controls at the point of data entry in the GP systems. The IM1 platform, which is used for data extraction, has been extensively tested by NHS England and organisations which use the platform ensuring data integrity. Data quality checks are also run on the resulting anonymised data to ensure data has transferred correctly.

Data Minimisation

QResearch IMOP–EMIS Specification provides an action, date range and proposed purpose for each table and associated fields. . The action is either “Drop”, “Keep”, “Hash” or “Transform”. Patient and event identifiers needed for linkage are hashed; date of birth is reduced to year; postcode is used for the deprivation lookup and then dropped; free text and contact details are not retained in the research extract. The inclusion of all ages supports longitudinal analyses across the approved population. QMUL has reviewed and approved the final necessity, period and transformation rationale, with technical verification that the implemented extract follows the schedule and that confidentiality flags are acted on before they are discarded.

Informing Data Subjects

GP practices display transparency notices in their waiting rooms and on their websites to inform patients their data is being used for research. There is also a transparency notice on the QResearch website.

https://www.qresearch.org/information/patient-information-and-privacy-notice/

Supporting Data Subjects’ Rights

National data opt-outs, Type 1 objections and QResearch-specific objections recorded in the available GP input are applied to subsequent IM1 transformations before research availability. A newly recorded opt-out is prospective, including in relation to historical events in the new extract; it does not automatically erase historical data.

Ensuring Compliance with GDPR

GDPR compliance is ensured by only limiting processing to the data necessary for the approved research purposes including special category health data; ensuring information is available through a privacy notice; ensuring users opt out of future data collections; ensuring there is a regular review of this policy at least annually. Only approved personnel with training can access data.

Safeguarding International Transfers

There are no international transfers of personal data. All the data reside in the UK at all times.

Risk Key					
Likelihood	**Impact**				
	1 - Negligible	2 - Minor	3 - Moderate	4 - Major	5 - Catastrophic
1 – Rare	**Low 1**	**Low 2**	**Low 3**	**Low 4**	**Low 5**
2 – Unlikely	**Low 2**	**Low 4**	**Low 6**	**Medium 8**	**Medium 10**
3 – Possible	**Low 3**	**Low 6**	**Medium 9**	**Medium 12**	**High 15**
4 – Likely	**Low 4**	**Medium 8**	**Medium 12**	**High 16**	**High 20**
5 – Almost Certain	**Low 5**	**Medium 10**	**High 15**	**High 20**	**High 25**

Risk Type	**Risk Score**
Low	1-7
Medium	8-14
High	15-25

Step 5: Identify and Assess Risks				
Describe source of risk and nature of potential impact on individuals. Include associated compliance and corporate risks as necessary.		**Likelihood** 1 - 5	**Impact** 1 - 5	**Risk Score** Likelihood x Impact = Risk Score
1	Personal data is kept for longer than needed. The risk of this is very low as personal data is processed only transiently within the controlled extraction/transformation stage, is effectively anonymised at the earliest opportunity before storage in QResearch and no identifiable personal data is stored in the QResearch database.	1	2	2
2	Key management / misuse: unauthorised or inappropriate access to the salt/keying material could increase the risk of linkage or re-identification. Existing controls include NHS England governance of the keying material; named, purpose-specific and time-limited salt access; prohibition on retaining/copying the salt; and separation of duties so that the individual with salt access does not have simultaneous access to the QResearch environment or NHS England datasets.	1	3	3
3	Unauthorised access: unauthorised access to source data during extraction or to the QResearch environment could result in inappropriate disclosure of health information. Existing controls include restricted authorised access, information governance/research training, technical access controls and use of the QResearch Trusted Research Environment for approved research access.	1	3	3
4	Re-identification: there is a residual risk that effectively anonymised records could be re-identified or linked inappropriately. Existing controls include removal of direct identifiers, one-way cryptographic hashing, separation of key/data access, data minimisation and a second project-specific hashing layer for research extracts.	1	2	2
5	Inappropriate onward disclosure: effectively anonymised data or research outputs could be disclosed outside the approved governance framework. Existing controls include project-specific approval, sublicensing where applicable, access within the QResearch Trusted Research Environment and controlled output checking/airlock arrangements.	1	3	3
6	Children and people with heightened confidentiality risks: rare or sensitive diagnoses, safeguarding and care histories, small cohorts, precise chronology and linkage could allow singling out or inference about a child or family even without names. Disclosure or inappropriate access could cause stigma, safety concerns or loss of trust. Existing baseline controls include the fact that QResearch only has a sample of patients, is effectively anonymised, limited access, project approval, study IDs and TRE output review.	1	3	3

Step 6: Identify Measures to Reduce Risk				
Options to reduce or eliminate risk: Identify additional measures you could take to reduce or eliminate risks identified as medium or high risk in step 5.		**Effect on Risk** (choose from Eliminated, Reduced or Accepted)	**Residual Risk**	**Measure Approved** Yes/No
1	Not applicable as there are no medium or high risks identified in step 5. The controls will continue to be maintained and reviewed through the existing QResearch, QMUL, and NHS England governance arrangements.			

Step 7: Sign Off and Record Outcomes			Notes
Approvals	Name - Position	Date	
Measures Approved by:	Julia Hippisley-Cox	01.10.2026	Integrate actions back into project plan, with date and responsibility for completion
Residual Risk Approved by:			If accepting any residual high risk, consult the ICO before going ahead
DPO Advice Provided by:	Paul Smallcombe – Records & Information Compliance Manager	02/10/2026	DPO should advise on compliance, step 6 measures and whether processing can proceed
Summary of DPO Advice: No medium or high risks. Processing can proceed.			
DPO Advice Accepted/Overruled by:			If overruled, you must explain your reasons
Comments:			
Consultation Responses Reviewed by:			If your decision departs from individuals' views, you must explain your reasons
Comments:			
DPIA kept under review by:			The DPO should also review ongoing compliance with DPIA

Integration of Risk Mitigation Measures			
Decide who is responsible for the integration of the risk mitigation measures which were decided and agreed upon in step 6. Record the completion status.		**Responsibility**	**Status** (RAG)
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			